

УНИВЕРЗИТЕТ У БЕОГРАДУ
ФАКУЛТЕТ ОРГАНИЗАЦИОНИХ НАУКА

Андрија А. Петровић

**Развој интегрисаног модела за предикцију
гужви и одређивање оптималног броја
активних канала у модулу**

ДОКТОРСКА ДИСЕРТАЦИЈА

Београд, 2019.

UNIVERSITY OF BELGRADE
FACULTY OF ORGANIZATIONAL SCIENCES

Andrija A. Petrović

**Development of an integrated model for
congestion prediction and determination of
optimal number of active channels in module**

DOCTORAL DISSERTATION

Belgrade, 2019.

Ментори:

1. др Борис Делибашић, редовни професор
Универзитета у Београду, Факултет организационих наука
2. др Угљеша Бугарић, редовни професор
Универзитета у Београду, Машински факултет

Чланови комисије:

1. др Милија Сукновић, редовни професор
Универзитета у Београду, Факултет организационих наука
2. др Драгана Макајић-Николић, ванредни професор
Универзитета у Београду, Факултет организационих наука
3. др Милош Јовановић, доцент
Универзитета у Београду, Факултет организационих наука
4. др Душан Петровић, редовни професор
Универзитета у Београду, Машински факултет
5. др Младен Николић, доцент
Универзитета у Београду, Математички факултет

Датум одбране: _____ 2019. godine

Захвалница

Имајући у виду да ми је при изради ове докторске дисертације помогао велики број људи, дужан сам да се истима захвалим.

Пре свега, најискреније се захваљујем свом ментору проф. др Борису Делибашићу који ме је саветодавно водио и пружио изузетну подршку при изради дисертације. Поред тога, изузетну захвалност дугујем и свом коментору проф. др Угљешу Бугарићу на великом стрпљењу и помоћи коју ми је пружио приликом израде ове дисертације. Проф. Делибашић и проф. Бугарић су се опходили, не само као ментори, већ као и добри људи, пружајући ми свакодневну помоћ у циљу савладавања препрека које сам имао при изради ове тезе.

Захвалност за дуге разговоре који су ми помогли да дефинишем тему дисертације, као и да исту реализујем, дугујем доценту др Милошу Јовановићу. Огромну захвалност дугујем и доценту др Младену Николићу без чијих идеја и помоћи ова дисертација не би добила облик у коме се сада налази. Без огромног знања доцента Јовановића и доцента Николића ова дисертација сигурно не би била реализована.

Имајући у виду да ме је у академски свет увео мој кум проф. др Јован Филиповић, дужан сам да му се захвалим на пружању подршке и мотивације при изради дисертације. Захваљујем се и својим колегама и пријатељима из кабинета, Сандру Радовановићу и доценту др Милану Вукићевићу.

Ову докторску тезу посвећујем својој породици, оцу проф. др Александру Петровићу, мајци Весни Петровић и сестри Ани Петровић који су ми несебично пружили подршку, љубав и разумевање свих ових година, а посебно у тренуцима када ми је то било најпотребније.

Развој интегрисаног модела за предикцију гужви и одређивање оптималног броја активних канала у модулу

Апстракт

Последњих година велики број истраживања је усмерен ка предикцији саобраћајних гужви. Различите статистичке методе нису показале значајан допринос у предиктивним перформансама предикције гужви. Стога се данас све чешће користе алгоритми машинског учења у циљу постизања задовољавајућих резултата предикције. У овој дисертацији, представљена је методологија за класификацију гужви на бази новоразвијеног модела Гаусових условних случајних поља за структурну бинарну предикцију (GCRFBC). Иста је успешно имплементирана на реалне проблеме предвиђања гужви. Методологија може бити успешно примењена на класификационе проблеме описане неусмереним графовима који се не могу ефективно решити стандардним условним случајним пољима (CRF). Новоразвијени модел, коришћен у методологији, је заснован на стандардним Гаусовим условним случајним пољима за регресију (GCRF) која су проширена латентним променљивим што даје бројне предности истом. Захваљујући латентној структури, учење и закључивање у модулу не захтева компликоване нумеричке процедуре, већ може бити решено аналитички. Поред тога, постојање латентне структуре омогућава да модел буде отворен ка даљим побољшањима. Три различита алгоритма су развијена: GCRFBCb (GCRFBC – Бајесовски), GCRFBCb-fast (GCRFBC – Бајесовски са апроксимацијом) и GCRFBCnb (GCRFBC – не-Бајесовски). Проширена метода локалне варијационе апроксимације сигмоидне функције коришћена је за решавање интеграла по латентним променљивим у Бајесовској верзији GCRFBC модела. У случају не-Бајесовског GCRFBC модела у учењу и закључивању је коришћена латентна променљива са максималном вредношћу функције густине вероватноће. Закључивање у GCRFBCb модулу је решено коришћењем Њутн-Котесовим формулама за једнодимензионалну интеграцију. Услед великог броја варијационих параметара, рачунски трошак учења је велики, стога је развијена брза верзија Бајесовског GCRFBC модела. Перформансе модела су евалуиране на синтетичким и реалним подацима. Показано је да се применом методологије остварују боље перформансе предвиђања гужви у поређењу са неструктурним моделима. Додатно су евалуирани рачунски и меморијски трошкови. Методологија је генерализована на примере из других домена. Детаљне предности и мане свих развијених модела су наглашене. У другом делу дисертације развијена је хибридна методологија за предвиђање индикатора саобраћаја који се заснива на комбинацији Гаусових условних случајних поља за регресију и класификацију. Услед коришћења структурних модела, методологија се користи за предвиђање индикатора саобраћаја на више излаза који су међусобно зависни. Поред тога, обезбеђује се учење из ретких података, односно података где многи излази немају никакаву вредност (ништа). Класификациони модел служи за елиминисање излаза са вредностима ништа, док регресиони модел служи за предвиђање индикатора саобраћаја на оним излазима који немају вредност ништа. Информације о индикаторима саобраћаја омогућавају ефикасан мониторинг саобраћаја, управљање, планирање као и доношење информација које учеснике у саобраћају могу да наведу на путање где гужве могу да се заобиђу. Предности и мане новоразвијене методологије приказане су на два примера. Први се тиче предвиђања гужви на ауто-путу E70-E75 који пролази кроз Србију, док је други проблем везан за предвиђање гужви на ски-центру Копаоник. У последњем делу дисертације развијена је методологија за

одређивања оптималног броја активних канала у будућности. Методологија је заснована на комбинацији рекурентних неуронских мрежа, теорије редова чекања и метахеуристика у циљу одређивања оптималног броја активних канала у будућности. Методологија се базира на предвиђању интензитета долазака и одређивању интензитета опслуживања у неком периоду у будућности. Коришћењем тих интензитета у моделима теорије редова чекања, поставља се функција циља која се оптимизује посредством избора броја активних канала у модулу. Приказана су два алгорита: први заснован на не-Бајесовском приступу одређивања броја активних канала у модулу и други заснован на Бајесовском приступу. На примеру одређивања оптималног броја наплатних рампи који треба да буде отворен у будућности на наплатној станици Врчин верификована је примена исте. Може се видети да у свим анализираним случајевима, резултати добијени новоразвијеном методологијом показују неупоредиво ниже очекиване укупне трошкове у поређењу са тренутном стратегијом отварања наплатних рампи.

Кључне речи: структурни пробабалистички модели, Гаусова условна случајна поља, класификација, регресија, предвиђање гужви, теорија редова чекања, нестационарни Марковљеви процеси, оптимизација.

Development of an integrated model for congestion prediction and determination of optimal number of active channels in module

Abstract

In the recent years, research committed in the field of congestion prediction present one of the most popular area of interest. A variety of novel methods for congestion prediction based on unstructured statistical (machine) learning have become the standard for congestion prediction. However, in this dissertation I argue that structured machine (statistical) learning algorithms can significantly improve congestion prediction performances. In this dissertation, a Gaussian conditional random field model for structured binary classification (GCRFBC) is proposed for solving problems of congestion prediction. The model is applicable to classification problems with undirected graphs, intractable for standard classification CRFs. The model representation of GCRFBC is extended by latent variables which yield some appealing properties. Thanks to the GCRF latent structure, the model becomes tractable, efficient and open to improvements previously applied to GCRF regression models. In addition, the model allows for reduction of noise, that might appear if structures were defined directly between discrete outputs. Three different forms of the algorithm are presented: GCRFBCb (GCRGBC - Bayesian), GCRFBCb-fast (GCRGBC - Bayesian approximation) and GCRFBCnb (GCRFBC - non-Bayesian). The extended method of local variational approximation of sigmoid function is used for solving empirical Bayes in Bayesian GCRFBCb variant, whereas MAP value of latent variables is the basis for learning and inference in the GCRFBCnb variant. The inference in GCRFBCb is solved by Newton-Cotes formulas for one-dimensional integration. Due to large numbers of variational parameters the computational costs of learning is significant, so fast version of GCRFBCb model is derived (GCRFBCb-fast). Models are evaluated on synthetic data and real data. It was shown that models achieve better congestion prediction performance than unstructured predictors. Furthermore, computational and memory complexity is evaluated. The generalization of the proposed models on other problems are discussed in details. Moreover, in the second part of this dissertation a hybrid model of two Gaussian Conditional Random Fields models (one recently proposed for classification, and one for regression) for inference of traffic speed, a relevant variable for traffic state estimation and travel information systems is proposed. It addresses two specifics of the problem - sparsity in traffic data and the fact that observations are not independent. It does so by combining a Gaussian conditional random field binary classification (GCRFBC) model (for gating of free-flow regimes and potentially congested traffic regimes) and a regression Gaussian conditional random field (GCRF) model with varying structure of nodes for prediction of traffic speed in dependent variables of potentially congested traffic regimes only. The information provided by the model can help in traffic monitoring, control, and planning, as well in congestion mitigation by providing information for avoiding congested routes. The proposed model is tested on two large-scale networks in Serbia, an arterial E70-E75 335km long highway stretch as well as in the ski resort Kopaonik with 55 km of ski slopes. The advantages and disadvantages of hybrid model is shown. In the last section of dissertation methodology for determination of optimal number of active channels in module is developed. Methodology is based on combination of recurrent neural networks, queuing theory and metaheuristics. Recurrent neural networks are used for prediction of arrival intensity and estimation of service intensity in some period in future. The predicted intensities are used in queuing theory models in order to develop objective function

that has to be minimized. Two different algorithms are presented: the first one is based on non-Bayesian and the second one is based on Bayesian approach. The application of methodology is presented on the example of pay toll ramp optimization on pay toll station Vrčin. In all analyzed cases the estimated total costs are significantly reduced compared to current policy.

Key words: structural probabilistic models, Gaussian conditional random fields, classification, regression, congestion prediction, queuing theory, non homogenous Markov process, optimization.

Садржај

1	Увод	4
1.1	Проблем и предмет истраживања	9
1.2	Циљеви истраживања	10
1.3	Полазне хипотезе.....	11
1.4	Методе научног истраживања.....	11
1.5	Структура рада	12
2	Осврт на релевантне библиографске изворе и остварене резултате	14
2.1	Преглед литературе у области класификације коришћењем условних случајних поља	14
2.2	Преглед литературе у области предвиђања индикатора саобраћаја	17
2.3	Преглед литературе у области примене математичког програмирања и модела теорије редова чекања у саобраћају	19
3	Теоријске основе развијених методологија	20
3.1	Гаусова условна случајна поља за регресију.....	20
3.2	Моделовање система опслуживања	23
3.2.1	Долазни ток клијената (извор).....	23
3.2.2	Ред	24
3.2.3	Дисциплина опслуживања	25
3.2.4	Механизам опслуживања	25
3.2.5	Расподеле вероватноћа стања система.....	26
3.2.6	Стохастички процеси.....	29
3.2.6.1	Случајни процес Марковљева	30
3.2.6.2	Нехомогени случајни процес Марковљева	32
3.2.7	Укупни трошкови система опслуживања.....	33
3.3	Рекурентне неуронске мреже	34
3.3.1	GRU мреже	36
3.3.2	LSTM мреже	36
3.3.3	LSTM мреже са унутрашњом стандарднизацијом слојева	37
3.4	Метахеуристике за одређивање оптималне полисе управљања отварања и затварања канала у модулу опслуживања	38
3.4.1	Диференцијална еволуција	39
3.5	Симулација нехомогеног Марковљевог ланца.....	40

4	Методологија за класификацију гужви помоћу Гаусових условних случајних поља за класификацију.....	42
4.1	Гаусова условна случајна поља за класификацију – репрезентација.....	42
4.2	Гаусова условна случајна поља за класификацију - закључивање	45
4.2.1	Закључивање у GCRFC _b моделу	45
4.2.2	Закључивање у GCRFC _{nb} моделу	46
4.3	Гаусова условна случајна поља за класификацију - учење.....	47
4.3.1	Учење у GCRFC _b моделу.....	47
4.3.2	Учење у GCRFC _{nb} моделу.....	52
4.4	Експерименти над синтетичким подацима.....	53
4.4.1	Опис експеримента	54
4.4.2	Предиктивне перформансе.....	54
4.4.3	Меморијска комплексност и рачунски трошак.....	56
4.5	Убрзање Бајесовских Гаусових условних случајних поља за класификацију ..	58
4.5.1	Алгоритам учења убрзане верзије Гаусових условних случајних поља за класификацију са кластеровањем инстанци и излаза	61
4.5.1.1	К-средина и мини беч К-средина	62
4.5.1.2	Мешавина Гаусијана	62
4.5.2	Алгоритам учења убрзане верзије Гаусових условних случајних поља за класификацију са одређивањем вероватноће припадања кластеру.....	63
5	Методологија за предвиђање индикатора саобраћаја -структурна предикција ретких зависних индикатора саобраћаја у великим мрежним системима	64
5.1	Комбинација регресионих и класификационих Гаусових условних случајних поља – методологија предвиђања индикатора саобраћаја	65
6	Методологија за одређивање оптималног броја активних канала у модулу	68
6.1	Опис методологије	68
7	Примена развијених методологија на скупу података са ауто-путева.....	73
7.1	Примена методологије за класификацију гужви коришћењем Гаусових условних случајних поља за класификацију	73
7.1.1	Дискусија	81
7.2	Примена комбинованог приступа регресије и класификације за предикцију просечне брзине возила на ауто-путевима	82
7.2.1	Дискусија	87

7.3	Одређивање оптималног броја отворених наплатних рампи на наплатној станици у неком периоду у будућности.....	88
7.3.1	Проналажење оптималне полисе.....	90
7.3.2	Припрема података	93
7.3.3	Тестирање времена опслуживања клијената.....	95
7.3.4	Учење и верификација рекурентних неуронских мрежа	98
7.3.5	Резултати добијени оптимизацијом	102
7.3.6	Симулација система опслуживања.....	111
7.3.7	Дискусија	114
8	Примена развијених методологија на реалним проблемима из других домена.....	115
8.1	Предвиђање гужви на скијалиштима	115
8.2	Предвиђање просечне брзине скијаша на скијалиштима.....	120
8.3	Класификација гена по функционалности.....	123
8.4	Семантичка класификација сцена	124
8.5	Класификација емоција из музике.....	126
8.6	Класификација болести	128
8.7	Класификација врсте птица уз помоћ аудио сигнала	130
9	Закључак.....	132
10	Листа слика	143
11	Листа табела.....	146
12	Литература.....	148

1 Увод

Одређивање оптималног броја канала код система опслуживања је проблем који треба да буде решен у неком периоду у будућности. На овај начин могу се исправно испланирати расположиви ресурси и ангажовати она количина истих како би укупни трошкови били минимални. Стога је неопходно предвидети како ће се мењати вероватноћа стања система у будућности, за различити број канала, а затим изабрати онај број који ће минимизовати очекиване укупне трошкове система за опслуживање. Развојем хардевера традиционални статистички модели се замењују моделима машинског учења. Самим тим показује се да се проблеми предикције гужви, предвиђања индикатора саобраћаја и одређивања оптималног броја активних канала у модулу у будућности могу на најефикаснији и најбољи начин решити коришћењем алгоритама машинског учења.

Машинско учење (енг. *Machine learning*) представља дисциплину рачунарских наука која за циљ има развијање модела реалних система на основу доступних података из постојећих система. Овакви модели омогућавају откривање законитости у подацима. Најчешће се овакви модели користе у сврху надгледаног учења, предвиђања, односно откривања вредности одређеног циљног атрибута (Y) на основу улазних атрибута (X). Надгледано учење обухвата регресионе моделе, где је циљни атрибут континуалан (реалан број) и класификационе моделе, где је циљни атрибут дискретна вредност (у случају бинарне класификације 1 и 0). Развијање модела за предвиђање обухвата неколико корака [1]:

1. Дефинисање форме модела;
2. Учење непознатих параметара модела из историјских података;
3. Закључивање, односно одређивање вредности циљног атрибута (Y) на основу познатих атрибута (X) и научених параметара.

Машинско учење данас представља једну од најбрже растућих дисциплина и ослања се на велики број научних дисциплина које покривају подручје математичке вероватноће и статистике, информационе теорије, операционих истраживања, теорије одлучивања и других инжењерских наука. Поред тога бројне научне дисциплине имплементирају моделе машинског учења у сопствене сврхе, па тако се на пример примене машинског учења срећу у: биомедицини и медицини [1], маркетингу [2], роботизи [3], механици флуида [4], транспорту [5] итд. Битно је напоменути да се

развојем алгоритама машинског учења омогућава налажење комплексних веза између атрибута које се не могу једноставно увидети. Међутим треба имати у виду да развијени алгоритми не могу представљати потпуну замену у процесу доношења одлуке, већ их треба посматрати као средство које ће доносиоцу одлуке знатно помоћи у процесу доношења одлука.

Велики број различитих алгоритама за машинско учење је развијен током последњих педесет година. На основу теореме „Непостојања бесплатног ручка“ (енг. „*No free lunch theorem*“) [6] не може се тврдити да постоји један најбољи алгоритам за прављење модела, већ сваки алгоритам има одређене предности које одговарају одређеним типовима података, а другима не одговарају. Повећањем количине и разноврсности података са корелисаним циљним атрибутима (излазним атрибутима), познатијим као структурираним циљним атрибутима, јавља се потреба за експлоатацијом додатних информација које постоје између инстанци зависног (излазног) атрибута, у циљу побољшања перформанси за предвиђање. Сумиран преглед великог броја структурираних класификатора за бинарну и вишеструку класификацију је приказан у [7].

Један од најбољих модела заснованих на вероватноћи за проблеме структуриране класификације су условна случајна поља (енг. „*Conditional random fields*“, скраћено CRF). Највећа предност условних случајних поља представља његову дискриминативну природу у предвиђању, која се резултује у релаксацији претпоставке о независности излазних атрибута који могу бити временски, просторно, или на други начин повезани (дефинисањем сличности) и смањивању велике системске пристрасности (енг. „*bias*“) која је присутна у многим графовским моделима. Поред тога могућност једноставне евалуације градијента и постојање информације о вероватноћи припадности класе, омогућили су условним случајним пољима велику популарност у различитим применама. Ипак, условна случајна поља имају и мане које се тичу сложености градијента, чија евалуација може бити изузетно рачунарски скупа посебно при учењу и закључивању. Понекад је чак и немогуће евалуирати градијент или закључивати о вредностима циљних атрибута без коришћења одређених апроксимативних метода. У циљу решавања ових недостатака развијен је велики број различитих алгоритама који су прилагођени различитим проблемима.

Регресиони модели који су базирани на условним случајним пољима су у много мањој мери развијени у односу на класификационе моделе. Један од најпопуларнијих регресионих модела, заснован на условним случајним пољима, има форму

мултиваријационе нормалне расподеле и познат је као Гаусова условна случајна поља (енг. „*Gaussian conditional random fields*“, скраћено GCRF). Основна претпоставка модела је да се „сличности“ (везе) између инстанци излаза моделују преко квадратне форме. Условна случајна поља у форми мултиваријационе нормалне расподеле имају низ предности који се тичу: конвексности параметара у простору, једноставне форме учења и ефикасног закључивања.

Данас постоји велики број научних радова на тему предвиђања гужви и протока саобраћајних возила. Услед недостатка довољно добрих доменских модела или превелике комплексности истих, поставља се потреба за коришћењем широког спектра алгоритама машинског учења. Ма и други [8] су приказали предности коришћења дубоког учења у предикцији саобраћајних гужви. На бази комбинације ограничених Болцманових машина (енг. „*Restricted Boltzmann machine*“) са рекурентним неуронским мрежама (енг. „*Recurrent Neural Network*“) и посредством историјских података добијених из ГПС сигнала такси возила, развијен је модел који има прецизност предвиђања од 88% за временски период од 6 минута у будућности. Сличан модел заснован на аутоенкодерима у комбинацији са дубоким неуронским мрежама приказан је у [9]. Већина ових модела, везе између корелисаних излаза моделује посредством неуронских мрежа, тако да сам модел проналази везе између излаза [10]. Иако наведени модели успешно раде посао при предвиђању протока возила у кратким временским интервалима, поставља се питање да ли се наведени проблем може приказати као класификациони и да ли је могуће користити неки други алгоритам машинског учења где се структуриране везе између излаза могу моделовати независно од начина учења алгорита. Тема ове докторске дисертације је усмерена на проблем предикције гужви у комбинацији са одређивањем оптималног броја канала у будућим временским тренуцима у циљу минимизације укупних трошкова, помоћу унапређених алгоритама Гаусових условних случајних поља.

Први део дисертације базиран је на развоју нове методологије за класификацију саобраћајних гужви која се заснива на новоразвијеном математичком моделу класификационог алгорита Гаусових условних случајних поља (енг. „*Gaussian conditional random fields for classification*“, скраћено GCRFC) у циљу структуриране предикције гужви на наплатним станицама. Поред тога развијена су два облика алгорита: први који је заснован на емпиријски-Бајесовском приступу учења и закључивања (енг. „*Gaussian conditional random fields for classification - Bayesian*“, скраћено GCRFCb) и други који је заснован на не-Бајесовском приступу (енг. „*Gaussian*

conditional random fields for classification – non Bayesian“, скраћено GCRFCnb). Услед високог рачунског трошка у фази учења алгоритма заснованог на Бајесовском приступу, неопходно је увести извесне апроксимације. Самим тим, у дисертацији је развијена и брза апроксимативна верзија Гаусових условних случајних поља за класификацију (енг. „*Gaussian conditional random fields for classification – Bayesian fast*“, скраћено GCRFCb-fast). Над синтетичким подацима одређен је рачунски и меморијски трошак коришћења новоразвијених математичких модела за структурну класификацију.

Услед различитих начина дефинисања периода када је била и када није била гужва, понекада је много корисније предвиђати одређене индикаторе на основу којих се може установити да ли постоји гужва или не. Један од таквих индикатора је просечна брзина возила која долазе на наплатну рампу у одређеном периоду у будућности. Везе између наплатних рампи на једној деоници пута јасно показују да су просечне брзине возила на различитим наплатним рампама међусобно корелисане, стога је коришћење структурираних алгоритама при предвиђању просечне брзине оправдано. Услед веома ретких података, односно периода времена када на одређеној наплатној рампи није нико прошао, у другом делу дисертације развијена је методологија за предвиђање индикатора саобраћаја заснована на комбинацији математичких модела регресионих и класификационих Гаусових условних случајних поља. Предност примене ове методологије у случају ретких података у односу на примену само регресије јасно је приказана у дисертацији.

Трећи део дисертације базира се на развијању методологије одређивања оптималног броја активних канала (наплатних рампи) у модулу (наплатној станици) у будућности. Овај процес се заснива на комбинацији рекурентних неуронских мрежа (енг. „*Recurrent neural networks*“, скраћено RNN), модела теорије редова чекања и оптимизације. Коришћене су различите архитектуре рекурентних неуронских мрежа: дугачка-кратка меморија [11] (енг. „*Long-short term memory*“, скраћено LSTM), дугачка-кратка меморија са унутрашњом стандардизацијом слојева [12] (енг. „*Long-short term memory with layer normalizaztion*“, скраћено LSTM-ln), закључаном рекурентном јединицом [13] (енг. „*Gated recurrent unit*“, скраћено GRU) и обичном рекурентном мрежом [14] (енг. „*Recurrent neural network*“, скраћено RNN). Посредством рекурентних мрежа предвиђани су средњи интензитети долазака клијената (возила) у систем који су коришћени као улазни параметри за одређивање оптималног броја активних канала (наплатних рампи) за опслуживање применом модела теорије редова чекања и метахеуристика за оптимизацију. Узет је у обзир утицај грешке у предвиђању

посредством нове мреже која предвиђа колика је грешка (стандардна девијација) усвојеног модела за предвиђање средњег интензитета долазака клијената. Одређивање оптималног броја активних канала извршено је на бази минимизације укупних трошкова који се састоје од трошкова чекања и трошкова опслуживања. Минимизација је извршена посредством Бајесовог и не-Бајесовог приступа.

Све развијене методологије су тестиране и верификоване на реалним подацима. На бази протока возила током 2017. године на ауто-путевима Републике Србије на правцима Београд-Адашевци и Београд-Ниш развијене методологије су валидиране посредством унакрсне валидације. Имајући у виду да су методологије за класификацију гужви и предвиђање индикатора саобраћаја наменски биле развијене у циљу решавања проблема гужви на ауто-путевима, показано је на подацима из других реалних домена (медицина, препознавање слика итд.) да се њихова примена може успешно генерализовати на многобројне области које захтевају класификацију повезаних (структурираних) излаза. Методологија за одређивање оптималног броја активних канала који треба да буде отворен у модулу је валидирана поређењем са тренутном стратегијом (полисом или политиком) отварања наплатних рампи на скупу података са ауто-путева. Верификација методологије је извршена симулацијом нехомогеног Марковљевог ланца.

Циљ ове дисертације је реализован кроз развој три методологије:

- 1) Методологија за класификацију гужви базирана на новоразвијеном моделу Гаусових условних случајних поља, која се састоји од:
 - i. Дефинисање појма гужви посредством агрегираних статистика из података за довољно дугачак временски период.
 - ii. Прављење атрибута из података, који се тичу статистика усмерених на брзину кретања возила, временске прилике у том тренутку итд.
 - iii. Дефинисање мрежних структура између система осплуживања (наплатних станица) посредством укључивања параметара који се тичу корелације, заједничке информације (енг. „*Mutual information*“) итд.
 - iv. Учење не-Бајесовског и Бајесовског основног и брзог модела заједно са одговарајућим неструктурним предикторима.
 - v. Валидација модела посредством методе унакрсне валидације и избора модела Гаусових условних случајних поља за класификацију који има најбоље перформансе предикције.
- 2) Методологије за предвиђање индикатора саобраћаја, која се састоји од:

- i. Прављења атрибута из података, који се тичу статистика усмерених на брзину кретања возила, временске прилике у том тренутку итд.
 - ii. Дефинисања структура регресионих и класификационих Гаусових условних случајних поља
 - iii. Учења не-Бајесовског и Бајесовског спророг и брзог класификационог модела у комбинацији са регресионим моделом, заједно са свим неструктурним предикторима.
 - iv. Валидације модела посредством методе унакрсне валидације и избора модела Гаусових условних случајних поља за класификацију који има најбоље перформансе предикције.
- 3) Методологија за одређивање оптималног броја активних канала у модулу у будућности, која се састоји од:
- i. Утврђивања карактера нестационарности Марковљевог ланца са континуалним временом и дискретним стањима на основу расположивог скупа података.
 - ii. Одређивања битних статистика које се тичу средњег интензитета долазака клијената, односно прављење атрибута релевантних за предвиђање средњег интензитета долазака клијената.
 - iii. Учење рекурентне мреже у циљу одређивања средњег интензитета долазака клијената у неком будућем периоду времена.
 - iv. Минимизације укупних трошкова у циљу налажења оптималног број канала опслуживања у будућности.
 - v. Верификације и валидације развијене методологије.

1.1 Проблем и предмет истраживања

Предмет истраживања дисертације је превасходно усмерен на развој методологија за предвиђање гужви, индикатора гужви и оптималног броја активних канала у модулу у будућности. Развијене методологије су коришћене за решавање проблема касификације гужви, предикције индикатора саобраћаја, као и за одређивање оптималног броја наплатних рампи које треба да буде отворене у будућности на наплатним станицама. Стога је предмет истраживања дисертације усмерен ка решавању проблема који се готово свакодневно срећу у саобраћају, а за које још увек не постоје методологије за добијање задовољавајућих решења.

На основу претходног описа може се видети да ова дисертација у потпуности покрива проблеме предикције гужви и дефинисања оптималног броја активних канала осплуживања у сложеним масовним и повезаним системима осплуживања.

1.2 Циљеви истраживања

Општи циљ истраживања подељен је на три дела:

1. Развој методологије за класификацију гужви на наплатним рампама базиране на новоформираном математичком моделу класификације помоћу Гаусових условних случајних поља, чији је циљ да побољша предикцију у односу на неструктурне предикторе.
2. Развој методологије за предвиђање индикатора саобраћаја који се заснива на комбинацији регресионих и класификационих математичких модела Гаусових условних случајних поља. Циљ ове методологије је да предвиди индикаторе саобраћаја у мрежним системима са великим бројем података који немају никакву вредност (вредност ништа).
3. Развој методологије за одређивање оптималног броја активних канала у модулу у будућности комбинацијом рекурентних неуронских мрежа и теорије редова чекања. Ова методологија има за циљ да одреди оптималну политику активирања канала у модулу у будућности, на основу унапред дефинисане функције трошкова.

Посебан циљ истраживања, који подразумева дефинисање праваца истраживања докторске дисертације, односно разраду општег циља, је:

1. Дефинисање метода за ненадгледано одређивање границе помоћу којег се стања могу класификовати као "гужва" или "није гужва".
2. Развијање алгоритама за убрзање Гаусових условних случајних поља заснованог на Бајесовском приступу.
3. Генерализација развијених алгоритама за предикцију гужви, помоћу тестирања истих на примерима из других домена.
4. Комбиновање рекурентне мреже у циљу предикције параметара дистрибуције времена долазака возила.
5. Моделовање система осплуживања и одређивање оптималног броја наплатних рампи који треба да буде отворен.

1.3 Полазне хипотезе

Општа хипотеза

$X(0)$ – Интеграција машинског учења и теорије редова чекања доприноси унапређењу методологије за решавање проблема предикције гужви и одређивања оптималног броја активних канала у модулу у будућности.

Посебне хипотезе

$X(1)$ – Креирање и развој математичких модела за класификацију заснованог на Гаусовим условним случајним пољима показује статистички боље резултате у односу на класичне моделе за предвиђање гужви. (Рађијена су три алгорита, два заснована на Бејесовском приступу закључивања и учења и један заснован на не-Бајесовском приступу.)

$X(2)$ – Комбинација регресионих и класификационих алгоритама у циљу одређивања битних статистика за предикцију гужви може додатно унапредити моделе за предвиђање гужви посебно у случајевима када су подаци ретки и излазни корелисани.

$X(3)$ – Комбинација предикције параметара дистрибуције односно средњег интензитета доласка возила у комбинацији са теоријом редова чекања (симулацијом) омогућава одређивање оптималног броја канала опслуживања који треба да буде отворен у будућности.

1.4 Методе научног истраживања

Основне методе истраживања које ће се користити при решавању постављеног проблема су следеће:

- Метода дескрипције која ће се користити за описивање појава и процеса од интереса, уз објашњења важних обележја описаних појава и процеса, уочавање законитости и узрочних веза и односа;
- Метода анализе ће се употребљавати кроз поступак научног истраживања ращлањивањем сложених појмова, судова и закључака на њихове једноставније саставне делове и елементе, односно кроз поступак мишљења од посебног ка општем;

- Примена метода синтезе ће се примењивати путем спајања једноставних судова у сложеније и кроз процес уопштавања, чиме ће се доћи до систематизованог знања, односно до изградње теоријског знања у правцу од посебног ка општем;
- Метода компилације која подразумева критичко разматрање туђих опажања, ставова и закључака, односно њиховог научно-истраживачког рада, а како би дисертација у највећој мери носила лични печат аутора, који ће коректно и на уобичајен начин цитирати сваки материјал који је преузео;
- Компаративна метода ће се користити кроз поступак упоређивања резултата модела, а ради утврђивања њихове сличности у понашању и разлика међу њима;
- Математичке методе ће се употребити кроз системски поступак који се састоји у примени математичке логике, релација, симбола и операција у научно-истраживачком раду;
- Метода моделовања која ће се применити кроз развој модела који треба да представља стварну појаву, а који се експериментално истражује са циљем да се добијени резултати и модели могу пренети и на реалну појаву;
- Метода мерења која ће се користити са циљем да се добију резултати предложених решења (модела), који ће служити за поређење квалитета различитих решења;
- Статистичке методе ће се примењивати ради утврђивања статистичке значајности перформанси различитих алгоритама.

Експерименти ће се спроводити на истој, јасно дефинисаној хардверској и софтверској инфраструктури чиме ће се омогућити транспарентност и поновљивост свих експеримената, а сва поређења ће бити фер и коректна. Сви експерименти ће се проверавати на што већем броју примера, како закључци не би били појединачни, већ генерални.

1.5 Структура рада

У поглављу 2 приказан је преглед литературе у области класификације коришћењем условних случајних поља, предвиђање индикатора саобраћаја и примене математичког програмирања и модела теорије редова чекања у саобраћају.

Теоријске основе развијених методологија приказане су у поглављу 3. У истом су представљени модели Гаусових условних случајних поља за регресију који су

коришћени као основа за развијање модела Гаусових условних случајних поља за класификацију. Поред тога, приказан је увод у моделовање система опслуживања, објашњене су различите архитектуре неуронских мрежа, представљена је метахеуристика диференцијалне еволуције и начин симулације нехомогеног Марковљевог ланца.

У поглављу 4 представљена је методологија за класификацију гужви коришћењем Гаусових условних случајних поља. Приказана је репрезентација модела Гаусових условних случајних поља за класификацију, објашњен је начин закључивања, учења Бајесовских и не-Бајесовских условних случајних поља заједно са апроксимацијама усвојеним у циљу смањења меморијског и рачунског трошка.

Предикција континуалних променљивих на ретким подацима помоћу методологије за предвиђање индикатора саобраћаја засноване на комбинацији регресионих и класификационих Гаусових условних случајних поља приказана је у поглављу 5.

Развој методологије за одређивање оптималног броја активних канала у модулу у будућности приказан је у поглављу 6. У истом поглављу детаљно је описан Бајесовски и не-Бајесовски приступ у оквиру опште методологије за одређивање оптималног броја активних канала у модулу у будућности.

Примена развијених методологија на скупу података са ауто-путева, заједно са дискусијом добијених резултата, приказана је у поглављу 7.

Генерализација развијених методологија на реалним проблемима из других домена представљена је у поглављу 8.

Научни и инжењерски допринос дисертације заједно са постигнутим резултатима и закључцима представљени су у поглављу 9.

2 Осврт на релевантне библиографске изворе и остварене резултате

Имајући у виду да је тема ове докторске дисертације усмерена ка развијању три различите методологије за решавање проблема класификације гужви, одређивање индикатора саобраћаја и одређивање оптималног броја активних канала у модулу, преглед литературе и значај истраживања дисертације је подељен у три целине. У поглављу 2.1 представљен је преглед литературе у области класификације коришћењем условних случајних поља, док је у поглављу 2.2 приказан преглед литературе области превиђања индикатора саобраћаја. Детаљан увид у област примене алгоритама математичког програмирања и модела теорије редова чекања у саобраћају представљен је у поглављу 2.3.

2.1 Преглед литературе у области класификације коришћењем условних случајних поља

Стохастичко моделовање реалних појава графовским моделима има низ предности које се пре свега тичу јасне репрезентације модела и тумачења зависних променљивих. У свим графовским моделима чворови представљају случајне променљиве, док везе између чворова представљају одређену врсту зависности између случајних променљивих. У односу на начин којим се исказују зависности између случајних променљивих разликујемо два основна типа графовских модела: усмерени графовски модели и неусмерени графовски модели (Марковљеви графови)[15].

Усмереним графовским моделима, посредством веза између чворова, директно се задају условне вероватноће. Приликом повезивања чворова (случајних променљивих) задаје се смер којим се означава да је расподела променљиве, ка којој је усмерена веза, директно условљена променљивом из које та веза креће. Многобројни модели машинског учења основу налазе у репрезентацији која је приказана усмереним графовским моделом. Неки од најзначајнијих су сакривени Марковљеви модели (енг. „*Hidden Markov models*“, скраћено НММ) [16], Калманови филтери (енг. „*Kalman filters*“) [17], латентни Дирхлетов локацијски модел (енг. „*Latent dirchlet allocation*“, скраћено LDA) [18]. Сви претходно набројани модели су генеративни графовски модели. Генеративним моделом директно се моделује заједничка расподела свих променљивих

које се користе у неком проблему који се решава (односи се на атрибуте, латентне факторе и излазе). У случају да се решава проблем надгледаног учења, применом правила Бајесовске вероватноће [19] може се одредити расподела излаза за задате атрибуте. Поред генеративних модела постоје дискриминативни модели код којих се директно моделује расподела циљне променљиве условљене атрибутима. Дискриминативни модели се користе код надгледаног учења. Једна од најважнијих предности дискриминативних модела у односу на генеративне моделе тиче се постизања бољих перформанси предикције код проблема надгледаног учења. Компаративно поређење генеративних и дискриминативних модела приказано је детаљно у [20].

За разлику од усмерених графовских модела, Марковљеви графови омогућавају моделовање зависности између случајних променљивих непосредно, без директног дефинисања смера везе између променљивих. Овим приступом моделовања омогућава се коришћење широког дијапазона различитих расподела заједно са могућношћу укључивања доменског знања у исто [21]. Једна од класа најчешће коришћених алгоритама машинског учења, заснованих на репрезентацији представљеној у облику неусмерених графовских модела, су условна случајна поља [22].

Условна случајна поља су успешно примењена на низу различитих структурираних проблема као што су: препознавање ентитета имена [23], сегментација слике [24], сегментација речи [25] итд. Мешавина условних случајних поља је један од модела који је успешно примењен у многим реалним задацима везаним за моделовање више различитих излаза и приказан је у [26]. Овај модел је повезан са добро познатим сакривеним условним случајним пољима (енг. „*Hidden unit Conditional Random Fields*“, скраћено HUCRF). Учење HUCRF модела, коришћењем методе минимизације очекивања (енг. „*Expectation Minimization*“, скраћено EM) приказано је у [27]. Мешавина условних случајних поља је успешно имплементирана на многим реалним примерима, где је верификована примена истих. Услед постојања све већег броја података битно је експлоатисати исте на најбољи могућ начин. Повећањем комплексности модела велики број података може бити искоришћен у циљу побољшања перформанси предикција. Стога се многобројна истраживања фокусирају на комбинацији стандардних условних случајних поља са моделима дубоког учења [28]. Дубока условна случајна поља показују боље резултате у односу на стандардну верзију условних случајних поља у случајевима када је скуп података из којих се учи велики. Ово се најбоље може видети на проблему класификације врсте микроорганизама [29], при чему је узета у обзир ретка структурна веза која постоји између различитих врста микроорганизама.

Већина савремених алгоритама, који спадају у класу условних случајних поља, немају једноставну форму при закључивању и учењу, стога се морају користити многе апроксимационе технике које нису једноставне и релативно су рачунски захтевне [30]. Гаусова условна случајна поља имају низ предности у односу на остале алгоритме, што је наравно узроковано њиховом каноничком формом која се може приказати у облику мултиваријационе нормалне расподеле. Самим тим, могућност експлоатације неструктурираних предиктора заједно са временским и просторним везама између излаза омогућава овим алгоритмима добру подлогу да се у истој форми примене и на класификационе проблеме.

Гаусова условна случајна поља су први пут имплементирана за проблеме комјутерске визије [31]. Различите адаптације и апроксимације Гаусових условних случајних поља су биле предмет многих истраживања [32]. Поред мерења параметра сличности између различитих излаза, предложен је модел који узима у обзир и различитости између истих [33]. Модел је проширен са додатним интерагујућим потенцијалом који је представљен у форми која омогућава конвексност процеса учења. Полунадгледана маргинализована Гаусова условна случајна поља (енг. „*Marginalized Gaussian Conditional Random Fields*“, скраћено MGCRF) за решавање проблема са недостајућим променљивим величинама приказана су у [34]. Детаљан приказ великог броја регресионих Гаусових условних случајних поља представљен је у [35]. Ретка Гаусова условна случајна поља заснована на l_1 регуларизацији су моделована у [36], док је исти модел са латентном декомпозицијом променљивих и изведеним условима конвергенције предиктора у високо димензионом простору приказан у [37]. Једна од адаптација регресионих Гаусових условних случајних поља на класификационе проблеме представљена је у [35]. Модел се састоји од латентног слоја променљивих, које су моделоване Гаусовим условним случајним пољем, док је расподела дискретних излаза везана овим променљивим величинама. Параметри модела су добијени ЕМ алгоритмом. С обзиром да су y и z непознате променљиве, закључивање је изведено помоћу очекиване вредности латенте променљиве.

Радови попут [38, 39] су већ показали да постоје различити начини и могућности предикције гужви у системима масовног опслуживања, међутим претраживањем литературе није приказана могућност искоришћења структурираних стохастичких модела ради одређивања постојања гужви у системима масовног опслуживања који су просторно и временски повезани.

2.2 Преглед литературе у области предвиђања индикатора саобраћаја

Предвиђање индикатора саобраћаја је један од веома важних савремених проблема, који је све присутнији услед константног повећања броја возила и ограничености постојеће инфраструктуре. Један од често коришћених индикатора саобраћаја је просечна брзина возила. Последице смањења просечне брзине возила (последично повезане са повећањем укупног броја возила) најчешће је повезано са губитком продуктивности услед саобраћајних гужви, повећањем загађености животне средине, потрошње енергената (најчешће фосилних горива) [40] или незадовољства услед плаћене услуге али предугачког чекања у реду. Индикатори саобраћаја обезбеђују корисне информације за планирање ресурса или операција. Успрешно процењене вредности индикатора саобраћаја, у неком периоду у будућности, обезбеђују информације које се могу користити у одређивању политике цена, смањењу саобраћајних гужви и у стратегијском одлучивању (прављењу нових путева, постављању нових наплатних рампи итд.) [41].

Дисциплина која се тиче предикције индикатора саобраћаја је један од истраживанијих праваца. Постоји велики број праваца у овом научном пољу, а тиче се два основна правца: традиционалног доменског моделовања саобраћаја у циљу предикције индикатора саобраћаја и одређивање индикатора саобраћаја методама машинског учења посредством доступних историјских података [41]. Други приступ заснован на постојању велике количине података, постао је један од најзаступљенијих услед бољих перформанси предикције постигнутим моделима машинског учења у односу на традиционалне моделе. Најчешће коришћени алгоритми машинског учења у циљу предикције индикатора саобраћаја су Калманови филтри [42]. Поред тога бројни радови се ослањају на коришћење других модела машинског учења као што су: кернел регресија (енг. „*Kernel regression*“) [43], К најближих суседа (енг. „*K nearest neighbour*“, скраћено KNN) [44] и модели дубоког учења (енг. „*Deep learning*“) [45]. Један од примера модела за предвиђање брзине возила заснованих на моделу мешавине експерата (енг. „*Mixture of experts*“), приказан је у [46].

Индикатори саобраћаја су такође важна тема код ски-центра, иако истраживања у овом домену нису превише одмакла, ипак одређени радови говоре о важности овог проблема исказаном у вези постојања гужви и задовољству скијаша у ски-центрима [47,

48]. Истраживање у коме се наводи да су скијашаи генерално вољни да плате већу цену ски паса за мање редове чекања и мањи број људи на стази приказано је у [49]. Ски-центар Вејл у Сједињеним Америчким државама је први увео могућност информисања скијаша о евентуалним гужвама које могу настати на ски-лифтовима. Овај сервис се заснива на одређивању очекиваног броја мобилних уређаја који се налазе испред улаза на ски-лифт. Ипак овакав начин предикције гужви захтева постављање озбиљне и скупе инфраструктуре. Моделовање гужви на ски-лифтовима у циљу бољег разумевања феномена и узрока настанка гужви приказано је посредством симулационог модела заснованог на агентима (назван DynaSki) и приказано у [50]. На сличан начин, предложен је модел мулти агента (назван Ski-Optim) који покушава да симулира гужве на улазу у ски-лифт [51]. Коришћење GPS података у предикцији гужви на улазима у ски-лифт приказано је у [52]. Као што је напоменуто, приступи предикцији индикатора саобраћаја заснованим на моделима машинског учења имају боље резултате у односу на традиционалне моделе и експертске системе из разлога што омогућавају коришћење латентних фактора које је тешко исказати, односно пронаћи традиционалним моделима. Ови латентни фактори се показују као веома важни за перформансе предиктора [53, 54]. Више различитих студија користе статистике из података у циљу предикције индикатора саобраћаја. Генерално коришћењем историјских података добијају се статистике неопходне за предвиђање перформанси. Најчешће се ове статистике тичу брзине возила [55, 56] и варијансе брзине возила [57]. Предикција индикатора саобраћаја је најчешће рађена у циљу да обезбеде информације о рутама у којима нема гужве. Начин проналажења рута у којима нема гужве за већ познате проблеме представља НП-тежак проблем (енг. „*NP-hard problem*“). Стога се и проблем налажења оптималних рута најчешће решава метахеуристикама као што су генетски, еволутивни алгоритми, рој честица [40] итд.

Ниједан од приказаних алгоритама не узима у обзир структуру између излаза у циљу побољшања перформанси предикције. Стога се и хибридна методологија који је развијена у овом раду разликује од осталих, по томе што узима у обзир и структуру између излаза при томе користећи и информације из неструктурних предиктора.

2.3 Преглед литературе у области примене математичког програмирања и модела теорије редова чекања у саобраћају

У свакодневном животу срећу се различити системи опслуживања који се могу моделовати теоријом редова чекања. Неки од примера система опслуживања су: редови на каси у самопослузи [58], редови на наплатним станицама на ауто-путевима [59, 60], редови на граничним прелазима итд.

У многим системима који се могу моделовати теоријом редова чекања, на различите начине коришћене су оптимизационе методе у циљу одређивање оптималне вредности управљачке променљиве. Посебно су различите оптимизационе и предиктивне технике присутне у области различитих врста саобраћаја где се јављају проблеми везани са гужвама. У раду [62] приказан је начин оптимизације потрошње електричне енергије на мобилним мрежама у неком периоду у будућности. Услед велике потрошње електричне енергије репетиторских станица развијен је систем који ће предвидети који је тренутно степен оптерећења мреже и одредити која репетиторска станица може да се угаси. На овај начин смањени су трошкови потрошње електричне енергије и смањена емисија CO₂ гасова. Модел за оптималну локацију возила у саобраћајној траци коришћењем предикције помоћу модела временских серија и развијањем оптимизационе методологије приказан је у [63]. Показано је да у просеку овај модел може смањити време путовања за око 6% у односу на тренутно стање. Вишекритеријумска оптимизација саобраћајних семафора приказана је у [64]. Коришћењем еволутивног алгоритма показано је да оптималан план паљења семафора може драматично да смањи време путовања током периода када су највеће гужве. Преглед начина на који алгоритми машинског учења могу значајно да помогну при оптималном управљању протоком 5G мрежи приказан је у [65]. Велики број стратегија и метода коришћених за управљање саобраћајем приказан је у [66], док су на сличан начин најмодернији начини предикције у саобраћају, представљени у [67].

Прегледом литературе нигде није пронађена методологија где се модели теорије редова чекања повезују са алгоритмима предикције у циљу оптимизације броја отворених активних канала опслуживања (наплатних рампи) у модулу (наплатна станица). Стога ће ова дисертације допринети унапређивању дефинисаних проблема.

3 Теоријске основе развијених методологија

У овом поглављу приказане су теоријске основе које су коришћене у циљу развијања претходно напоменуте три методологије. Дат је преглед модела Гаусових условних случајних поља за регресију, модела теорије редова чекања, различитих архитектура рекурентних неуронских мрежа, метахеуристика за одређивање оптималне полисе и поступак симулације Марковљевог нехомогеног ланца. Гаусова условна случајна поља за регресију су коришћења као основ за развијање модела Гаусових условних случајних поља за класификацију у оквиру методологије за класификацију гужви и предвиђање индикатора саобраћаја. Остали модели приказани у овом поглављу су коришћени у циљу развијања методологије за одређивање оптималног броја активних канала опслуживања у модулу.

3.1 Гаусова условна случајна поља за регресију

Гаусова условна случајна поља су дискриминативни графовски регресиони модел [68]. Чворови графа су излазне променљиве $\mathbf{y} = (y_1, y_2, \dots, y_N)$, које је потребно предвидети помоћу задатог скупа атрибута \mathbf{x} . У општем случају атрибути $\mathbf{x} = (x_1, x_2, \dots, x_N)$, интерагују са сваким излазом y_i , независно један од другог, док се релације између излаза одржавају, посредством функција интеракције између излаза. У циљу учења параметара модела, неопходно је дефинисати скуп за учење, који се састоји од атрибута \mathbf{X} и реалних вредности излазних променљивих \mathbf{Y} . Генерална форма условне расподеле $P(\mathbf{y} | \mathbf{x}, \boldsymbol{\alpha}, \boldsymbol{\beta})$ је:

$$P(\mathbf{y} | \mathbf{x}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{1}{Z(\mathbf{x}, \boldsymbol{\alpha}, \boldsymbol{\beta})} \exp\left(\sum_{i=1}^N A(\alpha, y_i, \mathbf{x}_i) + \sum_{i \neq j} I(\beta, y_i, y_j)\right) \quad (3.1)$$

Две различите функције атрибута су коришћене у претходном изразу:

- Потенцијал асоцијативности $A(\alpha, y_i, \mathbf{x})$ - који моделује релације између излаза y_i и одговарајућих улазних атрибута \mathbf{x}_i .
- Потенцијал интеракције $I(\beta, y_i, y_j)$ - који моделује релације (структуру) између излаза.

Вектори $\boldsymbol{\alpha}$ и $\boldsymbol{\beta}$ су параметри потенцијала асоцијативности и потенцијала интеракције. Потенцијал асоцијативности је дефинисан као:

$$A(\alpha, y_i, \mathbf{x}_i) = -\sum_{k=1}^K \alpha_k (y_i - R_{ik}(\mathbf{x}))^2 \quad (3.2)$$

где $R_{ik}(\mathbf{x})$ представља неструктурни предиктор за сваки излаз y_i у графу. Ови неструктурни предиктори су независни једни од других, односно независно предвиђају y_i за задате атрибуте \mathbf{x} , док је K укупан број неструктурних предиктора. Потенцијал интеракције је дефинисан као:

$$I(\beta, y_i, y_j) = -\sum_{l=1}^L \sum_{k=1}^K \beta_l S_{ij}^l (y_i - y_j)^2 \quad (3.3)$$

где је S_{ij}^l вредност којом се изражава сличност између излаза i и излаза j у графу l . L је укупан број задатих графова, односно функција сличности.

Један граф може да представља произвољно задату функцију сличности између излаза која може да карактерише временско-просторне везе. На основу овога условна расподела вероватноће приказаног модела може се представити као:

$$P(\mathbf{y} | \mathbf{x}, \alpha, \beta) = \frac{1}{Z(\mathbf{x}, \alpha, \beta)} \exp\left(-\sum_{i=1}^N \sum_{k=1}^K \alpha_k (y_i - R_{ik}(\mathbf{x}))^2 - \sum_{l=1}^L \sum_{k=1}^K \beta_l S_{ij}^l (y_i - y_j)^2\right) \quad (3.4)$$

Квадратна форма потенцијала интеракције и потенцијала асоцијативности омогућава да се наведена условна вероватноћа прикаже у облику који је представљен као мултиваријациона Гаусова расподела. Стога се каноничка форма Гаусових условних случајних поља може представити као [35]:

$$P(\mathbf{y} | \mathbf{x}, \alpha, \beta) = \frac{1}{(2\pi)^{\frac{N}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{y} - \boldsymbol{\mu})\right) \quad (3.5)$$

где је матрица прецизности $\Sigma^{-1} = 2Q$, а очекивање $\boldsymbol{\mu} = \Sigma \mathbf{b}$, дефинисано као:

$$Q = \begin{cases} \sum_k \alpha_k + \sum_h \sum_l \beta_l S_i^h, & \text{if } i=j \\ -\sum_l \beta_l S_{ij}^l, & \text{if } i \neq j \end{cases} \quad (3.6)$$

$$b_i = 2 \left(\sum_{k=1}^K \alpha_k R_{ik}(\mathbf{x}) \right) \quad (3.7)$$

Услед конвексности мултиваријационе Гаусове расподеле у зависности од очекивања и матрице прецизности, проблем закључивања $\operatorname{argmax}_{\mathbf{y}} P(\mathbf{y} | \mathbf{x}, \alpha, \beta)$ је врло једноставан.

Максимална вредност \mathbf{y} у расподели је очекивање расподеле $\boldsymbol{\mu}$. Циљ учења је

оптимизација параметара α и β максимизовањем условне функције логаритамске веродостојности $\operatorname{argmax}_{\alpha, \beta} \sum_y \log P(y | \mathbf{x}, \alpha, \beta)$. Један од начина да се ово постигне је обезбеђивањем позитивне дефинитности коваријационе матрице модела, која се обезбеђује посредством довољног услова да сви параметри α и β буду већи од 0. Функција веродостојности за приказани модел може се представити као:

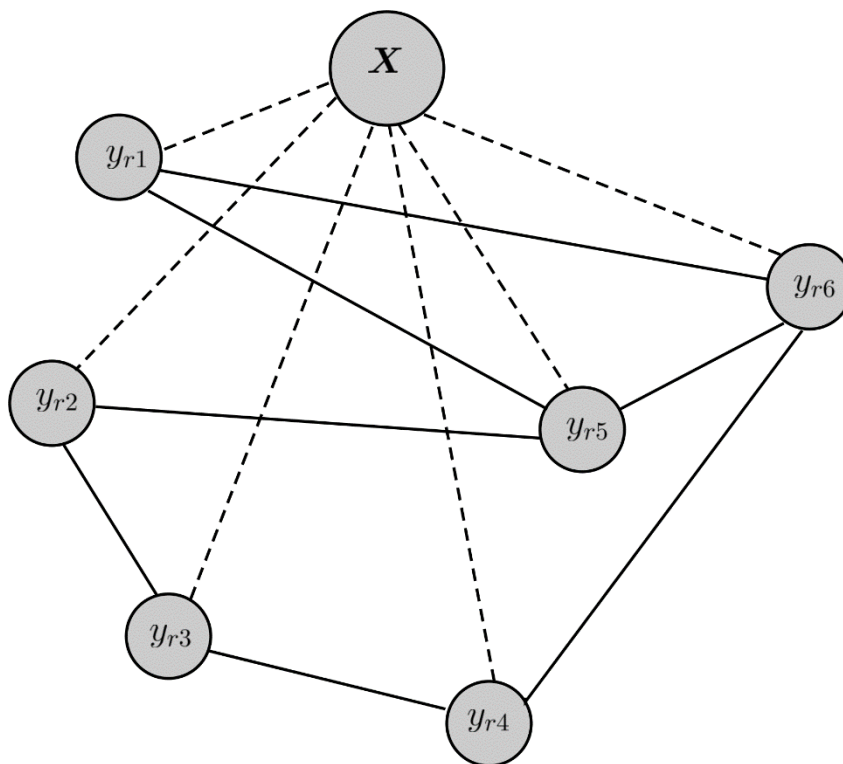
$$\log P = -\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{y} - \boldsymbol{\mu}) - \frac{1}{2} \log |\Sigma| \quad (3.8)$$

где је матрица Σ коваријациона матрица, а симбол $|\cdot|$ означава детерминанту.

Диференцијал функције веродостојности може се написати као:

$$d \log P = -\frac{1}{2} \operatorname{Tr} \left(d \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})^T - 2 \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu}) d \boldsymbol{\mu}^T \right) - \frac{1}{2} \operatorname{Tr} \left(\Sigma^{-1} \cdot d \Sigma^{-1} \right) \quad (3.9)$$

где је са Tr означен траг матрице.



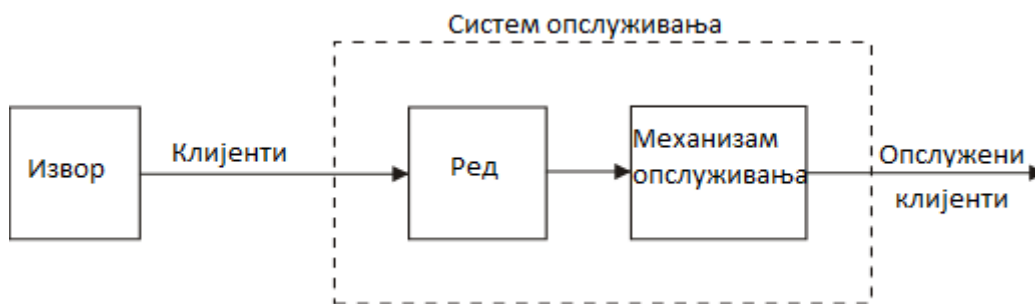
Слика 1 Гаусова условна случајна поља за регресију

Оптималне вредности параметара α и β се могу наћи коришћењем алгорита градијентног спуста, са ограничењима да вредности параметара α и β морају бити веће

од 0. Лако се може показати да је наведени оптимизациони проблем конвексан [35], проналажењем параметара за коју функција веродостојности има максималну вредност представља глобални максимум. Изглед Гаусових условних случајних поља за регресију, представљен у форми графа, приказан је на слици 1.

3.2 Моделовање система опслуживања

Најчешћи приказ процеса опслуживања приказан у великом броју модела теорије редова чекања приказан је на слици 2 [69].



Слика 2 Процес опслуживања

На случајан начин, клијенти се генеришу у времену из извора. Клијенти захтевају опслуживање и улазе у систем. Клијенти из реда се прихватају на опслуживање по некој дисциплини опслуживања или дисциплини у реду. На пример, уколико је канал за опслуживање заузет клијент стаје у ред, у супротном клијент попуњава канал и бива опслуживан. Након обављеног опслуживања клијент напушта систем за опслуживање.

3.2.1 Долазни ток клијената (извор)

Основна особина долазног тока клијената је његова величина, односно укупан број клијената који захтевају опслуживање или укупан број потенцијалних клијената који ће захтевати исто. Важно је напоменути да је у реалности најчешће овај долазни ток ограничен. Међутим, у већини реалних примера овај број је довољно велики тако да се може усвојити да је извор клијената неограничен, односно бесконачан. Популација потенцијалних клијената који захтевају опслуживање се назива долазна популација. У моделима теорије редова чекања се неограничен извор клијената усваја још из разлога

зато што модели постављени на овакав начин се могу много лакше математички формулисати. Системи опслуживања који се могу моделовати моделима са неограниченим извором су они где клијенти не припадају организацији – систему, чији је подсистем посматрани систем опслуживања тј. клијенти су „аутсајдери“ који доносе посао организацији.

С обзиром на претходно и у случају моделовања долазака јединица на наплатну станицу на ауто-путу, претпоставља се да је извор јединица неограничен, или да клијенти (возила) доносе посао организацији или механизму за опслуживање (наплатној рампи). Расподела вероватноће времена између два узастопна доласка клијента у систем опслуживања мора бити унапред дефинисана при моделовању процеса опслуживања. Односно расподела вероватноће времена између два узастопна доласка у систем дефинисана је као:

$$A(t) = P[\text{време између два узастопна доласка} \leq t] \quad (3.10)$$

Најчешћа претпоставка је да су времена између два узастопна доласка клијената у систем опслуживања случајне променљиве које су независне и које потичу из исте фамилије расподеле. Односно, конкретно у примеру моделовања времена долазака возила на наплатну рампу претпоставља се да је време између два узастопна доласка генерисано из експоненцијалне расподеле.

Елементарни модели теорије редова чекања који су коришћени у овој дисертацији подразумевају да клијенти у систем опслуживања долазе појединачно, било да је извор клијената ограничен или неограничен.

3.2.2 Ред

Ред представља место где клијенти чекају на опслуживање, односно његова основна особина је број клијената који може да стане у исти. Сваки клијент који не може да стане у ред (уколико је ред већ заузет клијентима) је одбачен односно неће бити примељен у систем и последично неће бити опслужен. Редови могу бити коначни и бесконачни. Уобичајена претпоставка код модела теорије редова чекања је да је ред бесконачан, чак и у ситуацијама где постоји (релативно велика) коначна горња граница могућег броја клијената у реду, из разлога што се анализа система опслуживања додатно компликује ако је горња граница могућег броја клијената у реду ограничена.

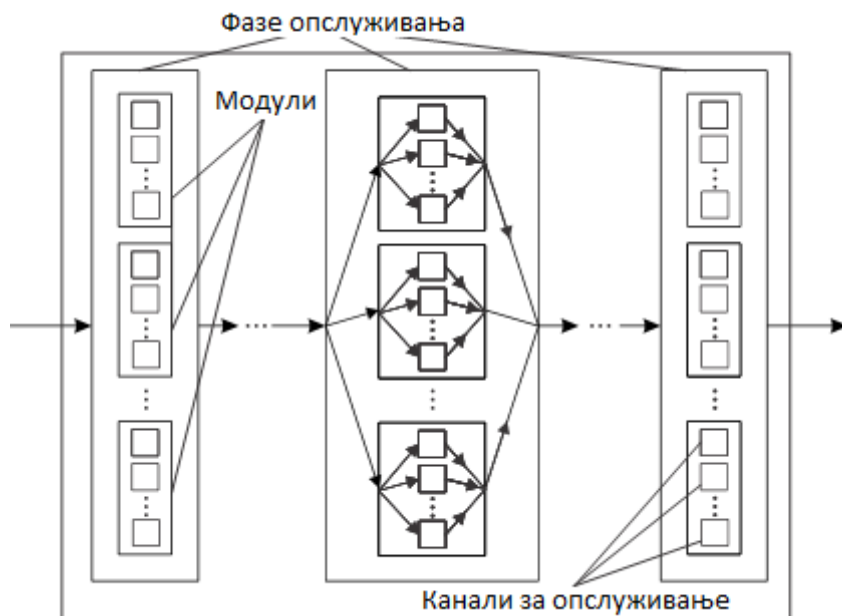
У реалним случајевима клијент који долази у ред не мора да буде човек. У случају моделовања система опслуживања наплатних станица, клијент који чека на

опслуживање је возило. Код елементарних модела теорије редова чекања, који су и коришћени у овој дисертацији, јединствени ред се формира испред сваке наплатне рампе.

3.2.3 Дисциплина опслуживања

Дисциплина опслуживања се односи на начин на који се клијенти који се налазе у реду опслужују. Постоји много начина на који клијенти могу да се опслуже, најчешће коришћен начин који је примењен и на примеру моделовања опслуживања на наплатним рампама јесте „*први приспели први опслужени*“ (енг. „*First In First Out*“, скраћено FIFO). FIFO дисциплина подразумева да ће клијент који је први доспео у ред бити и први опслужен. Односно, у случају модела система опслуживања на наплатној рампи, возило које је прво доспело у ред ће бити опслужено пре возила које се налази иза њега у реду. Додатно се претпоставља да не постоји могућност прескакања реда или одустајања јединице услед предугачког времена чекања у реду.

3.2.4 Механизам опслуживања



Слика 3 Механизам опслуживања

У најопштијем случају модели опслуживања се састоје од више фаза за опслуживање. Завршетком једне фазе клијент прелази у наредну фазу у којој се

опслужује. Клијент је потпуно опслужен проласком кроз све фазе опслуживања. Поред тога у свакој од фаза опслуживања може постојати један или више модула за опслуживање. У сваком од модула се налази један или више паралелних канала за опслуживање. Стога се у сваком моделу за опслуживање мора дефинисати број фаза, број модула у свакој фази и број канала по модулу. Уласком клијента у систем опслуживања, а на основу дефинисане дисциплине опслуживања, исти приступа једном од модула у првој фази, а затим једном од паралелних канала. Након завршетка прве фазе, наступа друга фаза све док клијент није у потпуности опслужен. Шематски приказ најопштијег модела за опслуживање представљен је на слици 3. У најопштијем случају време опслуживања представља случајну променљиву, а расподела вероватноће времена опслуживања се дефинише као:

$$B(t) = P[\text{време опслуживања} \leq t]. \quad (3.11)$$

Фазе опслуживања се у најопштијем случају разликују, стога се претпоставља да се време опслуживања у свакој фази разликује. На исти начин за сваки канал за опслуживање времена за опслуживање се могу разликовати. Поред тога, могу се разликовати и расподеле времена опслуживања за сваку врсту клијента у оквиру једног канала за опслуживање. Ипак у циљу поједностављења модела, најчешће се претпоставља да су расподеле времена за опслуживање у каналима истог модула исте за све клијенте који приступају опслуживању.

Елементарни модели теорије редова чекања, какав је и модел примењен у оквиру ове дисертације, претпостављају да се опслуживање обавља у једној фази са једним модулом. У оквиру тог једног модула може постојати један или више паралелних канала за опслуживање. Усваја се да је расподела времена опслуживања клијената међусобно независна и расподељена по истој теоријској расподели. Као и у случају клијената, канал за опслуживање не мора да буде једна особа или човек. Канал за опслуживање може представљати машину или групу особа. У случају наплатних рампи, опслуживање не мора да врши човек већ аутомат који наплаћује путарину.

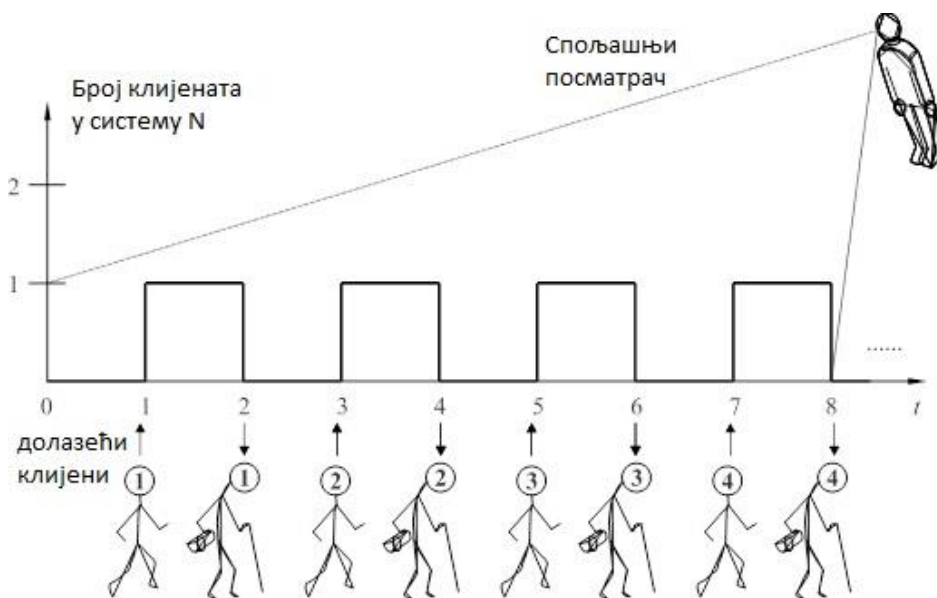
3.2.5 Расподеле вероватноћа стања система

Стање система опслуживања се дефинише на основу броја клијената који се налазе у систему односно броја јединица које захтевају опслуживање. Како су време између долазака клијената у систем тј. време између испостављања захтева за опслуживањем и време трајања опслуживања случајне величине, могу се одредити само

вероватноће да ће систем опслуживања у датом временском тренутку бити у неком од могућих стања. Другим речима, могу се одредити само расподеле вероватноћа стања система опслуживања. У наредном тексту биће приказана једна од најважнијих особина система опслуживања. Посматра се систем опслуживања код кога су вероватноће стања расподељене по расподели $p_k(t)$, $t \geq 0, k = 0, 1, \dots, n$, где је $p_k(t)$ вероватноћа да се систем опслуживања, у временском тренутку t , нађе у стању $N(t) = k$.

Другим речима, $p_k(t)$ је вероватноћа са којом ће спољашњи посматрач (спољашњи посматрач се налази изван система и види просечне вредности у времену на датом интервалу), који посматра систем у временском интервалу $[0, t]$, у тренутку t , наћи систем у стању $N(t) = k$. Са друге стране разматра се расподела вероватноћа стања система које види (затиче) долазећи клијент, где $q_k(t)$, $t \geq 0, k = 0, 1, \dots, n$, представља вероватноћу да се систем у временском тренутку t налази у стању $N(t) = k$ непосредно пред долазак клијента у систем. Расподела вероватноћа $q_k(t)$ представља гледиште долазећег клијента у временском тренутку t . У општем случају ове две расподеле вероватноћа стања система, расподела спољашњег посматрача и расподела долазећег клијента, се разликују, то јест:

$$p_k(t) \neq q_k(t), t \geq 0, k = 0, 1, \dots, n \quad (3.12)$$



Слика 4 Систем опслуживања D/D/1

Ради илустрације претходне тврдње, размотриће се систем опслуживања $D/D/1$ - једноканални систем опслуживања са константним временом између долазака клијената у систем и константним временом опслуживања, слика 4.

Систем је празан у временском тренутку $t=0$, а доласци клијената се одигравају у временским тренуцима $t_1 = 1, t_2 = 3, t_3 = 5, t_4 = 7$, док трајање опслуживања износи $t_{ops} = 1$. У овом случају долазак клијената у систем је одређен детерминистички тј. интервал између доласка два узастопна клијената је константан и износи $t_{dol} = 2$ и периодично се понавља. Како је константно време трајања опслуживања мање од константног времена између доласка два узастопна клијената у систем, то се у систему никада неће формирати ред. На основу претходног, сваки долазећи клијент затиче празан систем. Такође, просечна заузетост канала за опслуживање (како је види спољашњи посматрач) то јест однос дела времена када систем није празан према времену између два узастопна доласка клијената у систем или према дужини интервала посматрања, је једнака $1/2$. Другим речима, расподела вероватноћа стања спољашњег посматрача овако дефинисаног система $D/D/1$ је $p_0(t) = 1/2, p_1(t) = 1/2$. Док је расподела вероватноћа стања долазећег клијента $q_0(t) = 1, q_1(t) = 0$.

У циљу одређивања односа између расподела вероватноћа стања система опслуживања спољашњег посматрача и долазећег клијента, потребно је дефинисати догађај $C(t, t + \Delta t)$ - клијент долази у систем у интервалу $(t, t + \Delta t)$. Расподела долазећег клијента се сада може дефинисати као условна вероватноћа:

$$q_k(t) = \lim_{\Delta t \rightarrow 0} P(N(t) = k | C(t, t + \Delta t)), t \geq 0, k = 0, 1, \dots, n \quad (3.13)$$

Претходан израз означава вероватноћу да се систем у временском тренутку t налази у стању $N(t) = k$ под условом да ће се долазак клијента у систем остварити у $(t, t + \Delta t)$ интервалу времена. Са $N(t)$ је означен број клијената у систему опслуживања у временском тренутку t , док је одговарајућа вероватноћа $p_k(t) = P(N(t) = k)$. Важно је напоменути да одигравање догађаја $C(t, t + \Delta t)$ не подразумева да је долазећи клијент нужно и ушао у систем тј. обухваћен је и случај када долазећи клијент одмах и одлази и не проузрокује промену стања система. Расподела вероватноће долазећег клијента $q_k(t)$, користећи Бајесову формулу, може се написати у облику:

$$q_k(t) = \lim_{\Delta t \rightarrow 0} \frac{P\{C(t, t + \Delta t) | N(t) = k\} \cdot p_k(t)}{\sum_{j=0}^n P\{C(t, t + \Delta t) | N(t) = j\} \cdot p_j(t)} \quad (3.14)$$

где вероватноће $q_k(t)$ представљају апостериорне вероватноће догађаја B_k , вероватноће $p_k(t)$ представљају априорне вероватноће догађаја B_k , док догађај $C(t, t + \Delta t)$ представља догађај A .

У случају када је број места у реду у систему опслуживања ограничен вероватноћа која је од значаја за одређивање расподеле времена боравка клијента у систему опслуживања је вероватноћа стања долазећег клијента, односно $q_k(t)$ ¹.

3.2.6 Стохастички процеси

Скоро све физичке величине имају стохастички карактер, али многе од њих се довољно добро могу описати функционалним зависностима. У највећем броју случајева посматране величине неког сложеног система које дефинишу стање тог система се не могу довољно добро описати детерминистичким приступом. Односно, те величине имају стохастички карактер и самим тим представљају случајне величине. Вероватноћа реализације тих величина се понаша по одређеном статистичком закону који омогућава да се њихов карактер одреди.

С обзиром да систем мења стања кроз време, параметар од интереса за моделовање стохастичких процеса је време. Протицањем времена систем мења стања која су детерминисана случајном променљивом величином. Последишно стања система се мењају у времену на случајан начин који је описан одговарајућим стохастичким законом. Случајни процеси се могу поделити на основу три показатеља:

1. Простора стања случајног процеса.
2. Параметра случајног процеса.
3. Статистичке зависности између случајних величина за различите вредности параметра случајног процеса.

Простор стања дефинише скуп вредности стања које систем може да узме. Овај скуп може да буде дискретан или континуалан. У случају да је дискретан скуп стања, систем се може представити ланцем у чијем чворовима се налазе реализације истих.

¹ У наставку поглавља треба разликовати вероватноћу стања долазећег клијента која је означена са q_k и елементе транзиционе матрице који су обележени са q_{ij} .

Прелазак система из једног стања у друго стање описује се случајном променљивом величином која је временски зависна.

На потпуно исти начин, параметар случајног процеса представља величину која може да узима дискретне или континуалне вредности. Ова величина може бити и описана вектором са n димензија. Параметар случајног процеса не мора да има физичку интерпретабилност. С обзиром да у случају теорије редове посматрамо промене стања система током времена усваја се да је параметар случајног процеса време (t), док се фиксна вредност параметра случајног процеса назива временски тренутак. Уколико се промене у простору стања одигравају само за дискретне временске тренутке онда се дефинише случајан процес са дискретним параметром, у супротном када се простор стања може променити у било ком временском тренутку дефинише се случајан процес са континуалним параметром.

Статистичке зависности између случајних величина дефинишу и карактер случајног процеса. У великом броју случајева случајни процеси се јако тешко описују постојећим математичким апаратима, односно није могуће успоставити све зависности које постоје. Стога се и уводе одређене апроксимације и упроштења која омогућавају једноставније моделовање одређеног проблема са губитком на прецизности. Врло је битно изабрати оне зависности између случајних величина које ће у многеме поједноставити поступак решавања, без значајног губитка на тачности у односу на реалан случај. Из напред изнетог произилази да је за описивање стварног случајног процеса потребно, од постојећих типова случајних процеса, изабрати онај који га у потпуности или најверније описује. Најчешће коришћени случајни процеси су: случајни Марковљев процес, и процес "рађања и умирања" као специјални случај случајног Марковљевог процеса.

3.2.6.1 Случајни процес Марковљева

Елементарни модели теорије редова чекања заснивају се на коришћењу случајног процеса Марковљева, стога је неопходно и дефинисати исти. Случајан процес, који је коришћен у методологији одређивања оптималног броја активних канала у модулу у будућности, је дефинисан на дискретном простору стања са континуалним временом [70].

Случајни процес $X(m)$ дефинисан на дискретном простору стања, формира случајни процес Марковљева са континуалним временом (ланац Марковљева са

континуалним временом) уколико за све целе бројеве n и било који низ t_1, t_2, \dots, t_{n+1} , ($t_1 < t_2 < \dots < t_{n+1}$) важи следећа условна вероватноћа:

$$\begin{aligned} P[X(t_{n+1}) = j \mid X(t_1) = i_1, X(t_2) = i_2, \dots, X(t_n) = i_n] = \\ = P[X(t_{n+1}) = j \mid X(t_n) = i_n]. \end{aligned} \quad (3.15)$$

Условна вероватноћа из претходног израза представља најважнију особину Марковљевог процеса - одсуство памћења. Односно, она говори да је вероватноћа реализације догађаја условљена само реализацијом претходног догађаја. Овим се одбацује могућност да догађај буде условљен реализацијама несуседних претходних догађаја. У најопштијем случају прелазак из стања у стање у систему може се реализовати у било ком тренутку времена. У случају Марковљевог процеса може се показати да је случајна променљива која одређује период времена између прелазака из стања у стање расподељена по експоненцијалној расподели. Ова експоненцијална расподела може бити са константном или променљивом вредношћу параметра расподеле током времена. У циљу доказа претходно реченог, усваја се да τ_i буде случајна променљива која представља укупно време које процес проведе у стању i . С обзиром на чињеницу да претходна стања процеса утичу на наредна стања само посредством тренутног стања процеса, може се закључити да је преостало време које процес проведе у стању i независно од времена колико је процес провео у том стању, већ зависи само од њега. Претходно се може представити у математичком облику као:

$$P[\tau_i > s+t \mid \tau_i > s] = h(t) \quad (3.16)$$

где је s време које је процес већ провео у стању i , t је преостали део времена који ће процес провести у стању i и $h(t)$ је функција која зависи од времена t .

На основу Бајесове формуле, условна вероватноћа $P[\tau_i > s+t \mid \tau_i > s]$ се може представити у облику:

$$P[\tau_i > s+t \mid \tau_i > s] = \frac{P[\tau_i > s+t; \tau_i > s]}{P[\tau_i > s]} = \frac{P[\tau_i > s+t]}{P[\tau_i > s]} \quad (3.17)$$

Односно $P[\tau_i > s+t]$ се може записати као:

$$P[\tau_i > s+t] = P[\tau_i > s] \cdot h(t) \quad (3.18)$$

Време s може да узме било коју вредност, па се стога усваја да је ова вредност једнака 0 ($s=0$). Следи да је:

$$P[\tau_i > t] = h(t) \quad (3.19)$$

Додељивањем вредности 0 времену s могуће је због почетних услова тј. није специфицирано колико је процес већ провео у стању i што значи да време s може узети произвољну вредност. Имајући у виду чињеницу да је време s независно од времена t . Израз [4.8] се може представити као:

$$P[\tau_i > s+t] = P[\tau_i > s] \cdot P[\tau_i > t] \quad (3.20)$$

Диференцирањем овог израза следи да је:

$$\frac{dP[\tau_i > t+s]}{ds} = \frac{dP[\tau_i > s]}{ds} \cdot P[\tau_i > t] \quad (3.21)$$

Односно може се записати да је:

$$\frac{dP[\tau_i > s]}{ds} = \frac{d}{ds} (1 - P[\tau_i \leq s]) = -f_{\tau_i}(s) \quad (3.22)$$

где је $f_{\tau_i}(s)$ густина расподеле за τ_i .

Усвајањем да је вредност ($s=0$) и сређивањем претходног израза добија се да је:

$$\frac{dP[\tau_i > t]}{P[\tau_i > t]} = -f_{\tau_i}(0) \cdot ds \quad (3.23)$$

Интеграцијом горњег израза у границама од 0 до t добија се:

$$P[\tau_i > t] = e^{f_{\tau_i}(0) \cdot t} \quad (3.24)$$

Проналажењем густине вероватноће из овог израза добија се да је:

$$f_{\tau_i}(t) = f_{\tau_i}(0) \cdot e^{f_{\tau_i}(0) \cdot t} \quad (3.25)$$

где је $f_{\tau_i}(t)$ густина расподеле времена које процес проведе у стању i , чиме је доказ завршен.

На основу претходног, закључује се да је време преласка из стања у стање код Марковљевог процеса генерисано из експоненцијалне расподеле, односно време зависи само од текућег стања у коме се процес налази. Ова чињеница важи без обзира на то да ли се говори о хомогеним или нехомогеним Марковљевим процесима.

3.2.6.2 Нехомогени случајни процес Марковљева

У случају нехомогеног случајног процеса Марковљева време које протекне између промене стања система генерисано је експоненцијалном расподелом чији је параметар расподеле временски зависан. Односно, вероватноћа промене стања је случајна величина која је временски детерминисана (зависна од положаја на временској оси). У

најопштијем случају, вероватноћа да систем пређе из стања i у тренутку s у стање j у тренутку t дефинише се као условна вероватноћа дата следећим изразом:

$$p_{ij}(s, t) = P[X(t) = j \mid X(s) = i] \quad (3.26)$$

Пратећи стандардну нотацију, вероватноћа да се процес у тренутку t налази у стању j дефинише се као:

$$p_j(t) = P[X(t) = j] \quad (3.27)$$

Може се доказати да систем диференцијалних једначина приказан у матричној форми дефинише промену вероватноће да се систем налази у стању j у тренутку времена t [69]. Матрична диференцијална једначина има следећу форму:

$$\frac{dp(t)}{dt} = Q(t)^T p(t) \quad (3.28)$$

где је $p(t) = [p_0(t), p_1(t), p_2(t), \dots, p_n(t)]^T$ вектор који садржи вероватноће да се систем у тренутку t налази у стању j , $Q(t)$ је матрица транзиције односно интензитета преласка из стања у стање која је детерминисана временом.

Елементи матрице транзиције су дефинисани као вредности q_{ij} . Ови елементи се могу и физички интерпретирати. Може се рећи да $-q_{ii}$ представља интензитет при коме процес напушта стање i у случају да се налази у том стању, док $-q_{ij}$ представља интензитет при коме процес прелази из стања i у коме се тренутно налази у стање j . У циљу решавања система диференцијалних једначина неопходно је да се у почетном тренутку задају вероватноће стања система. Односно најчешће се претпоставља да је у почетном тренутку систем празан односно $p_0(t) = 1$, док су све остале вредности вероватноћа једнаке 0.

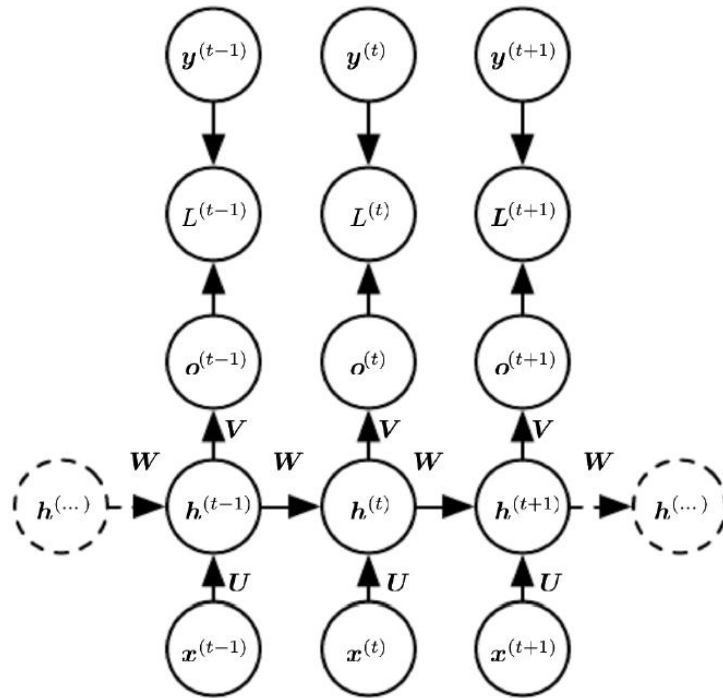
3.2.7 Укупни трошкови система опслуживања

С обзиром да су системи масовног опслуживања базирани на стохастичким процесима који се тичу долазака клијената и њиховог опслуживања, самим тим усвојени модел опслуживања може обезбедити информације које се тичу вероватноће стања система, односно вероватноће колико се тренутно клијената налази у систему. Информације о овим вероватноћама могу бити изузетно корисне зато што се посредством њих могу донети управљачке одлуке које ће минимизовати очекиване укупне трошкове [61]. Минимизација очекиваних укупних трошкова подразумева

минимизацију збира очекиваних трошкова чекања и очекиваних трошкова опслуживања. Очекивани трошкови чекања подразумевају трошкове настале услед чекања клијената. Односно ови трошкови подразумевају опортунитетне трошкове настале услед незадовољства клијената, односно чекања у реду и остале материјалне трошкове који су последица чекања (потрошња горива, изгубљено време итд.). Очекивани трошкови опслуживања односе се на трошкове који настају услед опслуживања. Додатним ангажманом нових канала за опслуживање ови трошкови расту, док очекивани трошкови чекања опадају зато што се убрзава процес опслуживања. Самим тим битно је нагласити да ова два трошка имају супротну тенденцију то јест: растом трошкова чекања опадају трошкови опслуживања и обрнуто. Један од најћешћих начина управљања овим трошковима је избор броја канала за опслуживање који треба да буде доступан. Односно, циљ је наћи број канала који ће минимизовати укупне трошкове опслуживања у датом временском интервалу уколико је модел опслуживања нехомоген.

3.3 Рекурентне неуронске мреже

Рекурентне неуронске мреже представљају посебан тип неуронских мрежа који се користи за моделовање секвенци (временских серија, реченица итд) [73]. Постоје и различита уопштења истих на домене који су везани са графовима. Основни циљ и задатак рекурентне неуронске мреже ослања се на проналажење сакривене структуре помоћу које се моделују зависности између секвенци (временских тренутака, речи у реченици и слично). Односно, информација у претходном временском тренутку сигурно детерминише излаз у наредном тренутку времена. Мрежа не покушава директно да моделује везе између излаза као у случају условних случајних поља, већ уводи велики број нелинеарних трансформација које из почетних атрибута генеришу информацију која је корисна за наредну секвенцу. Јасније речено, рекурентне неуронске мреже одустају од идеје да се параметрима модела дефинише предвиђање на основу текућих улаза мреже. Уместо тога, елементи улазне секвенце се обрађују у корацима, мрежа има скривено стање које акумулира информацију о елементима секвенце обрађеним у претходним корацима, а параметри одређују на који начин се то стање мења из корака у корак на основу претходног стања и текућих улаза и како се генерише излаз мреже у зависности од текућег стања.



Слика 5 Илустрација рекурентне неуронске мреже

Постоји велики број различитих модела неуронских мрежа, начин приказа овог модела је у сагласности са начином записа приказаном у [74]. Модел са једним скривеним слојем може се формулисати као:

$$h^{(0)} = 0 \quad (3.29)$$

$$h^{(t)} = g(b_h + Wh^{(t-1)} + Ux^{(t)}) \quad (3.30)$$

$$o^{(t)} = g(b_o + Vh^{(t)}) \quad (3.31)$$

где је са t обележен редни број секвенце која је дужине T , $x^{(t)}$ представља вектор улаза у тренутку t , $h^{(t)}$ је вектор вредности скривеног слоја, g је нелинеарна активациона функција, $o^{(t)}$ је вектор излаза мреже, W , V и U су параметри мреже, а b_o и b_h су слободни чланови. На слици 5 [74] приказана је илустрација рекурентне неуронске мреже.

Тренирање, односно учење параметара рекурентних неуронских мрежа спроводи се пропацијом у прошлости (енг. „*backpropagation through time*“). Више о методологији учења може се пронаћи у [75]. У зависности од типа проблема који се решава, приликом учења могуће је поставити различите функције циља које је потребно минимизовати променом вредности параметара мреже. У случају регресије користи се функција квадратне грешке између излаза које мрежа даје и стварних вредности излаза

из скупа података. Функција циља представља осредњену вредност средње квадратне грешке излаза свих инстанци.

Два основна проблема се јављају код рекурентних неуронских мрежа. Први проблем се односи на учење, а тиче се нестабилности процеса учења услед експлодирајућих или нестајућих градијената. Појава оваквих градијената јавља се услед коришћења пропагације уназад, односно множења парцијалних извода. Многе мреже оваквог типа су изузетно дубоке, стога је овај проблем неопходно решити. Други проблем, који се често среће, тиче се немогућности пропагације информације из далеких секвенци. Услед сталног множења са параметрима у оквиру једне секвенце временом утицај информације из далеких секвенци губи утицај и потпуно нестаје. Стога, није могуће дугорочно чувати информације. Ови проблеми се решавају употребом неких од следећих метода.

3.3.1 GRU мреже

GRU мреже представљају унапређену верзију стандардних рекурентних неуронских мрежа. За разлику од стандардних рекурентних неуронских мрежа код GRU мрежа уводи се GRU ћелија уместо неурона у скривеном слоју. GRU ћелија садржи такозване капије: капију за ажурирање (енг. „*update gate*”) и капију за ресетовање (енг. „*reset gate*”). Посредством ове две капије решава се проблем нестајућег градијента, односно ова два вектора одлучују која информација треба да буде допремљена до излаза. Највећа предност је могућност да мрежа буде истренирана тако да информација из далеких секвенци буде пропагирана, односно да иста не нестане и постане релевантна за предвиђање тренутног излаза. Тренирањем GRU мрежа чува и филтрира релевантне информације коришћењем ове две капије у оквиру једне ћелије. Важно је напоменути да је потребно пажљиво изабрати мета параметре при конструисању мреже (број слојева и број GRU ћелија у оквиру једног слоја), како би иста дала жељене перформансе предикције. Више о GRU мрежама и њиховој математичкој формулацији може се наћи у [76].

3.3.2 LSTM мреже

За разлику од обичних рекурентних мрежа прва архитектура мрежа која је допринела процвату рекурентних неуронских мрежа заснива се на коришћењу LSTM ћелије. LSTM ћелија је сложенија јединица у односу на GRU ћелију и њена структура

омогућава контролу читања и уписа у јединицу. LSTM ћелија садржи одговарајуће активационе функције које омогућавају да се сачувају претходне информације скривених стања уз контролу писања, читања и заборављања које се врши на основу низа научених правила.

Односно као и код GRU ћелија, ова ћелија садржи капије посредством којих се релевантне информације из претходних стања задржавају, делимично или потпуно заборављају. На пример, улазна капија контролише да ли ће улазна информација проћи ка ћелији, тако што на основу вредности улаза, свог стања у претходном кораку рачуна коефицијент који проласком кроз одговарајућу нелинеарну трансформацију контролише утицај улаза на стање ћелије. На сличан начин капија за заборављање на аналоган начин контролише утицај претходног стања на стање нове ћелије. Више о LSTM ћелији и формулацији математичког модела може се пронаћи у [77].

3.3.3 LSTM мреже са унутрашњом стандардизацијом слојева

Један од начина смањивања високог рачунског трошка при тренирању стандардних неуронских мрежа је стандардизацијом излаза из скривених слојева. Техника која се назива унутрашња стандардизација [78] (енг. „*batch normalization*”) користи улазе у неурон (из података за учење) као вредности случајних променљивих на основу којих се израчунава очекивана вредност и варијанса у циљу стандардизације истих током тренирања. Ова техника се примењује за сваки неурон понаособ и на тај начин значајно смањује целокупно време тренирања неуронске мреже, омогућавајући коришћење већих корака приликом учења и смањујући опасност од експлодирајућих градијената током тренинга. На врло сличан начин, унутрашња стандардизација се може уопштити на рекурентне неуронске мреже, прецизније рекурентне неуронске мреже са LSTM ћелијама. За разлику од стандардног приступа код рекурентних неуронских мрежа стандардизују се читави слојеви, користећи сумиране улазе у неуроне као променљиве на основу којих се одређује очекивана вредност и варијанса за стандардизацију истих.

У случају унутрашње стандардизације обичних неуронских мрежа, након стандардизације уводе се нови параметри очекивања и варијансе које се уче како мрежа не би изгубила експресивну моћ. Исто је примењено и на унутрашњу стандардизацију слојева у оквиру LSTM мреже. Може се показати да је исти број операција потребан при унутрашњој стандардизацији током тренирања и тестирања мреже. Утицај оваквог вида

стандардизације делује да се градијенти пропагирају стабилније у смислу тога да не долази до нестајања или експлодирања истих. Емпиријски је показано да LSTM мрежа са унутрашњом стандарднизацијом слојева постиже много боље резултате у односу на обичну LSTM мрежу. Детаљније о овом моделу може се пронаћи у [12].

3.4 Метахеуристике за одређивање оптималне полисе управљања отварања и затварања канала у модулу опслуживања

Велики број различитих алгоритама математичког програмирања је развијан у двадесетом веку [79]. Многи од ових алгоритама су специјализовани за решавање одређене врсте проблема као на пример, Симплекс алгоритама који се користи за решавање проблема линеарног програмирања [80] или градијентни метод који се користи за решавање проблема нелинеарног програмирања без ограничења [81]. Методе и технике за оптимизацију које су формулисане на начин где се итеративно прилази оптималној вредности функције циља (минималној или максималној вредности) не могу да гарантују да ће постигнута вредност функције циља бити и глобални оптимум. Међутим већина ових метода поседује начин на основу кога се може евалуирати колико је тренутна вредност функције циља далека од неког глобалног или локалног оптимума. Велики број данашњих проблема се не може једноставно решити са оваквим приступом. Проблеми због којих традиционалне технике математичког програмирања нису корисне тичу се превасходно комплексности дефинисања функције циља из које се не могу пронаћи аналитички градијенти, док је евалуација нумеричких вредности градијената превише рачунски скупа.

Стога се велики број проблема математичког програмирања може успешно решити коришћењем хеуристика, односно метахеуристика. Метахеуристике представљају алгоритме, то јест процедуре које се заснивају на тражењу, генерисању и одабиру хеуристика које теже да пронађу решење оптимизационог проблема.

У поређењу са стандардним итеративним методама метахеуристике не могу да гарантују да ће решење које је добијено бити локални или глобални оптимум. Само у одређеним случајевима се може доказати конвергенција метахеуристика. Поред тога, метахеуристике су изузетно рачунски скупе у односу на итеративне методе где је могуће одредити аналитичку вредност градијента или Хесијана. Метахеуристике се најчешће заснивају на коришћењу стохастике, односно крајња вредност функције циља зависи од генерисаних вредности случајних променљивих. Метахеуристике се стога могу успешно

применити на проблемима комбинаторне оптимизације где могу са много мање компјутерских ресурса да пронађу задовољавајуће решење у поређењу са другим алогитмима. Велики број метахеуристика је инспирисан неком биолошком појавом стога многи од њих носе и називе по истим.

Неке од познатијих метахеуристика су: генетски алгоритми [82], алгоритми мравоједа [83], алгоритми мрава [84] итд. У овој дисертацији коришћена је метахеуристика диференцијалне еволуције (енг. „*differential evolution*“).

3.4.1 Диференцијална еволуција

Алгоритам диференцијалне еволуције је метахеуристика чији је циљ да проласком кроз већи број итерација нађе решење које минимизује (максимизује) задату функцију циља [85]. Диференцијална еволуција омогућава налажење вектора решења који минимизује задату функцију циља. Принцип рада алгоритма се заснива на неколико корака:

1. Иницијализација N агената представља почетни избор N могућих решења. Ова решења треба да покрију целокупну област претраге. Најчешће се иницијализација врши генерисањем решења из униформне расподеле која је ограничена на област претраге решења. Међутим уколико постоји одређено претходно знање о томе где се налази оптимум (минимум или максимум) функције циља, решења се могу генерисати у тој околини.
2. Све док је број итерација мањи од максималног броја итерација који се задаје унапред, у оквиру сваке итерације се обављају следеће операције:
 - a. За сваког агента у популацији неопходно је случајним избором доделити три агента a , b и c . Генеришу се нова решења посредством отежане линеарне комбинације вредности изабраних агента, а затим се врши мутација одређених вредности унутар добијеног агента на случајан начин.
 - b. Након добијених нових агената, врши се поређење старих и нових агената и N агената са најбољим (најнижим или највишим) вредностима функције циља користи се у наредној итерацији претраге оптималних решења.
3. Након завршетка претраге бира се агент са најбољом (најнижом или највишом) вредношћу функције циља која представља решење оптимизационог проблема. Детаљан опис алгоритма заједно са математичком формулацијом приказан је у [85].

3.5 Симулација нехомогеног Марковљевог ланца

У циљу верификације резултата, добијеног решавањем система матричних диференцијалних једначина за добијање веорватноћа стања система нехомогеног Марковљевог ланца, извршена је симулација система. Општи алгоритам који је коришћен за симулацију система опслуживања, при претпоставци да су времена долазака клијената и времена опслуживања генерисана из експоненцијалне расподеле са константним параметром, детаљно је приказан у [69]. Важно је напоменути да при симулацији процеса са неограниченим редом, неопходно је увести претпоставку да систем има ограничен ред али да је број места у реду довољно велики како резултати симулације не би значајно одступили од резултата добијених теоријом редова чекања. Ова претпоставка је уведена узевши у обзир да није могуће нумерички решити систем са бесконачним бројем диференцијалних једначина.

У случају када је време које протекне између долазака клијената генерисано из експоненцијалне расподеле са параметром који је временски зависан (интензитет долазака није константа), постоји неколико различитих начина симулације стохастичког процеса [71]. Начин који је коришћен у овом раду детерминисан је узајамном повезаношћу експоненцијалне и Пуасонове расподеле. Односно, док Пуасонова расподела детерминише број долазака клијената у неком интервалу времена, експоненцијалном расподелом се моделује време које протекне између два суседна доласка. Стога, симулација стохастичког Пуасоновог процеса са параметром расподеле који је временски зависан, извршава се према следећем алгоритму:

1. Генерисати број долазака клијената у систем из Пуасонове расподеле чија је

$$\text{функција густине дата изразом } P_N(n) = \frac{\left(\int_{\mathbb{R}} \lambda(s) ds \right)^n}{n!} e^{-\int_{\mathbb{R}} \lambda(s) ds}.$$

2. За сваку од n реализација потребно је генерисати време када се клијент појавио, та времена се могу добити узимањем узорка из расподеле која је дефинисана као

$$P_x(s) = \frac{\lambda(s)}{\int_{\mathbb{R}} \lambda(s) ds}.$$

Из ове расподеле се узорци могу извлачити посредством

алгоритма узорковања прихвати-одбиј (енг. „*Acceptance-rejection sampling method*“) [72].

3. Вектор добијених времена долазака клијента поређати у растућем редоследу.

Добијена времена долазака клијената се потом могу користити у општем алгоритму за симулацију нехомогеног Марковљевог процеса. Независно од тога да ли је систем једноканални или вишекканални, овако добијена времена долазака клијената се даље могу користити при симулацији истог. Важно је напоменути да у случају када је и опслуживање генерисано из експоненцијалне расподеле са временски зависним параметром, користи се исти поступак при генерисању времена опслуживања клијената.

4 Методологија за класификацију гужви помоћу Гаусових условних случајних поља за класификацију

У овом поглављу је представљена је новоразвијена методологија за класификацију гужви коришћењем новог модела Гаусових условних случајних поља за класификацију.

Методологија за класификацију гужви се заснива на дефинисању појма гужви посредством агрегираних статистика из података, прављењу скупа података за учење, дефинисању мрежних структура, учењу Бајесовских и не-Бајесовских модела Гаусових условних случајних поља за класификацију, валидацији научених модела и избору модела са најбољим предиктивним перформансама.

Три верзије математичког модела Гаусових условних случајних поља за класификацију су развијене Бајесовска, Бајесовска апроксимативна и не-Бајесовска верзија у зависности од усвојених претпоставки. Модел је успешно имплементиран на проблему предвиђања гужви на ауто-путевима и скијалиштима. Поред тога, имплементација истог је даље генерализована на различите врсте реалних проблема.

Репрезентација модела Гаусових условних поља за класификацију приказана је у поглављу 4.1. Начин на који се закључује у Бајесовском и не-Бајесовском моделу Гаусових условних случајних поља за класификацију приказан је у поглављу 4.2, док је учење истих приказано у поглављу 4.3. Утицај различитих оптимизационих параметара на перформансе предикције Гаусових условних случајних поља приказан је у поглављу 4.4. На самом крају у поглављу 4.5 приказана је убрзана апроксимативна верзија алгоритма за учења Бајесовског модела.

4.1 Гаусова условна случајна поља за класификацију – репрезентација

Један од начина адаптације Гаусових условних случајних поља за проблем класификације је покушај замене континуалног излаза, дискретном вредношћу. У случају бинарне класификације овај излаз може имати вредност 0 или 1. Услед губитака својстава Гаусове расподеле настале овом трансформацијом, лакши начин решавања истог проблема је апроксимацијом дискретног излаза континуалним излазом. Овај континуални излаз може се затим сигмоидном трансформацијом свести на вредности између 0 и 1, чиме се проблем класификације може решити аналогно логистичком

регресијом. На овај начин је веза између регресионог и класификационог модела Гаусових условних случајних поља слична вези која постоји између линеарне и логистичке регресије. Овакав начин представљања модела има низ предности:

1. Модел се може применити на класификационим структурним проблемима који се не могу аналитички решити стандардним класификационим условним случајним пољима. Услед постојања латентне структуре код овако дефинисаних Гаусових условних случајних поља за класификацију, обезбеђује се могућност даљих унапређења модела и решавање реалних проблема ефикасно без коришћења скуних нумеричких метода.
2. Дефинисање директне корелације између дискретних излаза може да унесе у модел непотребни шум [86]. Увођењем латентних континуалних променљивих у модел, посредством којих се дефинише структура, овај проблем се превазилази.
3. У случају непоузданости неструктурних предиктора који показују различите резултате предикције, што је изражено високом нормом варијансе дијагоналних елемената коваријационе матрице нормалне расподеле латентних променљивих, проблем може бити решен маргинализацијом латентних променљивих, што побољшава предиктивну моћ истог.

Претпоставља се да су y_i дискретни бинарни излази, а z_i континуалне латентне променљиве које су додељене сваком дискретном бинарном излазу y_i . Сваки излаз y_i је условно независан од других излаза, за познату вредност одговарајуће латентне променљиве z_i . Илустрација ових зависности приказана је на слици 6.

Условна расподела $P(y_i | z_i)$ је дефинисана као Бернулијева расподела облика:

$$P(y_i | z_i) = \text{Ber}(y_i | \sigma(z_i)) = \sigma(z_i)^{y_i} (1 - \sigma(z_i))^{1-y_i} \quad (4.1)$$

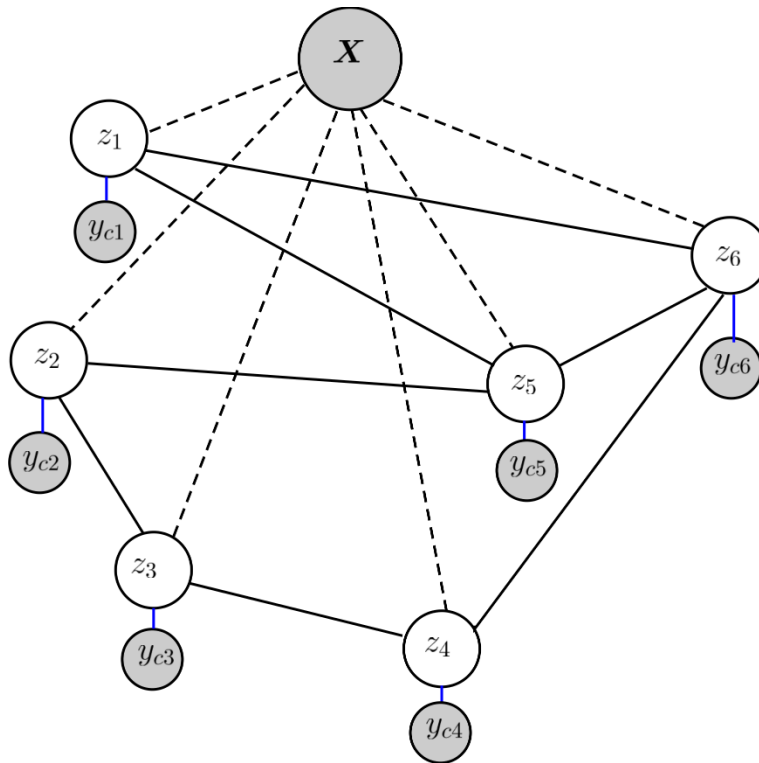
где је $\sigma(\cdot)$ сигмоидна функција. Услед услова о независности расподела, заједничка расподела излаза y_i , може се изразити као:

$$P(y_1, y_2, \dots, y_N | \mathbf{z}) = \prod_{i=1}^N \sigma(z_i)^{y_i} (1 - \sigma(z_i))^{1-y_i} \quad (4.2)$$

Условна расподела латентних променљивих представља модел Гаусових условних случајних поља за регресију, који се у каноничкој форми може дефинисати мултиваријационом Гаусовом расподелом. Стога је заједничка расподела латентних променљивих и излаза дефинисана као:

$$P(\mathbf{y}, \mathbf{z} | \mathbf{x}, \boldsymbol{\theta}) = \prod_{i=1}^N \sigma(z_i)^{y_i} (1 - \sigma(z_i))^{1-y_i} \cdot \frac{1}{(2\pi)^{N/2} |\Sigma|^{1/2}} \cdot \exp\left(-\frac{1}{2}(\mathbf{z} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{z} - \boldsymbol{\mu})\right) \quad (4.3)$$

где је $\boldsymbol{\theta} = (\alpha_1, \dots, \alpha_K, \beta_1, \dots, \beta_L)$.



Слика 6 Гаусова условна случајна поља за класификацију

Заједничка расподела латентних променљивих и излаза представља општу форму математичког модела Гаусових условних случајних поља за класификацију. Имајући у виду форму модела, дефинишу се два основна начина закључивања и учења Гаусових условних случајних поља за класификацију:

1. Гаусова условна случајна поља за класификацију – Бајесовска (енг. „*Gaussian conditional random fields for classification - Bayesian*“, скраћено GCRFCb) – са условном расподелом облика $P(\mathbf{y} | \mathbf{x}, \boldsymbol{\theta})$, која је добијена маргинализацијом латентних променљиве.
2. Гаусова условна случајна поља за класификацију – не-Бајесовска (енг. „*Gaussian conditional random fields for classification – non Bayesian*“, скраћено GCRFCnb) –

са условном расподелом облика $P(\mathbf{y}|\mathbf{x},\boldsymbol{\theta},\boldsymbol{\mu}_z)$, која је добијена заменом латентних променљивих њиховим очекивањима.

4.2 Гаусова условна случајна поља за класификацију - закључивање

4.2.1 Закључивање у GCRFCb моделу

Предикција дискретних излаза \mathbf{y} за задате вредности атрибута \mathbf{x} и параметара $\boldsymbol{\theta}$ се не може решити аналитички, услед немогућности маргинализације латентних променљивих у оквиру заједничке расподеле $P(\mathbf{y},\mathbf{z}|\mathbf{x},\boldsymbol{\theta})$. Ипак услед условне независности излаза за задату одговарајућу вредност латенте променљиве, могуће је једноставније евалуирати расподелу $P(y_i = 1 | \mathbf{x}, \boldsymbol{\theta})$.

$$P(y_i | \mathbf{x}, \boldsymbol{\theta}) = \int_{\mathbf{z}} P(y_i | \mathbf{z}) P(\mathbf{z} | \mathbf{x}, \boldsymbol{\theta}) d\mathbf{z} \quad (4.4)$$

$$P(y_i = 1 | \mathbf{x}, \boldsymbol{\theta}) = \int_{\mathbf{z}} \sigma(z_i) P(\mathbf{z} | \mathbf{x}, \boldsymbol{\theta}) d\mathbf{z} \quad (4.5)$$

Као резултат условне независности излаза за задате латенте променљиве $P(y_i = 1 | \mathbf{z}) = P(y_i = 1 | z_i)$, могуће је маргинализовати $P(\mathbf{z} | \mathbf{x}, \boldsymbol{\theta})$ по вредностима латентних променљивих задатих скупом $\mathbf{z}' = (z_1, \dots, z_{i-1}, z_{i+1}, \dots, z_N)$:

$$P(y_i = 1 | \mathbf{x}, \boldsymbol{\theta}) = \int_{z_i} \sigma(z_i) \left(\int_{\mathbf{z}'} P(\mathbf{z}', z_i | \mathbf{x}, \boldsymbol{\theta}) d\mathbf{z}' \right) dz_i \quad (4.6)$$

где је $\int_{\mathbf{z}'} P(\mathbf{z}', z_i | \mathbf{x}, \boldsymbol{\theta}) d\mathbf{z}'$ нормална расподела очекивања $\mu = \mu_i$ и варијансе $\sigma_i^2 = \Sigma_{ii}$.

Стога се условна расподела $P(y_i = 1 | \mathbf{x}, \boldsymbol{\theta})$ може приказати као:

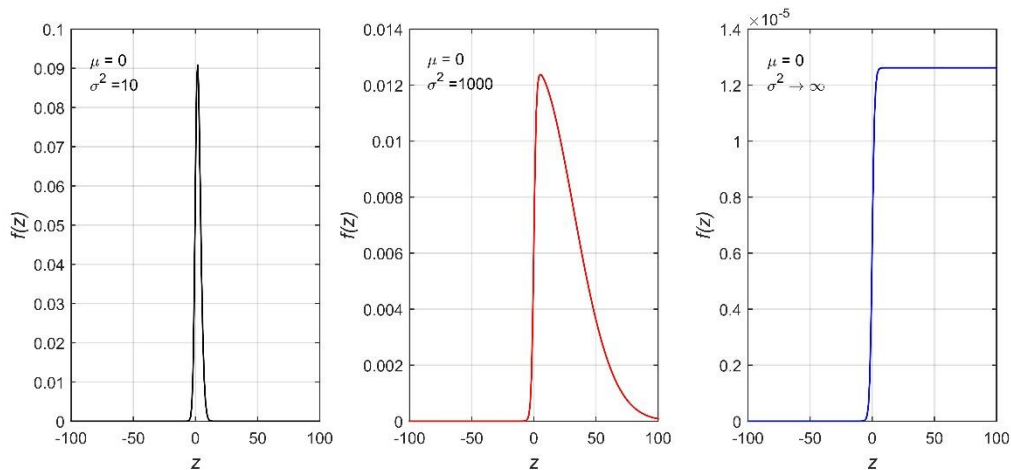
$$P(y_i = 1 | \mathbf{x}, \boldsymbol{\theta}) = \int_{-\infty}^{+\infty} \sigma(z_i) N(z_i | \mu_i, \sigma_i^2) dz_i \quad (4.7)$$

Евалуација вредности $P(y_i = 0 | \mathbf{x}, \boldsymbol{\theta})$ је очигледна и може се изразити као:

$$P(y_i = 0 | \mathbf{x}, \boldsymbol{\theta}) = 1 - P(y_i = 1 | \mathbf{x}, \boldsymbol{\theta}) \quad (4.8)$$

Једнодимензиону интеграцију израза за условну расподелу није могуће решити аналитички стога се примењују методе једнодимензионе нумеричке интеграције. Поред тога површина испод функције, која је изражена производом сигмоидне функције и једнодимензионе нормалне расподеле, највише је концентрисана у области око очекивања, осим у случајевима када је варијанса нормалне расподеле велика. Изглед функције $\sigma(z)N(z_i | \mu_{ii}, \sigma^2)$ је приказан на слици 7. На основу претходног може се

закључити да се границе интеграла $(-\infty, +\infty)$ могу са великом тачношћу апроксимирати интервалом $(\mu - 10\sigma_i, \mu + 10\sigma_i)$. Ова апроксимација у многоме олакшава начин решавања једноструке интеграције Њутн-Котесовим формулама и повећава тачност израчунатог интеграла коришћењем истих [87]. Представљен начин закључивања може се ефикасно користити са релативно ниским трошком рачунања услед брзе евалуације једноструких интеграла нумеричким методама.



Слика 7 Облик функције $\sigma(z)N(z_i | \mu_{z_i}, \sigma^2)$ за три различите варијансе нормалне расподеле

4.2.2 Закључивање у GCRFCnb моделу

Закључивање код GCRFBCnb модела је много једноставније, зато што није потребно маргинализовати латентне променљиве. Предвиђање излаза \mathbf{y} се евалуира посредством налажења максималне вредности латентне променљиве $\mathbf{z}_{\max} = \underset{\mathbf{z}}{\operatorname{argmax}} P(\mathbf{z} | \mathbf{x}, \boldsymbol{\theta})$. Условна расподела $P(\mathbf{z} | \mathbf{x}, \boldsymbol{\theta})$, која је мултиваријациона Гаусова расподела, има максималну вредност у очекивању исте, то јест $\mathbf{z}_{\max} = \boldsymbol{\mu}_z$. Стога се условна расподела $P(y_i = 1 | \mathbf{x}, \boldsymbol{\mu}_{z,i}, \boldsymbol{\theta})$ може изразити као:

$$P(y_i = 1 | \mathbf{x}, \boldsymbol{\mu}_{z,i}, \boldsymbol{\theta}) = \sigma(\mu_{z,i}) = \frac{1}{1 + \exp(-\mu_{z,i})} \quad (4.9)$$

где је $\mu_{z,i}$ очекивање латентне променљиве z_i .

4.3 Гаусова условна случајна поља за класификацију - учење

4.3.1 Учење у GCRFCb моделу

У поређењу са закључивањем, процедура учења GCRFCb модела је компликованија. Маргинализацију латентних променљивих у заједничкој расподели излаза и латентних променљивих није могуће аналитички одредити, самим тим је и немогуће извести условну функцију веродостојности. Условна логаритамска функција веродостојности представљена је као:

$$\begin{aligned} L(\mathbf{Y} | \mathbf{X}, \boldsymbol{\theta}) &= \log \left(\int_{\mathbf{Z}} P(\mathbf{Y}, \mathbf{Z} | \mathbf{X}, \boldsymbol{\theta}) d\mathbf{Z} \right) = \sum_{j=1}^M \log \left(\int_{z_j} P(y_j, z_j | \mathbf{x}_j, \boldsymbol{\theta}) dz_j \right) \\ &= \sum_{j=1}^M L_j(y_j | \mathbf{x}_j, \boldsymbol{\theta}) \end{aligned} \quad (4.10)$$

Односно логаритамска функција веродостојности за једну инстанцу може се приказати као:

$$L_j(y_j | \mathbf{x}_j, \boldsymbol{\theta}) = \log \int_{z_j} \prod_{i=1}^N \sigma(z_{ji})^{y_{ji}} (1 - \sigma(z_{ji}))^{1-y_{ji}} \frac{\exp(-\frac{1}{2}(\mathbf{z}_j - \boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_j^{-1}(\mathbf{z}_j - \boldsymbol{\mu}_j))}{(2\pi)^{N/2} |\boldsymbol{\Sigma}_j|^{1/2}} dz_j \quad (4.11)$$

где је $\mathbf{Y} \in \mathbb{R}^{M \times N}$ је скуп излаза у подацима који су намењени учењу модела; $\mathbf{X} \in \mathbb{R}^{M \times N \times A}$ је скуп атрибута у подацима који су намењени учењу модела; M је укупан број инстанци; A је укупан број атрибута. Треба поменути да је свака инстанца структурирана, тако да су инстанце међусобно независне али су излази у оквиру једне инстанце међусобно зависни.

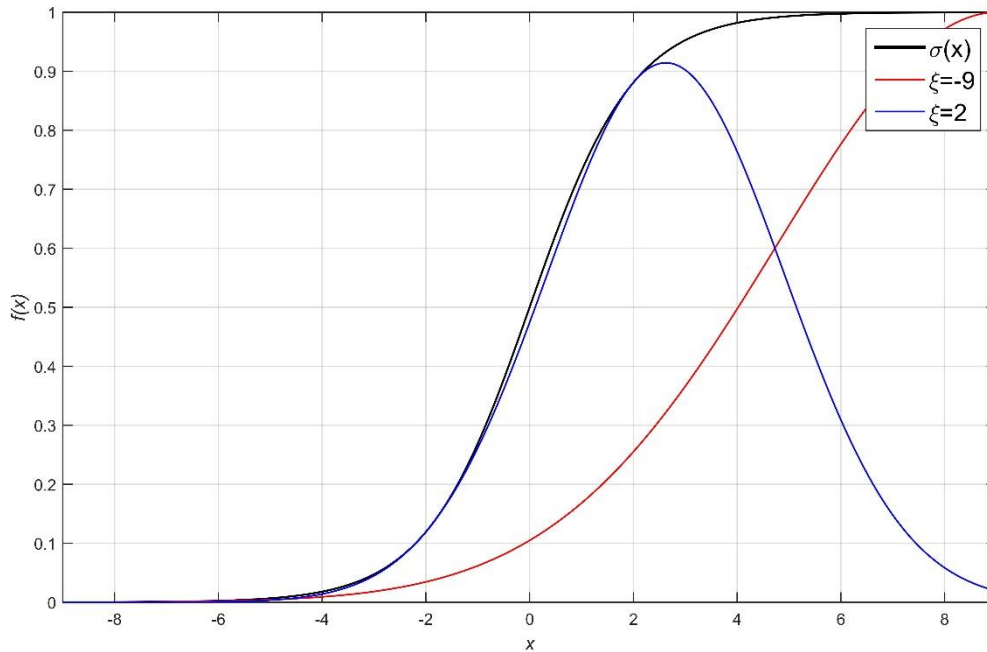
Један од начина апроксимације интеграла у условној логаритамској функцији веродостојности је посредством локалне варијационе апроксимације. Локална варијациона апроксимација изведена за сигмоидну функцију, представља функцију која у свакој тачки ограничава сигмоиду са доње стране и њен облик се може приказати као [88]:

$$\sigma(x) \cdot \sigma(\xi) \exp\{(x - \xi) / 2 - \lambda(\xi)(x^2 - \xi^2)\} \quad (4.12)$$

где је $\lambda(\xi) = -\frac{1}{2\xi} \cdot \left[\sigma(\xi) - \frac{1}{2} \right]$ и ξ је варијациони параметар.

Овакав начин ограничавања сигмоидне функције назива се ξ трансформација сигмоиде. Вредност доње границе једнака је стварној вредности сигмоидне функције

само у тачки у којој је извршена апроксимација исте. Изглед доње границе сигмоидне функције за различите вредности параметра ξ приказан је на слици 8.



Слика 8 Сигмоидна функција са доњим границама

Ова апроксимација може се успешно применити у циљу маргинализације условне функције веродостојности за једну инстанцу $P(\mathbf{y}_j, \mathbf{z}_j | \mathbf{x}_j, \boldsymbol{\theta})$ и може се приказати у следећем облику:

$$P(\mathbf{y}_j, \mathbf{z}_j | \mathbf{x}_j, \boldsymbol{\theta}) = P(\mathbf{y}_j | \mathbf{z}_j)P(\mathbf{z}_j | \mathbf{x}_j, \boldsymbol{\theta}) \geq \underline{P}(\mathbf{y}_j, \mathbf{z}_j | \mathbf{x}_j, \boldsymbol{\theta}, \boldsymbol{\xi}_j) \quad (4.13)$$

$$\underline{P}(\mathbf{y}_j, \mathbf{z}_j | \mathbf{x}_j, \boldsymbol{\theta}, \boldsymbol{\xi}_j) = \prod_{i=1}^N \sigma(\xi_{ji}) \exp\left(z_{ji} y_{ji} - \frac{z_{ji} + \xi_{ji}}{2} - \lambda(\xi_{ji})(z_{ji}^2 - \xi_{ji}^2)\right) \cdot \frac{1}{(2\pi)^{N/2} |\boldsymbol{\Sigma}_j|^{1/2}} \exp\left(-\frac{1}{2} (\mathbf{z}_j - \boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_j^{-1} (\mathbf{z}_j - \boldsymbol{\mu}_j)\right) \quad (4.14)$$

Упрошћена форма доње границе заједничке расподеле за једну инстанцу, добијене увођењем варијационе апроксимације, може се приказати у простијем облику као:

$$\underline{P}(\mathbf{y}_j, \mathbf{z}_j | \mathbf{x}_j, \boldsymbol{\theta}, \boldsymbol{\xi}_j) = T(\boldsymbol{\xi}_j) \exp\left(\mathbf{z}_j^T (\mathbf{y}_j - \frac{1}{2} \mathbf{I}) - \lambda \mathbf{z}_j^T \mathbf{z}_j - \frac{1}{2} \mathbf{z}_j^T \boldsymbol{\Sigma}_j^{-1} \mathbf{z}_j + \mathbf{z}_j^T \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\mu}_j\right) \quad (4.15)$$

где је:

$$T(\xi_j) = \frac{1}{(2\pi)^{N/2} |\Sigma_j|^{1/2}} \prod_{i=1}^N \sigma(\xi_{ji}) \exp\left(-\frac{1}{2} \boldsymbol{\mu}_j^T \Sigma_j^{-1} \boldsymbol{\mu}_j - \frac{\xi_{ji}}{2} + \lambda(\xi_{ji}) \xi_{ji}^2\right) \quad (4.16)$$

Доња граница функције веродостојности (условне расподеле) $\underline{P}(\mathbf{y}_j | \mathbf{x}_j, \boldsymbol{\theta}, \xi_j)$ за једну променљиву, може се добити маргинализацијом латентних променљивих \mathbf{z}_j као:

$$\begin{aligned} \underline{P}(\mathbf{y}_j | \mathbf{x}_j, \boldsymbol{\theta}, \xi_j) &= \int \underline{P}(\mathbf{y}_j, \mathbf{z}_j | \mathbf{x}_j, \boldsymbol{\theta}, \xi_j) d\mathbf{z}_j \\ &= T(\xi_j) \int \exp\left(\mathbf{z}_j^T (\mathbf{y}_j - \frac{1}{2} \mathbf{I}) - \Lambda_j \mathbf{z}_j^T \mathbf{z}_j - \frac{1}{2} \mathbf{z}_j^T \Sigma_j^{-1} \mathbf{z}_j + \mathbf{z}_j^T \Sigma_j^{-1} \boldsymbol{\mu}_j\right) d\mathbf{z}_j \\ &= T(\xi_j) \int \exp\left(\begin{aligned} &-\frac{1}{2} \mathbf{z}_j^T (\Sigma_j^{-1} + 2\Lambda_j) \mathbf{z}_j + \\ &\mathbf{z}_j^T (\Sigma_j^{-1} + 2\Lambda_j) (\Sigma_j^{-1} + 2\Lambda_j)^{-1} ((\mathbf{y}_j - \frac{1}{2} \mathbf{I}) + \Sigma_j^{-1} \boldsymbol{\mu}_j) \end{aligned}\right) d\mathbf{z}_j \end{aligned} \quad (4.17)$$

Простија форма доње границе функције веродостојности $\underline{P}(\mathbf{y}_j | \mathbf{x}_j, \boldsymbol{\theta}, \xi_j)$ је:

$$\begin{aligned} \underline{P}(\mathbf{y}_j | \mathbf{x}_j, \boldsymbol{\theta}, \xi_j) &= T(\xi_j) \int \exp\left(-\frac{1}{2} (\mathbf{z}_j - \mathbf{m}_j)^T S_j^{-1} (\mathbf{z}_j - \mathbf{m}_j) + \frac{1}{2} \mathbf{m}_j^T S_j^{-1} \mathbf{m}_j\right) d\mathbf{z}_j \\ &= T(\xi_j) \exp\left(\frac{1}{2} \mathbf{m}_j^T S_j^{-1} \mathbf{m}_j\right) \int \exp\left(-\frac{1}{2} (\mathbf{z}_j - \mathbf{m}_j)^T S_j^{-1} (\mathbf{z}_j - \mathbf{m}_j)\right) d\mathbf{z}_j \end{aligned} \quad (4.18)$$

Интеграција овог израза се може евалуирати аналитички имајући у виду да наведена форма подинтегралне функције одговара вредности ненормализоване мултиваријационе Гаусове расподеле. Стога, следи да је доња граница логаритамске функције веродостојности $\underline{L}_j(\mathbf{y}_j | \mathbf{x}_j, \boldsymbol{\theta}, \xi_j)$ приказана као:

$$\begin{aligned} \underline{L}_j(\mathbf{y}_j | \mathbf{x}_j, \boldsymbol{\theta}, \xi_j) = \log \underline{P}(\mathbf{y}_j | \mathbf{x}_j, \boldsymbol{\theta}, \xi_j) &= \sum_{i=1}^N \left(\log \sigma(\xi_{ji}) - \frac{\xi_{ji}}{2} + \lambda(\xi_{ji}) \xi_{ji}^2 \right) - \\ &\frac{1}{2} \boldsymbol{\mu}_j^T \Sigma_j^{-1} \boldsymbol{\mu}_j + \frac{1}{2} \mathbf{m}_j^T S_j^{-1} \mathbf{m}_j + \frac{1}{2} \log |S_j| \end{aligned} \quad (4.19)$$

где су коришћени следећи изрази:

$$S_j^{-1} = \Sigma_j^{-1} + 2\Lambda_j \quad (4.20)$$

$$\mathbf{m}_j = \Sigma_j \left((\mathbf{y}_j - \frac{1}{2} \mathbf{I}) + \Sigma_j^{-1} \boldsymbol{\mu}_j \right) \quad (4.21)$$

$$\Lambda_j = \begin{bmatrix} \lambda(\xi_{j1}) & 0 & 0 & \dots & 0 \\ 0 & \lambda(\xi_{j2}) & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \lambda(\xi_{jN}) \end{bmatrix} \quad (4.22)$$

У циљу оптимизације доње границе условне функције веродостојности, односно налажење њене максималне вредности у зависности од параметара α , β и матрице варијационих параметара $\xi \in \mathbb{R}^{M \times N}$ коришћен је алгоритам градијентног спуста. Ипак, у циљу очувања позитивне дефинитности коваријационе матрице у нормалној расподели, потребан, али не и довољан услов, је обезбедити да коефицијенти α и β буду позитивни односно $\alpha > 0$ и $\beta > 0$. Парцијални извод условне логаритамске функције

веродостојности по једној инстанци $\frac{\partial \underline{L}_j(\mathbf{y}_j | \mathbf{x}_j, \theta, \xi_j)}{\partial \alpha_k}$ је одређен као:

$$\begin{aligned} \frac{\partial \underline{L}_j(\mathbf{y}_j | \mathbf{x}_j, \theta, \xi_j)}{\partial \alpha_k} = & -\frac{1}{2} \text{Tr} \left(S_j \frac{\partial S_j^{-1}}{\partial \alpha_k} \right) + \frac{\partial \mathbf{m}_j^T}{\partial \alpha_k} S_j^{-1} \mathbf{m}_j + \frac{1}{2} \mathbf{m}_j^T \frac{\partial S_j^{-1}}{\partial \alpha_k} \mathbf{m}_j - \\ & - \frac{\boldsymbol{\mu}_j^T}{\partial \alpha_k} \Sigma_j^{-1} \boldsymbol{\mu}_j - \frac{1}{2} \boldsymbol{\mu}_j^T \frac{\partial \Sigma_j^{-1}}{\partial \alpha_k} + \frac{1}{2} \text{Tr} \left(\Sigma_j \frac{\partial \Sigma_j^{-1}}{\partial \alpha_k} \right) \end{aligned} \quad (4.23)$$

где су коришћени следећи изрази:

$$\frac{\partial S_j^{-1}}{\partial \alpha_k} = \frac{\partial \Sigma_j^{-1}}{\partial \alpha_k} = \begin{cases} 2, & \text{ако је } i=j \\ 0, & \text{ако је } i \neq j \end{cases} \quad (4.24)$$

$$\frac{\partial \mathbf{m}_j^T}{\partial \alpha_k} = - \left(\mathbf{y}_j - \frac{1}{2} \mathbf{I} + \boldsymbol{\mu}_j^T \Sigma_j^{-1} \right) S_j \frac{\partial S_j^{-1}}{\partial \alpha_k} S_j + \frac{\partial \boldsymbol{\mu}_j^T}{\partial \alpha_k} \Sigma_j^{-1} S_j + \boldsymbol{\mu}_j^T \frac{\partial \Sigma_j^{-1}}{\partial \alpha_k} S_j \quad (4.25)$$

$$\frac{\partial \boldsymbol{\mu}_j^T}{\partial \alpha_k} = \left(2\alpha_k R_{jk}(\mathbf{x}) - \frac{\partial \Sigma_j^{-1}}{\partial \alpha_k} \boldsymbol{\mu}_j \right)^T \Sigma_j^T \quad (4.26)$$

На сличан начин су дефинисани парцијални изводи у односу на параметар β :

$$\begin{aligned} \frac{\partial \underline{L}_j(\mathbf{y}_j | \mathbf{x}_j, \theta, \xi_j)}{\partial \beta_l} = & -\frac{1}{2} \text{Tr} \left(S_j \frac{\partial S_j^{-1}}{\partial \beta_l} \right) + \frac{\partial \mathbf{m}_j^T}{\partial \beta_l} S_j^{-1} \mathbf{m}_j + \frac{1}{2} \mathbf{m}_j^T \frac{\partial S_j^{-1}}{\partial \beta_l} \mathbf{m}_j - \\ & - \frac{\boldsymbol{\mu}_j^T}{\partial \beta_l} \Sigma_j^{-1} \boldsymbol{\mu}_j - \frac{1}{2} \boldsymbol{\mu}_j^T \frac{\partial \Sigma_j^{-1}}{\partial \beta_l} + \frac{1}{2} \text{Tr} \left(\Sigma_j \frac{\partial \Sigma_j^{-1}}{\partial \beta_l} \right) \end{aligned} \quad (4.27)$$

где су коришћени следећи изрази:

$$\frac{\partial S_j^{-1}}{\partial \beta_l} = \frac{\partial \Sigma_j^{-1}}{\partial \beta_l} = \begin{cases} \sum_{n=1}^N e_{in}^l S_{in}^l(x), & \text{ако је } i=j \\ -e_{ij}^l S_{ij}^l(x), & \text{ако је } i \neq j \end{cases} \quad (4.28)$$

$$\frac{\partial \mathbf{m}_j^T}{\partial \beta_l} = - \left(\mathbf{y}_j - \frac{1}{2} \mathbf{I} + \boldsymbol{\mu}_j^T \Sigma_j^{-1} \right) S_j \frac{\partial S_j^{-1}}{\partial \beta_l} S_j + \frac{\partial \boldsymbol{\mu}_j^T}{\partial \beta_l} \Sigma_j^{-1} S_j + \boldsymbol{\mu}_j^T \frac{\partial \Sigma_j^{-1}}{\partial \beta_l} S_j \quad (4.29)$$

$$\frac{\partial \boldsymbol{\mu}_j^T}{\partial \beta_l} = \left(-\frac{\partial \Sigma_j^{-1}}{\partial \beta_l} \boldsymbol{\mu}_j \right)^T \Sigma_j^T \quad (4.30)$$

Парцијални изводи у односу на варијационе параметре ξ_{ji} се могу изразити као:

$$\begin{aligned} \frac{\partial \underline{L}_j(\mathbf{y}_j | \mathbf{x}_j, \boldsymbol{\theta}, \boldsymbol{\xi}_j)}{\partial \xi_{ji}} &= -\frac{1}{2} \text{Tr} \left(2S_j \frac{\partial \Lambda_j}{\partial \xi_{ji}} \right) - \left[2 \left(\mathbf{y}_j - \frac{1}{2} \mathbf{I} \right) S_j \frac{\partial \Lambda_j}{\partial \xi_{ji}} S_j \right] S_j^{-1} \mathbf{m}_j + \\ &+ \mathbf{m}_j^T \frac{\partial \Lambda_j}{\partial \xi_{ji}} \mathbf{m}_j + \sum_{i=1}^N \left(\left(\frac{1}{\sigma(\xi_{ji})} + \frac{1}{2} \xi_{ji} \right) \frac{\partial \sigma(\xi_{ji})}{\partial \xi_{ji}} + \frac{1}{2} \left(\sigma(\xi_{ji}) - \frac{3}{4} \right) \right) \end{aligned} \quad (4.31)$$

где су коришћени следећи изрази:

$$\frac{\partial \Lambda_j}{\partial \xi_{ji}} = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \frac{\partial \lambda(\xi_{ji})}{\partial \xi_{ji}} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix} \quad (4.32)$$

$$\frac{\partial \sigma(\xi_{ji})}{\partial \xi_{ji}} = \sigma(\xi_{ji})(1 - \sigma(\xi_{ji})) \quad (4.33)$$

$$\frac{\partial \lambda(\xi_{ji})}{\partial \xi_{ji}} = \frac{1}{2\xi_{ji}} \frac{\partial \sigma(\xi_{ji})}{\partial \xi_{ji}} - \frac{1}{2} \left(\sigma(\xi_{ji}) - \frac{1}{2} \right) \frac{1}{\xi_{ji}^2} \quad (4.34)$$

С обзиром да је у питању оптимизација са ограничењима, алгоритам градијентног спуста се не може директно применити. Један од начина примене алгоритма директног спуста је коришћењем логаритамске трансформације која је представљена као [35]:

$$\alpha_k = \exp(u_k) \quad (4.35)$$

$$\beta_l = \exp(v_l) \quad (4.36)$$

Овим су вредности кофицијената $\boldsymbol{\alpha}$ и $\boldsymbol{\beta}$ ограничени да буду позитивне за било коју вредност u и v . Стога је циљ наћи вредности u и v који максимизују логаритамску функцију условне веродостојности као:

$$u_k = \log(\alpha_k) \quad (4.37)$$

$$v_l = \log(\beta_l) \quad (4.38)$$

$$u_k^{novo} = u_k^{staro} + \eta \frac{\partial \underline{L}_j(\mathbf{y}_j | \mathbf{x}_j, \boldsymbol{\theta}, \boldsymbol{\xi}_j)}{\partial u_k}, \quad \frac{\partial \underline{L}_j(\mathbf{y}_j | \mathbf{x}_j, \boldsymbol{\theta}, \boldsymbol{\xi}_j)}{\partial u_k} = \frac{\partial \underline{L}}{\partial \log(\alpha_k)} = \alpha_k \frac{\partial \underline{L}}{\partial \alpha_k} \quad (4.39)$$

$$v_l^{novo} = v_l^{staro} + \eta \frac{\partial L_j(\mathbf{y}_j | \mathbf{x}_j, \boldsymbol{\theta}, \xi_j)}{\partial v_l}, \quad \frac{\partial L_j(\mathbf{y}_j | \mathbf{x}_j, \boldsymbol{\theta}, \xi_j)}{\partial v_l} = \frac{\partial L}{\partial \log(\beta_l)} = \beta_l \frac{\partial L}{\partial \beta_l} \quad (4.40)$$

где је η параметар учења.

Ова процедура може даље бити генерализована на било који алгоритам који се заснива на градијентном спусту, а узима у обзир адаптивну вредност параметра учења. Конкретно у овом случају могу се применити и друге технике нелинеарног програмирања са ограничењима. Једна од њих која је посебно добра услед оптимизације великог броја варијационих параметара, а која је и коришћена у овом раду је зарубљен Њутнов алгоритам (енг. „*Truncated Newtons constrained algorithm*“, скраћено TNC) који се може пронаћи у [89] и [90]. Потребно је нагласити да проблем оптимизације логаритамске вредности дође границе функције веродостојности у односу на параметре $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ и ξ није конвексан као у случају Гаусових условних случајних поља за регресију, стога се исти мора више пута иницијализовати и понављати све док се не добију прихватљиви резултати, односно дође до одговарајућег оптимума (локалног или глобалног).

4.3.2 Учење у GCRFCnb моделу

Учење у GCRFCnb моделу је много једноставније у поређењу са GCRFCb алгоритмом, зато што се уместо маргинализације латентних променљивих, користи модус расподеле латентних променљивих условљеним атрибутима. Модус се може евалуирати директно тако да апроксимационе технике нису потребне. Условна логаритмска функција веродостојности може се исказати као:

$$\begin{aligned} L(\mathbf{Y} | \mathbf{X}, \boldsymbol{\theta}, \boldsymbol{\mu}) &= \log P(\mathbf{Y} | \mathbf{X}, \boldsymbol{\theta}, \boldsymbol{\mu}) = \sum_{j=1}^M \sum_{i=1}^N \log P(y_{ji} | \mathbf{x}_j, \boldsymbol{\theta}, \mu_{ji}) \\ &= \sum_{j=1}^M \sum_{i=1}^N L_{ji}(y_{ji} | \mathbf{x}_j, \boldsymbol{\theta}, \mu_{ji}) \end{aligned} \quad (4.41)$$

$$L_{ji}(y_{ji} | \mathbf{x}_j, \boldsymbol{\theta}, \mu_{ji}) = y_{ji} \log \sigma(\mu_{ji}) + (1 - y_{ji}) \log (1 - \sigma(\mu_{ji})) \quad (4.42)$$

Парцијални изводи условне логаритамске функције веродостојности у односу на параметре $\boldsymbol{\alpha}$ и $\boldsymbol{\beta}$ дефинисани су као:

$$\frac{\partial L_{ji}(y_{ji} | \mathbf{x}_j, \boldsymbol{\theta}, \mu_{ji})}{\partial \alpha_k} = (y_{ji} - \sigma(\mu_{ji})) \frac{\partial \mu_{ji}}{\partial \alpha_k} \quad (4.43)$$

$$\frac{\partial L_{ji}(y_{ji} | \mathbf{x}_j, \boldsymbol{\theta}, \mu_{ji})}{\partial \beta_l} = (y_{ji} - \sigma(\mu_{ji})) \frac{\partial \mu_{ji}}{\partial \beta_l} \quad (4.44)$$

где су $\frac{\partial \mu_{ji}}{\partial \alpha_k}$ и $\frac{\partial \mu_{ji}}{\partial \beta_l}$ елементи вектора $\boldsymbol{\mu}_j$ и $\frac{\partial \boldsymbol{\mu}_j}{\partial \beta_l}$.

На сличан начин услед постојања ограничења да параметри $\boldsymbol{\alpha}$ и $\boldsymbol{\beta}$ буду позитивни, може се користити неки од предложених метода нелинеарног програмирања. Услед малог броја оптимizacionих параметара може се користити и секвенцијално квадратно програмирање за проблеме нелинеарног програмирања са ограничењима (енг. „*Sequential Quadratic Programming for constrained nonlinear optimization*“, скраћено SQP) [91].

4.4 Експерименти над синтетичким подацима

У циљу поређења перформанси презентованих класификатора коришћена је метрика површина испод криве (скраћено ПИК) (енг. „*Area Under Curve*“, скраћено AUC). Крива (енг. „*Receiver operating characteristic Curve*“, скраћено ROC) испод које се тражи површина је крива у координатном систему удела лажно позитивних инстанци у скупу свих позитивних инстанци (x оса) и удела стварно позитивних инстанци у скупу свих позитивних инстанци (y оса) и она се може представити као квадрат са страницама 1. За сваку вредност прага на скоровима које класификатор даје, дефинисана је једна тачка у овом координатном систему, а мењајући овај скор од минималног (на датим инстанцама) до максималног, добија се крива која спаја координатни почетак са тачком (1, 1). Скор 1 означава да класификатор ради перфектно, док скор од 0,5 означава да класификатор погађа потпуно случајно.

Поред ПИК мере у циљу што бољег разумевања предности Бајесовског у односу на не-Бајесовски приступ класификације, код експеримената над синтетичким подацима коришћена је и доња граница условне логаритмаске функције веродостојности $\underline{L}(\mathbf{Y} | \mathbf{X}, \boldsymbol{\theta}, \boldsymbol{\mu})$ (у случају GCRFBCb модела) и условна функција веродостојности $L(\mathbf{Y} | \mathbf{X}, \boldsymbol{\theta})$ (у случају GCRFBCnb модела).

Сви модели и алгоритми су имплементирани у програмском језику Python, док су експерименти покретани на Ubuntu серверу са 128 GB радне меморије и два процесора Intel Xeon 2.9 GHz CPU.

4.4.1 Опис експеримента

Основни циљ експеримента над вештачким скупом података је да се тестирају развијени класификациони модели и да се прикажу предности и мане сваког од модела понаособ. У свим експериментима са синтетичким подацима коришћена су два графа ($\beta \in \mathbb{R}^2$) и два неструктурна предиктора ($\alpha \in \mathbb{R}^2$). Униформном расподелом генерисани су и маркирани излази у графу, генерисане су тежине између веза које повезују излазе S и генерисани су неструктурни предиктори R . За различите вредности параметара α и β , извршени су експерименти у циљу процене осетљивости модела. Веће вредности параметра α представљају индикатор да је модел више сигуран у перформансе неструктурних предиктора, док веће вредности β означавају да модел већу тежину ставља на корелисаност излаза у односу на вредност добијену неструктурним предикторима. Ниско поверење у неструктурне предикторе означава и да је норма варијансе дијагоналих елемената коваријационе матрице латентних променљивих висока.

За генерисане вредности S , R , и задате параметре α и β , добијене су вероватноће излаза, док су исти маркирани на основу граничне вредности 0,5. Од укупно генерисаних података 20% је коришћено у тест скупу података док је 80% коришћено у скупу за учење модела.

4.4.2 Предиктивне перформансе

Основни циљ експеримента је утврђивање утицаја параметара α и β , у процесу генерисања података, на перформансе предикције модела GCRFBC α и GCRFBC β . Шест различитих вредности параметара α и β су коришћени. Вредности параметара су подељени у три различите групе:

1. Прва група, у којој параметри α и β имају сличне вредности, стога утицај неструктурних предиктора и везе између излаза имају подједнаку важност.
2. Друга група, у којој параметар α има већу вредност од параметра β , односно на излазну вредност више утичу неструктурни предиктори у односу на везе између истих.
3. Трећа група, у којој параметар β има вишу вредност у односу на параметар α сходно томе и везе између излаза знатно више утичу на вредност истих у односу на предикције неструктурних предиктора.

Резултати добијени експериментима приказани су у табели 1. Заједно са ПИК скором и условном функцијом логаритамске веродостојности (у случају GCRFBCb модела њеном доњом границом), у табели 1 приказана је и норма варијансе латентних променљивих (дијагоналних елемената коваријационе матрице). Може се приметити да у случају када је норма варијанси латентних променљивих мала оба модела имају једнаку вредност ПИК скорa и једнаке вредности условне логаритамске функције веродостојности $\underline{L}(\mathbf{Y} | \mathbf{X}, \theta)$.

Ово је везано за случај када су вредности параметра α велике у односу на параметар β . Самим тим, условна расподела $P(\mathbf{z} | \mathbf{x}, \theta)$ је концентрована око очекиване вредности, стога је апроксимација латентних променљивих очекивањем задовољавајућа.

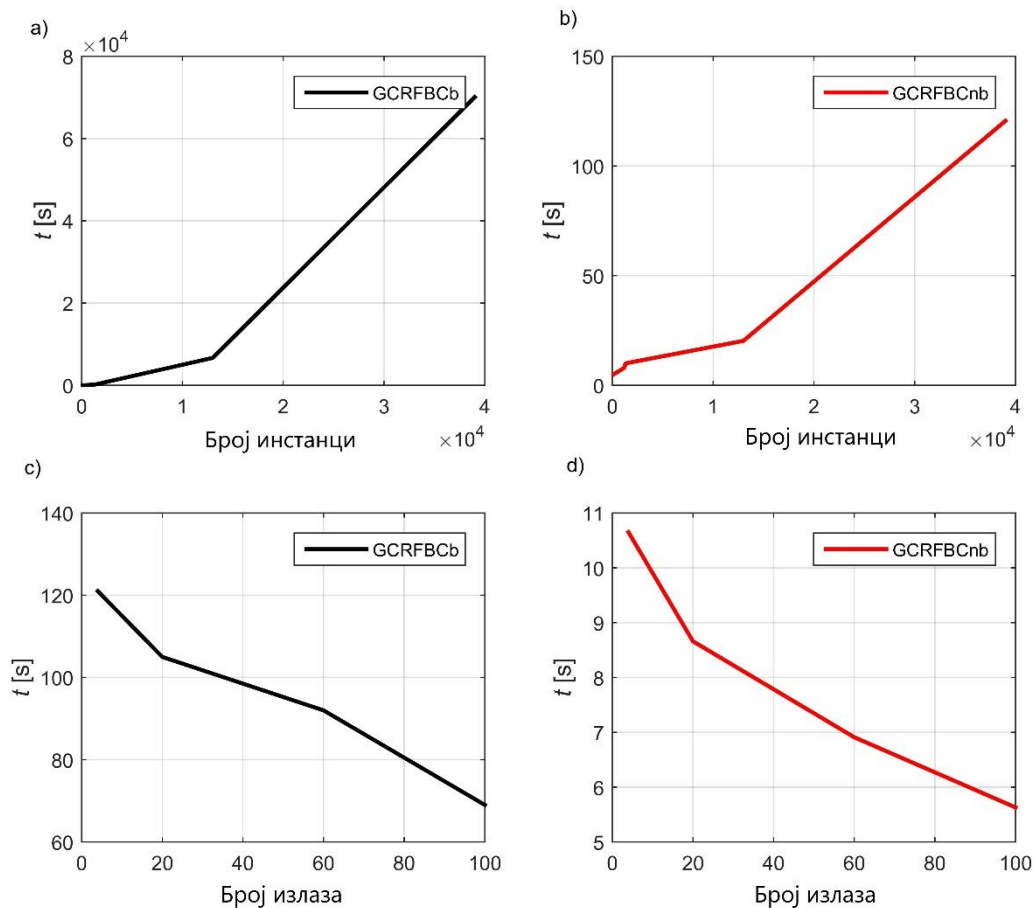
За разлику од тога, у случајевима када је параметар β значајно већи у односу на параметар α , GCRFBCb модел показује боље перформансе предикције у односу на GCRFBCnb. Ово је изазвано тиме што је расподела $P(\mathbf{z} | \mathbf{x}, \theta)$ широко распрострањена око очекиване вредности и самим тим апроксимација латентних променљивих посредством очекиваних вредности не даје задовољавајуће резултате. Може се закључити да GCRFBCb модел има предиктивне перформансе које су у најгорем случају исте као код GCRFBCnb модела. Такође је показано да су оба модела у могућности да експлоатишу већину информација која им је на располагању из атрибута или структуре између излаза, што је видљиво посредством ПИК скорa у табели 1.

Табела 1 Поређење предиктивних перформанси GCRFBC β и GCRFBC $n\beta$ модела за различите вредности параметара α и β , измерених помоћу ПИК скорa, условне логаритамске функције веродостојности и нормe дијагоналних елемената коваријационе матрице латентних променљивих

Број	Параметри					
		ПИК	$\underline{L}(\mathbf{Y} \mathbf{X}, \theta, \mu)$	$\ \text{Var}\ ^2$	ПИК	$L(\mathbf{Y} \mathbf{X}, \theta)$
1	$\alpha = [5,4]$	0,812	-71,150	0,000	0,812	-71,151
	$\beta = [5,22]$					
2	$\alpha = [1,18]$	0,903	-75,033	0,001	0,902	-75,033
	$\beta = [1,18]$					
3	$\alpha = [22,21]$	0,988	-83,957	0,000	0,988	-83,957
	$\beta = [5,22]$					
4	$\alpha = [22,21]$	0,866	-83,724	0,000	0,886	-83,466
	$\beta = [0.1,0.67]$					
5	$\alpha = [0.8,0.5]$	0,860	-83,353	34,827	0,817	-84,009
	$\beta = [5,22]$					
6	$\alpha = [0.2,0.4]$	0,931	-70,692	35,754	0,821	-70,391
	$\beta = [1,18]$					

4.4.3 Меморијска комплексност и рачунски трошак

Рачунски и меморијски трошак GCRFBC $n\beta$ модела током учења и закључивања је исти као и код стандардног GCRF модела. Ако учење траје T итерација укупни трошак рачунања GCRFBC $n\beta$ модела је $O(TN^3)$. У најбољем случају када је матрица прецизности ретка овај трошак може да се смањи на вредност $O(TN^2)$. Меморијска комплексност овог алгоритма је занемарљива с обзиром на мали број оптимизационих параметара и матрица које нису великих димензија.



Слика 9 Време учења GCRFBCb и GCRFBCnb модела у зависности од броја инстанци и броја излаза

У случају GCRFBCb модела меморијска комплексност током учења је $O(M)$, услед великог броја варијационих параметара. Трошак рачунања овог модела је реда $O(TMN^3)$ и може се такође редуковати у случају ретке матрице прецизности на $O(TMN^2)$.

Тестови брзине учења GCRFBCb и GCRFBCnb модела су спроведени са синтетички генерисаним подацима са различитим бројем излаза и инстанци. Под инстанцом се подразумева скуп података који садржи вредности атрибута (x) свих излаза (y) заједно са класом истих у посматраном тренутку. На слици 9(a) и 9(b) приказана су времена рачунања оба модела у зависности од броја инстанци. Може се закључити да већи број инстанци има значајан утицај на време израчунавања алгоритма. На сликама 9(c) и 9(d) може се уочити зависност времена израчунавања од броја излаза, док се производ броја излаза и броја инстанци одржава константним. Овим се закључује

да се време израчунавања брже повећава са повећањем броја инстанци у односу на број излаза.

4.5 Убрзање Бајесовских Гаусових условних случајних поља за класификацију

С обзиром на висок рачунски и меморијски трошак учења Бајесовских Гаусових условних случајних поља за класификацију, развијена је апроксимативна верзија алгоритма за учење истих.

Може се приметити да је компонента која највише утиче на висок меморијски трошак везана за парцијални извод доње границе логаритамске вредности функције веродостојности по варијационим параметрима. Овај парцијални извод је различит за сваку инстанцу и сваки излаз, с обзиром да је број варијационих параметара једнак производу броја инстанци и броја излаза. Самим тим, неопходно је током сваке итерације обезбедити меморијски простор у коме се смештају ови парцијални изводи. Поред тога, у случају већег броја инстанци или излаза, врло велики број варијационих параметара знатно отежава учење, а самим тим и једноставну конвергенцију параметара ка некој тачки која представља максимум (локални или глобални) функције веродостојности. Огроман број варијационих параметара ствара низ проблема који се тичу појаве прелазних тачака и експлодирање градијента (посебно се ово примећује при коришћењу логаритамске трансформације и коњугованог градијентног спуста). Рачунски трошак стандарних алгоритама нелинеарног програмирања расте са повећањем броја параметара по којима се тражи оптимално решење. Стога се може и закључити да велики број параметара знатно отежава учење и онемогућава примену алгоритма на реалне проблеме са великим бројем излаза и инстанци.

Ulaz: Skup podataka (\mathbf{H}, \mathbf{Y}), Matrica sličnosti S , broj_klastera
metod klasterovanja,
Izlaz: α, β
 $\mathbf{R}(X) \leftarrow$ Nestruturni prediktori učenje (\mathbf{H}, \mathbf{Y})
 $(\alpha, \beta, \xi_{VEC}) \leftarrow$ inicijalizuj vrednosti
dok epoha < maks_epoha **uradi**
 $\mu \leftarrow$ odredi($\mathbf{H}, \mathbf{Y}, S, \alpha, \beta$)
 $\mu_{klas} \leftarrow$ promeni_oblik(μ ili $\sigma(\mu)$)
 $klasteri \leftarrow$ klasteruj(μ_{klas})
 $centri \leftarrow$ odredi_centre(*кластеру*)
za $ind \leftarrow 1$ do broj_klastera **uradi**
 $rep_{ind} \leftarrow$ nađi_najbliži($\mu_{klas}, centri_{ind}$)
 $\frac{\partial L_{ind}(\mathbf{y}_{rep_{ind}} | \mathbf{x}_{rep_{ind}}, \theta, \xi_{VEC})}{\partial \xi_{ind}} \leftarrow$ odredi($\mathbf{x}_{rep_{ind}}, \mathbf{y}_{rep_{ind}}, S, \alpha, \beta$)
za $j \leftarrow 1$ do broj_ins **uradi**
 $\frac{\partial L_j(\mathbf{y}_j | \mathbf{x}_j, \theta, \xi_j)}{\partial \alpha} \leftarrow$ odredi($\mathbf{x}_j, \mathbf{y}_j, S, \alpha, \beta, \xi_{VEC}$)
 $\frac{\partial L_j(\mathbf{y}_j | \mathbf{x}_j, \theta, \xi_j)}{\partial \beta} \leftarrow$ odredi($\mathbf{x}_j, \mathbf{y}_j, S, \alpha, \beta, \xi_{VEC}$)
za $i \leftarrow 1$ do broj_izlaza **uradi**
 $klase_{ji} \leftarrow$ odredi_klasu_instanci($klasteri, \mu_{ji}$ ili $\sigma(\mu_{ji})$)
 $\frac{\partial L_j(\mathbf{y}_j | \mathbf{x}_j, \theta, \xi_j)}{\partial \xi_{ji}} \leftarrow$ odredi($\frac{\partial L_{ind}(\mathbf{y}_{rep_{ind}} | \mathbf{x}_{rep_{ind}}, \theta, \xi_{VEC})}{\partial \xi_{ind}}, klase_{ji}$)
gradijent $\leftarrow \left[\frac{\partial L_j(\mathbf{y}_j | \mathbf{x}_j, \theta, \xi_j)}{\partial \alpha}, \frac{\partial L_j(\mathbf{y}_j | \mathbf{x}_j, \theta, \xi_j)}{\partial \beta}, \frac{\partial L_j(\mathbf{y}_j | \mathbf{x}_j, \theta, \xi_j)}{\partial \xi_{ji}} \right]$
 $[\alpha, \beta, \xi_{VEC}] \leftarrow [\alpha, \beta, \xi] + \eta \cdot$ gradijent
vрати α, β

Слика 10 Алгоритам учења убрзане верзије Гаусових условних случајних поља за класификацију са кластеровањем инстанци и излаза

Један од начина на основу кога се може решити наведени проблем је покушајем да се велики број варијационих параметара сведе на мањи. Овим видом апроксимације смањује се број варијационих параметара и самим тим сви проблеми који су везани за

процес учења могу бити знатно олакшани, док се меморијски и рачунски трошак смањује.

С обзиром на то да се члан $\frac{\partial L_j(\mathbf{y}_j | \mathbf{x}_j, \boldsymbol{\theta}, \xi_j)}{\partial \xi_{ji}}$ одређује за сваку инстанцу и сваки

излаз понаособ неопходно је покушати на неки начин груписати сличне инстанце и излазе у одговарајуће групе које су међусобно сличне. Односно циљ је да се током сваке епохе учења, за тренутне вредности $\boldsymbol{\alpha}$ и $\boldsymbol{\beta}$, одреди за сваку инстанцу очекивана вредност латентних променљивих (као у случају не-Бајесовских Гаусових условних случајних поља за класификацију). Потом, све инстанце и одговарајуће излазе у оквиру инстанце треба кластеровати на основу очекиване вредности латентних променљивих (или сигмоидне трансформације очекиваних вредности) у групе. За сваку групу односно кластер, одређује се центроид кластера, а затим налази елемент из групе који је најближи овом центру и он се усваја као репрезент групе. Потом се за сваки репрезент кластера евалуирају вредности $\frac{\partial L_j(\mathbf{y}_j | \mathbf{x}_j, \boldsymbol{\theta}, \xi_j)}{\partial \xi_{ji}}$, а затим се на основу ове вредности одређују

парцијални изводи за све остале вредности инстанци и одговарајућих излаза.

На овај начин је знатно умањен број варијационих параметара и смањен је знатно меморијски трошак учења. Рачунски трошак алгоритма превасходно зависи од усвојеног броја кластера, као и од усвојеног алгоритма кластеровања. У зависности од начина груписања инстанци и потом усвајања одговарајућих вредности парцијалних извода за евалуиране вредности истих, разликују се два развијена апроксимативна алгоритма:

- 1) Алгоритам учења убрзане верзије Гаусових условних случајних поља за класификацију са кластеровањем инстанци и излаза. Алгоритам је приказан на слици 10.
- 2) Алгоритам учења убрзане верзије Гаусових условних случајних поља за класификацију са одређивањем вероватноће припадања кластеру. Алгоритам је приказан на слици 11.

Ulaz: Skup podataka (\mathbf{H}, \mathbf{Y}), Matrica sličnosti S , broj_klastera
 metod klasterovanja,
Izlaz: α, β
 $\mathbf{R}(X) \leftarrow$ Nestruturni prediktori učenje (\mathbf{H}, \mathbf{Y})
 $(\alpha, \beta, \xi_{VEC}) \leftarrow$ inicijalizuj vrednosti
dok epoha < maks_epoha **uradi**
 $\mu \leftarrow$ odredi($\mathbf{H}, \mathbf{Y}, S, \alpha, \beta$)
 $\mu_{klas} \leftarrow$ promeni_oblik(μ ili $\sigma(\mu)$)
 klasteri \leftarrow klasteruj(μ_{klas})
 centri \leftarrow odredi_centre(кластеру)
 za ind $\leftarrow 1$ do broj_klastera **uradi**
 $rep_{ind} \leftarrow$ nađi_najbliži($\mu_{klas}, centri_{ind}$)
 $\frac{\partial L_{ind}(\mathbf{y}_{rep_{ind}} | \mathbf{x}_{rep_{ind}}, \theta, \xi_{VEC})}{\partial \xi_{ind}} \leftarrow$ odredi($\mathbf{x}_{rep_{ind}}, \mathbf{y}_{rep_{ind}}, S, \alpha, \beta$)
 za j $\leftarrow 1$ do broj_ins **uradi**
 $\frac{\partial L_j(\mathbf{y}_j | \mathbf{x}_j, \theta, \xi_j)}{\partial \alpha} \leftarrow$ odredi($\mathbf{x}_j, \mathbf{y}_j, S, \alpha, \beta, \xi_{VEC}$)
 $\frac{\partial L_j(\mathbf{y}_j | \mathbf{x}_j, \theta, \xi_j)}{\partial \beta} \leftarrow$ odredi($\mathbf{x}_j, \mathbf{y}_j, S, \alpha, \beta, \xi_{VEC}$)
 za i $\leftarrow 1$ do broj_излаза **uradi**
 $verov_{ji} \leftarrow$ odredi_verovatnoću_instanci(klasteri, μ_{ji} ili $\sigma(\mu_{ji})$)
 $\frac{\partial L_j(\mathbf{y}_j | \mathbf{x}_j, \theta, \xi_j)}{\partial \xi_{ji}} \leftarrow$ odredi($\frac{\partial L_{ind}(\mathbf{y}_{rep_{ind}} | \mathbf{x}_{rep_{ind}}, \theta, \xi_{VEC})}{\partial \xi_{ind}}, verov_{ji}$)
 gradijent $\leftarrow \left[\frac{\partial L_j(\mathbf{y}_j | \mathbf{x}_j, \theta, \xi_j)}{\partial \alpha}, \frac{\partial L_j(\mathbf{y}_j | \mathbf{x}_j, \theta, \xi_j)}{\partial \beta}, \frac{\partial L_j(\mathbf{y}_j | \mathbf{x}_j, \theta, \xi_j)}{\partial \xi_{ji}} \right]$
 $[\alpha, \beta, \xi_{VEC}] \leftarrow [\alpha, \beta, \xi] + \eta \cdot$ gradijent
vрати α, β

Слика 11 Алгоритам учења убрзане верзије Гаусових условних случајних поља за класификацију са одређивањем вероватноће припадања кластеру

4.5.1 Алгоритам учења убрзане верзије Гаусових условних случајних поља за класификацију са кластеровањем инстанци и излаза

Код овог типа алгоритма се на основу израчунатих вредности парцијалних извода логаритамске функције веродостојности по варијационом параметру, вредности парцијалних извода осталих инстанци и излаза одређују изједначавањем истих са одговарајућим парцијалним изводом репрезента који се налази у истом кластеру.

Алгоритми кластеровања који су коришћени су: К-средина (енг. „*K-means*“), мини беч К-средина (енг. „*Mini batch K-means*“) и мешавина Гаусијана (енг. „*Mixture of Gaussian*“ скраћено „*Gaussian Mix*“).

4.5.1.1 К-средина и мини беч К-средина

Алгоритам К-средина проналази К кластера у подацима које представља помоћу К центроида тих кластера, од којих се свака добија упросечавањем елемената датог кластера [72]. Ова претпоставка чини алгоритам примењивим само на податке који се могу упросечавати, попут вектора. Под одређеним условима, постоје уопштења алгоритма и на другачије врсте података. Полазних К центроида се бира насумично (ако корисник зна нешто о структури својих података, могу бити и унапред дате), а потом се понављају следећи кораци:

- 1) Распоредити све инстанце у нове кластере тако што се свака инстанца придружи најближем центруиду.
- 2) Израчунати нове центроиде као просек инстанци које су им придружене.
- 3) Ови кораци се извршавају све док се центроиди мењају. Када су центроиди исти у две узастопне итерације, алгоритам се зауставља.

Идеја мини беч К-средине је у репрезентацији скупа података мањим подскупом података. Репрезентација модела је таква да се исти може врло добро поредити са стандардним алгоритмом К-средина без губљења на генерализацији, а са много мањим временом израчунавања.

Алгоритам користи мале подскупове произвољно изабраних података. За сваку инстанцу на основу претходно одређених центроида додељује се класа припадности. Затим се посредством градијентног спуста одређују нови центроиди кластера, што је неупоредиво брже у односу на класичан начин одређивања центроида. Детаљније о алгоритму се може наћи у [92].

4.5.1.2 Мешавина Гаусијана

Мешавина Гаусијана је напреднија верзија алгоритма К-средина иако се може показати да постоји директна веза између два модела. Модел се заснива на претпоставци да су подаци генерисани из расподеле која се представља у облику:

$$p(\mathbf{x}) = \sum_z p(z) \cdot p(\mathbf{x} | z) = \sum_{k=1}^K p(z = k) \cdot p(\mathbf{x} | z = k) \quad (4.45)$$

где је z случајна дискретна латентна променљива која је генерисана из мултиноминалне расподеле као:

$$p(z) = \prod_{k=1}^K \pi_k^{z_k} \quad (4.46)$$

при чему морају да буду задовољени услови:

$$0 \leq \pi_k \leq 1 \quad (4.47)$$

$$\sum_{k=1}^K \pi_k = 1 \quad (4.48)$$

Условна вероватноћа $p(\mathbf{x}|z)$ је нормално расподељена и може записати у следећем облику:

$$p(\mathbf{x}|z=k) = N(\mathbf{x} | \mu_k, \Sigma_k) \quad (4.49)$$

$$p(\mathbf{x}|z) = \prod_{k=1}^K N(\mathbf{x} | \mu_k, \Sigma_k)^{z_k} \quad (4.50)$$

Латентне променљиве престављају случајне променљиве које означавају припадност одређеном кластеру, односно условна вероватноћа $p(z=k|\mathbf{x})$ представља вероватноћу да дата инстанца припада кластеру k . С обзиром да је тешко оптимизовати логаритамску функцију веродостојности, користи се алгоритам ЕМ [93].

4.5.2 Алгоритам учења убрзане верзије Гаусових условних случајних поља за класификацију са одређивањем вероватноће припадања кластеру

У овом алгоритму се вредности парцијалних извода инстанци и одговарајућих излаза не одређују усвајањем вредности истих за одговарајући репрезент, већ се евалуирају вероватноће припадности кластера за сваку инстанцу и одговарајући излаз. Затим се парцијални извод за сваку инстанцу налази као отежана сума парцијалних извода репрезентата посредством вероватноће припадања одговарајућем кластеру. Односно парцијални извод једне инстанце и излаза представља линеарну комбинацију парцијалних извода репрезентата отежаних вероватноћом припадања кластера. С обзиром да је неопходно евалуирати вероватноћу припадности кластеру, коришћена је мешавина Гаусијана (овакав начин учења са мешавином Гаусијана је у наставку обележен скраћеницом Gaussian Mix prob).

5 Методологија за предвиђање индикатора саобраћаја - структурна предикција ретких зависних индикатора саобраћаја у великим мрежним системима

За разлику од већ постојећих приступа предвиђања индикатора саобраћаја у овој дисертацији је предложена методологија који се заснива на синхронном предвиђању саобраћајних индикатора на великој саобраћајној мрежи који су географски расути. Предложена методологија је у могућности да предвиди највероватније вредности саобраћајних индикатора (просечна брзина возила на ауто-путу или просечно време које је потребно да се скијаш спусти низ стазу), што представља средство на основу кога се може одредити појава гужве или оценити утицај индикатора на потенцијалне промене на саобраћајној инфраструктури.

Просечна брзина може представљати јасан индикатор постојања гужви, иако и други потенцијално недоступни атрибути могу утицати на њих, као што су стање на путевима, односно стање на стазама. Међутим, коришћење ових индикатора уз доступност доменског знања може да наведе менаџмент да лакше дође до одлука које се тичу планирања ресурса у будућности. На пример информација о очекиваним просечним брзинама на ауто-путу у неком периоду у будућности, може да има изузетно велики значај при доношењу одлука о потребном броју наплатних рампи које треба да буду отворене или повећању нивоа сигурности у циљу смањења просечне брзине возила. На сличан начин, коришћењем информације о очекиваном просечном времену које ће бити потребно скијашу да направи круг на некој стази може истовремено бити корисно менаџменту ски-центра који променом брзине ски-лифтова могу делимично контролисати проток скијаша, и скијашима (корисницима) који могу да се усмере ка оним локацијама где ће просечно време спуштања скијаша бити краће.

Развијена методологија за предвиђање очекиваних индикатора саобраћаја заснована је на посматрању података који се тичу саобраћаја уопште. Примећено је да индикатори саобраћаја на различитим местима која су просторно повезана нису међусобно независни. На пример, просечна брзина возила на једној наплатној станици може бити у великој мери корелисана са истом на некој блиској станици (просторно блиској или станици са сличним бројем наплатних рампи итд.). Слично, просечно време које је потребно да скијаш направи круг на некој стази може бити корелисано са истим на некој другој стази. С обзиром на зависност излаза, информације о начину на који су

исти зависни може бити искоришћена у циљу побољшања перформанси предикције [94]. Стога за разлику од стандардних модела и методологија који се користе у предвиђању новоразвијена методологија узима у обзир и информације садржане у везама које постоје између излаза.

Поред наведеног, у проблемима који се тичу предикције индикатора саобраћаја многи подаци су често ретки, односно не постоје никакве информације у одређеном временском периоду. На пример, у одређеним интервалима времена може се десити да нико није присутан на скијашкој стази или нико није прошао кроз одређену наплатну станицу. Ови подаци не говоре о томе да је просечна брзина возила у овом временском интервалу 0, нити говоре о томе да је просечно време које је скијашу потребно да направи круг на одређеној стази бесконачно или једнако 0. Стога било какво маркирање временских интервала у овом периоду представља само додавање додатног шума у податке што последично доводи и до лоших перформанси предиктора. У односу на претходно коришћене методологије, развијена методологија је у могућности да научи да препозна да ли на је на одређеном месту (стази или наплатној станици) у временском интервалу где се врши предвиђање прошао неко од возила. Уколико је неко прошао, на наведеном месту се предвиђа стварна вредност очекиваног индикатора саобраћаја.

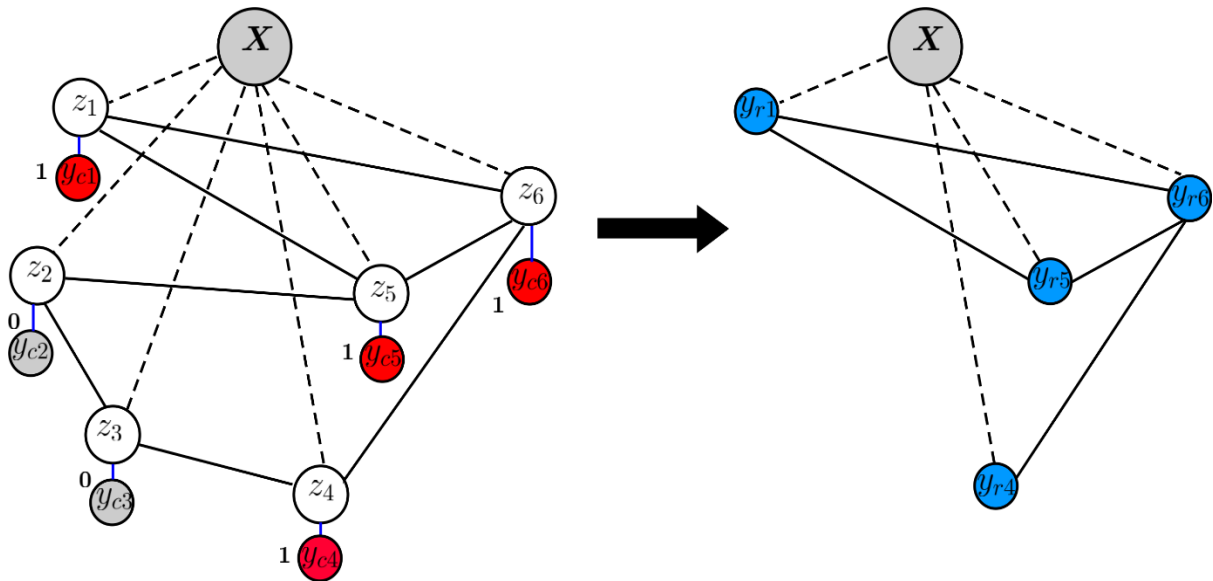
На основу овога може се закључити да развијена методологија за одређивање очекиваних индикатора саобраћаја узима у обзир структурну повезаност излаза и елиминацију места на којима нема никаквих информација. Методологија је заснована на комбинацији Гаусових условних случајних поља за класификацију и Гаусових условних случајних поља за регресију. Поред тога, развијена методологија користи сва три типа развијених модела Гаусових условних случајних поља за класификацију (Бајесовски модел, апроксимативну верзију Бајесовског модела и не-Бајесовски модел) евалуира њихове перформансе унакрсном валидацијом и бира онај који даје најбоље резултате као претходницу регресионог модела.

5.1 Комбинација регресионих и класификационих Гаусових условних случајних поља – методологија предвиђања индикатора саобраћаја

У случају када је циљ предвидети реалне излазе y , који су просторно и временски корелисани, помоћу задатог скупа атрибута x , Гаусова условна случајна поља за

регресију се могу успешно применити. Ипак у подацима који су ретки, директна примена Гаусових условних случајних поља за регресију подразумева замену вредности ништа са неком другом вредношћу (при предикцији просечних брзина возила вредност ништа се може заменити са 0). С обзиром на уношење шума (неисправних података) у скуп података за учење, перформансе предиктора неће бити толико сјајне, стога се и користи комбинација регресионих и класификационих алгоритама. Процес учења параметара регресионих и класификационих Гаусових условних случајних поља решен је симултаним учењем оба модела заједно. Прво су за све инстанце и излазе учени параметри Гаусових условних случајних поља за класификацију. Затим су за сваку инстанцу, само излази који имају вредност која се разликује од ништа, коришћени у процесу учења Гаусових условних случајних поља за регресију. На тај начин током тренирања само излази који немају вредност ништа служе за максимизацију функције веродостојности Гаусових условних случајних поља за регресију. Односно вредности параметара α и β нису учене излазима са вредностима ништа и на тај начин није додат додатни шум у модел. Важно је нагласити да током процеса учења граф стално мења чворове над којима учи параметре.

На сличан начин током закључивања, прво су класификовани излази на оне који ће имати неку вредности (класа 1) и оне који неће имати вредност (класа 0). Затим су излази са вредношћу 1 коришћени у циљу одређивања жељене реалне вредности. Закључивање у хибридном моделу приказано је на слици 12, док је на слици 13 дат псеудо код учења у хибридном моделу. На слици 13 ознакама $X_{ništa}$ и $Y_{ništa}$ приказан је скуп атрибута и излаза за класификацију, који се састоји од инстанци са излазима који имају вредност ништа. Аналогно томе ознакама X_{ima} и Y_{ima} приказан је скуп атрибута излаза за класификацију, који се састоји од инстанци које имају реалне вредности индикатора саобраћаја. Ознака Y_{ima_reg} односи се на вредности индикатора саобраћаја инстанци које имају реалну вредност истих. R_{reg} и R_{klas} представљају неструктурне предикторе за регресију и класификацију, док S_{reg} и S_{klas} представљају матрице сличности, коришћене при учењу Гаусових условних случајних поља за регресију и класификацију.



Слика 12 Шематски приказ модела заснованог на комбинацији регресионих и класификационих Гаусових условних случајних поља

Ulaz: Skup podataka (\mathbf{H}, \mathbf{Y}), Matrice sličnosti $\mathbf{S}_{\text{klas}}, \mathbf{S}_{\text{reg}}$
Izlaz: Regresioni (α, β), Klasifikacioni (θ)
 $\mathbf{R}_{\text{klas}}(\mathbf{X}) \leftarrow$ Nestruturni_klasifikacija_učenje (\mathbf{H}, \mathbf{Y})
 $(\alpha, \beta, \theta) \rightarrow$ inicijalizuj vrednosti
 $\theta \leftarrow$ GCRFBC_učenje($\mathbf{R}_{\text{klas}}(\mathbf{X}), \mathbf{Y}, \theta, \mathbf{S}_{\text{klas}}(\mathbf{X})$)
 $(\mathbf{Y}_{\text{ništa}}, \mathbf{H}_{\text{ništa}}, \mathbf{Y}_{\text{ima}}, \mathbf{H}_{\text{ima}}) \leftarrow$ GCRFBC_zaključivanje(\mathbf{H}, θ)
 $\mathbf{R}_{\text{reg}}(\mathbf{H}_{\text{ima}}) \leftarrow$ Nestruturni_regresija_učenje ($\mathbf{H}_{\text{ima}}, \mathbf{Y}_{\text{ima_reg}}$)
 $(\alpha, \beta) \leftarrow$ GCRF_učenje($\mathbf{R}_{\text{reg}}(\mathbf{H}_{\text{ima}}), \mathbf{Y}_{\text{ima}}, \alpha, \beta, \mathbf{S}_{\text{reg}}(\mathbf{H}_{\text{ima}})$)
 vrati (α, β, θ)

Слика 13 Учење хибридног модела

6 Методологија за одређивање оптималног броја активних канала у модулу

У овој дисертацији коришћењем комбинације рекурентних неуронских мрежа заједно са моделима теорије редова чекања и оптимизационим методама развијена је методологија за одређивање оптималног броја активних канала за опслуживање у модулу. Методологија се може успешно применити на било који реални проблем који се може представити моделима теорије редова чекања, где је потребно одредити оптималну политику промене управљачке величине (отварање и затварање канала) у неком периоду у будућности.

Методологија се заснива на учењу рекурентне неуронске мреже у циљу предвиђања промене интензитета долазака и интензитета опслуживања клијената. На основу одређених интензитета долазака применом модела теорије редова чекања може се одредити оптимална број канала у модулу који треба да буде активан у будућности у циљу минимизације укупних трошкова система опслуживања. Развијена методологија за одређивање оптималног броја активних канала у модулу представљена је у облику две верзије са Бајесовским и не-Бајесовским приступом.

6.1 Опис методологије

С обзиром да је циљ наћи оптималан број активних канала у модулу у неком периоду у будућности неопходно је описати развијену методологију. Методологија која је развијена у овом раду заснована је на комбинацији рекурентних неуронских мрежа, теорије редова чекања и метахеуристика за одређивање оптималне политике. Имајући у виду да је развијена методологија намењена системима за опслуживање иста се може применити на различитим доменима где су присутни овакви системи. Методологија има за циљ да избором оптималне политике управљања у неком периоду у будућности минимизује укупне трошкове система масовног опслуживања. Као што је речено, укупни трошкови се састоје од очекиваних трошкова чекања и очекиваних трошкова опслуживања. Избором облика функције укупних трошкова и дефинисањем истих директно се утиче на промену броја активних канала у модулу у будућности. Моделовањем промене стања система у будућности добијају се битне информације на основу којих се могу минимизовати трошкови рада истог.

Пре свега, уколико се посматра одређени систем опслуживања неопходно је одредити о каквом се стохастичком процесу ради. Под претпоставком да се систем може моделовати нехомогеним Марковљевим процесом потребно је одредити динамику промене параметара експоненцијалне расподеле времена долазака клијената и времена опслуживања. У најопштијем случају расподеле времена долазака клијената и опслуживања су променљиве током времена, односно параметар расподеле је временски зависан. Међутим у већини реалних система најчешће је само параметар експоненцијалне расподеле времена долазака клијената временски зависан, стога је потребно показати да је расподела времена опслуживања константна. Посматрањем система сакупљају се подаци о временима опслуживања и временима долазака клијената. Велики број података омогућава да се времена трајања опслуживања тестирају на припадност експоненцијалној расподели.

Без обзира на исход теста подаци који су добијени снимањем времена долазака клијената и времена опслуживања, користе се за учење параметара расподеле. Уколико се покаже да је време опслуживања расподела која је временски независна, врши се оцена параметра експоненцијалне расподеле из скупа података. За разлику од тога параметре расподеле који нису временски независни није могуће оценити. Стога је потребно научити рекурентну неуронску мрежу како се параметри ове расподеле мењају током времена за задате улазне атрибуте од којих су исти зависни.

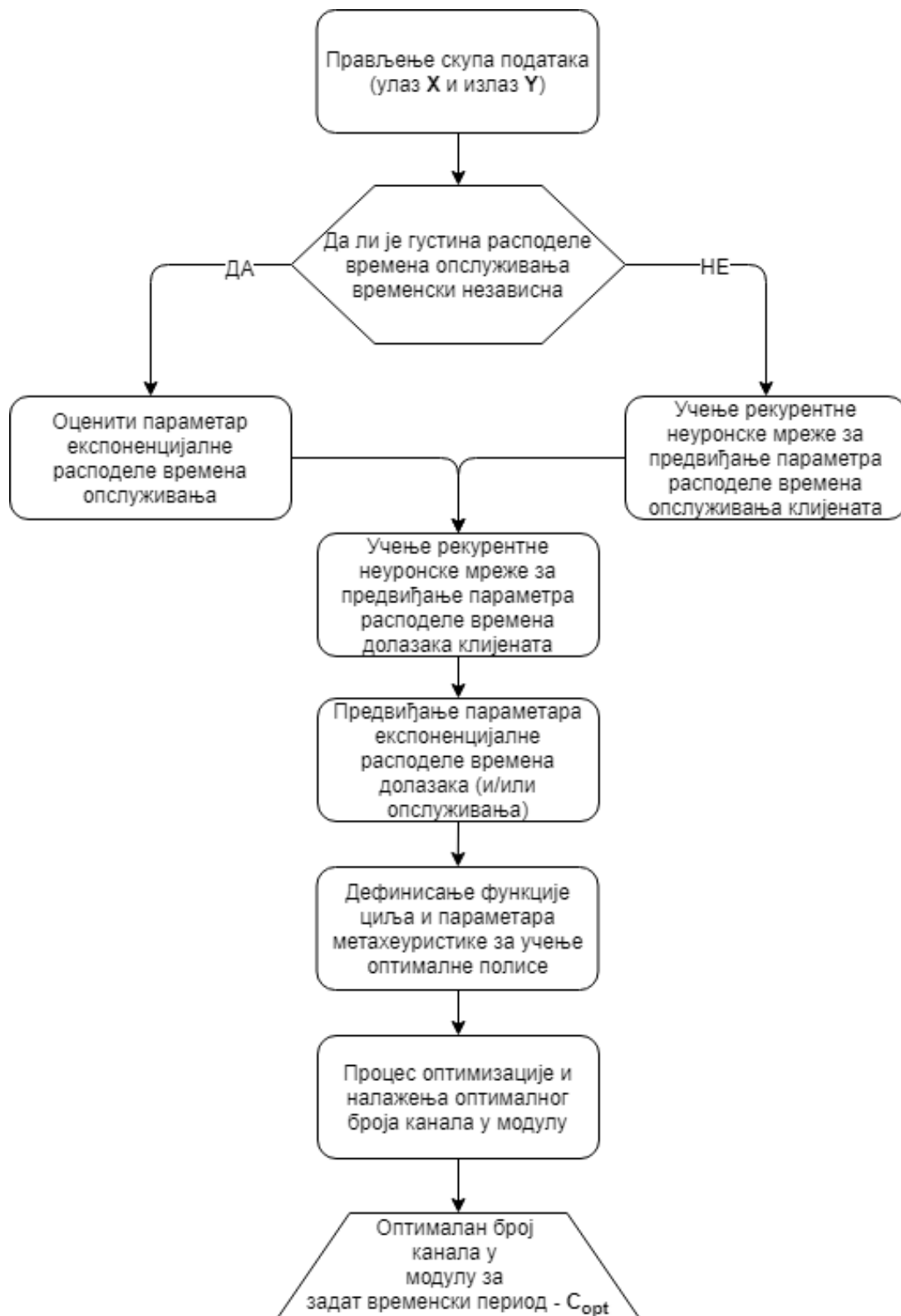
С обзиром да је за учење мреже неопходно обезбедити улазе и излазе у мрежу, потребно је извршити оцену параметара расподеле током времена. Један од начина на који се врши оцена параметара расподеле током времена је коришћењем померајућег прозора произвољно задате дужине. Циљ је задати дужину тако да не долази до наглих скоковитих промена параметара расподеле током кратких временских тренутака. Овај параметар треба мењати тако да што веродостојније симулира стварно стање система опслуживања. Скуп података који је доступан се дели на два дела: скуп за учење мреже и скуп за тестирање мреже. Поткрепљивањем мреже скупом података за учење, уче се тежине исте при чему је број ћелија, изабрана врста активационих ћелија и број латентних слојева произвољан параметар који је потребно проценити валидацијом резултата исте. Најчешће се користи већи број различитих параметара и усваја онај са најбољим резултатима добијеним на тест скупу података. С обзиром да није могуће научити мрежу да са апсолутном тачношћу предвиђа параметре експоненцијалне расподеле, могуће је оценити колико заиста мрежа греша при предвиђању параметара мреже. У том случају целокупан скуп података подељен је на три дела: скуп података за учење вредности

параметара експоненцијалне расподеле, скуп за предвиђање грешке коју мрежа прави при предвиђању параметара експоненцијалне расподеле и скуп за тестирање. Битно је поменути да мрежа као излаз из једне секвенце има вектор параметара експоненцијалне расподеле у неком временском периоду. У зависности од типа проблема број подинетравала у којима се врши оптимизација се може произвољно бирати.

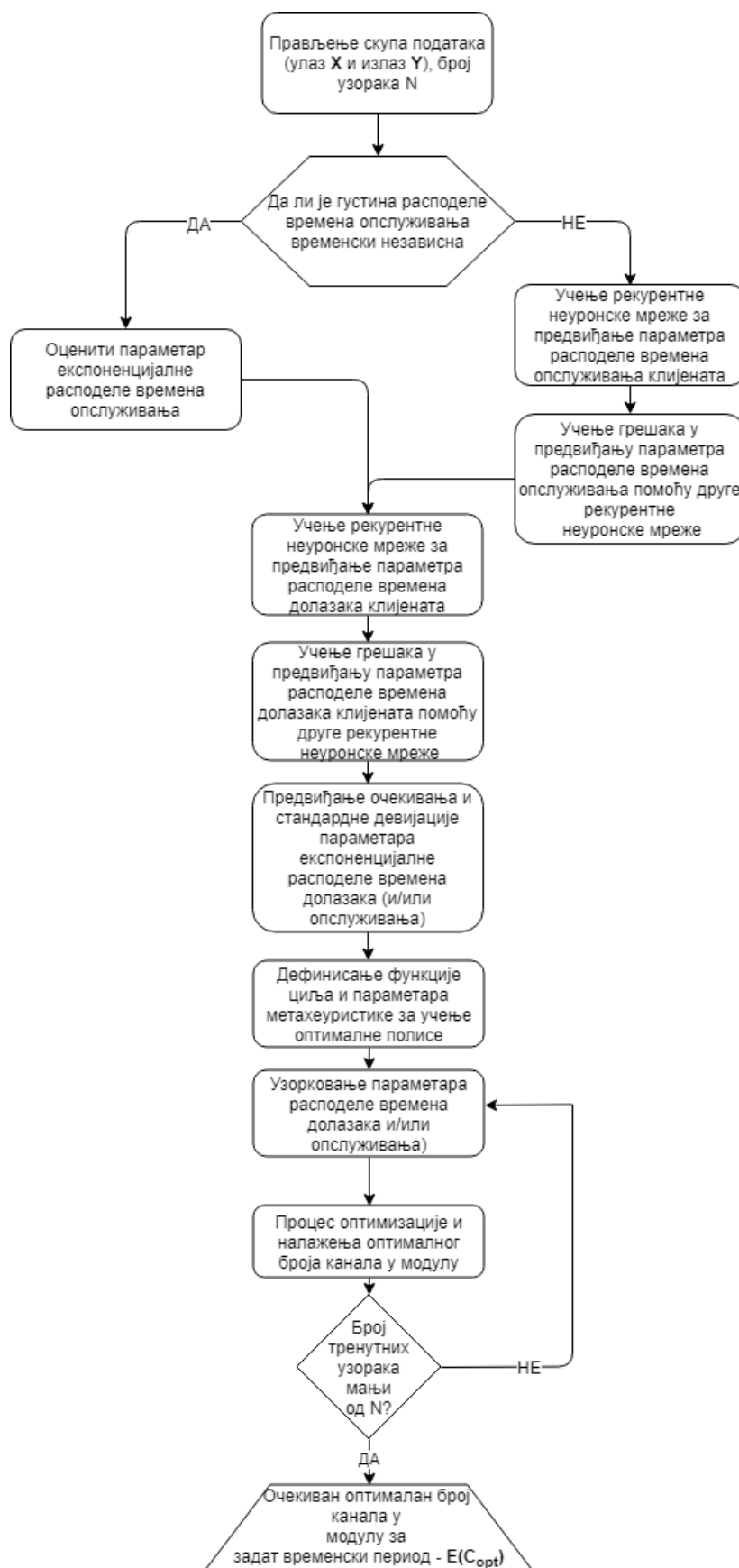
Након научене рекурентне мреже и изабраних мета параметара, односно архитектуре исте, коришћењем модела теорије редова чекања у комбинацији са еволутивним алгоритмима, одређује се оптимална полиса, односно оптималан број канала који треба да буде отворен у модулу. У зависности од типа проблема овај број канала се може дискретизовати на ситније или крупније временске интервале. За задате улазне атрибуте, мрежа предвиђа вредност излаза који се користи као вектор улаза у моделе теорије редова чекања. Променом броја отворених канала мењају се вредности транзиционе матрице, а самим тим се мењају и вероватноће стања система. Променом вероватноћа стања система мењају се и укупни трошкови.

Оптимизациони алгоритам има за циљ да нађе полису односно број канала који треба да буде отворен како би ови трошкови у посматраном периоду времена били минимални. У случају када се врши и оцена грешке мреже, претпоставља се да једна мрежа оцењује очекивање параметара експоненцијалне расподеле, док друга мрежа оцењује стандардну девијацију параметара експоненцијалне расподеле. Односно усваја се претпоставка да ове две оцене представљају параметре нормалне расподеле случајне променљиве која представља параметар експоненцијалне расподеле доласка возила. Потом да би се узела у обзир и грешка мреже, врши се узорковање случајних променљивих из нормалне расподеле описане вредностима добијеним за једну секвенцу из претходно научених мрежа. За сваки од узорака врши се процес оптимизације, односно проналажење оптималног броја активних канала за опслуживање, а затим се након тога налази очекивана вредност оптималне полисе. Ова два различита приступа у суштини представљају начин проналажења оптималног броја активних канала у модулу Бајесовском и не-Бајесовском методологијом. На сликама 14 и 15 приказни су алгоритми, развијене методологије, за оба описана приступа.

Детаљнији опис алгоритама и верификација њиховог рада за домен одређивања броја отворених наплатних рампи на наплатној станици представљен је у поглављу 7.



Слика 14 Алгоритам 1 за налажење оптималног броја активних канала у модулу у неком временском периоду у будућности



Слика 15 Алгоритам 2 за налажење оптималног броја активних канала у модулу у неком временском периоду у будућности

7 Примена развијених методологија на скупу података са ауто-путева

Ауто-пут Е70-Е75 је најважнији транзитни пут који пролази кроз Републику Србију. Укупна дужина деоница Ниш – Београд (Е70) и Београд – Адашевци (Е75) је 335 км. Ове две деонице спајају северо-запад и југ Републике Србије, повезујући Батровце (гранични прелаза на граници са Републиком Хрватском) и Прешево (гранични прелаз са Републиком Северном Македонијом).

У циљу валидације рада методологија приказаних у претходним поглављима, у овом делу дисертације исте ће бити примењене на проблем класификације гужви на наплатним станицама, предвиђање индикатора просечне брзине возила која долазе на наплатну станицу и одређивање оптималног броја отворених наплатних рампи на наплатној станици у неком периоду у будућности. Резултати добијени новоразвијеним методологијама поређени су са тренутно примењеним моделима који се користе при решавању наведених проблема.

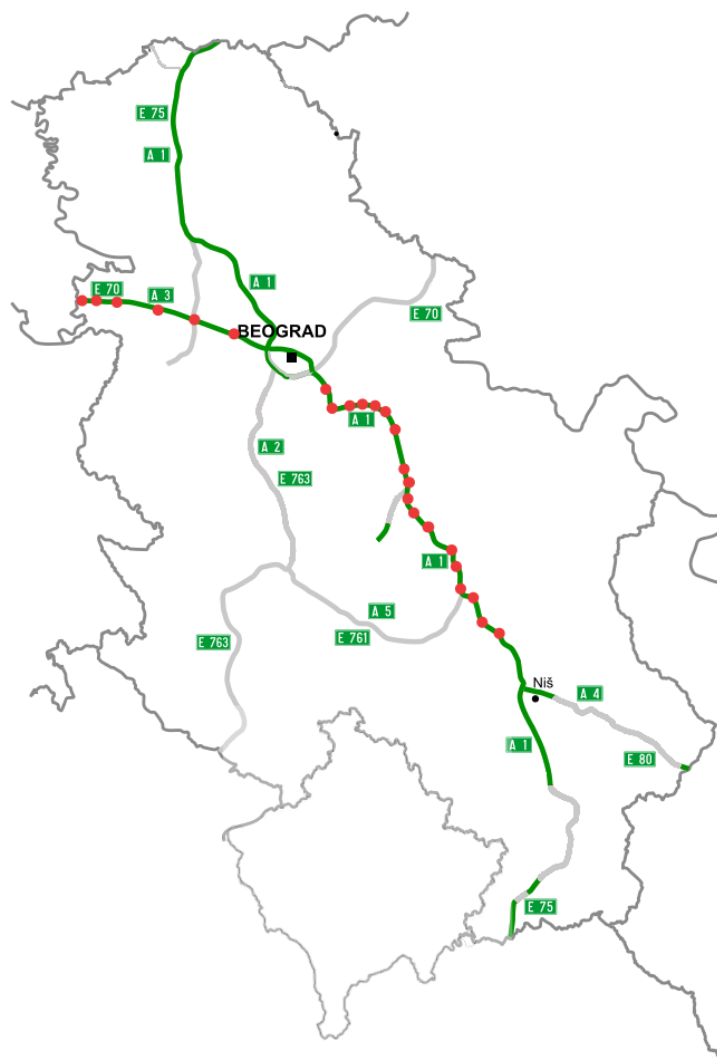
Методологија за класификацију гужви применом Гаусових условних случајних поља за класификацију приказана је у поглављу 7.1, док је у поглављу 7.2 приказана примена методологије за предвиђање индикатора брзине. Детаљан опис примене методологије за одређивање оптималног броја отворених наплатних рампи на наплатној станици у неком периоду у будућности приказан је у поглављу 7.3.

7.1 Примена методологије за класификацију гужви коришћењем Гаусових условних случајних поља за класификацију

У овом поглављу је приказана примена Гаусових условних случајних поља за класификацију на примеру предвиђања гужви на ауто-путевима. У циљу реалне имплементације модела неопходно је обезбедити инфраструктуру која омогућава предвиђање гужви у реалном времену. Ово се може постићи изградњом фибер – оптичких комуникационих система. Једном имплементиран систем може да помогне возачима на ауто-путевима на два начина:

- 1) Информише кориснике о потенцијалним гужвама у будућности и усмери их на друге наплатне станице.

- 2) Информације менаџмент компаније о највероватнијем стању на наплатним станицама, тако да исти буде у могућности да донесе одређене одлуке како би покушао да смањи трошкове или незадовољство корисника.



Слика 16 Приказ анализираних деоница Београд – Адашевци и Ниш – Београд.

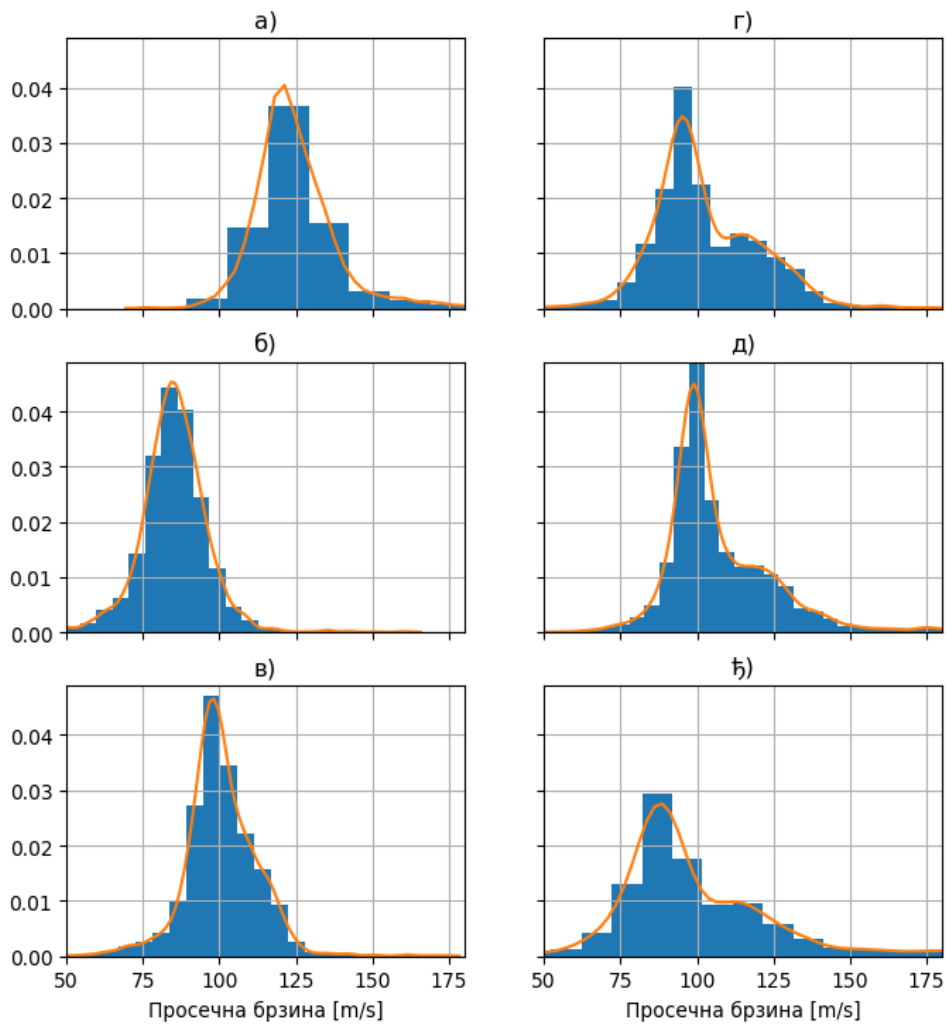
Црвеним тачкама означене су наплатне рампе на којима се предвиђала гужва

Један од највећих проблема на ауто-путу E70-E75 је честа појава гужви. Један од разлога зашто се гужве појављују јесте услед недовољног броја отворених наплатних рампи. Један од начина како да се реше свакодневни проблеми настанка гужви је предвиђањем да ли ће се у одређеном периоду у будућности десити гужва. У случају предикције да ће доћи до гужве, менаџмент има довољно времена да на правилан начин алоцира доступне ресурсе како би спречио настанак гужве или умањио последице појаве истих. Подаци коришћени у истраживању обухватају информације аутомобилских

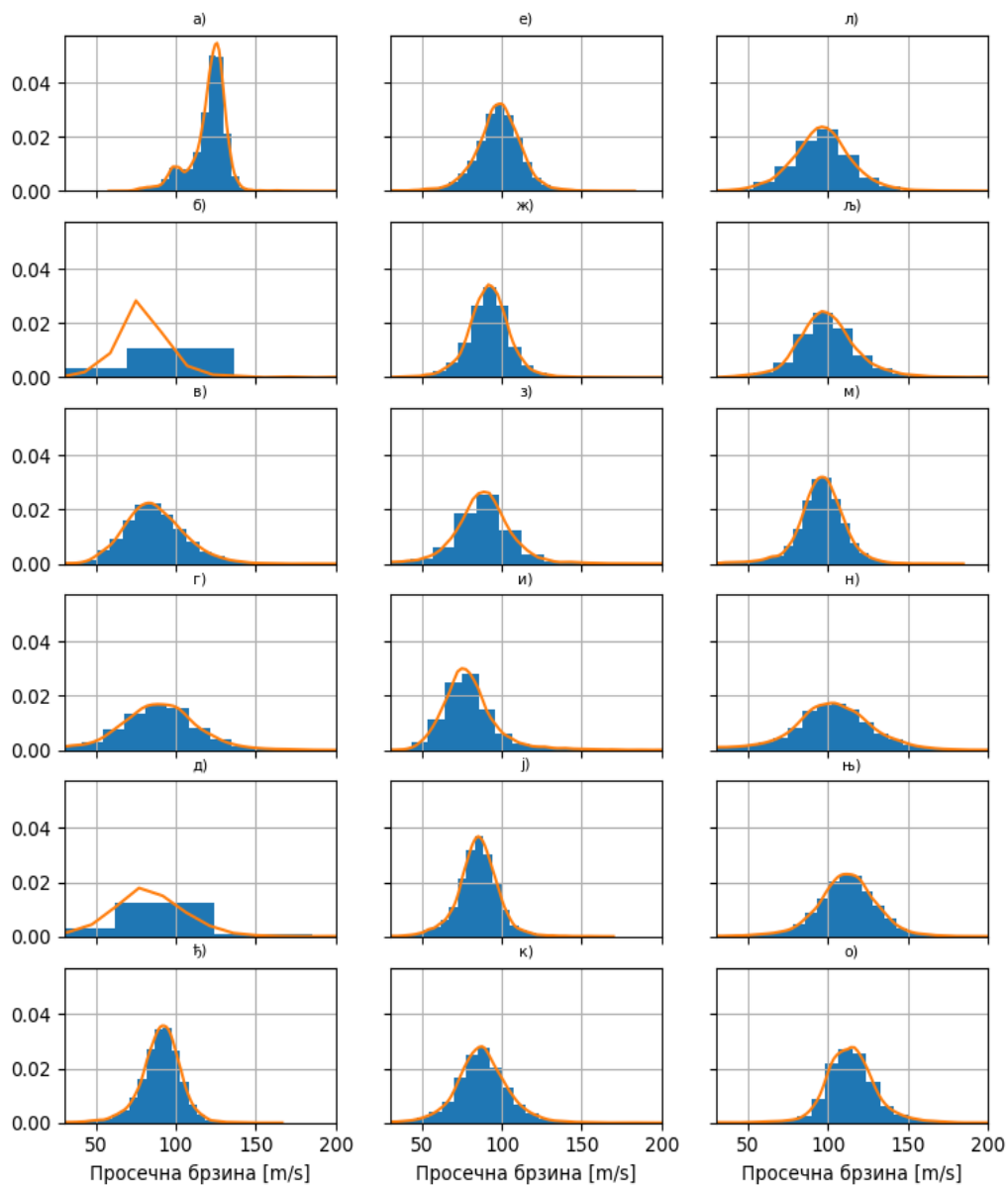
улазака и излазака у 2017. години. На две различите деонице су предвиђане гужве: Ниш – Београд и Београд – Адашевци. Деоница Београд – Адашевци је анализирана у периоду јануар 2017. године, док је деоница Ниш – Београд анализирана у периоду април – јули 2017. године. На слици 16 приказане су анализирани деонице, док су наплатне рампе на којима се предвиђају гужве приказане црвеним кружићима.

Укупан број сирових података из којих се правио скуп података за верификацију модела, на деоници Ниш – Београд је 5,132,918, док је на деоници Београд – Адашевци 487,767. Нови скуп података за деоницу Ниш - Београд је оформљен посматрајући временске прозоре у временским интервалима од 10 минута, док су на деоници Београд – Адашевци посматрани временски прозори у интервалима од 20 минута. Укупан број инстанци и наплатних станица на којима се врши предвиђање на деоници Београд – Адашевци је 50,964 и 6, док је на деоници Ниш – Београд 235,872 и 18.

Емиријска расподела средње брзине возила за свих шест наплатних станица на деоници Београд – Адашевци приказана је на слици 17, док је на слици 18 приказана емпиријска расподела средње брзине возила на деоници Ниш - Београд. Наранџастом бојом обележена је расподела добијена помоћу модела Кернеловог одређивање густина (енг. „*Kernel density estimation*“ скраћено „*KDE*“) са Гаусовим кернелом. На основу добијених резултата и посредством модуса расподеле брзине добијеном Кернеловим одређивањем густине, могу се све инстанце маркирати са 1 (гужва) или 0 (није гужва). Инстанца се маркира са 1 уколико је просечна брзина возила мања од модуса расподеле брзине, односно брзине кретања возила су мање од највероватније вредности којим се возила крећу што је јасан индикатор застоја у саобраћају. За разлику од овог случаја у случају када је просечна брзина већа од модуса расподеле брзине инстанца се обележава 0, односно нема гужве у саобраћају.



Слика 17 Емпиријска расподела средње брзине возила која долазе на наплатне рампе на деоници Београд-Адашевци



Слика 18 Емпиријска расподела средње брзине возила која долазе на наплатне рампе на деоници Ниш-Београд

Атрибути који су коришћени за предвиђање подељени су у три групе: дескриптивни, статистички и временски атрибути. Детаљан опис свих коришћених атрибута приказан је у табели 2.

Табела 2 Групе и опис атрибута коришћених при предвиђању гужви на ауто-путевима

Група атрибута	Опис атрибута
Дескриптивни	Број возила
	Број наплатних рампи кроз које су прошла возила
Статистички	Средњи брзина возила
	Максималан брзина возила
	Десети перцентил брзине возила
	Први квантил брзине возила
	Медијана брзине возила
	Други квантил брзине возила
	Деведесети перцентил брзине возила
Временски	Температура у °C
	Влажности ваздуха у %
	Тачка росе у °C
	Брзина ветра у m/s
	Притисак у mbar
	Покривеност облацима у %

Формиране су четири различите матрице сличности (односно коришћена су четири различита графа за дефинисање структуре између излаза) при предвиђању:

1. Корелациона матрица између излаза претрениране неуронске мреже.
2. Матрица међусобне информације (енг. „*Mutual information*“, скраћено MI).
3. Матрица са вредностима χ^2 статистика.
4. Матрица која се састоји од норме разлике између вектора састављених од историје маркера.

Коришћена су четири различита неструктурирана предиктора: риџ логистичка регресија (енг. „*Ridge logistic regression*“, скраћено Ridge-logistic), ласо логистичка регресија (енг. „*LASSO logistic regression*“, скраћено LASSO-logistic), неуронска мрежа за класификацију (енг. „*Neural network classifier*“, скраћено NN) и случајна шума (енг. „*Random forrest classifier*“, скраћено RF). 80% података је коришћено за учење неструктурираних предиктора, док је остатак коришћен за тренирање Гаусових условних случајних поља за класификацију. Регуларизациони параметар риџ и ласо логистичке регресије вариран је у распону од 0,1 до 10 и валидиран је посредством унакрсне валидације са 15 делова. Неуронска мрежа се састоји од 3 слоја са 30, 20 и 14 неурона по скривеном слоју респективно. Активационе функције коришћене у скривеном слоју су променљене линеарне јединице (енг. „*Rectified Linear Unit*“,

скраћено RELU). Функција циља се односи на минимизацију унакрсне ентропије (енг. „*Cross entropy*“), док је стохастички градијентни спуст коришћен као оптимизациони алгоритам у циљу одређивања тежина неуронске мреже (енг. „*Stochastic gradient descent*“). С обзиром да је неструктурни предиктор Гаусових условних случајних поља за класификацију број који пролази кроз сигмоидну функцију и пропуштен кроз сигмоиду, па самим тим представља вероватноћу да инстанца припада класи 1. Елиминисањем излазног слоја са сигмоидном функцијом активације добија се облик излаза из мреже који одговара облику неструктурних предиктора неопходног при учењу и закључивању модела Гаусових условних случајних поља за класификацију.

На сличан начин облик неструктурног предиктора потребног за тренирање Гаусових условних случајних поља добија се применом функције логита на вероватноће припадања класи добијене класификатором случајне шуме, риц и ласо логистичке регресије. Поређени су резултати Бајесовских Гаусових условних случајних поља са убрзањем, не-Бајесовских Гаусових условних случајних поља и брзом верзијом Бајесовских Гаусових условних случајних поља са различитим начином кластерованја.

Резултати класификације су валидирани посредством унакрсне валидације са 10 делова, и приказани су у табели 3 за деоницу Београд – Адашевци и табели 4 за деоницу Ниш – Београд. GCRFBCnb модел је постигао највиши ПИК скор на деоници Београд – Адашевци, док је на деоници Ниш – Београд ПИК скор GCRFBCb највећи. Примећује се да је ПИК скор Бајесовских Гаусових условних случајних поља на деоници Београд – Адашевци нижи од вредности добијених најбољим структурним предиктором. Такође, може се закључити да брза верзија Бајесовских Гаусових условних случајних поља не постиже боље резултате од неструктурних предиктора. Време учења GCRFBCnb модела је много мање од времена учења GCRFBCb и GCRFBCb-fast. Осетљивост перформанси модела у односу на број и врсту кластера код GCRFBCb-fast модела није значајна. Односно само приметну разлику прави GCRFBCb-fast модел са кластерованњем помоћу вероватноћа (*Gaussian Mix prob*). Брзина израчунавања GCRFBCb-fast модела варира и зависи од усвојеног начина кластерованја и усвојеног броја кластера. У случају обе деонице време израчунавања брзих верзија GCRFBCb-fast модела је мало за случај мањег броја кластера односно мањег броја варијационих параметара.

Табела 3 Перформансе предикције и време израчунавања модела за проблем предивђања гужви на деоници Београд - Адашевци

Београд - Адашевци					
Р. бр.	Модел	Метод кластеровања	Број кластера	ПИК	Време учења по једној итерацији унакрсне валидације у секундама
1	GCRFBCnb	-	-	0,968	13,685
2	GCRFBCnb	-	-	0,953	40,053
3	GCRFBCb-fast	K-means	5	0,931	126,570
4	GCRFBCb-fast	K-means	50	0,932	173,687
5	GCRFBCb-fast	K-means	150	0,933	246,148
6	GCRFBCb-fast	K-means	250	0,934	298,752
7	GCRFBCb-fast	Mini batch K-means	5	0,933	153,278
8	GCRFBCb-fast	Mini batch K-means	50	0,932	210,207
9	GCRFBCb-fast	Mini batch K-means	150	0,930	278,278
10	GCRFBCb-fast	Mini batch K-means	250	0,93	377,970
11	GCRFBCb-fast	Gaussian Mix	5	0,933	112,768
12	GCRFBCb-fast	Gaussian Mix	50	0,934	151,763
13	GCRFBCb-fast	Gaussian Mix	150	0,935	158,191
14	GCRFBCb-fast	Gaussian Mix	250	0,930	176,711
15	GCRFBCb-fast	Gaussian Mix prob	5	0,932	155,413
16	GCRFBCb-fast	Gaussian Mix prob	50	0,933	233,220
17	GCRFBCb-fast	Gaussian Mix prob	150	0,933	897,518
18	GCRFBCb-fast	Gaussian Mix prob	250	0,932	1750,971
19	Ridge-logistic	-	-	0,912	-
20	LASSO-logistic	-	-	0,931	-
21	NN	-	-	0,955	-
22	RF	-	-	0,956	-

Табела 4 Перформансе предикције и време израчунавања модела за проблем предивђања гужви на деоници Ниш-Београд

Ниш – Београд					
Р. бр.	Модел	Метод кластеровања	Број кластера	ПИК	Време учења по једној итерацији унакрсне валидације у секундама
1	GCRFBCnb	-	-	0,740	28,788
2	GCRFBCb	-	-	0,746	1318,097
3	GCRFBCb-fast	K-means	5	0,733	293,744
4	GCRFBCb-fast	K-means	50	0,735	433,445
5	GCRFBCb-fast	K-means	150	0,738	1306,514
6	GCRFBCb-fast	K-means	250	0,735	997,533
7	GCRFBCb-fast	Mini batch K-means	5	0,736	382,973
8	GCRFBCb-fast	Mini batch K-means	50	0,740	479,099
9	GCRFBCb-fast	Mini batch K-means	150	0,741	651,413
10	GCRFBCb-fast	Mini batch K-means	250	0,733	659,778
11	GCRFBCb-fast	Gaussian Mix	5	0,739	350,508
12	GCRFBCb-fast	Gaussian Mix	50	0,739	384,566
13	GCRFBCb-fast	Gaussian Mix	150	0,739	702,489
14	GCRFBCb-fast	Gaussian Mix	250	0,739	787,316
15	GCRFBCb-fast	Gaussian Mix prob	5	0,739	454,569
16	GCRFBCb-fast	Gaussian Mix prob	50	0,741	1245,749
17	GCRFBCb-fast	Gaussian Mix prob	150	0,741	4185,990
18	GCRFBCb-fast	Gaussian Mix prob	250	0,741	5217,043
19	Ridge-logistic	-	-	0,716	-
20	LASSO-logistic	-	-	0,716	-
21	NN	-	-	0,722	-
22	RF	-	-	0,740	-

7.1.1 Дискусија

На претходним примерима закључено је да методологија за класификацију гужви заснована на Гаусовским условним случајним пољима даје задовољавајуће резултате. Примењени новоразвијени модел за структурну класификацију успешно решава проблем предикције гужви и даје резултате који су бољи од тренутно коришћених модела. Анализом Бајесовског и не-Бајесовског приступа у учењу и закључвању показано је да оба модела могу бити имплементирана на примерима предивђања гужви.

Не-Бајесовска Гаусова условна случајна поља за класификацију у најбољем случају могу да остваре исте резултате предикције као Бајесовска Гаусова условна случајна поља за класификацију. Иако у реалним примерима услед отежаног тренирања (услед великог броја варијационих параметара) ово није случај. Брза верзија Бајесовских Гаусових условних случајних поља олакшава процес учења и показује боље резултате у односу на остале неструктурне моделе. У зависности од проблема који се решава иста може да буде боља и од не-Бајесовских Гаусових условних случајних поља за класификацију.

Неколико битних карактеристика Гаусових условних случајних поља за класификацију може се извести на основу спроведених експеримената:

1. Сви анализирани модели имају боље перформансе предикције у поређењу са неструктурним предикторима.
2. GCRFBCb модел има боље предиктивне перформансе, узимајући у обзир ПИК, у случајевима када је норма варијанси латентних променљивих висока. Када је ова норма ближа вредностима 0, оба модела имају добре предиктивне перформансе.
3. GCRFBCb-fast није превише осетљив на варијационе параметре што се показује и на реалним примерима где број кластера не утиче битно на крајње предиктивне перформансе класификатора.
4. Услед високих меморијских и рачунских трошкова учења и великог броја варијационих параметара оправдано је користити GCRFBCb-fast уместо GCRFBCb. Ово је оправдано посебно у случајевима када је присутан велики број инстанци и излаза.
5. Повећањем броја инстанци у скупу података за тренирање повећава се и број варијационих параметара, па самим тим тачност GCRFBCb модела расте и предности истог у односу на примену GCRFBCnb модела.

7.2 Примена комбинованог приступа регресије и класификације за предикцију просечне брзине возила на ауто-путевима

Као што је речено потребно је предвидети индикаторе саобраћаја односно просечну брзину возила на ауто-путу. С обзиром да велики број променљивих утиче на перформансе предикције, многе од њих се не могу директно увидети, стога се могу моделовати посредством веза између излаза. Индикатори саобраћаја у оквиру једног проблема моделују се као зависне величине. Поред тога саобраћајни подаци су у већини случајева ретки, што значи да постоји велики број инстанци са вредностима ништа за

излазе. У циљу елиминације ових променљивих, стандардни регресиони модели се не могу директно користити. Предложена методологија која је заснована на комбинацији регресионих и класификационих Гаусових условних случајних поља омогућава предикцију већег број зависних индикатора саобраћаја из ретких података. Предложена методологија је поређена са неколико базичних модела као што су риџ регресија (енг. „*Ridge regression*“, скраћено Ridge), ласо регресија (енг. „*LASSO regression*“, скраћено LASSO) [95], неуронска мрежа за регресију (енг. „*Neural network regression*“, скраћено NN-reg) [96] и случајна шума за регресију (енг. „*Random forrest regression*“, скраћено RF-reg) [97]. Додатно је комбиновани модел поређен са резултатима Гаусових условних случајних поља за регресију без претходно елиминисаних података са излазима који имају вредност ништа.

Предвиђање индикатора саобраћаја (просечне брзине возила) урађено је на истом скупу података на коме је приказана примена методологије за класификацију гужви. Периоди у коме су предвиђане брзине кретања возила, интервали временских прозора и укупан број записа коришћен при формирању скупа података на деоницама Ниш – Београд и Београд - Адашевци су исти као и при класификацији гужви. У оба случаја, индикатори су предвиђани 30 минута унапред. Атрибути који су коришћени при предвиђању су приказани у табели 2. У случајевима када није било возила у временском пролазу излазна вредност просечне брзине возила је ништа. Укупан број инстанци и број излаза на којима се вршило предвиђање индикатора брзине је исти за обе деонице као и у случају примене методологије за класификацију гужви. Укупан број излазних величина са вредностима ништа је 16,523 у случају деонице Београд – Адашевци и 30,938 у случају деонице Ниш – Београд.

Исто као у случају прогнозирања гужви на ауто-путевима, перформансе модела су оцењене унакрсном валидацијом са 10 делова. Укупно 80% података је коришћено за учење неструктурних предиктора, док је остатак коришћен за учење Гаусових условних случајних поља за класификацију. Као што је напоменуто четири различита неструктурна класификатора су коришћена: риџ логистичка регресија, ласо логистичка регресија, неуронска мрежа за класификацију и случајна шума. Регуларизациони параметар риџ и ласо логистичке регресије вариран је у распону од 0,1 до 10 и валидиран је посредством унакрсне валидације са 15 делова. Неуронска мрежа се састоји од 3 слоја са 30, 20 и 14 неурона по скривеном слоју редом. Активационе функције коришћене у скривеном слоју су RELU. Функција циља се односи на минимизацију унакрсне

ентропије, док је стохастички градијентни спуст коришћен као оптимизациони алгоритам у циљу одређивања тежина неуронске мреже.

Коришћени неструктурни регресори су: риц регресија, ласо регресија, неуронска мрежа за регресију и случајна шума за регресију. Слично као и у регресионом случају мета параметри риц и ласо регресије добијени су унакрсном валидацијом са 5 делова варирајући параметре у границама 0,1 до 10. неуронска мрежа за регресију користи 2 сакривена RELU слоја са 10 и 5 неурона по слоју. Максимални број епоха за тренирање мреже је 1200, а стохастички градијентни спуст са по 200 инстанци по итерацији је употребљен као оптимизациони алгоритам. Као регуларизациона процедура имплементирано је рано заустављање (енг. „*Early stopping*“). Средња апсолутна грешка коришћена је као функција која се минимизује током учења. Коришћене су исте матрице сличности које су описане у проблему класификације гужви.

У циљу валидације модела коришћене су следеће метрике [98]:

- 1) ПИК скор за евалуацију перформанси класификатора. Значење ПИК скорa објашњено је у поглављу 3.
- 2) Коефицијент детерминације (R^2) за евалуацију регресора. Коефицијент детерминације представља удео варијабилитета зависне променљиве која се може предвиђати независним променљивим величинама.

R^2 скор и укупно време учења Гаусових условних случајних поља за регресију, у случају када су вредности излаза ништа третиране као 0, приказане су у табели 5 за обе деонице. За разлику од тога вредности ПИК и R^2 скорa, заједно са временима учења комбинованог модела и одговарајућих неструктурних предиктора представљене су у табели 6 за деоницу Београд – Адашевци и табели 7 за деоницу Ниш – Београд.

Табела 5 Перформансе предикције и време израчунавања регресионих модела за проблем одређивања просечне брзине на деоницима Београд – Адашевци и Ниш - Београд

Регресија ауто-пут					
		Ниш - Београд		Београд - Адашевци	
Р. бр.	Модел	R²	Време учења по једној итерацији унакрсне валидације у секундама	R²	Време учења по једној итерацији унакрсне валидације у секундама
1	GCRF	0,509	52,494	0,947	17,832
2	Ridge	0,449	-	0,928	-
3	LASSO	0,452	-	0,928	-
4	NN-reg	-0,06	-	0,891	-
5	RF-reg	0,507	-	0,946	-

Табела 6 Перформансе предикције и време израчунавања комбиноване методологије за проблем одређивања просечне брзине на деоници Београд-Адашевци

Класификација Београд - Адашевци					
Р. бр.	Модел	Метод кластеровања	Број кластера	ПИК/R²	Време учења по једној итерацији унакрсне валидације у секундама
1	GCRFBCnb	-	-	0,648	9,137
2	GCRFBCb	-	-	0,999	251,723
3	GCRFBCb-fast	K-means	50	0,626	93,570
4	Ridge-logic	-	-	0,995	-
5	LASSO-logic	-	-	0,995	-
6	NN	-	-	0,997	-
7	RF	-	-	0,998	-
Регресија					
1	GCRF	-	-	0,981	10,524
2	Ridge	-	-	0,970	-
3	LASSO	-	-	0,971	-
4	NN-reg	-	-	0,961	-
5	RF-reg	-	-	0,980	-

Табела 7 Перформансе предикције и време израчунавања комбиноване методологије за проблем одређивања просечне брзине на деоници Ниш - Београд

Класификација Ниш - Београд					
Р. бр.	Модел	Метод кластеровања	Број кластера	ПИК/R²	Време учења по једној итерацији унакрсне валидације у секундама
1	GCRFBCnb	-	-	0,893	38,230
2	GCRFBCb	-	-	0,878	1255,530
3	GCRFBCb-fast	K-means	50	0,921	649,877
4	Ridge-logistic	-	-	0,853	-
5	LASSO-logistic	-	-	0,854	-
6	NN	-	-	0,832	-
7	RF	-	-	0,858	-
Регресија					
1	GCRF	-	-	0,859	62,948
2	Ridge	-	-	0,826	-
3	LASSO	-	-	0,828	-
4	NN-reg	-	-	0,796	-
5	RF-reg	-	-	0,834	-

У табели 5 може се видети да је структурни GCRF постигао боље резултате од свих осталих неструктурних регресора. Поред тога, примећује се да структурни регресор, без претходно елиминисаних излаза са вредностима ништа, постиже R² скорове који су задовољавајући на деоници Београд – Адашевци и не разликују се значајно од резултата добијених алгоритмом случајна шума. Међутим, R² скор на деоници Ниш – Београд је низак самим тим неопходно је елиминисати инстанце са излазима које имају вредност ништа. ПИК скор класификатора, који елиминише излазе са вредностима ништа, на деоници Београд – Адашевци је јако близак 1 за GCRFBCb модел, односно класификатор је скоро идеално научио да разликује оне инстанце са вредностима излаза ништа. GCRFBCb-fast и GCRFBCnb модел имају ниске вредности ПИК скорa и лошији су у односу на неструктурне класификаторе. ПИК скор класификатора на деоници Ниш – Београд показује високе вредности за GCRFBCb-fast модел са 50 кластера. Време учења овог алгоритма је дупло краће у поређењу са класичним GCRFBCb моделом. На овој деоници сви структурни класификатори

показују боље перформансе предикције у поређењу са неструктурним класификаторима. R^2 скор GCRF модела, са претходно елиминисаним излазима са вредностима ништа, је неупоредиво бољи у односу на модел где ове вредности нису елиминисане. Посебно је ова разлика приметна на деоници Ниш – Београд где комбиновани модел показује преко 30% боље ПИК скорове у поређењу са GCRF моделом без елиминисаних вредности ништа. На сличан начин се виде побољшања перформанси предикције на деоници Београд – Адашевци.

Закључује се да се коришћењем комбиноване методологије, у случају предикције индикатора брзине на више различитих повезаних излаза и у ситуацијама када су подаци ретки, знатно побољшавају перформансе предикције.

7.2.1 Дискусија

На претходним примерима јасно се види да предложена методологија за предвиђање индикатора саобраћаја, која се заснива на комбинацији класификационих и регресионих Гаусових условних случајних поља, може бити успешно примењена у решавању реалних проблема. Поред тога, новоразвијена методологија може бити успешно имплементирана у предвиђању индикатора брзине у реалном времену. Две битне опсервације се могу приметити:

- Индикатори саобраћаја се боље могу моделовати ако се предвиђају заједно као зависне променљиве, односно структура између излаза битно утиче на крајње перформансе предикције.
- Индикатори саобраћаја најчешће представљају ретке подате са великим бројем немаркираних инстанци.

Методологија која је базирана на комбинацији Гаусових условних случајних поља за класификацију и регресију има боље резултате предикције у односу на методологију која се односи на регресију, где су немаркиране инстанце обележене вредностима 0. GCRFBC алгоритам (Бајесовски и не-Бајесовски модел) је коришћен у предикцији излаза које имају вредност ништа, док је GCRF коришћен у предикцији континуланих излазних атрибута (индикатора саобраћаја), при чему су елиминисани сви излази без вредности. Приказана методологија у свим случајевима показује боље резултате у односу на резултате добијене неструктурним предикторима.

Неколико битних закључака се може извести:

1. Структура између излаза битно утиче на перформансе предикције и показано је да неструктурни предиктори у најбољем случају имају сличне перформансе предикције као структурни предиктори. Ово се односи само на случајеве када је структура излаза небитна, односно проблем регресије или класификације је лак.
2. Комбиновани приступ класификације и регресије који узима у обзир излазе са вредностима ништа, има већи R^2 скор у односу на регресиони приступ без претходне класификације.
3. Боље перформансе предикције су добијене на рачун већег рачунског трошка. С обзиром на то, неопходно је применити одговарајући приступ у проблемима који нису лаки и где нису добијене задовољавајуће перформансе класификатора.

Предложена методологија се успешно може генерализовати и на примере који имају сличну структуру података као у претходним случајевима. Закључује се да предложени приступ даје значајне доприносе у односу на анализирани алтернативе и да се може успешно користити при предикцији ретких зависних индикатора саобраћаја у оквиру великих мрежа.

7.3 Одређивање оптималног броја отворених наплатних рампи на наплатној станици у неком периоду у будућности

Методологија за одређивање оптималног броја активних канала у модулу је примењена на проблему одређивања оптималног броја наплатних рампи на наплатној станици Врчин. Наплатна станица Врчин на ауто-путу Е70 у смеру Ниш-Београд је једна од најпосећенијих наплатних станица са великим бројем свакодневних пролазака возила. Наплатна станица садржи 13 наплатних рампи које се користе за пролазак возила која немају уређај за електронску наплату путарине. Велики проблем на овој наплатној станици се везује за појаву гужви. Стога је одлуком менаџмента на наплатној станици примењена стратегија отварања великог броја наплатних рампи како би се евентуалне гужве избегле или смањиле. Примећено је да ова стратегија ипак није оптимална, с обзиром на то да у одређеним временским тренуцима наплатне рампе не опслужују никога, односно систем је потпуно празан.

Стога је неопходно применити одговарајућу оптималну полису, која треба менаџменту да каже који број наплатних рампи треба да буде отворен у будућности како би укупни трошкови који се тичу трошкова чекања и трошкова опслуживања били минимални. Наплатна станица је посматрана као систем опслуживања који се састоји од

више једноканалних система за опслуживање, односно наплатних рампи, са ограниченим редом. Разумно је претпоставити да долазеће возило на случајан начин (униформна расподела) бира која ће наплатна рампа да га опслужи. Пратећи методологију представљену у поглављу 4.5, прво је потребно формирати скуп података за утврђивање параметара расподеле. Након тога је потребно усвојити претпоставку да су времена долазака возила на наплатну рампу генерисана из експоненцијалне расподеле. Поред тога посредством померајућег временског прозора треба утврдити да ли је вероватноћа времена долазака возила временски зависна, односно да ли су параметри експоненцијалне расподеле променљиви током времена.

Коришћењем статистичких тестова и информационих критеријума утврђује се да ли је функција густине расподеле времена опслуживања експоненцијална расподела са константним параметром расподеле. Потом се за задати скуп података и дате улазне атрибуте уче параметри расподеле долазака возила на наплатну рампу. Након завршеног учења, тестирају се научени модели и бира онај који има најбоље предиктивне перформансе. Задавањем улазних атрибута, модел предвиђа излазне параметре који се могу даље користити за улаз у процесу проналажења оптималног броја канала у модулу. Битно је нагласити да излази мреже представљају интензитете долазака возила на наплатну станицу, док се интензитети долазака возила на наплатну рампу добијају када се интензитет долазака возила на наплатну станицу подели са бројем тренутно отворених наплатних рампи. Стога је потребно пронаћи онај број наплатних рампи који за задату функцију циља даје минималан збир очекиваних трошкова чекања и трошкова опслуживања.

Модел теорије редова чекања који је усвојен за моделовање система опслуживања (наплатне рампе) је једноканални модел са неограниченим редом. Имајући у виду да се ради о нехомогеном Маркољевом процесу, а да су предикције интензитета долазака добијене рекурентним неуронским мрежама, модел нема аналитичко решење. Односно, једини начин да се добију вероватноће стања система је коришћењем нумеричких метода. Стога је потребно увести апроксимацију да се систем наплатне рампе моделује као једноканални систем са ограниченим редом, где је број места у реду довољно велики тако да се вероватноће стања стварног и апроксимативног модела битно не разликују.

Детаљан опис дефинисања оптимизационог проблема заједно са објашњењем функције укупних трошкова приказано је у поглављу 7.3.1. Начин на који су припремљени подаци и тестиране хипотезе о промени параметара експоненцијалне

расподеле долазака клијената у времену приказан је у поглављу 7.3.2 и 7.3.3, док су резултати добијени тренирањем неуронских мрежа и резултати оптимизације приказани у поглављима 7.3.4 и 7.3.5. Дискусија добијених резултата је представљена у поглављу 7.3.6.

7.3.1 Проналажење оптималне полисе

Циљ примене методологије на примеру одређивања броја отворених наплатних рампи на наплатној станици у будућности се заснива на проналажењу оптималне полисе отварања наплатних рампи за унапред дефинисану функцију циља. Предвиђањем интензитета долазака возила на наплатну станицу могуће је предвидети појаву гужви за изабрану стратегију отварања наплатних рампи. Имајући у виду да се наплатна станица може моделовати као више једноканалних система опслуживања, где један канал представља једну наплатну рампу, могуће је дефинисати функцију циља на основу које се одређује оптималан број активних канала на истој. Напоменуто је да се претпоставља да је интензитет долазака возила на сваку рампу исти и једнак количнику интензитета долазака возила на наплатну станицу и тренутног броја отворених наплатних рампи. Шематски приказ расподеле интензитета долазака возила на наплатним рампама приказан је на слици 19.



Слика 19 Шематски приказ расподеле интензитета долазака возила на наплатним рампама

На основу претходно наведеног могуће је пре свега поставити функцију циља на основу које ће бити пронађен оптимални број наплатних рампи коришћењем алгоритма диференцијалне еволуције. С обзиром да се ради о нестационарном систему, односно систему који је променљив у времену укупни трошкови, које је потребно минимизовати, се представљају као:

$$E(UT) = \int_0^t (TC(t) + TO(t)) dt \quad (7.1)$$

где је $E(UT)$ - очекивани укуни трошак, $TC(t)$ - трошак чекања у тренутку t и $TO(t)$ трошак опслуживања у тренутку t . Трошак опслуживања у тренутку t се може записати као:

$$TO(t) = c(t) \cdot C_c \quad (7.2)$$

где је C_c - укупни трошак наплатне рампе изражен у EUR/min, а $c(t)$ је број отворених наплатних рампи у тренутку t . Укупни трошак наплатне рампе једнак је збиру следећих трошкова:

- Трошкова ел. енергије C_{el} који износе за просечну месечну потрошњу наплатне рампе 1200 kWh 0,0027 EUR/min.
- Трошкова радне снаге C_{rad} који се састоје од радника на наплатној станици који има бруто месечну плату од 800 EUR и радника на присмотри наплатне рампе који просечно прати 9 наплатних рампи и има исту бруто плату. Овај укупан трошак износи 0,0972 EUR/min,
- Трошкови грејања и хлађења $C_{gr,hl}$ који износе 50% од трошка ел. енергије односно 0,0013 EUR/min.
- Осталих трошкова C_{ostl} који се односе на одржавање и материјалне трошкове и износе 60% од збира трошкова радне снаге, ел. енергије и хлађења и грејања и износе 0,0607 EUR/min.

Укупни трошак наплатне рампе, добијени као збир свих набројаних трошкова $C_c = C_{el} + C_{rad} + C_{gr,hl} + C_{ostl}$, износи 0,162 EUR/min.

Трошкови чекања у тренутку t се могу представити као производ функције трошкова чекања по клијенту $h(t)$ у зависности од времена t и густине расподеле вероватноће времена проведеног у систему $f(t)$, и интензитета долазака клијената $\lambda(t)$ у времену t . Односно ови трошкови се могу записати као:

$$TC(t) = \lambda(t) \cdot h(t) \cdot f(t) \quad (7.3)$$

Уколико се функција трошкова чекања по клијенту напише као линеарна зависност у функције времена, онда следи да је:

$$h(t) = C_w \cdot t \quad (7.4)$$

где је C_w трошак чекања изражен у EUR/(min·клијенту). Овај трошак је рачунат као збир следећих трошкова:

- Трошкова горива C_{gor} за просечну потрошњу горива од 1,3 l/h износи 0,0325 EUR/min.
- Трошак изгубљеног времена C_{vr} изражен у просечно бруто заради која износи 700 EUR месечно по особи у возилу. Уколико је просечно 1,5 особа у возилу овај трошак износи 0,109 EUR/min.
- Остали трошкови чекања $C_{w,ostl}$ се тичу одржавања возила, трошкова потенцијалних губитака, трошкова изазваних умором и последицом истог. Ови трошкови износе 10% од збира претходно набројаних трошкова.

Укупни трошкови чекања по клијенту добијени као збир $C_w = C_{gor} + C_{vr} + C_{w,ostl}$ износе 0,155 EUR/(min·клијент).

Густина расподеле вероватноће времена се може исказати у следећем облику:

$$f(t) = \sum_{k=0}^m q_k \cdot \frac{\mu \cdot (\mu \cdot t)^k \cdot e^{-\mu t}}{k!} \quad (7.5)$$

где је m број места у реду, а q_k претставља вероватноћу долазећег клијента која се може изразити преко вероватноће стања p_k као:

$$q_k = \begin{cases} \frac{p_k}{1 - p_{k+m}} & k = 0, 1, 2, \dots, m \\ 0 & k = m + 1 \end{cases} \quad (7.6)$$

Вероватноћа стања система се израчунава из матричне диференцијалне једначине као:

$$\frac{dp}{dt} = Q^T p \quad (7.7)$$

где p представља вектор вероватноћа стања система $p = [p_0(t), p_1(t), \dots, p_{m+1}(t)]$, а q је вектор вероватноћа долазећег клијента $q = [q_0(t), q_1(t), \dots, q_m(t)]$.

Елементи транзиционе матрице Q се одређују као:

$$Q = \begin{bmatrix} -\frac{\lambda(t)}{c(t)} & \frac{\lambda(t)}{c(t)} & 0 & \dots & 0 \\ \mu & -(\mu + \frac{\lambda(t)}{c(t)}) & \frac{\lambda(t)}{c(t)} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mu & -\mu \end{bmatrix} \quad (7.8)$$

Функција циља очекиваних укупних трошкова се стога може представити у форми:

$$E(UT) = \int_0^t (c(t) \cdot C_c + \lambda(t) \cdot C_w \cdot t \cdot f(t)) dt \quad (7.9)$$

Имајући у виду да су познате вредности интензитета долазака возила на наплатну станицу $\lambda(t)$, интензитета опслуживања μ , укупног броја наплатних рампи на којима је могуће обавити опслуживање c_{\max} , могуће је одредити полису отварања и затварања канала као:

$$c_{opt}(t) = \arg \min_{c(t)} E(UT) = \arg \min_{c(t)} \int_0^t (c(t) \cdot C_c + \lambda(t) \cdot C_w \cdot t \cdot f(t)) dt \quad (7.10)$$

При чему важи услов да се број наплатних рампи креће у следећем интервалу:

$$1 \leq c(t) \leq c_{\max} \quad (7.11)$$

Уколико број канала дискретизујемо на подинтервале могуће је проблем решити коришћењем метахеуристике диференцијалне еволуције. Као критеријум конвергенције метахеуристика у овом примеру коришћен је максимални број итерација који је ограничен на 100 епоха, док је број јединки у једној популацији 150.

7.3.2 Припрема података

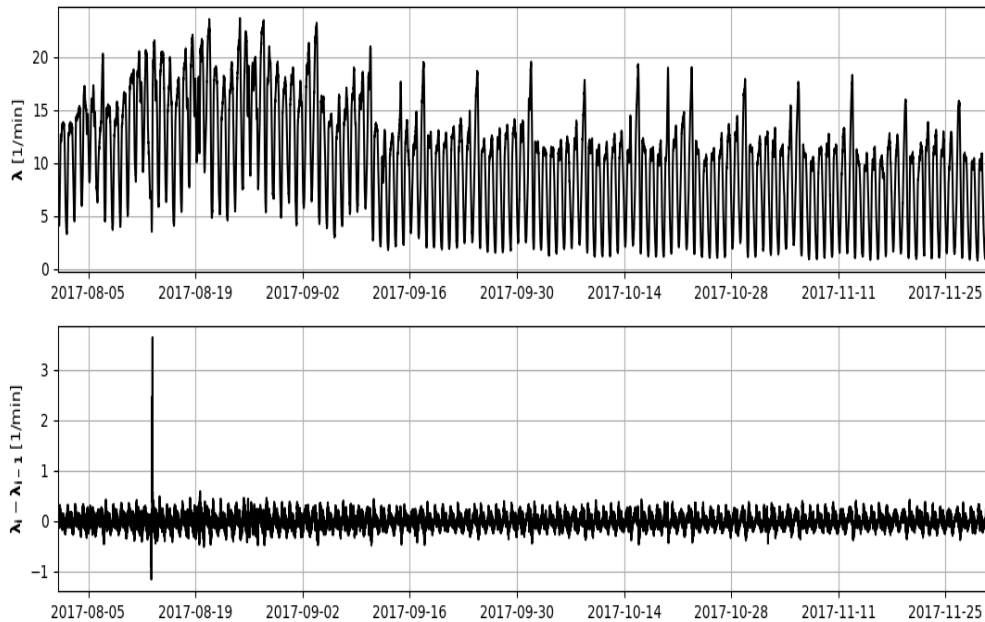
Подаци о проласцима возила кроз наплатну станицу Врчин у периоду од августа до септембра 2017. године коришћени су у циљу формирања скупа података на основу којег је учена и тестирана рекурентна мрежа. Такође, на овом скупу података верификован је предложени модел, односно поређен је укупан трошак праћењем стратегије која је добијена применом развијене методологије, са укупним трошком који је добијен тренутно усвојеним планом отварања и затварања наплатних рампи. Излазне вредности параметара су формиране коришћењем временског прозора са интервалом од 180 минута. Интензитети долазака возила су прогнозирани на сваких 5 минута. Самим тим, сви атрибути који су коришћени као улаз у рекурентну мрежу, представљају изведене статистике возила које су евалуиране на сваких 5 минута са истим временским

прозором посредством кога су тражени интензитети долазака. У табели 8 дат је преглед свих коришћених атрибута. Поред коришћења временских атрибута у Врчину, коришћени су и временски атрибути у Нишу. Опис времена је приказан као вектор са 14 елемената. С обзиром на то да се предвиђају интензитети долазака 300 минута унапред, излаз представља вектор са 60 елемената.

Поред коришћених атрибута као улаз у мрежу коришћени је вектор излаза из претходне секвенце стога је укупан број улазних атрибута који је формиран у скупу података 92. У случају предвиђања интензитета долазака возила, као улазни атрибут који је у наредној секвенци коришћен као улаз је вектор интензитета долазака возила, док је у случају учења грешке, као улазни атрибут у наредној секвенци коришћена претходна вредност грешке предвиђања. Укупан број инстанци коришћен за тестирање и тренирање мреже је 35013. Укупно 18040 инстанци коришћено је за тренирање рекурентне мреже која предвиђа очекивану вредност интензитета доласка, док је 9714 инстанци коришћено за тренирање грешке која се прави при прогнозирању стварне вредности интензитета долазака возила, а осталих 6939 инстанци је коришћено за предвиђање.

Табела 8 Коришћени атрибути при учењу рекурентне неуронске мреже

Групе атрибута	Опис атрибута
Дескриптивни	Број возила
	Време у часовима и минутима
Статистички	Средњи брзина возила
	Максималан брзина возила
	Десети перцентил брзине возила
	Први квантил брзине возила
	Медиана брзине возила
	Други квантил брзине возила
	Деведесети перцентил брзине возила
Временски	Температура у °C
	Влажности ваздуха у %
	Тачка росе у °C
	Брзина ветра у m/s
	Опис времена
	Видљивост %



Слика 20 Интензитети долазака возила и прва диференца интензитета долазака возила

Интензитети долазака возила добијени померањем временског прозора заједно са првом диференцом приказани су на слици 20. Прва диференца коришћена је како би проблем учења постао лакши за рекурентну неурноску мрежу. Из података види се да је присутна периодичност током дана то јест, интензитети долазака возила су ниски током ноћи, а високи током дана. Прва диференца делимично елиминише ову периодичност и олакшава мрежи процес учења (слично као и при нормализацији података).

7.3.3 Тестирање времена опслуживања клијената

Први наредни корак који прати прављење података, односи се на проверу да ли времена опслуживања клијената могу бити моделована експоненцијалном расподелом са временски независним параметром расподеле. Односно, потребно је проверити да ли се густина расподеле времена опслуживања мења током истог. Рационално је претпоставити да су времена опслуживања клијената генерисана из експоненцијалне расподеле, имајући у виду чињеницу да време опслуживања једног клијента ни на који начин не утиче на време опслуживања наредног. Односно, у конкретном случају код наплатних рампи, време које је потребно да се једном возачу (возилу) наплати путарина

и врати кусур није корелисано са временом опслуживања наредног клијента то јест возача. Такође, претпоставка која се тестира, а тиче се да је расподела којим се генеришу времена опслуживања непроменљива током времена, је оправдана у случају наплатних рампи. С обзиром на то да је посао који се обавља релативно једноставан, сви радници са потпуно сличном ефикасношћу наплаћују путарину, па се и расподела времена опслуживања не мења током времена.

Један од начина да се ово провери је коришћењем традиционлних статистичких тестова. Најчешће коришћени тестови који служе за проверу да ли су задати емпиријски подаци генерисани из неке теоријске расподеле су χ^2 тест и Колмогоров-Смирнов тест [99]. Ови тестови су намењени тестирању у случају када је доступан мањи број података и изузетно су осетљиви када је број података велики. Односно, примећује се да без обзира на то да ли су подаци заиста генерисани из експоненцијалне расподеле, у случају великог броја података оба теста увек одбацују могућност да су подаци генерисани из исте. Стога је један од начина да се ови тестови користе у случају великог броја података заснован на узимању случајног узорка малог обима.

Други начин провере да ли одговарајућа расподела долази из претпостављене теоријске расподеле је коришћењем информационих критеријума. Кулбак-Лајблер дивергенција (енг. „*Kullback–Leibler divergence*“ скраћено KL-divergence) представља меру која говори колико је једна функција вероватноће блиска другој функцији вероватноће [72]. Кулбак-Лајблер дивергенција се дефинише за дискретне расподеле као:

$$D_{KL}(P \parallel Q) = -\sum_{x \in \mathcal{X}} P(x) \log \left(\frac{Q(x)}{P(x)} \right) \quad (7.12)$$

где је $D_{KL}(P \parallel Q)$ Кулбак-Лајблер дивергенција између вероватноћа $P(x)$ и $Q(x)$, док је \mathcal{X} домен вредности x . Важно је нагласити да вредности Кулбак-Лајблер дивергенције нису симетричне, односно $D_{KL}(P \parallel Q) \neq D_{KL}(Q \parallel P)$. У циљу стварања симетричне мере могуће је дефинисати следећу меру:

$$f(P, Q) = 0,5 \cdot D_{KL}(P \parallel Q) + 0,5 \cdot D_{KL}(Q \parallel P) \quad (7.13)$$

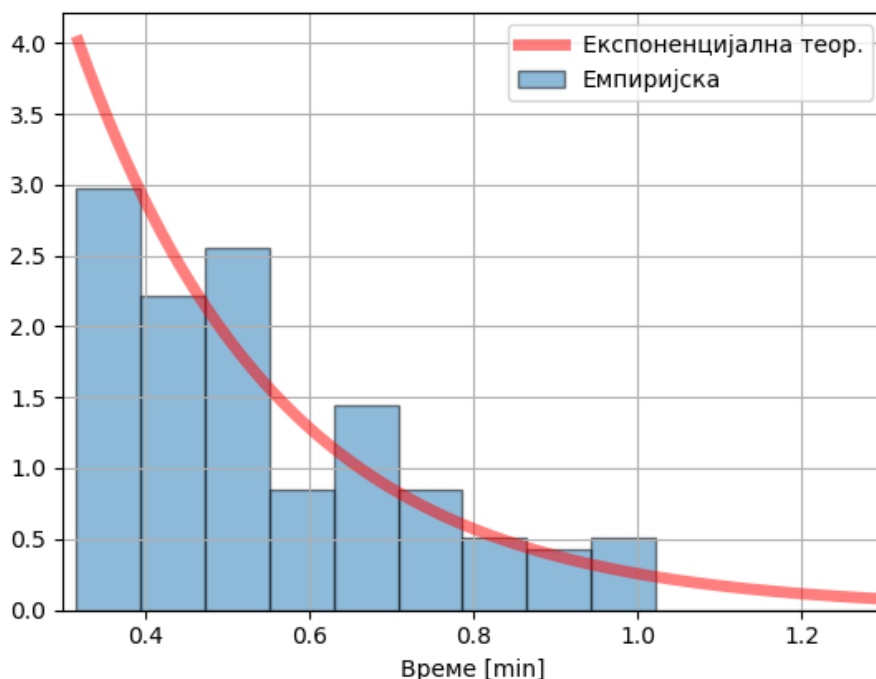
У случају када су две расподеле потпуно исте $D_{KL}(P \parallel Q) = 0$, док је у случају када се расподеле разликују $D_{KL}(P \parallel Q) > 0$. У циљу ограничавања вредности Кулбак-Лајблер

дивергенције на домен $[0,1]$ (0 када су расподеле исте 1 када се у потпуности разликују) могуће је извршити следећу нелинеарну трансформацију Кулбак-Лајблер дивергенције:

$$D_{KL_norm}(P \parallel Q) = 1 - \exp^{-D_{KL}(P \parallel Q)} \quad (7.14)$$

У овом поглављу, с обзиром да се врши поређење емпиријских података са теоријском расподелом, теоријска расподела је дискретизована како би могао да се успешно користи Кулбак-Лајблер информациони критеријум.

На основу прикупљених података за случај тестирања да ли време опслуживања клијената на наплатној рампи потиче из померене експоненцијалне расподеле, прво се врши оцена параметара исте. Након извршене оцене параметара померене експоненцијалне расподеле, на случајни начин се бира 100 узорака из расподеле и тестира статистичким тестовима. На слици 21 приказан је хистограм емпиријских података и померена експоненцијална расподела.



Слика 21 Хистограм емпиријских података и померена експоненцијална расподела

Табела 9 Вредности Кулбак-Лајблер дивергенције и тестова

Мера	р вредност	Вредност информационог критеријума
Колмогоров-Смирнов два узорка	0,312	-
Колмогоров-Смирнов један узорак	0,324	-
χ^2 тест	0,059	-
$D_{KL}(P \parallel Q)$	-	0,057
$D_{KL_{norm}}(P \parallel Q)$	-	0,055
$D_{KL}(Q \parallel P)$	-	0,056
$D_{KL_{norm}}(Q \parallel P)$	-	0,054
$f(P, Q)$	-	0,056
$f_{norm}(P, Q)$	-	0,055

Вредности Кулбак-Лајблер дивергенције и статистике добијене χ^2 и Колмогоров-Смирнов тестовима приказане су у табели 9. Из табеле 9 јасно се види да су р вредности више од критичне р вредности која износи 0,05, самим тим може се прихватити хипотеза да су емпиријски подаци генерисани из померене експоненцијалне расподеле. Стога, у даљем току рада није потребно тренирати рекурентну мрежу да оцењује параметре експоненцијалне расподеле времена опслуживања зато што је овај параметар константан и није временски зависан.

7.3.4 Учење и верификација рекурентних неуронских мрежа

На основу припремљеног скупа података различите неуронске мреже са различитим мета параметрима су учене у циљу предвиђања интензитета долазака возила. У свим случајевима вариран је број слојева и број ћелија по слоју. Коришћено је 8, 16, 32 и 64 ћелија по слоју, док се број слојева мења у распону од 1 до 3. У свим архитектурама укупан број секвенци у оквиру једне инстанце је 320. Средња квадратна грешка је коришћена као функција коју је потребно минимизовати при учењу рекурентне неуронске мреже. Након учења мреже су валидиране на тест скупу података. Укупно је научено 96 различитих модела по 24 модела за RNN, GRU, LSTM и LSTM-In архитектуру. Почетни параметар учења који је коришћен при тренирању је 0.1, наредних 7 епоха овај параметар је коришћен за учење, а након тога је исти опадао са степеном 0.92. Број епоха тренирања свих архитектура је 100.

Резултати перформанси предикције мрежа приказани су у табелама 10, 11, 12 и 13. Из табела се може видети да унутрашња стандардизација слојева LSTM мреже

значајно доприноси побољшању резултата предикције односно LSTM-ln мреже генерално имају боље вредности R^2 скорa у односу на остале архитектуре. Са становништва регуларизације примећује се да мреже код којих се одбацују ћелије (енг. „*dropout regularization*“) немају значајно бољи резултат у односу на архитектуре код којих се ћелије не одбацују. Усвојено је да фактор избацивања ћелија износи 0,8 односно 80%. Такође се види да у случају када је број секвенци 320, RNN мреже могу да ухвате дугорочну зависност између скривених слојева, иако се не користе капије. Од свих приказаних типова архитектура најслабији резултати су добијени за GRU мрежу.

LSTM-ln мрежа са 3 слоја, 32 LSTM ћелије по слоју и регуларизацијом постигнутом техником избацивањем ћелија, приказује најбоље перформансе предикције на тест скупу података. Односно R^2 скор ове мреже је 0,879, средња квадратна грешка је 2,012, док је средња апсолутна грешка 0,964. На основу овог запажања у алгоритму 1 методологије, ова мрежа је коришћена за предвиђање интензитета долазака возила, а у алгоритму 2 иста је коришћена за предвиђање очекивања интензитета долазака возила на наплатну станицу.

Табела 10 Перформансе предикције RNN мреже

R^2 скор са избацивањем ћелија					R^2 скор без избацивања ћелија			
Број слојева	Број ћелија по слоју				Број ћелија по слоју			
	8	16	32	64	8	16	32	64
1	0,484	0,453	0,555	0,646	0,328	0,382	0,523	0,551
2	0,446	0,501	0,385	0,383	0,596	-0,389	0,427	0,544
3	0,382	0,275	0,28	-0,193	0,626	-0,151	0,239	0,133
Средње квадратно одступање са избацивањем ћелија					Средње квадратно одступање без избацивања ћелија			
Број слојева	Број ћелија по слоју				Број ћелија по слоју			
	8	16	32	64	8	16	32	64
1	9,017	9,563	7,779	6,18	11,743	10,794	8,333	7,843
2	9,691	8,729	10,743	10,79	7,051	24,306	10,009	7,962
3	10,8	12,679	17,003	38,363	6,541	20,119	13,307	15,161
Средња апсолутна грешка са избацивањем ћелија					Средња апсолутна грешка без избацивања ћелија			
Број слојева	Број ћелија по слоју				Број ћелија по слоју			
	8	16	32	64	8	16	32	64
1	2,278	2,368	2,128	1,918	2,627	2,455	2,161	2,135
2	2,328	2,253	2,463	2,484	1,963	3,733	2,369	2,121
3	2,447	2,604	3,093	5,021	1,919	3,432	2,681	2,959

Табела 11 Перформансе предикције GRU мреже

R² скор са избацавањем ћелија					R² скор без избацавања ћелија			
Број слојева	Број GRU ћелија по слоју				Број GRU ћелија по слоју			
	8	16	32	64	8	16	32	64
1	0,594	0,644	0,413	0,212	-0,534	-2,423	-5,596	-0,584
2	0,084	-8,209	-14,333	-1,254	-7,663	0,157	-0,288	0,412
3	0,213	0,142	-3,623	0,231	-1,248	0,324	0,147	0,212
Средње квадратно одступање са избацавањем ћелија					Средње квадратно одступање без избацавања ћелија			
Број слојева	Број GRU ћелија по слоју				Број GRU ћелија по слоју			
	8	16	32	64	8	16	32	32
1	6,752	5,923	9,475	12,342	25,577	57,077	110,002	21,999
2	15,263	153,531	255,652	65,232	29,395	14,049	21,481	9,444
3	12,587	10,325	72,244	12,674	17,325	9,876	10,021	12,122
Средња апсолутна грешка са избацавањем ћелија					Средња апсолутна грешка без избацавања ћелија			
Број слојева	Број GRU ћелија по слоју				Број GRU ћелија по слоју			
	8	16	32	64	8	16	32	64
1	1,799	1,697	2,088	3,665	3,549	4,567	6,274	42,222
2	2,742	8,061	10,037	7,231	3,575	2,446	2,689	2,001
3	3,668	2,344	4,564	3,894	5,642	2,677	2,384	3,665

Табела 12 Перформансе предикције LSTM мреже

R² скор са избацавањем ћелија					R² скор без избацавања ћелија			
Број слојева	Број LSTM ћелија по слоју				Број LSTM ћелија по слоју			
	8	16	32	64	8	16	32	64
1	-0,095	-0,1	-0,377	-4,205	0,14	-0,319	-4,361	-1,031
2	0,267	0,218	-1,97	-2,253	-0,936	0,071	0,516	-2,278
3	0,681	0,633	0,391	0,483	0,632	0,578	0,619	-0,004
Средње квадратно одступање са избацавањем ћелија					Средње квадратно одступање без избацавања ћелија			
Број слојева	Број LSTM ћелија по слоју				Број LSTM ћелија по слоју			
	8	16	32	64	8	16	32	64
1	18,264	18,353	79,58	86,784	14,331	21,991	89,377	188,64
2	12,222	13,024	49,628	17,048	32,284	15,486	8,058	54,657
3	5,317	6,108	10,142	8,613	6,124	7,032	6,347	16,747
Средња апсолутна грешка са избацавањем ћелија					Средња апсолутна грешка без избацавања ћелија			
Број слојева	Број LSTM ћелија по слоју				Број LSTM ћелија по слоју			
	8	16	32	64	8	16	32	64
1	2,74	2,722	4,986	4,759	2,37	2,887	4,57	5,552
2	2,028	2,055	2,561	2,508	2,614	2,148	1,812	2,977
3	1,448	1,598	1,995	1,98	1,542	1,802	1,645	2,282

Табела 13 Перформансе предикције LSTM-ln мреже

R² скор са избацивањем ћелија					R² скор без избацивања ћелија			
Број слојева	Број LSTM ћелија по слоју				Број LSTM ћелија по слоју			
	8	16	32	64	8	16	32	64
1	0,726	0,703	0,624	0,692	0,679	0,517	0,431	0,77
2	0,812	0,834	0,803	0,817	0,855	0,803	0,868	0,856
3	0,781	0,772	0,879	0,653	0,824	0,849	0,843	0,849
Средње квадратно одступање са избацивањем ћелија					Средње квадратно одступање без избацивања ћелија			
Број слојева	Број LSTM ћелија по слоју				Број LSTM ћелија по слоју			
	8	16	32	64	8	16	32	64
1	4,558	4,942	6,253	5,125	5,348	8,04	9,481	3,822
2	3,119	2,761	3,27	3,044	2,408	3,27	2,194	2,391
3	3,698	3,797	2,012	5,784	2,931	2,515	2,609	2,503
Средња апсолутна грешка са избацивањем ћелија					Средња апсолутна грешка без избацивања ћелија			
Број слојева	Број LSTM ћелија по слоју				Број LSTM ћелија по слоју			
	8	16	32	64	8	16	32	64
1	1,412	1,449	1,576	1,483	1,439	1,763	1,645	1,347
2	1,204	1,149	1,191	1,184	1,025	1,162	1,013	1,029
3	1,338	1,352	0,964	1,524	1,182	1,053	1,073	1,035

Имајући у виду да је у алгоритму 2 методологије потребно предвиђати грешку коју мрежа прави при предвиђању, коришћена је друга рекурентна мрежа за учење грешке. С обзиром на резултате која је постигла LSTM-ln мрежа, иста је коришћена при предвиђању грешке предвиђања интензитета долазака. У табели 14 приказане су перформансе предикције LSTM-ln мреже за предвиђање грешака при предикцији интензитета долазака возила са 8, 16, 32 и 64 ћелије по слоју и 1, 2 и 3 слоја. Резултати показују да најбоље перформансе предикције има мрежа са 3 слоја и 64 ћелије по сваком слоју и имплементираном регуларизацијом са избацивањем ћелија (фактор избацивања 0,8).

На основу приказаних резултата у даљем процесу оптимизације у оквиру алгоритма 1 развијене методологије, коришћена је LSTM-ln рекурентна мрежа са 3 слоја, 32 LSTM ћелије по слоју и регуларизацијом постигнутом техником избацивањем ћелија. У оквиру алгоритма 2 иста мрежа је коришћена за предвиђање очекивања нормалне расподеле, док је LSTM-ln мрежа са 3 слоја и 54 ћелије по слоју коришћена при предвиђању стандардне девијације нормалне расподеле интензитета долазака возила.

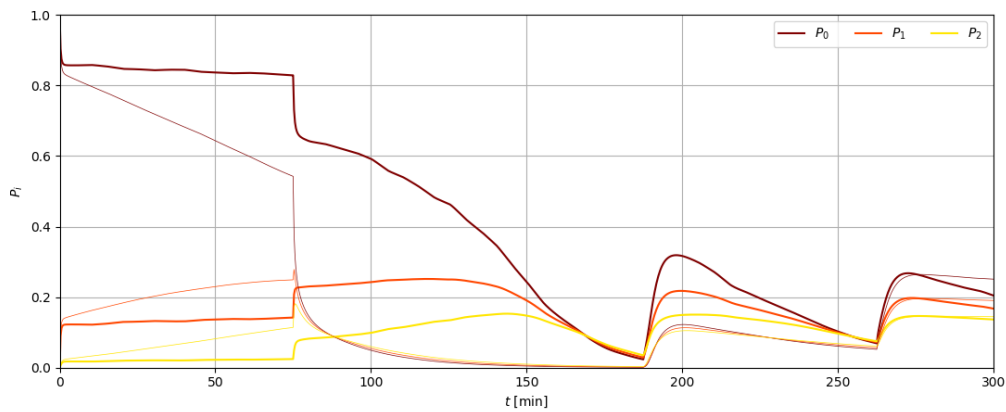
Табела 14 Перформансе предикције LSTM-In мреже за предвиђање грешке предикције параметара расподеле

R² скор са избацивањем ћелија					R² скор без избацивања ћелија			
Број слојева	Број LSTM ћелија по слоју				Број LSTM ћелија по слоју			
	8	16	32	64	8	16	32	64
1	0.614	0.629	0.642	0.659	0.594	0.550	0.608	0.634
2	0.635	0.673	0.650	0.662	0.642	0.432	0.596	0.594
3	0.634	0.691	0.674	0.692	0.684	0.651	0.619	0.675
Средње квадратно одступање са избацивањем ћелија					Средње квадратно одступање без избацивања ћелија			
Број слојева	Број LSTM ћелија по слоју				Број LSTM ћелија по слоју			
	8	16	32	64	8	16	32	64
1	0.870	0.854	0.839	0.821	0.892	0.944	0.877	0.849
2	0.816	0.806	0.832	0.817	0.840	1.020	0.893	0.892
3	0.847	0.785	0.802	0.784	0.792	0.829	0.865	0.802
Средња апсолутна грешка са избацивањем ћелија					Средња апсолутна грешка без избацивања ћелија			
Број слојева	Број LSTM ћелија по слоју				Број LSTM ћелија по слоју			
	8	16	32	64	8	16	32	64
1	0.712	0.705	0.695	0.690	0.716	0.736	0.716	0.699
2	0.692	0.684	0.691	0.688	0.711	0.755	0.701	0.705
3	0.705	0.682	0.688	0.676	0.676	0.697	0.704	0.679

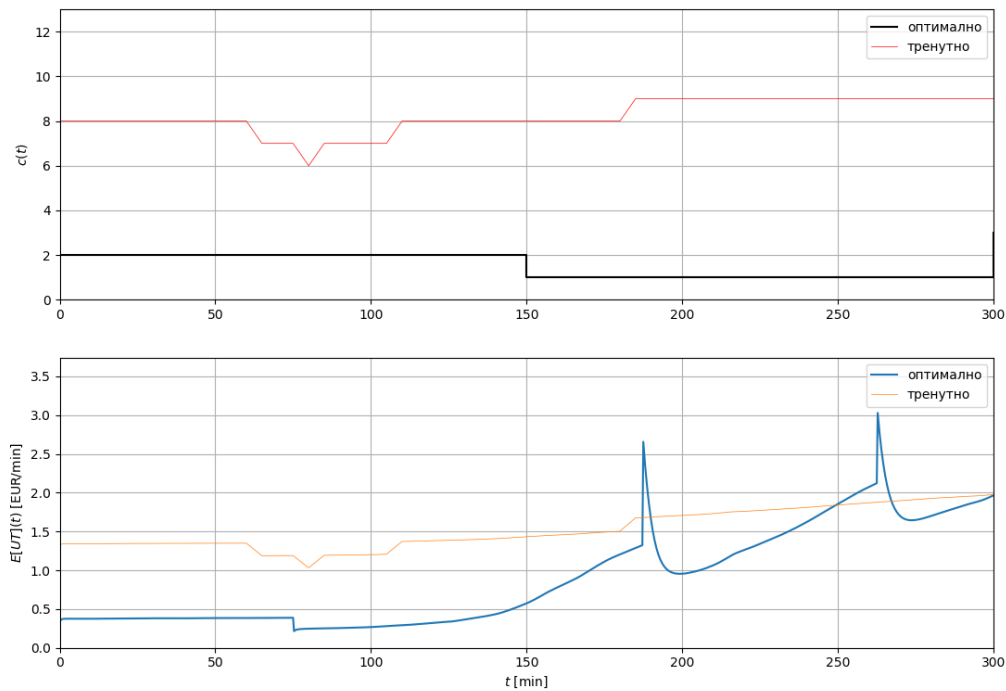
7.3.5 Резултати добијени оптимизацијом

На слици 22 може се видети промена вероватноће стања система за стварне и предвиђане интензитете долазака возила у случају дискретизације временске осе на два подинтервала на коме су нађене вредности оптималног броја наплатних рампи које треба да буду отворене². Одступање предвиђане и стварне вероватноће стања система је највише изражено у случају када је систем празан (p_0), док су одступања за остала стања безначајна. На слици 23 приказана је оптимална промена броја канала и промена укупних трошкова у случају дискретизације временске осе на два подинтервала и у случају тренутне стратегије управљања. Види се да оптимизација значајно утиче на смањење укупних трошкова.

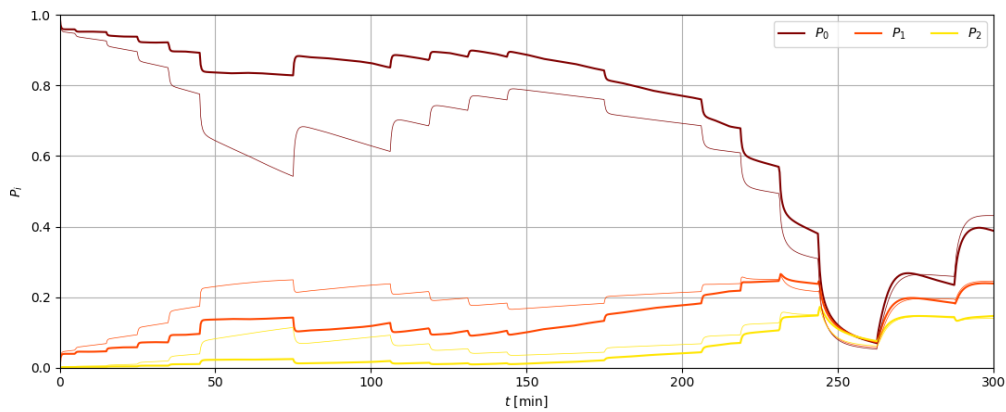
² У циљу лакшег разумевања слика, на истим су приказане само вероватноће стања система када у систему нема никога, када се налази једно и два возила. Примећује се да је у свим случајевима збир вероватноћа претходно набројаних стања система блиска 1.



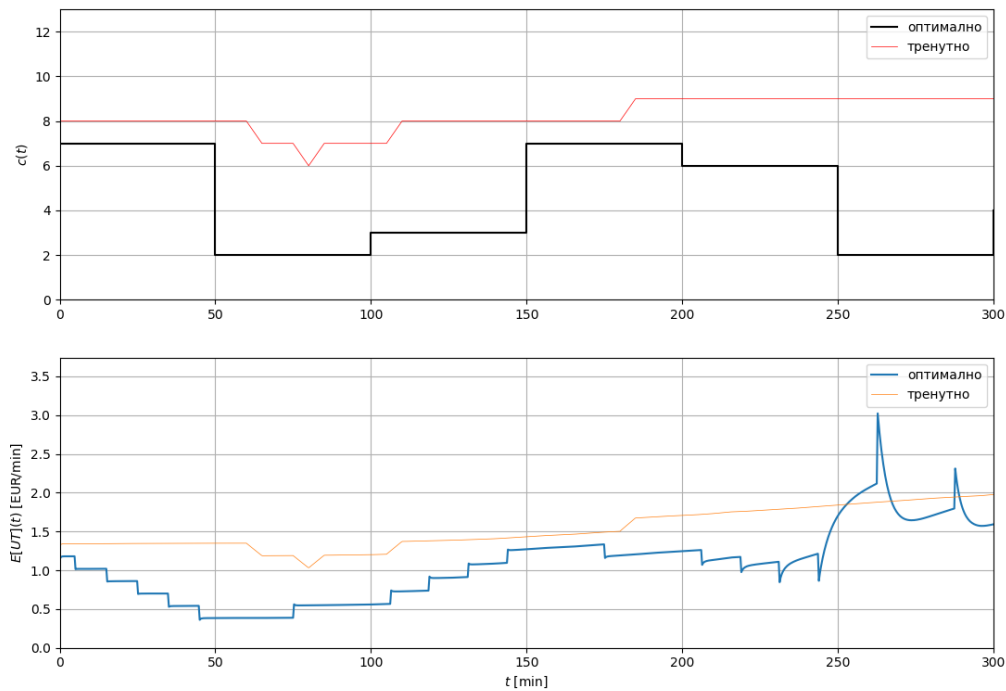
Слика 22 Промена вероватноће стања система за предвиђане вредности интензитета долазака (задебљана линија) и стварне вредности интензитета долазака возила (танка линија)) у случају одређивања оптималног броја отворених наплатних рампи за два подинтервала



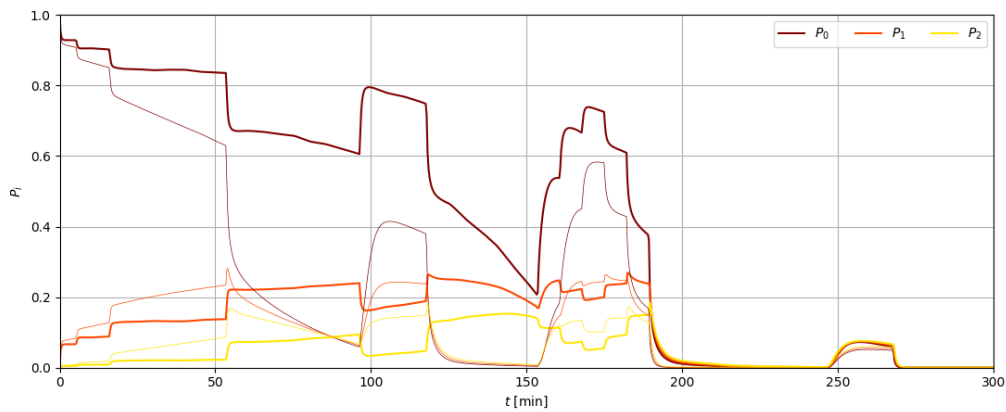
Слика 23 Промена броја канала у времену (горња слика) и промена вредности функције укупних трошкова током времена (доња слика)) у случају одређивања оптималног броја отворених наплатних рампи за два подинтервала



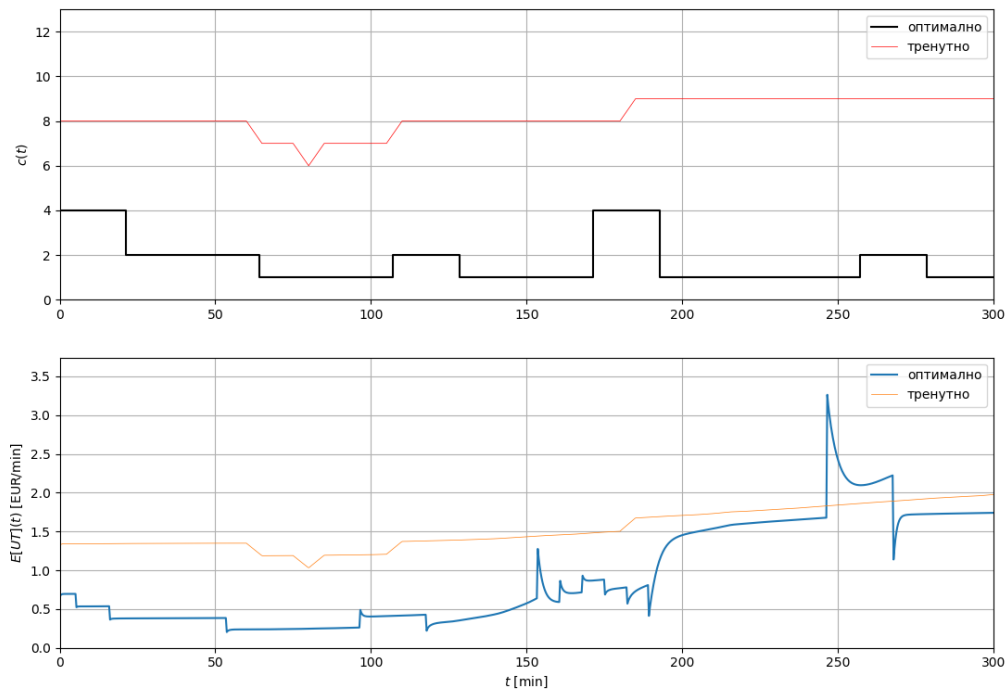
Слика 24 Промена вероватноће стања система за предвиђане вредности интензитета долазака (задебљана линија) и стварне вредности интензитета долазака возила (танка линија)) у случају одређивања оптималног броја отворених наплатних рампи за шест подинтервала



Слика 25 Промена броја канала у времену (горња слика) и промена вредности функције укупних трошкова током времена (доња слика)) у случају одређивања оптималног броја отворених наплатних рампи за шест подинтервала



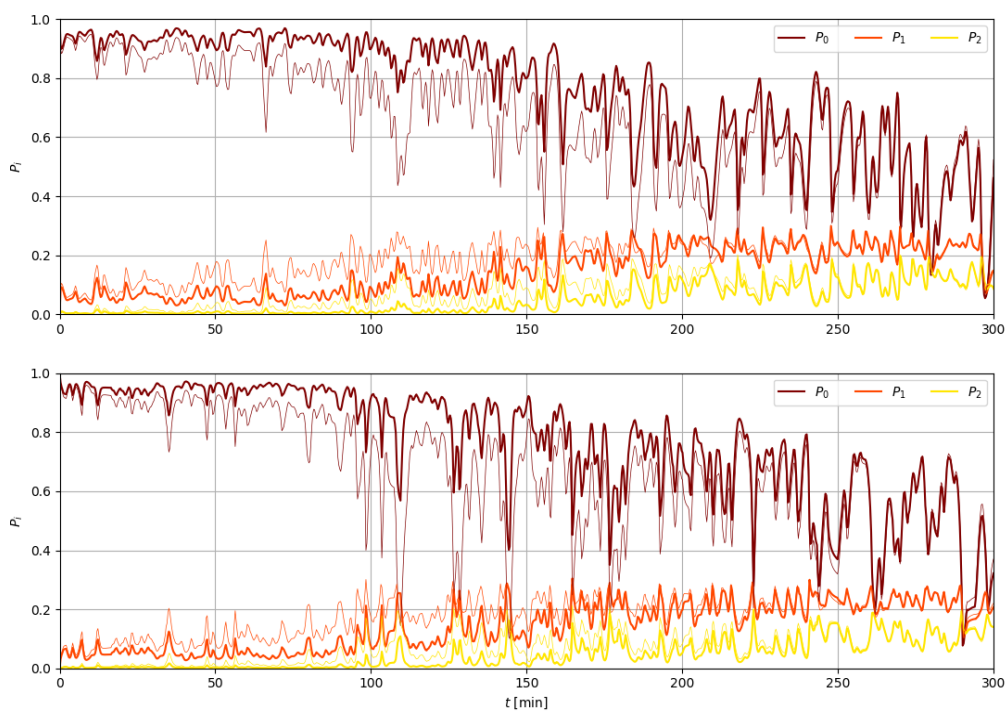
Слика 26 Промена вероватноће стања система за предвиђане вредности интензитета долазака (задебљана линија) и стварне вредности интензитета долазака возила (танка линија) у случају одређивања оптималног броја отворених наплатних рампи за петнаест подинтервала



Слика 27 Промена броја канала у времену (горња слика) и промена вредности функције укупних трошкова током времена (доња слика) у случају одређивања оптималног броја отворених наплатних рампи за петнаест подинтервала

На сличан начин на сликама 24 и 25 се виде промене вероватноће стања система за стварне и предвиђане интензитете долазака возила у случају дискретизације временске осе на шест подинтервала на коме су нађене вредности оптималног броја наплатних рампи које треба да буду отворене. Поред тога на сликама 26 и 27 су приказане промене оптималног броја наплатних рампи које треба да буду отворене у случају дискретизације времена на петнаест под интервала.

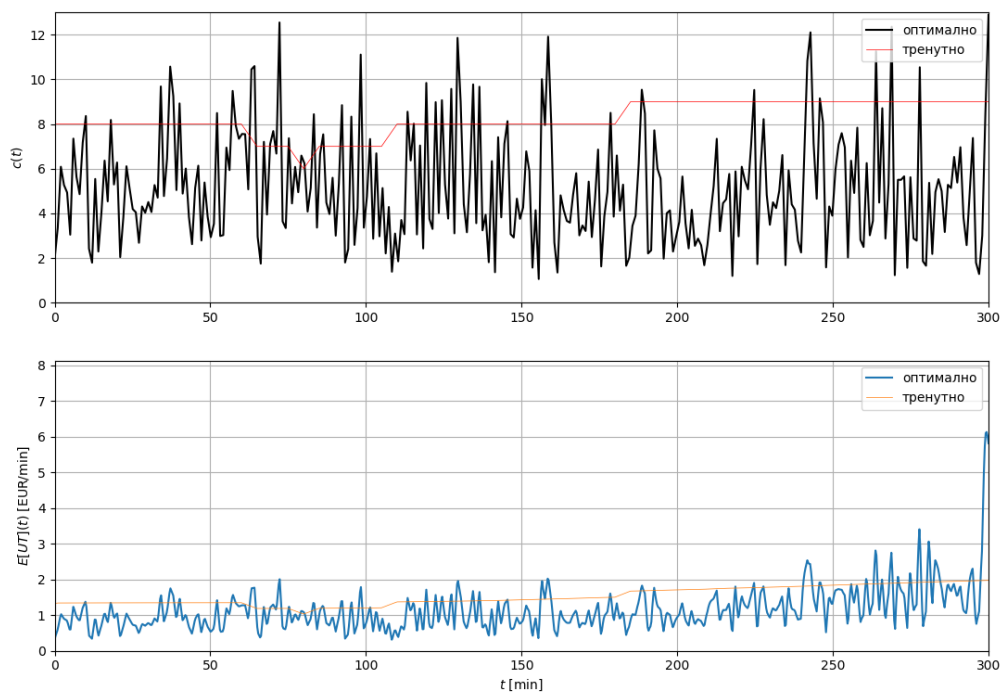
Одступање вероватноћа стања система је нешто мање у случају дискретизације на 6 интервала у односу на дискретизацију на 15 подинтервала. У оба случаја се може видети да нова стратегија отварања оптималног броја канала даје ниже укупне трошкове у односу на тренутну стратегију отварања наплатних рампи.



Слика 28 Промена вероватноће стања система за предвиђане вредности интензитета долазака (задебљана линија) и стварне вредности интензитета долазака возила (танка линија) у случају одређивања оптималног броја отворених наплатних рампи за три стотине подинтервала са континуалним вредностима (горња слика) и заокруженим (доња слика) управљања.

На слици 28 приказане су промене вероватноћа стања система за предвиђане вредности интензитета долазака и стварне вредности интензитета долазака возила у

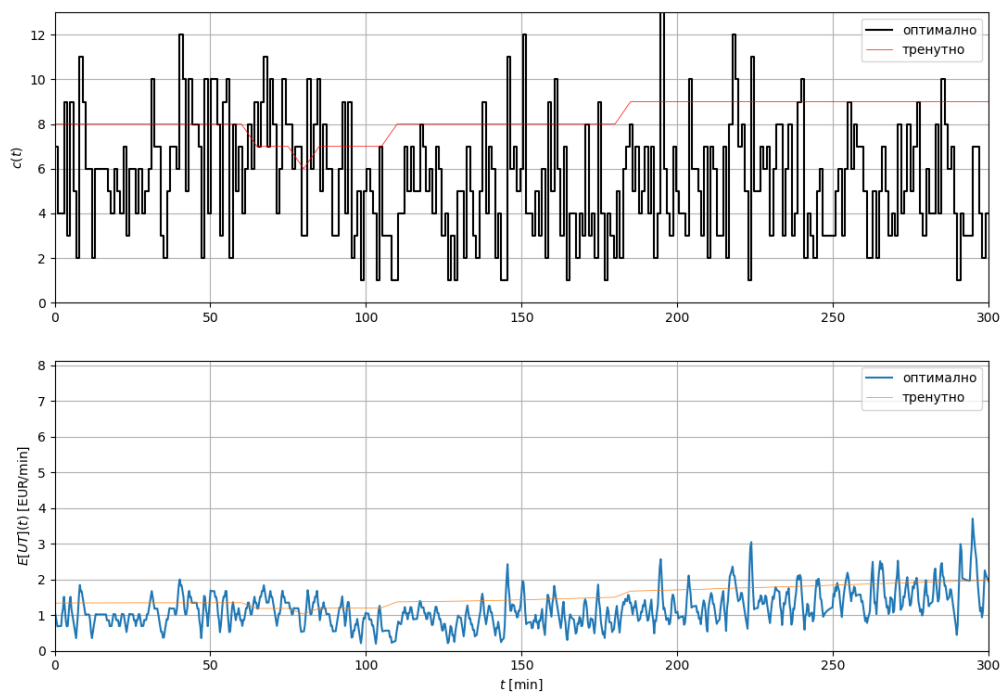
случају одређивања оптималног броја отворених наплатних рампи са три стотине подинтервала са континуалним вредностима (горња слика) и целобројним вредностима управљања (доња слика). Види се да у случају већег броја интервала утицај грешке у предвиђању мање утиче на одступање стварне и предвиђане промене вероватноће стања система.



Слика 29 Промена броја канала у времену (горња слика) и промена вредности функције укупних трошкова током времена (доња слика) у случају одређивања оптималног броја отворених наплатних рампи за три стотине подинтервала – континуалне вредности

Промене оптималног броја наплатних рампи приказане реалним и целобројним вредностима на дискретизованим подинтервалима приказане су на сликама 29 и 30. У оба случаја види се да повећањем броја подинтервала оптимизациони проблем постаје комплекснији па самим тим да би се достигла решења блиска оптималним неопходан је већи број итерација. Ипак и за максимални број итерација који износи 100, решења добијена оптимизацијом дају ниже вредности очекиваних трошкова у поређењу са тренутном стратегијом отварања и затварања наплатних рампи. Очекиване вероватноће

стања система добијене применом алгоритма 2 на 20 узорака из нормалне расподеле приказане су на слици 31 за континуалне (горња слика) и целобројне (доња слика) вредности броја отворених наплатних рампи, у случају дискретизације на шест подинтервала. Промене вероватноћа стања система добијене за овако одређен оптималан број отворених наплатних рампи (целобројне вредности) показују мала одступања у случају коришћења предвиђених и стварних интензитета долазака возила.

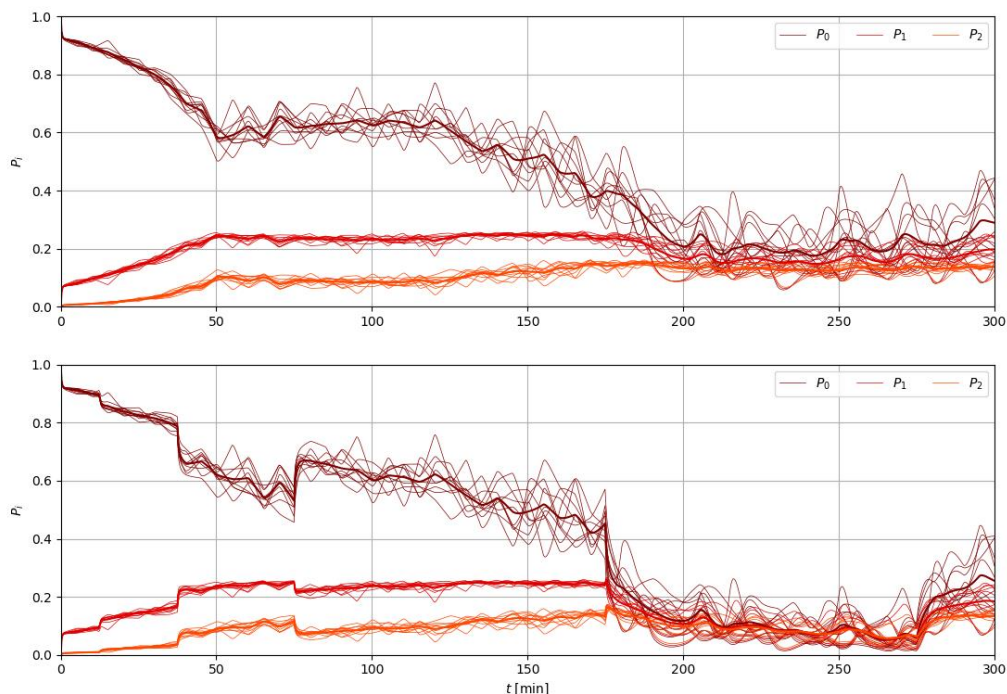


Слика 30 Промена броја канала у времену (горња слика) и промена вредности функције укупних трошкова током времена (доња слика) у случају одређивања оптималног броја отворених наплатних рампи за три стотине подинтервала – заокружене вредности

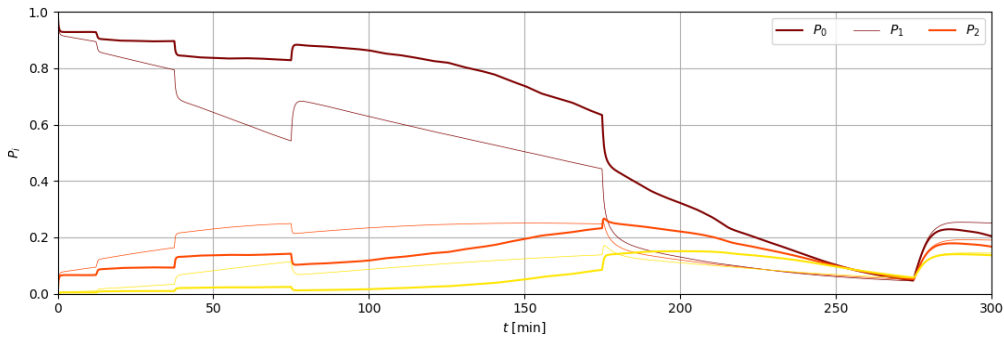
Промена вероватноћа стања система, броја канала и промена укупних трошкова током времена приказана је на слици 32 и 33. Види се да као и у претходним случајевима оптималан број отворених наплатних рампи, добијен применом алгоритма 2, има ниже вредности очекиваних трошкова у односу на тренутну стратегију отварања наплатних рампи.

Сумиране вредности укупних очекиваних трошкова за различити број дискретизованих подинтервала приказан је у табели 15. Може се видети да је

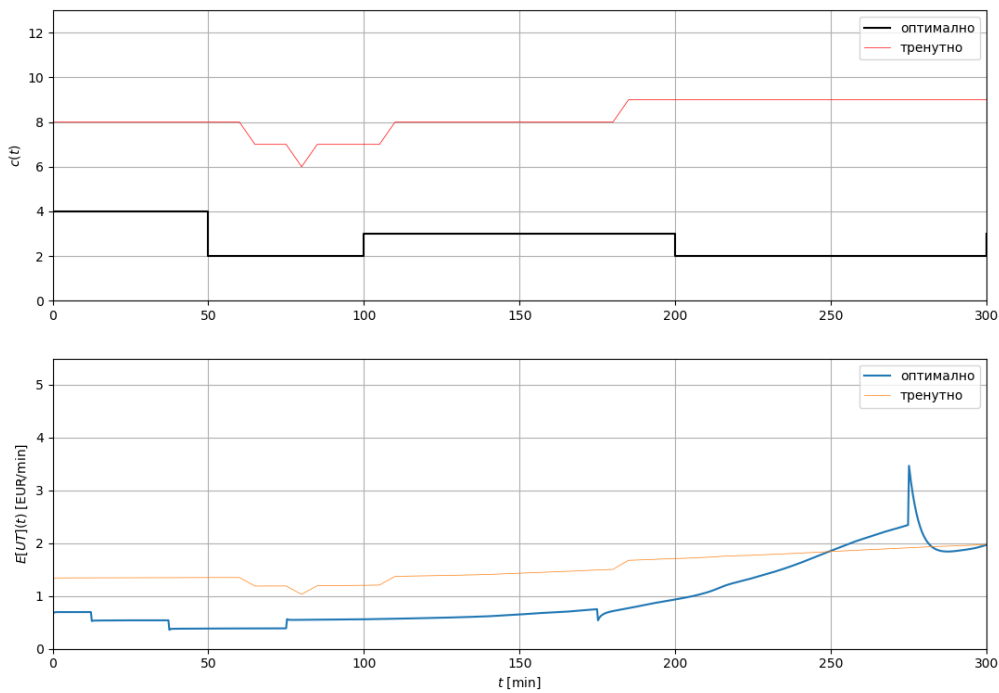
коришћењем метахеуристике у свим случајевима укупан очекивани трошак мањи од трошка постигнутог тренутном стратегијом отварања и затварања наплатних рампи на наплатној станици. Минимални трошак је постигнут за случај дискретизације временске осе на 5 подинтервала. Такође се примећује да поред већег броја степени слободе, при дискретизацији на већи број интервала, веће су вредности очекиваних трошкова добијене у поређењу са решењима добијеним дискретизацијом на мањи број подинтервала. Стога је потребно повећати или број јединки у популацији или максималан број итерација ради добијања бољих решења.



Слика 31 Промена вероватноће стања система за предвиђане вредности интензитета долазака у случају одређивања оптималног броја отворених наплатних рампи за шест подинтервала одређених Бајесовским приступом са континуалним вредностима (горња слика) и заокруженим вредностима (доња слика) управљања.



Слика 32 Промена вероватноће стања система за предвиђане вредности интензитета долазака (задебљана линија) и стварне вредности интензитета долазака возила (танка линија) у случају одређивања оптималног броја отворених наплатних рампи за шест подинтервала одређених Бајесовским приступом



Слика 33 Промена броја канала у времену (горња слика) и промена вредности функције укупних трошкова током времена (доња слика) у случају одређивања оптималног броја отворених наплатних рампи за шест подинтервала одређене-Бајесовским приступом

Табела 15 Очекивани укупни трошкови за разичите промене броја канала

	Број дискретизованих подинтервала							Тренутно управљање
	2	5	6	6 - Бајесовски	15	300	300 - континуално	
$E(UT)$ [EUR]	267,36	258,52	319,95	288,79	275,61	354,31	345,42	457,56

7.3.6 Симулација система опслуживања

У циљу верификације резултата добијених решавањем система диференцијалних једначина за различити број отворених наплатних рампи, извршена је симулација једноканалног система опслуживања за усвојену стратегију отварања и затварања наплатних рампи на наплатној станици. На слици 34 приказане су промене вероватноћа стања система добијене симулацијом и нумеричким решавањем система диференцијалних једначина, за различити број узорака у случају оптималног броја отворених наплатних рампи на шест дискретизованих подинтервала времена. Аналогно томе на слици 35 приказане су промене вероватноћа стања система добијене симулацијом и нумеричким решавањем система диференцијалних једначина за оптимални број отворених наплатних рампи добијен на три стотине подинтервала.

Може се са слика уочити да у оба случају вероватноће стања система добијене решавањем система диференцијалних једначина прате резултате добијене симулацијом. Такође се уочава да су, повећавањем броја узорака при симулацији, одступања између вероватноћа стања добијених симулацијом и нумеричким решавањем система диференцијалних једначина мање уочљива. У циљу квантификације одступања броја возила у систему у тренутку t , евалуираном за вероватноће добијене симулацијом и нумеричким решењима, добијене су вредности апсолутних и релативних грешака. Поред тога оцењена је и грешка симулације. У табели 16 приказане су апсолутна грешка, релативна грешка и оцењена вредност грешке. Средњи број возила у оквиру једног канала је одређен као:

$$L(t) = \sum_{i=0}^{m+1} i \cdot p_i(t) \quad (7.15)$$

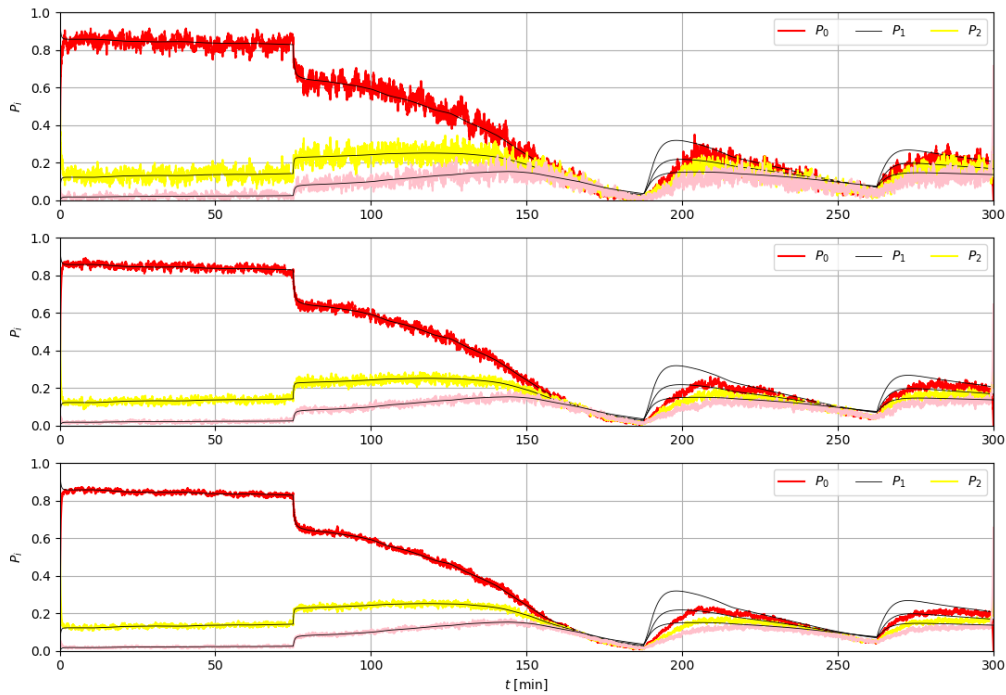
Апсолутна грешка је одређена као аритметичка средина апсолутне разлике између броја возила у систему одређеног симулацијом ($L_{sym}(t)$) и нумеричким решењима система диференцијалних једначина ($L(t)$).

$$Abs_{gr} = \frac{1}{N_t} \sum_t |L_{sym}(t) - L(t)| \quad (7.16)$$

где је N_t број подинтервала усвојених у иницијализацији симулације.

Аналогно претходном релативна грешка се одређује као:

$$rel_{gr} = \frac{1}{N_t} \sum_t \left| \frac{L_{sym}(t) - L(t)}{L(t)} \right| \quad (7.17)$$



Слика 34 Поређење вероватноћа промене стања система добијеног симулацијом (обојене линије) и решавањем система диференцијалних једначина (црне линије) у случају одређивања оптималног броја отворених наплатних рампи за шест подинтервала са 200 (горњи дијаграм), 1000 (средњи дијаграм) и 3000 (доњи дијаграм) узорака у симулацији

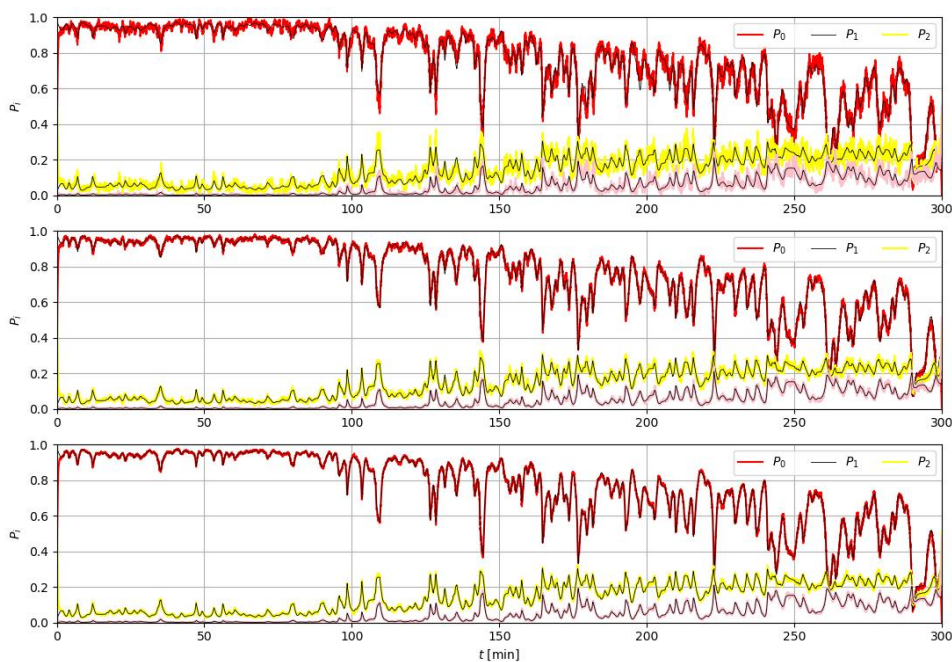
Оцена грешке симулације се процењује као количник између стандардне девијације симулације средњег броја возила у систему и квадратног корена броја симулација (N_{sym}).

$$ose_{gr} = \frac{\sigma_{sym}}{\sqrt{N_{sym}}} \quad (7.18)$$

Може се приметити да грешке опадају са порастом броја симулација, што је већ и закључено посредством приказаних слика 34 и 35.

Табела 16 Апсолутна, релативна грешка и оцена грешке

Бр. подинтервала	Број узорака у симулацији				
	Апсолутна грешка				
	200	1000	3000	5000	10000
300	0,044	0,021	0,017	0,014	0,013
15	0,055	0,025	0,022	0,018	0,016
	Релативна грешка				
300	0,237	0,142	0,119	0,108	0,095
15	0,169	0,121	0,106	0,091	0,085
	Оцена грешке				
300	0,051	0,022	0,013	0,01	0,007
15	0,072	0,032	0,019	0,014	0,010



Слика 35 Поређење вероватноћа промене стања система добијеног симулацијом (обојене линије) и решавањем система диференцијалних једначина (црне линије) у случају одређивања оптималног броја отворених наплатних рампи за три стотине подинтервала са 200 (горњи дијаграм), 1000 (средњи дијаграм) и 3000 (доњи дијаграм) узорака у симулацији

7.3.7 Дискусија

На основу изложене методологије одређивања оптималног броја канала у модулу у неком периоду у будућности и на основу спроведених експеримената на примеру оптимизације отворених наплатних рампи на наплатној станици Врчин, може се извести неколико битних закључака.

Пре свега показано је да примена методологије на реалним проблемима за усвојену адекватну функцију трошкова може да исте умањи у односу на стратегију управљања каналима која није експлицитно заснована на применама математичких модела. Такође је показано да рекурентне мреже могу успешно да предвиде средње интензитете долазака возила на наплатну рампу, односно да су перформансе предикције постигнуте применом истих задовољавајуће. Самим тим, оптимизација заснована на овако предвиђаним интензитетима долазака клијената резултира у стратегију која даје боље резултате у односу на тренутне политике управљања. Поред тога, иако се вероватноће стања система за предвиђане интензитете долазака клијената разликују у односу на вероватноће стања система за стварне интензитете долазака клијената, показује се да оптимална стратегија управљања добијена применом предвиђања показује боље резултате у односу на тренутно стање.

Може се видети да примена алгоритма 2 може да доведе до бољих резултата у односу на алгоритам 1 у случају када рекурентна мрежа предвиђа интензитете долазака са лошим перформансама. Стога је битно узети у обзир грешке које иста прави и применом Бајесовског приступа наћи оптималан број канала у модулу. Иако је ова процедура рачунски много скупља, оправданост њеног коришћења може бити доведена у питање када су варијансе предикције ниске односно када су перформансе предикције мреже одличне.

Повећањем броја подинтервала на којим се врши оптимизација даје се већи степен слободе, који као последицу има повећање максималног броја итерација ради добијања бољих решења. Стога се у зависности од домена на који се методологија примењује бира и број подинтервала како би резултати били што ближи оптималном решењу. Односно на примеру управљањем наплатним рампама рационално је мењати број истих на минималном периоду од 60 минута, све испод тога није могуће реализовати у стварности.

8 Примена развијених методологија на реалним проблемима из других домена

Представљене методологије могу да буду успешно примењене и на реалним проблемима који се не односе на саобраћај. У овом поглављу приказана је примена методологије за класификацију гужви коришћењем Гаусових условних случајних поља на скупове података које се односе на: предвиђање гужви на скијалиштима, класификацију гена по функционалности, семантичку класификацију сцена, класификацију емоција из музике, класификацију болести, класификацију врсте птица уз помоћ аудио сигнала. Поред тога, на различитим доменима приказана је предност коришћења Гаусових условних случајних поља за класификацију у односу на стандардне неструктурне предикторе. Додатно, методологија за предикцију индикатора саобраћаја је примењена у циљу прогнозирања брзине скијаша на скијалиштима. Наведене су предности исте у поређењу са моделима где се претходно не елиминишу вредности ништа.

У поглављу 8.1 приказана је примена Гаусових условних случајних поља за класификацију на примеру предвиђања гужви на скијалишту, док је у поглављу 8.2 приказана примена методологије за предвиђање просечне брзине скијаша на скијалишту. Примена Гаусових условних случајних поља за класификацију на примеру класификације гена по функционалности приказана је у поглављу 8.3, семантичке класификације сцена у поглављу 8.4, класификације емоција из музике у поглављу 8.5, класификације болести у поглављу 8.6, док је класификација врсте птица уз помоћ аудио сигнала приказана у поглављу 8.7.

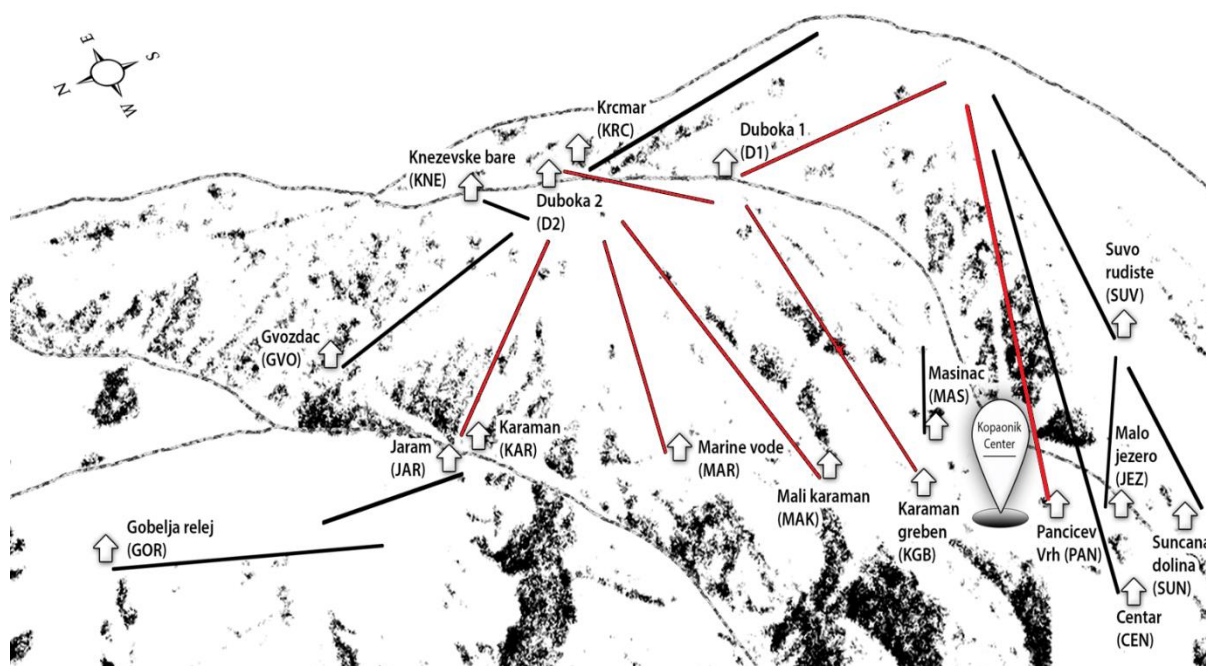
8.1 Предвиђање гужви на скијалиштима

Копаоник је највећи скијашки центар у Републици Србији, са око 25 ски-лифтова и више од 55 км скијашких стаза. Један од највећих проблема на Копаонику се односи на појаву великих гужви на ски-лифтовима. Копаоник углавном поседује ски-лифтове са 4 или 6 седишта у корпи и спада у једно од најсавременијих скијалишта у југоисточној Европи. Услед неравномерне заступљености скијаша на ски стазама, врло често долази до великих гужви и редова чекања на улазима на ски-лифтове. Најчешће се гужве јављају на 7 најпосећенијих ски-лифтова који су приказани на слици 36. У циљу решавања проблема гужви неопходно је предвидети када ће доћи до гужве на ски-

лифтовима. На овај начин се скијашаши могу алоцирати на скијашке стазе на којима се исте не јављају. Стога је и потребно предвидети да ли ће доћи до гужве 40 минута унапред, што је и довољно дугачак временски период како би скијаш могао да оде са једног на други крај скијалишта.

Подаци коришћени у овом раду односе се на информације о улазу скијаша на ски-лифтове. Посредством ових информација могу се пратити путање кретања скијаша. Подаци коришћени при предвиђању гужви се односе на период од 15. до 30. марта током сезона у интервалу од 2006 до 2011 године. Укупан број инстанци добијених трансформацијом сирових података је 4,850 за сваки ски-лифт што је укупно 33,950 инстанци. Временски прозор у којима се посматрају статистике везане за скијаше као и просечно време потребно да се направи круг на ски-лифту је 5 минута.

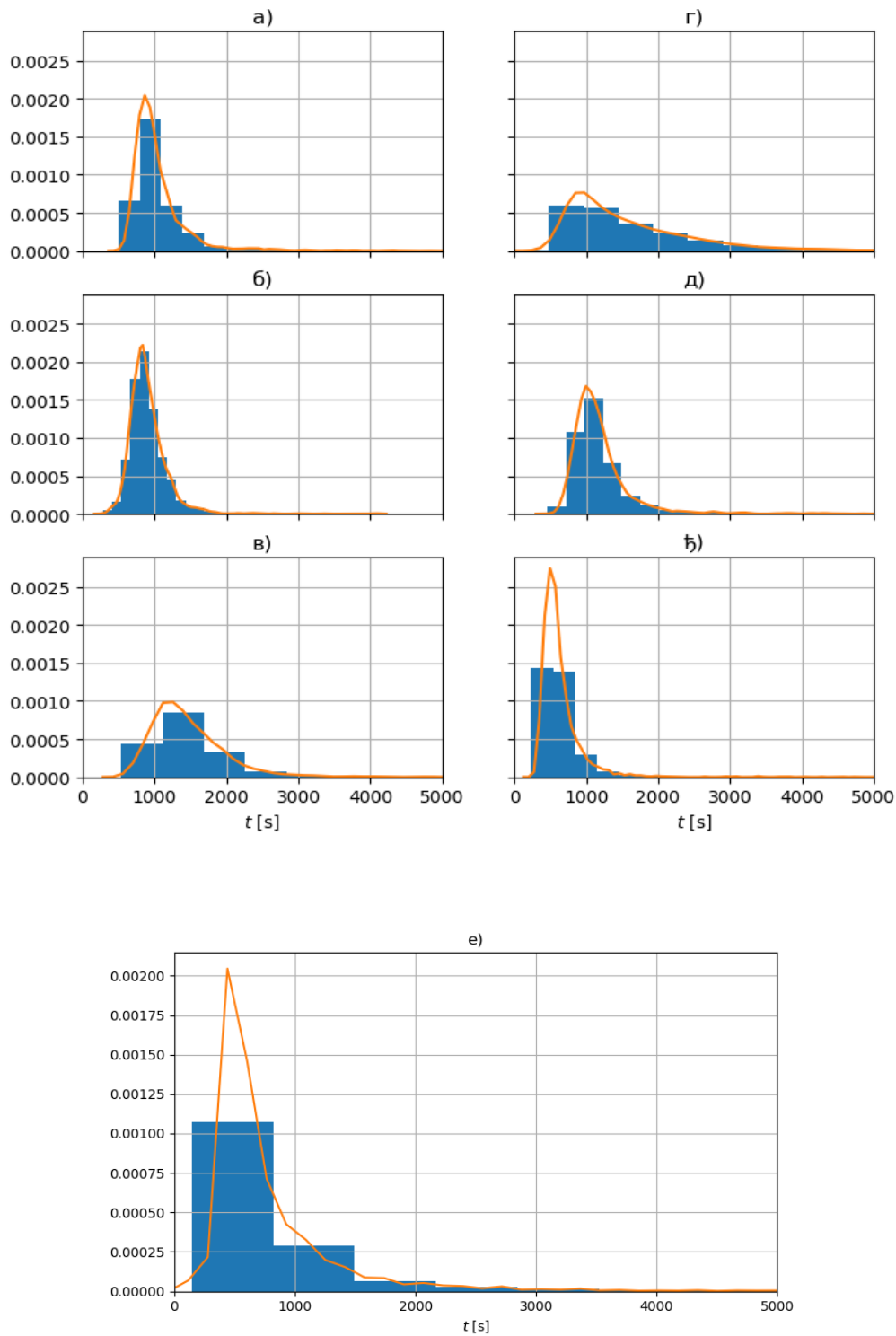
Претпоставља се да је гужва на ски-лифтовима детерминисана просечним временом које је потребно скијашу да направи круг на ски-лифту. Расподела просечног времена потребног да скијаш направи круг на стази заједно са Кернеловим одређивањем густина приказана је на слици 37.



Слика 36 Мапа стаза ски-центра Копеоиик – црвеном су приказане стазе на којима се предвиђала гужва

На основу просечног времена које је потребно да скијаш направи круг на једној стази, одређено је постојање гужви на ски-лифту. Односно уколико гужва постоји

просечно време које је потребно да скијаш направи круг на ски стази је ниже од модуса расподеле просечног времена које је потребно да скијаш направи круг на ски стази.



Слика 37 Емпиријска расподела времена спуштања скијаша на анализираним стазама ски-центра Копаоник

У циљу формирања скупа података на основу кога се врши предвиђање коришћени су додатни атрибути. Укупно је коришћено 18 различитих атрибута који су подељени у 3 групе. У табели 17 приказане су групе и описи коришћених атрибута при предвиђању гужви на скијалиштима.

Табела 17 Групе и опис атрибута коришћених при предвиђању гужви на скијалишту

Група атрибута	Опис атрибута
Дескриптивни	Укупан број свих пролазака на ски-лифту
	Време изражено у минутима
	Укупан број скијаша на ски-лифту
Статистички	Средња брзина скијаша
	Минимална брзина скијаша
	Десети перцентил брзине скијаша
	Први квантил брзине скијаша
	Медиана брзине скијаша
	Други квантил брзине скијаша
	Максимална брзина скијаша
	Спљоштеност расподеле брзине
	Издуженост расподеле брзине
	Деведесети перцентил брзине скијаша
Временски	Температура у °C
	Влажности ваздуха у %
	Тачка росе у °C
	Брзина ветра у m/s
	Притисак у mbar

Као и у случају предвиђања гужви на ауто-путевима, коришћени су следећи неструктурни предиктори: риц логистичка регресија, ласо логистичка регресија, неуронска мрежа за класификацију и случајна шума. неуронска мрежа се састоји од 3 сакривена слоја са по 30, 20 и 14 неурона по слоју. Активационе функције коришћене у овим слојевима су RELU. Исто као у случају прогнозирања гужви на ауто-путевима, перформансе модела су оцењене унакрсно валидацијом са 10 делова. Укупно 80% података за учење је коришћено за тренирање неструктурних предиктора, док је остатак коришћен за учење Гаусових условних случајних поља за класификацију.

Шест различитих матрица сличности (структурних графова између излаза) је коришћено:

1. Корелациона матрица између излаза претрениране неуронске мреже.
2. Матрица међусобне информације (енг. „*Mutual information*“, скраћено MI).

3. Матрица са вредностима χ^2 статистика.
4. Матрица која се састоји од норме разлике између вектора састављених од историје маркера.
5. Матрица која се састоји од норме разлике између вектора састављених од маркера.
6. Матрица која се састоји од норме разлике између вектора састављених од историје просечних времена скијаша.

Табела 18 Перформансе предикције и време израчунавања модела за проблем предивђања гужви на ски-центру Копаоник

Копаоник					
Р. бр.	Модел	Метод кластеровања	Број кластера	ПИК	Време учења по једној итерацији унакрсне валидације у секундама
1	GCRFCnb	-	-	0,836	18,809
2	GCRFCb	-	-	0,837	378,864
3	GCRFBCb-fast	K-means	5	0,830	146,144
4	GCRFBCb-fast	K-means	50	0,830	154,006
5	GCRFBCb-fast	K-means	500	0,830	600,907
7	GCRFBCb-fast	Mini batch K-means	5	0,830	118,237
8	GCRFBCb-fast	Mini batch K-means	50	0,830	185,917
9	GCRFBCb-fast	Mini batch K-means	500	0,830	507,045
11	GCRFBCb-fast	Gaussian Mix	5	0,830	139,294
12	GCRFBCb-fast	Gaussian Mix	50	0,830	176,018
13	GCRFBCb-fast	Gaussian Mix	500	0,830	306,330
15	GCRFBCb-fast	Gaussian Mix prob	5	0,829	109,867
16	GCRFBCb-fast	Gaussian Mix prob	50	0,830	203,290
17	GCRFBCb-fast	Gaussian Mix prob	500	0,831	1489,861
19	Ridge-logistic	-	-	0,793	-
20	LASSO-logistic	-	-	0,794	-
21	NN	-	-	0,791	-
22	RF	-	-	0,783	-

Елементи четврте и шесте матрице сличности дефинисани су као:

$$S_{ij} = \exp(-c\mathbf{y}_i^h - \mathbf{y}_{j2}^h) \quad (8.1)$$

где је c мета параметар, који се одређује на основу унакрсне валидације посредством ПИК метрике, а y_i^h је вредност историјских података (маркера или просечног времена) инстанце i .

Перформансе предикције структурних и неструктурних предиктора, заједно са временом учења структурних предиктора приказане су у табели 18. Може се видети да структурни предиктори постижу боље резултате у односу на неструктурне предикторе, односно да утицај структуре значајно утиче на перформансе предикције. Најбоље резултате је постигао модел GCRFBCb, иако је ПИК скор занемарљиво бољи у односу на GCRBCnb модел. Брзе верзије Бајесовских Гаусових условних случајних поља за класификацију показују лошије перформансе у односу на небајесовски алгоритам. Такође се може видети да број кластера и избор модела кластерованја не утичу нешто значајно на резултате предикције.

Време учења не-Бајесовских Гаусових условних случајних поља је много мање у односу на време учења Бајесовских модела. На исти начин се види да је у случају мањег броја кластера учење брзог Бајесовског алгоритма знатно брже у поређењу са стандардним моделом.

8.2 Предвиђање просечне брзине скијаша на скијалиштима

Једна од могућности решавања проблема гужви на Копаонику повезан је са предикцијом индикатора саобраћаја на ски стазама који се могу користити као једно од помоћних средстава у предвиђању постојаности гужве. У овом конкретном случају као индикатор саобраћаја на скијалишту узето је просечно време које је скијашу потребно да направи круг на једној ски стази. Циљ је предвидети просечно време које је скијашу потребно да направи круг на стази 25 минута у будућности. Аналогно претходном примеру, методологија је тестирана и учена из скупа података за период од 15. до 30. марта, за шест сезона од 2006 до 2011 године. Интервал временског прозора за формирање атрибута и укупан број инстанци коришћен при тренирњу је исти као у случају класификације гужви. Укупан број инстанци са вредностима излаза ништа је 8,468. Укупно је коришћено 18 атрибута за предвиђање и они су приказани у табели 17.

Као и у претходном проблему, исти неструктурни предиктори су коришћени при тренирању класификационих и регресионих структурних модела. Предиктивне перформансе структурних и неструктурних предиктора евалуиране су унакрсном

валидацијом са 15 делова. Четири различита неструктурна класификатора су коришћена: риц логистичка регресија, ласо логистика регресија, неуронска мрежа за класификацију и случајна шума за класификацију. Поред тога, коришћена су и четири различита регресора: риц регресија, ласо регресија, неуронска мрежа за регресију и случајна шума за регресију. Укупно 80% података је коришћено за учење неструктурних предиктора, док је остатак коришћен за учење Гаусових условних случајних поља за класификацију. Потпуно исти однос података коришћен је и при тренирању регресионог модела Гаусових условних случајних поља и одговарајућих неструктурних регресора. Аналогно проблему класификације гужви, шест различитих матрица сличности (структурних графова између излаза) је коришћено код регресионог и класификационог структурног модела.

Табела 19 Перформансе предикције и време израчунавања регресионих модела за проблем одређивања просечног времена скијаша, потребног да направи круг на ски стази на ски-центру Копаоник

Регресија скијаша			
Р. бр.	Модел	R²	Време учења по једној итерацији унакрсне валидације у секундама
1	GCRF	0,397	74,28
2	Ridge	-0,421	-
3	LASSO	-0,346	-
4	NN-reg	0,1	-
5	RF-reg	0,185	-

У табели 19 и табели 20 приказане су перформансе предикције и времена израчунавања регресионих и комбинованих модела за проблем одређивања просечног времена скијаша, потребног да направи круг на ски стази, на ски-центру Копаоник. Може се приметити да вредности излаза на којима није било скијаша, а које су маркиране са 0, битно утичу на лоше перформансе предикције исказане ниском вредношћу R² скорa. Поред тога што GCRF има више вредности R² скорa у односу на неструктурне предикторе, овај скор је и даље релативно лош. Елиминацијом вредности ништа из излаза посредством класификатора резултати регресије су побољшани за око 16% у односу на R² скор. Посматрајући ПИК скорове GCRFBCb-fast модела за класификацију може се закључити да су исти високи, то јест проблем класификације

вредности излаза у категорије 0 (ништа) и категорије 1 (различито од ништа) је решен са задовољавајућом тачношћу. Такође, види се да инстанце, са вредностима излаза ништа, битно утичу на крајњи исход регресије како код структурних тако и код неструктурних предиктора.

Табела 20 Перформансе предикције и време израчунавања комбиноване методологије за проблем одређивања просечног времена скијаша, потребног да направи круг на ски стази, на ски-центру Копаоник

Класификација скијаша					
Р. бр.	Модел	Метод кластеровања	Број кластера	ПИК/R²	Време учења по једној итерацији унакрсне валидације у секундама
1	GCRFBCnb	-	-	0,837	15,842
2	GCRFBCb	-	-	0,904	494,486
3	GCRFBCb-fast	K-means	50	0,911	151,199
4	Ridge-logistic	-	-	0,897	-
5	LASSO-logistic	-	-	0,896	-
6	NN	-	-	0,897	-
7	RF	-	-	0,893	-
Регресија					
1	GCRF	-	-	0,551	68,06
2	Ridge	-	-	0,042	-
3	LASSO	-	-	0,484	-
4	NN	-	-	0,443	-
5	RF	-	-	0,460	-

Важно је нагласити да у свим случајевима структурни регресиони модели имају боље перформансе у односу на неструктурне моделе. Ова констатација је последица значаја структуре у везама између излаза. Ипак, предност структурних модела у односу на неструктурне је повезана са дужим временом учења и закључивања то јест већим рачунским трошком.

8.3 Класификација гена по функционалности

Скуп података у овом проблему формиран је од низова експресија гена и филогенетских профила. Укупан број инстанци који је коришћен при тренирању и тестирању је 2417.

Табела 21 Перформансе предикције и време израчунавања модела за проблем класификације гена

Класификација гена					
Р. бр.	Модел	Метод кластеровања	Број кластера	ПИК	Време учења по једној итерацији унакрсне валидације у секундама
1	GCRFBCnb	-	-	0,768	4,879
2	GCRFBCb	-	-	0,797	299,920
3	GCRFBCb-fast	K-means	5	0,794	77,083
4	GCRFBCb-fast	K-means	50	0,794	84,043
5	GCRFBCb-fast	K-means	150	0,794	252,909
6	GCRFBCb-fast	K-means	250	0,794	311,587
7	GCRFBCb-fast	Mini batch K-means	5	0,794	100,447
8	GCRFBCb-fast	Mini batch K-means	50	0,795	102,561
9	GCRFBCb-fast	Mini batch K-means	150	0,794	199,251
10	GCRFBCb-fast	Mini batch K-means	250	0,794	266,148
11	GCRFBCb-fast	Gaussian Mix	5	0,794	77,093
12	GCRFBCb-fast	Gaussian Mix	50	0,793	89,404
13	GCRFBCb-fast	Gaussian Mix	150	0,794	150,340
14	GCRFBCb-fast	Gaussian Mix	250	0,794	157,579
15	GCRFBCb-fast	Gaussian Mix prob	5	0,794	99,447
16	GCRFBCb-fast	Gaussian Mix prob	50	0,794	218,299
17	GCRFBCb-fast	Gaussian Mix prob	150	0,794	383,490
18	GCRFBCb-fast	Gaussian Mix prob	250	0,794	1248,968
19	Ridge-logistic	-	-	0,575	-
20	LASSO-logistic	-	-	0,575	-
21	NN	-	-	0,571	-
22	RF	-	-	0,598	-

Укупан број атрибута коришћен за предвиђање је 103. Сваки ген је повезан са скупом функционалних класа чија максимална димензија може бити 103. Стога овај

проблем класификације спада у домен вишеизлазне класификације (енг. „*multilabel classification*“). Овај скуп података је већ анализиран у радовима [100, 101] и познат је као тежак проблем класификације. Као и у случају предвиђања гужви на ауто-путевима исти неструктурни предиктори и исти графови су коришћени и у овом класификационом проблему. Од података који су коришћени за учење 60% је коришћено за тренирање неструктурних предиктора, док је остатак коришћен за тренирање структурних предиктора. Резултати перформанси неструктурних и структурних предиктора валидирани унакрсном валидацијом са 10 делова приказани су у табели 21.

Може се приметити да су Бајесовски и не-Бајесовски структурни модели остварили значајно боље резултате у поређењу са неструктурним предикторима. На основу овога може се закључити да структура има изузетно значајну улогу у побољшању перформанси предикције. Најбољи Бајесовски модел Гаусових условних случајних поља за класификацију је бољи од најбољег неструктурног предиктора (алгоритам Случајне шуме) за око 35%. GCRFBCb-fast даје исте резултате као и GCRFBCb стандардни модел али са мањим временом учења за мањи број кластера. Стога се може закључити да ја на овом скупу података узимајући у обзир ПИК скор и време учења најбоље вредности показао GCRFBCb-fast са K-means алгоритмом за кластеровање и 5 кластера.

8.4 Семантичка класификација сцена

Семантичка класификација сцена је реални проблем који обухвата анализу и организацију слика, препознавање делова слика итд. Оригинални рад из којег су искоришћени подаци за семантичку класификацију сцена може се видети у [102]. Исто као и у случају функционалне класификације гена, овај тип класификације спада у домен вишеизлазне класификације. Односно циљ је да се на основу слике препознају семантичке сцене које се на истој налазе. При томе, треба имати у виду да сцене нису међусобно искључиве, то јест на једној слици се може наћи више различитих сцена. Пре почетка семантичке класификације сцена, генерисани су атрибути за предвиђање који представљају просторне боје пиксела у Лув простору. Детаљни опис начина на који су генерисани атрибути из оригиналних слика и њихова интерпретација у Лув простору може се наћи у оригиналном раду [102]. Укупан број инстаци који је коришћен је 2407, док је укупан број атрибута 294. Шест различитих сцена је обележено као могућ излаз: плажа, излазак сунца, падање лишћа, поље, планина и градски део. Исто као и у претходним случајевима коришћене су четири различите матрице сличности (графа):

корелациона матрица између излаза претрениране неуронске мреже, матрица међусобне информације, матрица са вредностима χ^2 статистика и матрица која се састоји од норме разлике између вектора састављених од историје маркера. Такође исти неструктурни предиктори су коришћени: риц и ласо логистичка регресија, неуронске мреже и случајна шума. Од података који су коришћени за учење 60% је коришћено за тренирање неструктурних предиктора, док је остатак коришћен за тренирање структурних предиктора. Резултати унакрсне валидације приказани су у табели 22.

Може се видети да је ПИК скор GCRFBCb модела виши од ПИК скор неструктурних предиктора. Исто тако закључује се да је ПИК скор GCRFBCnb модела нижи од резултата који су постигли неструктурни предиктори, односно GCRFBCnb модел није постигао задовољавајуће резултате. GCRFCb-fast са Gaussian Mix prob и бројем кластера 50 постигао је боље резултате у односу на стандардни GCRFBCb модел. Поред тога време тренирања истог модела је мање за око 10% у односу на време тренирања стандардног GCRFBCb модела.

Табела 22 Перформансе предикције и време израчунавања модела за проблем класификације сцена

Семантичка класификација сцена					
Р. бр.	Модел	Метод кластеровања	Број кластера	ПИК	Време учења по једној итерацији унакрсне валидације у секундама
1	GCRFCnb	-	-	0,859	4,298
2	GCRFCb	-	-	0,931	89,542
3	GCRFCb-fast	K-means	5	0,934	64,146
4	GCRFCb-fast	K-means	50	0,934	83,891
5	GCRFCb-fast	K-means	500	0,763	162,998
6	GCRFCb-fast	K-means	1200	0,850	309,563
7	GCRFCb-fast	Mini batch K-means	5	0,934	73,330
8	GCRFCb-fast	Mini batch K-means	50	0,934	93,323
9	GCRFCb-fast	Mini batch K-means	500	0,765	153,892
10	GCRFCb-fast	Mini batch K-means	1200	0,852	297,192
11	GCRFCb-fast	Gaussian Mix	5	0,934	79,797
12	GCRFCb-fast	Gaussian Mix	50	0,935	78,385
13	GCRFCb-fast	Gaussian Mix	500	0,934	181,666
14	GCRFCb-fast	Gaussian Mix	1200	0,934	443,772
15	GCRFCb-fast	Gaussian Mix prob	5	0,935	72,009
16	GCRFCb-fast	Gaussian Mix prob	50	0,935	140,327
17	GCRFCb-fast	Gaussian Mix prob	500	0,934	1256,819
18	GCRFCb-fast	Gaussian Mix prob	1200	0,934	2705,995
19	Ridge-logistic	-	-	0,898	-
20	LASSO-logistic	-	-	0,917	-
21	NN	-	-	0,922	-
22	RF	-	-	0,924	-

8.5 Класификација емоција из музике

Скуп података који је коришћен у овом проблему класификације састоји се од 100 песама за сваку од 7 различитих категорија: класична музика, реге, рок, поп музика, хип хоп, техно и џез музика [103]. Скуп података је креиран избором 3 песме са сваког од 233 различита музичка албума. Атрибути који су коришћени при класификацији подељени су у две целине: ритмички и тонски атрибути. Укупан број атрибута је 72: од тога 8 ритмичких атрибута и 64 тонска атрибута. Музика је маркирана у 6 различитих

категорија: изненађујућа, смирујућа, музика која изазива срећу, опуштајућа, тужна и музика која изазива страх. Детаљно објашњење о начину маркирања инстанци и генерисању атрибута може се наћи у [103]. Укупан број инстанци у скупу података је 593. Исти неструктурни предиктори и структурне везе између излаза су коришћене као и у претходним проблемима.

Потребно је поменути да једна инстанца може бити означена са више класа (нпр. може да буде смирујућа и опуштајућа итд.). Од података који су коришћени за учење 50% је коришћено за тренирање неструктурних предиктора док је остатак коришћен за тренирање структурних предиктора. Перформансе структурних и неструктурних предиктора и времена учења модела приказана су у табели 23.

Може се видети да су Бајесовски модели Гаусових условних случајних поља за класификацију много бољи у поређењу са не-Бајесовским моделом и неструктурним предикторима. Поред тога GCRFBCb-fast, са K-means алгоритмом и 50 кластера, је постигао најбоље резултате у односу на све остале моделе. Време учења овог алгоритма је приближно исто времену учења осталих алгоритама. ПИК скор GCRFBCb-fast, са алгоритмом K-means и 50 кластера, има боље резултате у односу на најбољи модел (ПИК = 0,8237) приказан у оригиналном раду. Такође је приметно да GCRFBCnb модел има нижи ПИК скор од најбољег неструктурног предиктора (случајна шума).

Табела 23 Перформансе предикције и време израчунавања модела за проблем класификације емоција из музике

Класификација емоција из музике					
Р. бр.	Модел	Метод кластеровања	Број кластера	ПИК	Време учења по једној итерацији унакрсне валидације у секундама
1	GCRFBCnb	-	-	0,823	0,860
2	GCRFBCb	-	-	0,854	48,122
3	GCRFBCb-fast	K-means	5	0,819	20,594
4	GCRFBCb-fast	K-means	50	0,855	45,447
5	GCRFBCb-fast	K-means	150	0,851	85,049
6	GCRFBCb-fast	K-means	250	0,851	100,297
7	GCRFBCb-fast	Mini batch K-means	5	0,855	20,232
8	GCRFBCb-fast	Mini batch K-means	50	0,853	32,574
9	GCRFBCb-fast	Mini batch K-means	150	0,852	75,949
10	GCRFBCb-fast	Mini batch K-means	250	0,852	90,102
11	GCRFBCb-fast	Gaussian Mix	5	0,852	13,921
12	GCRFBCb-fast	Gaussian Mix	50	0,852	20,717
13	GCRFBCb-fast	Gaussian Mix	150	0,853	43,298
14	GCRFBCb-fast	Gaussian Mix	250	0,853	46,331
15	GCRFBCb-fast	Gaussian Mix prob	5	0,852	13,679
16	GCRFBCb-fast	Gaussian Mix prob	50	0,853	44,581
17	GCRFBCb-fast	Gaussian Mix prob	150	0,851	330,289
18	GCRFBCb-fast	Gaussian Mix prob	250	0,852	580,634
19	Ridge-logic	-	-	0,815	-
20	LASSO-logic	-	-	0,823	-
21	NN	-	-	0,800	-
22	RF	-	-	0,848	-

8.6 Класификација болести

Скуп података који је коришћен у овом проблему састоји се од болничких записа у Калифорнији у периоду од јануара 2009 до децембра 2011 године. Скуп података преузет је из државне базе података агенције за истраживање и квалитет у здравству (оригинално енг. „*State Inpatient Database (SID), Healthcare Cost and Utilization Project (HCUP), Agency for Healthcare Research and Quality*“).

Подаци се састоје од информација које су везане за пацијенте педијатрије, односно историју њихових болести. За сваког пацијента (инстанцу) дати су бинарни

атрибути који означавају да ли је пацијент имао одређену болест или није. Укупан број атрибута који је коришћен за сваку инстанцу је 930. Излази представљају векторе чији елементи означавају болести због којих се пацијент вратио у болницу у периоду од месец дана. Могуће је да се пацијент вратио због више различитих болести које су му након дијагностиковане. Десет различитих болести (излаза) је анализирано.

Табела 24 Перформансе предикције и време израчунавања модела за проблем класификације болести

Класификација болести					
Р. бр.	Модел	Метод кластеровања	Број кластера	ПИК	Време учења по једној итерацији унакрсне валидације у секундама
1	GCRFBCnb	-	-	0,808	32,512
2	GCRFBCb	-	-	0,834	1255,588
3	GCRFBCb-fast	K-means	5	0,830	472,180
4	GCRFBCb-fast	K-means	50	0,831	629,879
5	GCRFBCb-fast	K-means	500	0,827	2204,468
7	GCRFBCb-fast	Mini batch K-means	5	0,831	467,574
8	GCRFBCb-fast	Mini batch K-means	50	0,830	539,592
9	GCRFBCb-fast	Mini batch K-means	500	0,829	1139,841
11	GCRFBCb-fast	Gaussian Mix	5	0,831	575,889
12	GCRFBCb-fast	Gaussian Mix	50	0,830	615,286
13	GCRFBCb-fast	Gaussian Mix	500	0,830	1147,169
15	GCRFBCb-fast	Gaussian Mix prob	5	0,830	478,014
16	GCRFBCb-fast	Gaussian Mix prob	50	0,831	699,413
17	GCRFBCb-fast	Gaussian Mix prob	500	0,831	3248,991
19	Ridge-logistic	-	-	0,782	-
20	LASSO-logistic	-	-	0,797	-
21	NN	-	-	0,782	-
22	RF	-	-	0,811	-

Укупан скуп података састоји се од 20,589 инстанци. 80% од укупног броја инстанци за тренирање је коришћено за учење неструктурних предиктора, док је остатак коришћен за тренирање структурних предиктора. Исти неструктурни предиктори и исте

матрице сличности су коришћене као у претходним примерима. У табели 24 приказане су ПИК скорови структурних и неструктурних предиктора коришћени у овом проблему.

Може се видети да су структурни Бајесовски алгоритми остварили боље предиктивне перформансе у односу на неструктурне. Односно GCRFBCnb модел је остварио лошије резултате у односу на неструктурне предикторе. Стандардни Бајесовски GCRFBCb модел је остварио најбоље резултате, али и време учења истог по једној итерацији унакрсне валидације је високо. Имајући у виду да је разлика у ПИК скоровима између GCRFBCb и GCRFBCb-fast између 0,3% – 0,7% може се закључити да је време учења у овом случају битан репер за избор између ова два анализирана модела.

8.7 Класификација врсте птица уз помоћ аудио сигнала

С обзиром на то да су птице један од врло јаким индикатора за различита еколошка истраживања, посебно она истраживања која се тичу промена у екосистему, битно је познавати које врсте птица постоје на изабраном локалитету. Птице врло брзо реагују на промене, па се на основу њиховог понашања може закључити о каквој врсти промена се ради. Традиционалне методе за прикупљање података о птицама на одређеној локацији захтевају доста ресурса и времена. Стога се технике машинског учења могу користити у случајевима када се треба препознати које су све птице заступљене у одређеном еко систему. Један од начина на који се ово може проверити је снимањем звука и потом препознавањем која врста птица се чује на аудио сигналу. Овај проблем спада у вишеизлазну класификацију, с обзиром да се на једном снимку могу појавити звукови различитих птица. Постоји велики број предности зашто би се на овакав начин лоцирале које птице постоје у еко систему, а тиче се проширивања временске и просторне резолуције без улагања додатних ресурса, примена истог метода на местима која нису доступна. Скуп података који је коришћен је преузет из [104].

Скуп података је имао 19 различитих врста птица као излазе, док је укупан број инстанци који је коришћен при учењу и валидацији алгоритама 645. Укупан број атрибута добијен обрадом звучног сигнала је 260. Исти неструктурни предиктори и матрице сличности су коришћене, као и у претходним проблемима. Од података који су коришћени за учење 60% је коришћено за тренирање неструктурних предиктора док је остатак коришћен за тренирање структурних предиктора. Резултати унакрсне валидације приказани су у табели 25.

Табела 25 Перформансе предикције и време израчунавања модела за проблем класификације птица

Класификација птица					
Р. бр.	Модел	Метод кластеровања	Број кластера	ПИК	Време учења по једној итерацији унакрсне валидације у секундама
1	GCRFBCnb	-	-	0,786	1,387
2	GCRFBCb	-	-	0,791	55,631
3	GCRFBCb-fast	K-means	5	0,804	35,978
4	GCRFBCb-fast	K-means	50	0,795	53,826
5	GCRFBCb-fast	K-means	150	0,801	69,019
6	GCRFBCb-fast	K-means	250	0,796	86,964
7	GCRFBCb-fast	Mini batch K-means	5	0,795	35,236
8	GCRFBCb-fast	Mini batch K-means	50	0,796	47,680
9	GCRFBCb-fast	Mini batch K-means	150	0,799	66,029
10	GCRFBCb-fast	Mini batch K-means	250	0,798	75,111
11	GCRFBCb-fast	Gaussian Mix	5	0,797	32,442
12	GCRFBCb-fast	Gaussian Mix	50	0,795	78,267
13	GCRFBCb-fast	Gaussian Mix	150	0,806	161,915
14	GCRFBCb-fast	Gaussian Mix	250	0,797	186,074
15	GCRFBCb-fast	Gaussian Mix prob	5	0,795	27,733
16	GCRFBCb-fast	Gaussian Mix prob	50	0,798	63,320
17	GCRFBCb-fast	Gaussian Mix prob	150	0,797	236,504
18	GCRFBCb-fast	Gaussian Mix prob	250	0,800	350,787
19	Ridge-logistic	-	-	0,722	-
20	LASSO-logistic	-	-	0,774	-
21	NN	-	-	0,681	-
22	RF	-	-	0,803	-

Може се видети да неструктурни предиктори у овом скупу података не постижу резултате који су значајније бољи од перформанси структурних предиктора. Односно структурни предиктор са најбољим ПИК скором, GCRFBCb-fast са Gaussian Mix кластеровањем и 150 кластера, само је 0,2% бољи у односу на калсификатор случајна шума. Време израчунавања овог алгоритма је више од 3 пута веће од времена израчунавања стандардног GCRFBCb модела.

9 Закључак

У овој дисертацији представљене су три методологије чијом се применом значајно могу смањити проблеми гужви. Имајући у виду да је основни задатак при решавању проблема гужви процена да ли ће доћи до истих, први део дисертације заснива се на решавању проблема предикције саобраћајних гужви развојем методологије за класификацију гужви коришћењем Гаусових условних случајних поља.

Истраживање начина и метода за предвиђање гужви представља једну од веома битних области које је узело маха у претходном периоду. Велики број методологија заснованих на статистичком моделовању временских серија нису успешно решиле овај проблем стога имплементација и развој нових алгоритама из области машинског учења може значајно допринети развоју и проширењу нових сазнања у овој област. Многи од представљених модела за предикцију гужви покушавају да на основу датих атрибута предиктују да ли ће на одређеној локацији до гужве доћи или неће, не узимајући при томе информацију која је садржана у околним локацијама које директно или индиректно утичу на појаву гужве. Стога је у овој дисертацији представљена методологија за класификацију гужви која користи новоразвијени модел машинског учења који покушава да истовремено предвиђа појаву гужви на више локација. Развијени модел имплементиран је на основу структурног пробалистичког модела Гаусових условних случајних поља за регресију. С обзиром да је циљ предвидети да ли ће доћи до гужве или неће, проблем предвиђања гужви спада у домен бинарне класификације. Самим тим Гаусова условна случајна поља за регресију, модификована су у циљу имплементације на проблемима бинарне класификације.

Новоразвијени модел Гаусових условних случајних поља за регресију састоји се од два слоја случајних променљивих. Први слој дефинише условну расподелу латентне случајне променљиве над атрибутима за предвиђање. Овај слој представљен је у облику Гаусових условних случајних поља за регресију где су излази повезани на основу унапред задатих графова (веза између истих). Над латентним слојем се задаје нови слој који представља излазе из класификације. Вероватноће условних расподела излаз, за задату латентну случајну величину, су међусобно независне и расподељене по бинарној расподели. Сумирано заједничка расподела латентних случајних променљивих и излаза представља производ бинарних расподела и каноничке форме Гаусових условних случајних поља представљених нормалном расподелом. С обзиром да улазне величине у Гаусова условна случајна поља представљају неструктурне предикторе који су

претходно тренирани, циљ је научити колико је сваки од неструктурних предиктора значајан за дати проблем класификације. Поред тога, током процеса учења се одређује колико је сваки од унапред задатих графова значајан при предикцији излаза. Сложена репрезентација модела Гаусових условних случајних поља за класификацију онемогућава једноставно учење истог. Имајући у виду наведено, развијена су два модела за учење и закључивање.

Први модел је заснован на не-Бајесовском приступу где се очекивана вредност латентних променљивих користи као улаз у сигмоидну функцију. Коришћење оваквог приступа при учењу и закључивању поједностављује и убрзава примену предложеног модела. Међутим у случају када је норма дијагоналних елемената коваријационе матрице латентних променљивих велика, очекивана вредност није довољно добар репрезент нормалне расподеле. Односно коришћење очекиване вредности довешће до незадовољавајућих перформанси предикције. Стога је циљ покушати маргинализовати заједничку расподелу латентних променљивих величина и излаза по латентним променљивим величинама и исту користити као релевантан модел за предвиђање.

Самим тим други модел је заснован на Бајесовском приступу при учењу и закључивању. Маргинализацију заједничке расподеле латентних променљивих величина није могуће аналитички спровести, стога је и отежано учење и закључивање. Поступак закључивања је спроведен поједностављивањем подинтегралне функције производа нормалне расподеле и бинарних расподела на облик који је могуће евалуирати једноструком нумеричком интеграцијом. Приликом овога, узете су у обзир независности условних расподела излаза за задате одговарајуће латентне променљиве. За разлику од овог поступка при учењу је коришћена апроксимација сигмоидне функције коришћењем локаланих варијационих апроксимација. Сигмоидна функција је представљена доњом границом исте посредством увођења варијационих константни које су учене заједно са параметрима модела. Форма доње границе сигмоидне функције омогућава аналитичку интеграцију подинтегралне функције над латентним променљивим величинама. Након интеграције новодобијена форма представља доњу границу функције веродостојности. У циљу примене градијентних метода, са ограничењима над параметрима модела у циљу обезбеђивања позитивне полудефинисаности коваријационе матрице Гаусових условних случајних поља, евалуирани су парцијални изводи доње границе функције веродостојности над параметрима који се уче. Потпуно исти поступак је примењен у случају не-Бајесовских Гаусових условних случајних поља за класификацију.

У циљу показивања предности и мана једног и другог модела спроведени су експерименти над синтетичким подацима. Показано је да у најгорем случају Бајесовски модел има исте предиктивне перформансе као и не-Бајесовски модел. Међутим време учења и закључивања Бајесовског модела је значајно веће у односу на не-Бајесовски модел. Велика разлика у времену последица је огромног броја варијационих параметара које је потребно научити, што самим тим представља и отежавајућу околност при примени оптимизационих алгоритама.

Ради смањења рачунских трошкова при учењу и закључивању развијена је апроксимативна верзија Гаусових условних случајних поља за класификацију. Основа одлика овог брзог Бајесовског модела лежи у покушају смањивања броја варијационих параметара кластеровањем истих током процеса учења. Увођењем различитих модела кластеровања и избором броја кластера новоформирани принцип учења може да се убрзава или успорава у зависности од жељене тачности модела. Ради евалуације свих претходно развијених модела коришћена је скуп података са ауто-путева.

Проблем који је решаван се односио на предвиђање гужви на ауто-путу E70-E75 кроз Републику Србију на деоницама Ниш – Београд и Београд – Адашевци. Анализирани су сирови подаци о временима пролазака возила и посредством њих је направљен нови скуп података који је коришћен за тестирање модела. Атрибути који су коришћени при тренирању неструктурних предиктора се односе на различите статистике изведене из података и временских прилика на ауто-путу. Шест различитих графова задатих матрицом сличности је коришћено при учењу модела. Резултати су верификовани унакрсном валидацијом. Показано је да на примеру предвиђања гужви на ауто-путу новоразвијени модел постиже значајно боље резултате у поређењу са неструктурним предикторима, који не узимају у обзир информацију која је садржана у везама између излазних величина. Види се да је време учења и закључивања Бајесовског у односу на не-Бајесовских модел значајно веће иако се перформансе предикције превише не разликују.

На сличан начин методологија за класификацију гужви применом Гаусових условних случајних поља је додатно тестирани на проблему предикције гужви на ски-лифтовима ски-центра Копаоник. Коришћени атрибутити су на сличан начин формирано од почетног скупа података. Аналогно претходном Бајесовски, брзи Бајесовски са различитим моделима кластеровања и не-Бајесовски модели Гаусових условних случајних поља за класификацију постижу боље резултате предикције од неструктурних предиктора. Слично као и у случају предвиђања гужви на ауто-путевима Бајесовски

модел и поред најбољих перформанси, не постиже значајано боље резултате у односу на не-Бајесовски модел.

Предложена методолога за класификацију гужви коришћењем Гаусових условних случајних поља је додатно генерализована на велики број различитих домена: класификација гена по функционалности, семантичка класификација сцена, класификација емоција из музике, класификација болести и класификација врсте птица из аудио сигнала. У свим случајевима Гаусова условна случајна поља за класификацију су показала веће вредности ПИК скорова у односу на неструктурне предикторе. На примерима је показано да услед великог броја варијационих параметара брзи Бајесовски модел у одређеним примерима показује боље резултате од осталих модела. Показано је и да постоје случајеви када је и услед велике рачунске сложености оправдано користити Бајесовски модел Гаусових условних случајних поља за регресију.

На основу свега изложеног у првом делу рада показује се да методологија за класификацију гужви применом Гаусових условних случајних поља може бити успешно примењена на проблемима предикције гужви и проблемима из других домена. Односно коришћењем Гаусових условних случајних поља за класификацију, допринос у погледу побољшања перформанси предикције је значајан у поређењу са неструктурним предикторима. Избор одговарајућег модела Гаусових условних случајних поља за класификацију је детерминисан перформансама предикције као и рачунским трошком истих. У одређеним ситуацијама примена Бајесовских или брзих Бајесовских модела може имати значајно боље резултате у односу на не-Бајесовски модел. Најчешћи случај када се јавља наведена ситуација је када неструктурни предиктори нису сигурни у одлуку коју доносе, што је узрок високе норме дијагоналних елемената коваријационе матрице. Стога је најбоље при избору усвојеног модела пре тога исти тестирати на мањем скупу података, а затим изабрати онај који показује најбоље перформансе у погледу предикције и рачунског трошка учења и закључивања.

Даља истраживања на ову тему треба да буду усмерена ка проширивању модела, коришћеног у методологији, на временско – просторне зависности између излаза. Односно циљ би био да излазне променљиве нису повезане само у простору, посредством графа који детерминише просторне везе, већи и у времену. На тај начин се може експлоатисати информација садржана у излазима која тумачи временску и просторну зависност излазних величина. Многи проблеми из овог домена још увек нису решени са задовољавајућим резултатима, стога даља истраживања у овој области несумњиво могу да имају значајан утицај.

У другом делу дисертације представљена је методологија за предвиђање индикатора саобраћаја који се заснива на комбинацији класификационих и регресионих алгоритама машинског учења. Под индикаторима саобраћаја подразумевају се различите величине на основу којих менаџмент може да донесе одговарајућу одлуку у циљу смањења трошкова или спречавања гужви. Најчешћи показатељи саобраћаја су средња вредност брзине или средње време путовања. На основу ових вредности експерт или менаџмент може без обзира на статистичке податке да оцени да ли постоји гужва на ауто-путу или не, стога овај показатељ може итекако да помогне при тачној предикцији гужве. Поред тога је јако тешко на основу података одредити да ли је било гужве или није било гужве, стога индикатори саобраћаја могу да пруже информације које су много корисније менаџменту. На пример радови на путу могу да доведу до смањења средње брзине возила али то није јасан индикатор да је повећана гужва на наплатним рампама. Самим тим информација о стварној вредности индикатора саобраћаја је много вреднија узевши у обзир да агрегиране статистике на основу које се врши класификација у стања гужва и није гужва не гарантују прецизност. Стога се на основу информације о стварној вредности средње брзине уз експертско знање може одредити да ли се одређено стање може класификовати као гужва.

Методологија која је развијена се базира на коришћењу комбинације Гаусових условних случајних поља за класификацију и регресију. Велики број саобраћајних података нема вредности индикатора саобраћаја у одређеним инстанцама зато што није регистрован ни један пролазак возила. Важно је напоменути да ове излазе не треба мешати са излазима где је вредност излаза једнака нули, стога је неопходно елиминисати наведене индикаторе саобраћаја како перформансе регресора не би биле покварене. Један од начина елиминације оваквих података при закључивању је коришћењем класификатора који ће индикаторе класификовати у групе где су исти већи од нуле и где су једнаки нули. А затим на оним вредностима излаза применити регресиони алгоритам који ће предвидети њихове стварне вредности у будућности. Овако дефинисана методологија може бити генерализована на свим скуповима података који су ретки и имају структурну везу између излазних величина. На примерима проблема везаним за предикције средње вредности брзине возила која долазе на наплатне станице и средње вредности времена које је потребно скијашима да направе једну вожњу на одређеној ски стази показане су предности методологије. Показано је да значај структуре има изузетно велики утицај на перформансе предикције. Међутим елиминација излаза са вредностима ништа при учењу и закључивању има значајнији утицај на крајње резултате у поређењу

са предиктивним перформансама. Коначно на оба примера показано је да примена претходне елиминације излаза са вредностима ништа и коришћењем структуре, при предикцији индикатора саобраћаја, може да повећа R^2 скор за 0,4.

На свим приказаним примерима показује се да примена предложене методологије може значајно да утиче на перформансе предикције. Као и у првом делу рада примена одговарајућег класификационог модела Гаусових условних случајних поља зависи од перформанси предикције и од рачунских трошкова односно доступних ресурса.

Комбинација алгоритама машинског учења са неком од области из домена операционих истраживања није толико често коришћена. Велики број реалних система опслуживања моделује се теоријом редова чекања. У последњем делу дисертације приказана је методологија за одређивање оптималног броја активних канала у модулу у будућности. Методологија која је развијена представља комбинацију модела машинског учења, теорије редова чекања и алгоритама математичког програмирања. Методологија је развијена у циљу прогнозирања који број канала у модулу треба да буде отворен у неком периоду у будућности. Новоразвијена методологија базира се на коришћењу рекурентних неуронских мрежа за прогнозирање интензитета долазака клијената. Под претпоставком да је време долазака клијената расподељено по експоненцијалној расподели са временски зависним параметром расподеле, прогнозира се како ће се вредност истог мењати у будућности. Уколико се докаже да је време опслуживања клијената експоненцијална расподела са константним интензитетом, врши се оцена параметра расподеле. Оцењена вредност интензитета долазака клијената и предвиђана вредност интензитета опслуживања представља улаз у моделе теорије редова чекања. Вероватноће стања система се одређују посредством нумеричког решења матричне диференцијалне једначине, где интензитети представљају елементе транзиционе матрице. Имајући у виду да је интензитет долазака клијената зависан од броја активних канала у модулу, при чему се сматра да је овај интензитет униформно распоређен на активне канале, применом оптимизационих метода неопходно је одредити политику отварања канала. Ова политика отварања и затварања канала је детерминисана формом функције циља. У већини случајева код система масовног опслуживања функција циља представља очекиване укупне трошкове који се састоје од трошкова чекања и трошкова опслуживања, њихова минимизација треба да буде обезбеђена исправно изабраном политиком отварања и затварања канала у модулу. Као оптимизациона техника у дисертацији, предложено је коришћење метахеуристика зато што није могуће показати да је постављена функција циља конвексна. Поред тога немогућност експлицитне

евалуације функције вероватноће стања система у зависности од времена, захтева нумеричко решавање система диференцијалних једначина тако да примена било које од градијентних метода може бити спроведена само коришћењем нумеричких извода.

Предложена су два алгоритма у оквиру методологије за одређивање оптималног броја активних канала у модулу у будућности. Први алгоритам се заснива на не-Бајесовском одређивању оптималног броја активних канала у модулу, док је други алгоритам заснован на Бајесовском одређивању оптималног броја активних канала у модулу. Код Бајесовског алгоритма неопходно је тренирати нову рекурентну мрежу која ће предвиђати колико прва рекурентна мрежа греша при оцени параметра експоненцијалне расподеле. Примена Бајесовског приступа је оправдана само у ситуацијама када су перформансе предикције прве мреже лоше.

У циљу верификације рада методологије на примеру одређивања броја отворених наплатних рампи у будућности иста је примењена на наплатној станици Врчин. Првобитно су на основу сирових података извучене статистике коришћене као атрибути при предвиђању и формирану су излази помоћу померајућег прозора. Затим је тестирана расподела времена опслуживања помоћу стандардних статистичких тестова и са информационим критеријумима. Показано је да су времена опслуживања расподељена по експоненцијалној расподели са констатним интензитетом опслуживања. Укупно је научно и тестирано 96 различитих архитектура неуронских мрежа. Резултати су показали да LSTM мреже са унутрашњом стандардизацијом слојева имају најбоље резултате на скупу за тестирање и најбоља архитектура истих остварује R^2 скор од 0,879. У циљу примене другог алгоритма тренирана је још једна мрежа чији је циљ био да предвиђа грешке које прва прави при предикцији. Трениране су различите архитектуре LSTM мреже са унутрашњом стандардизацијом слојева за оцену грешке. Након завршеног тренирања за произвољно изабрану инстанцу, у тест скупу података, извршено је предвиђање интензитета долазака и њихово коришћење у циљу формирања функције циља засноване на примени теорије редова чекања. Након формирања функције циља очекивани укупни трошкови система опслуживања минимизовани су применом метахеуристике диференцијалне еволуције, за различити број дискретизованих подинтервала на временској оси. Верификација резултата спроведена је поређењем симулације са решењима добијеним нумеричким решавањем система матричних диференцијалних једначина. Верификација је квантификована посредством представљања апсолутне, релативне и оцењене грешке симулације.

Резултати показују да су без обзира на број усвојених подинтервала очекивани укупни трошкови, добијени оптимизацијом, много нижи у поређењу са истим трошковима добијеним тренутном стратегијом отварања и затварања наплатних рампи. Ови трошкови достижу вредности које су ниже до 50% у односу на тренутну стратегију управљања. Показано је да и примена Бајесовског приступа при проналажењу оптималне политике управљања може да смањи трошкове оптимизације у односу на не-Бајесовски приступ. Ово смањење може бити значајно посебно у случајевима када је грешка предикције висока. Вероватноће стања система, добијене на основу предвиђених и стварних интензитета, се разликују. Међутим ове разлике не утичу значајно на оптималне политике управљања. Примећено је да повећањем броја подинтервала времена при оптимизацији, разлике између предвиђаних и стварних вероватноћа стања система постепено опадају. Такође је показано да већи број подинтервала омогућава већу флексибилност при оптимизацији али захтева и већи број максимално дозвољених итерација оптимизационог алгоритма, што је спрегнуто са чињеницом да бољи резултати оптимизације изискују већи рачунски трошак. Може се на основу свега приказаног закључити да развијена методологија значајно умањује укупне трошкове у поређењу са тренутним стратегијама управљања отварањем наплатних рампи на наплатној станици Врчин.

Даље модификације методологије могу да буду усмерене ка моделовању времена долазака клијената процесом чији интензитети долазака представљају рекурентну мрежу која се учи директно из функције веродостојности стохастичког процеса. Последично, даља примена Марковљевих процеса није могућа, стога је неопходно оптималну политику управљања наћи или моделима комбинаторне оптимизације или применом учења поткрепљивањем (енг. „*Reinforcement learning*“).

На основу представљених развијених методологија може се закључити да је тачност опште хипотезе $X(0)$ дисертације доказана кроз велики број нумеричких примера. Односно, показује се да су развијене методологије, које су настала интеграцијом алгоритама машинског учења и модела теорије редова чекања, значајно допринеле унапређењу методологије за решавање проблема предикције гужви и одређивања оптималног броја активних канала у модулу у будућности. Поред тога, приказана методологија за класификацију коришћењем новоразвијеног модела Гаусових условних случајних поља показује статистички боље резултате у односу на класичне моделе за предвиђање гужви. Овим је утврђена тачност посебне хипотезе $X(1)$. Као што је напоменуто, методологија за предвиђање индикатора саобраћаја, базирана на

комбинацији регресионих и класификационих Гаусових условних случајних поља, значајно повећава перформансе предикције у случајевима када су подаци ретки и излази корелисани. На овај начин утврђена је тачност посебне хипотезе $X(2)$. Методологија за одређивање броја активних канала у модулу, представљена комбинацијом модела за предикцију параметара средњег интензитета долазака возила заједно са теоријом редова чекања, омогућава одређивање оптималног броја канала опслуживања у модулу који треба да буде отворен у будућности. Овим је доказана и тачност посебне хипотезе $X(3)$.

Општи циљеви истраживања дисертације су у потпуности испуњени, узевши у обзир да су све три методологије за класификацију гужви коришћењем Гаусових условних случајних поља, предикцију индикатора брзине и одређивање броја активних канала у модулу у будућности развијене. Поред тога, у оквиру методологије за класификацију гужви дефинисана је метода за одређивање границе помоћу које се стања могу класификовати као „гужва“ или „није гужва“. Такође у оквиру исте је развијена апроксимативна верзија Бајесовских Гаусових условних случајних поља за смањење меморијског и рачунског трошка. Модели коришћени за класификацију гужви и предикцију индикатора брзине су генерализовани на проблеме из других домена. Додатно, у оквиру методологије за одређивање оптималног броја активних канала у модулу у будућности примена рекурентних мрежа у циљу предикције интензитета долазака клијената показала је задовољавајуће перформансе предикције. Методологија се поред тога показала као изузетно успешна на примеру одређивања оптималног броја наплатних рампи на наплатној станици у будућности. На основу претходног закључује се да су посебни циљеви дисертације испуњени.

Најзначајнији допринос дисертације је развој три методологије које се тичу предикције гужве у комплексним системима масовног опслуживања и одређивања оптималног броја активних канала опслуживања у будућности. Прво су развијене методологије које су се заснивале на новоразвијеним структурним класификационим алгоритмима заснованим на Гаусовим условним случајним пољима. Затим се показала могућност комбинације регресионих и класификационих алгоритама у циљу предикције гужви. Евалуација рада развијених алгоритама заједно са предностима и манама је детаљно продискутована и емпиријски евалуирана. Методологија која се заснива на комбинацији машинског учења заједно са теоријом редова чекања у циљу одређивања оптималног броја канала опслуживања који треба да буде отворен у будућности је у трећем делу дисертације детаљно описана и њена примена је приказана на реалном проблему.

Научни и инжењерски допринос дисертације се огледа у следећем:

1. Развојем методологије за класификацију гужви коришћењем Гаусових условних случајних поља
 - a. Прегледом савремених начина за предикцију гужви у системима масовног опслуживања.
 - b. Развојем Бајесовског и не-Бајесовског математичког модела за структурну класификацију базираног на Гаусовим условним случајним пољима.
 - c. Развојем апроксимативне брзе верзије Бајесовског модела Гаусових условних случајних поља за класификацију,
 - d. Применом методологије на проблем предикције гужви на ауто-путевима.
 - e. Генерализацијом методологије на проблеме из других домена.
2. Развојем методе за комбинацију класификационих и регресионих структурираних предиктора у циљу одређивања индикатора саобраћаја.
 - a. Евалуација рада развијене методологије на проблем предвиђања индикатора саобраћаја на мрежном систему са ретким подацима.
 - b. Поређење развијене методологије са методологијама које се не заснивају на одстрањивању података без вредности.
 - c. Генерализацијом методологије на проблем из другог домена.
3. Развојем методологије за одређивање оптималног броја отворених канала опслуживања у будућности.
 - a. Повезивањем алгоритама машинског учења са теоријом редова чекања односно моделима система опслуживања
 - b. Развијање Бајесовског и не-Бајесовског алгорита.
 - c. Применом и валидацијом резултата на примеру одређивања оптималног броја активних канала у модулу.
 - d. Верификацијом резултата методологије за одређивање оптималног броја активних канала у модулу посредством симулације.

На основу доказивања тачности свих хипотеза постављених у овој дисертацији, испуњењем општих и посебних циљева дисертације и приказивању научног и инжењерског доприноса исте може се закључити да приказане методологије на задовољавајући начин решавају савремене саобраћајне проблеме.

Узевши у обзир да је све већи проблем са саобраћајним гужвама присутан у свакодневном животу и да исти има тенденцију раста током наредних година, засновано се претпоставља да је истраживање у овом правцу плодносно, како за научну, тако и за

стручну заједницу. Досадашњи резултати из области упућују да ће ова област бити интересантна и у годинама које следе. Посебно је интересантно напоменути да је предвиђање гужви и налажење оптималних величина којим се гужве регулишу, још увек нерешено и научно изазовно питање, стога се претпоставља да ће новоразвијени алгоритми и методологије бити од великог теоријског и практичног значаја у временима која долазе.

10 Листа слика

Слика 1 Гаусова условна случајна поља за регресију.....	22
Слика 2 Процес опслуживања	23
Слика 3 Механизам опслуживања	25
Слика 4 Систем опслуживања D/D/1	27
Слика 5 Илустрација рекурентне неуронске мреже.....	35
Слика 6 Гаусова условна случајна поља за класификацију	44
Слика 7 Облик функције $\sigma(z)N(z_i \mu_i, \sigma^2)$ за три различите варијансе нормалне расподеле	46
Слика 8 Сигмоидна функција са доњим границама.....	48
Слика 9 Време учења GCRFBCb и GCRFBCnb модела у зависности од броја инстанци и броја излаза.....	57
Слика 10 Алгоритам учења убрзане верзије Гаусових условних случајних поља за класификацију са кластеровањем инстанци и излаза.....	
Слика 11 Алгоритам учења убрзане верзије Гаусових условних случајних поља за класификацију са одређивањем вероватноће припадања кластеру	
Слика 12 Шематски приказ модела заснованог на комбинацији регресионих и класификационих Гаусових условних случајних поља	67
Слика 13 Учење хибридног модела	
Слика 14 Алгоритам 1 за налажење оптималног броја активних канала у модулу у неком временском периоду у будућности.....	71
Слика 15 Алгоритам 2 за налажење оптималног броја активних канала у модулу у неком временском периоду у будућности.....	72
Слика 16 Приказ анализираних деоница Београд – Адашевци и Ниш – Београд. Црвеним тачкама означене су наплатне рампе на којима се предвиђала гужва	74
Слика 17 Емпиријска расподела средње брзине возила која долазе на наплатне рампе на деоници Београд-Адашевци.....	76
Слика 18 Емпиријска расподела средње брзине возила која долазе на наплатне рампе на деоници Ниш-Београд	77
Слика 19 Шематски приказ расподеле интензитета долазака возила на наплатним рампама	90

Слика 20 Интензитети долазака возила и прва диференца интензитета долазака возила	95
Слика 21 Хистограм емпиријских података и померена експоненцијална расподела ..	97
Слика 22 Промена вероватноће стања система за предвиђане вредности интензитета долазака (задебљана линија) и стварне вредности интензитета долазака возила (танка линија)) у случају одређивања оптималног броја отворених наплатних рампи за два подинтервала	103
Слика 23 Промена броја канала у времену (горња слика) и промена вредности функције укупних трошкова током времена (доња слика)) у случају одређивања оптималног броја отворених наплатних рампи за два подинтервала.....	103
Слика 24 Промена вероватноће стања система за предвиђане вредности интензитета долазака (задебљана линија) и стварне вредности интензитета долазака возила (танка линија)) у случају одређивања оптималног броја отворених наплатних рампи за шест подинтервала	104
Слика 25 Промена броја канала у времену (горња слика) и промена вредности функције укупних трошкова током времена (доња слика)) у случају одређивања оптималног броја отворених наплатних рампи за шест подинтервала	104
Слика 26 Промена вероватноће стања система за предвиђане вредности интензитета долазака (задебљана линија) и стварне вредности интензитета долазака возила (танка линија)) у случају одређивања оптималног броја отворених наплатних рампи за петнаест подинтервала	105
Слика 27 Промена броја канала у времену (горња слика) и промена вредности функције укупних трошкова током времена (доња слика) у случају одређивања оптималног броја отворених наплатних рампи за петнаест подинтервала.....	105
Слика 28 Промена вероватноће стања система за предвиђане вредности интензитета долазака (задебљана линија) и стварне вредности интензитета долазака возила (танка линија) у случају одређивања оптималног броја отворених наплатних рампи за три стотине подинтервала са континуалним вредностима (горња слика) и заокруженим (доња слика) управљања.	106
Слика 29 Промена броја канала у времену (горња слика) и промена вредности функције укупних трошкова током времена (доња слика) у случају одређивања оптималног броја отворених наплатних рампи за три стотине подинтервала – континуалне вредности	107

Слика 30 Промена броја канала у времену (горња слика) и промена вредности функције укупних трошкова током времена (доња слика) у случају одређивања оптималног броја отворених наплатних рампи за три стотине подинтервала – заокружене вредности	108
Слика 31 Промена вероватноће стања система за предвиђане вредности интензитета долазака у случају одређивања оптималног броја отворених наплатних рампи за шест подинтервала одређених Бајесовским приступом са континуалним вредностима (горња слика) и заокруженим вредностима (доња слика) управљања.....	109
Слика 32 Промена вероватноће стања система за предвиђане вредности интензитета долазака (задебљана линија) и стварне вредности интензитета долазака возила (танка линија) у случају одређивања оптималног броја отворених наплатних рампи за шест подинтервала одређених Бајесовским приступом.....	110
Слика 33 Промена броја канала у времену (горња слика) и промена вредности функције укупних трошкова током времена (доња слика) у случају одређивања оптималног броја отворених наплатних рампи за шест подинтервала одређене-Бајесовским приступом.....	110
Слика 34 Поређење вероватноћа промене стања система добијеног симулацијом (обојене линије) и решавањем система диференцијалних једначина (црне линије) у случају одређивања оптималног броја отворених наплатних рампи за шест подинтервала са 200 (горњи дијаграм), 1000 (средњи дијаграм) и 3000 (доњи дијаграм) узорака у симулацији.....	112
Слика 35 Поређење вероватноћа промене стања система добијеног симулацијом (обојене линије) и решавањем система диференцијалних једначина (црне линије) у случају одређивања оптималног броја отворених наплатних рампи за три стотине подинтервала са 200 (горњи дијаграм), 1000 (средњи дијаграм) и 3000 (доњи дијаграм) узорака у симулацији.....	113
Слика 36 Мапа стаза ски-центра Копаоник – црвеном су приказане стазе на којима се предвиђала гужва	116
Слика 37 Емпиријска расподела времена спуштања скијаша на анализираним стазама ски-центра Копаоник.....	117

11 Листа табела

Табела 1 Поређење предиктивних перформанси GCRFBC _b и GCRFBC _{nb} модела за различите вредности параметара α и β , измерених помоћу ПИК скорa, условне логаритамске функције веродостојности и нормe дијагоналних елемената коваријационе матрице латентних променљивих	56
Табела 2 Групе и опис атрибута коришћених при предвиђању гужви на ауто-путевима	78
Табела 3 Перформансе предикције и време израчунавања модела за проблем предивђања гужви на деоници Београд - Адашевци	80
Табела 4 Перформансе предикције и време израчунавања модела за проблем предивђања гужви на деоници Ниш-Београд	81
Табела 5 Перформансе предикције и време израчунавања регресионих модела за проблем одређивања просечне брзине на деоницима Београд – Адашевци и Ниш - Београд	85
Табела 6 Перформансе предикције и време израчунавања комбиноване методологије за проблем одређивања просечне брзине на деоници Београд-Адашевци	85
Табела 7 Перформансе предикције и време израчунавања комбиноване методологије за проблем одређивања просечне брзине на деоници Ниш - Београд	86
Табела 8 Коришћени атрибути при учењу рекурентне неуронске мреже	94
Табела 9 Вредности Кулбак-Лајблер дивергенције и тестова	98
Табела 10 Перформансе предикције RNN мреже	99
Табела 11 Перформансе предикције GRU мреже	100
Табела 12 Перформансе предикције LSTM мреже	100
Табела 13 Перформансе предикције LSTM-ln мреже	101
Табела 14 Перформансе предикције LSTM-ln мреже за предвиђање грешке предикције параметара расподеле	102
Табела 15 Очекивани укупни трошкови за разичите промене броја канала	111
Табела 16 Апсолутна, релативна грешка и оцена грешке	113
Табела 17 Групе и опис атрибута коришћених при предвиђању гужви на скијалишту	118
Табела 18 Перформансе предикције и време израчунавања модела за проблем предивђања гужви на ски-центру Копаоник	119

Табела 19 Перформансе предикције и време израчунавања регресионих модела за проблем одређивања просечног времена скијаша, потребног да направи круг на ски стази на ски-центру Копаоник.....	121
Табела 20 Перформансе предикције и време израчунавања комбиноване методологије за проблем одређивања просечног времена скијаша, потребног да направи круг на ски стази, на ски-центру Копаоник.....	122
Табела 21 Перформансе предикције и време израчунавања модела за проблем класификације гена.....	123
Табела 22 Перформансе предикције и време израчунавања модела за проблем класификације сцена.....	126
Табела 23 Перформансе предикције и време израчунавања модела за проблем класификације емоција из музике	128
Табела 24 Перформансе предикције и време израчунавања модела за проблем класификације болести.....	129
Табела 25 Перформансе предикције и време израчунавања модела за проблем класификације птица	131

12 Литература

1. Foster, K.R., R. Koprowski, and J.D. Skufca, *Machine learning, medical diagnosis, and biomedical engineering research-commentary*. Biomedical engineering online, 2014. **13**(1): p. 94.
2. Cui, G., M.L. Wong, and H.-K. Lui, *Machine learning for direct marketing response models: Bayesian networks with evolutionary programming*. Management Science, 2006. **52**(4): p. 597-612.
3. Pelossof, R., et al. *An SVM learning approach to robotic grasping*. in *Robotics and Automation, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on*. 2004. IEEE.
4. Petrovic, A., et al., *Evaluating performances of 1-D models to predict variable area supersonic gas ejector performances*. Energy, 2018. **163**: p. 270-289.
5. Arentze, T.A. and H.J. Timmermans, *A learning-based transportation oriented simulation system*. Transportation Research Part B: Methodological, 2004. **38**(7): p. 613-633.
6. Murphy, K.P., *Machine Learning: A Probabilistic Perspective*. Adaptive Computation and Machine Learning. 2012, MIT press.
7. Su, H., *Multilabel Classification through Structured Output Learning-Methods and Applications*. 2015.
8. Ma, X., et al., *Large-scale transportation network congestion evolution prediction using deep learning theory*. PloS one, 2015. **10**(3): p. e0119044.
9. Lv, Y., et al., *Traffic flow prediction with big data: A deep learning approach*. IEEE Trans. Intelligent Transportation Systems, 2015. **16**(2): p. 865-873.
10. Kumar, K., M. Parida, and V. Katiyar, *Short term traffic flow prediction for a non urban highway using artificial neural network*. Procedia-Social and Behavioral Sciences, 2013. **104**: p. 755-764.
11. Gers, F.A., D. Eck, and J. Schmidhuber, *Applying LSTM to time series predictable through time-window approaches*, in *Neural Nets WIRN Vietri-01*. 2002, Springer. p. 193-200.
12. Lei Ba, J., J.R. Kiros, and G.E. Hinton, *Layer normalization*. arXiv preprint arXiv:1607.06450, 2016.
13. Chung, J., et al., *Empirical evaluation of gated recurrent neural networks on sequence modeling*. arXiv preprint arXiv:1412.3555, 2014.

14. Han, M., et al., *Prediction of chaotic time series based on the recurrent predictor neural network*. IEEE transactions on signal processing, 2004. **52**(12): p. 3409-3416.
15. Whittaker, J., *Graphical models in applied multivariate statistics*. 2009: Wiley Publishing.
16. Beal, M.J., Z. Ghahramani, and C.E. Rasmussen. *The infinite hidden Markov model*. in *Advances in neural information processing systems*. 2002.
17. Welch, G. and G. Bishop, *An introduction to the Kalman filter*. 1995.
18. Blei, D.M., A.Y. Ng, and M.I. Jordan, *Latent dirichlet allocation*. Journal of machine Learning research, 2003. **3**(Jan): p. 993-1022.
19. Howson, C. and P. Urbach, *Scientific reasoning: the Bayesian approach*. 2006: Open Court Publishing.
20. Ng, A.Y. and M.I. Jordan. *On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes*. in *Advances in neural information processing systems*. 2002.
21. Koller, D., N. Friedman, and F. Bach, *Probabilistic graphical models: principles and techniques*. 2009: MIT press.
22. Sutton, C. and A. McCallum, *An introduction to conditional random fields*. Foundations and Trends® in Machine Learning, 2012. **4**(4): p. 267-373.
23. Cotterell, R. and K. Duh. *Low-Resource Named Entity Recognition with Cross-Lingual, Character-Level Neural Conditional Random Fields*. in *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*. 2017.
24. Zhang, P., et al., *Hierarchical conditional random fields model for semisupervised SAR image segmentation*. IEEE Transactions on Geoscience and Remote Sensing, 2015. **53**(9): p. 4933-4951.
25. Zia, H.B., A.A. Raza, and A. Athar, *Urdu Word Segmentation using Conditional Random Fields (CRFs)*. arXiv preprint arXiv:1806.05432, 2018.
26. Kim, M., *Mixtures of conditional random fields for improved structured output prediction*. IEEE transactions on neural networks and learning systems, 2017. **28**(5): p. 1233-1240.
27. Van Der Maaten, L., M. Welling, and L. Saul. *Hidden-unit conditional random fields*. in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*. 2011.

28. Chen, G., Y. Li, and S.N. Srihari, *Word recognition with deep conditional random fields*. arXiv preprint arXiv:1612.01072, 2016.
29. Kosov, S., et al., *Environmental microorganism classification using conditional random fields and deep convolutional neural networks*. *Pattern Recognition*, 2018. **77**: p. 248-261.
30. Hoffman, M.D., et al., *Stochastic variational inference*. *The Journal of Machine Learning Research*, 2013. **14**(1): p. 1303-1347.
31. Tappen, M.F., et al. *Learning gaussian conditional random fields for low-level vision*. in *2007 IEEE Conference on Computer Vision and Pattern Recognition*. 2007. IEEE.
32. Radosavljevic, V., S. Vucetic, and Z. Obradovic. *Neural gaussian conditional random fields*. in *Joint European conference on machine learning and knowledge discovery in databases*. 2014. Springer.
33. Glass, J., et al. *Extending the modelling capacity of Gaussian conditional random fields while learning faster*. in *Thirtieth AAAI Conference on Artificial Intelligence*. 2016.
34. Stojanovic, J., et al. *Semi-supervised learning for structured regression on partially observed attributed graphs*. in *Proceedings of the 2015 SIAM International Conference on Data Mining*. 2015. SIAM.
35. Radosavljevic, V., S. Vucetic, and Z. Obradovic. *Continuous Conditional Random Fields for Regression in Remote Sensing*. in *ECAI*. 2010.
36. Wytock, M. and Z. Kolter. *Sparse Gaussian conditional random fields: Algorithms, theory, and application to energy forecasting*. in *International conference on machine learning*. 2013.
37. Frot, B., L. Jostins, and G. McVean, *Graphical model selection for Gaussian conditional random fields in the presence of latent variables*. *Journal of the American Statistical Association*, 2018: p. 1-12.
38. Fan, S.-K.S., et al., *Using machine learning and big data approaches to predict travel time based on historical and real-time data from Taiwan electronic toll collection*. *Soft Computing*, 2018. **22**(17): p. 5707-5718.
39. Zhao, J., et al., *Highway Travel Time Prediction Using Sparse Tensor Completion Tactics and-Nearest Neighbor Pattern Matching Method*. *Journal of Advanced Transportation*, 2018. **2018**.
40. Jabbarpour, M.R., et al., *Applications of computational intelligence in vehicle traffic congestion problem: a survey*. *Soft Computing*, 2018. **22**(7): p. 2299-2320.

41. Seo, T., et al., *Traffic state estimation on highway: A comprehensive survey*. Annual Reviews in Control, 2017. **43**: p. 128-151.
42. Okutani, I. and Y.J. Stephanedes, *Dynamic prediction of traffic volume through Kalman filtering theory*. Transportation Research Part B: Methodological, 1984. **18**(1): p. 1-11.
43. Sun, S. and X. Xu, *Variational inference for infinite mixtures of Gaussian processes with applications to traffic flow prediction*. IEEE Transactions on Intelligent Transportation Systems, 2010. **12**(2): p. 466-475.
44. Yu, B., et al., *k-Nearest neighbor model for multiple-time-step prediction of short-term traffic condition*. Journal of Transportation Engineering, 2016. **142**(6): p. 04016018.
45. Polson, N.G. and V.O. Sokolov, *Deep learning for short-term traffic flow prediction*. Transportation Research Part C: Emerging Technologies, 2017. **79**: p. 1-17.
46. Coric, V., Z. Wang, and S. Vucetic. *Traffic speed forecasting by mixture of experts*. in *2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*. 2011. IEEE.
47. Li, K., C. Yuen, and S. Kanhere. *Senseflow: An experimental study of people tracking*. in *Proceedings of the 6th ACM Workshop on Real World Wireless Sensor Networks*. 2015. ACM.
48. Wu, F.-J. and G. Solmaz. *Are you in the line? RSSI-based queue detection in crowds*. in *2017 IEEE International Conference on Communications (ICC)*. 2017. IEEE.
49. Fonner, R.C. and R.P. Berrens, *A hedonic pricing model of lift tickets for US alpine ski areas: examining the influence of crowding*. Tourism Economics, 2014. **20**(6): p. 1215-1233.
50. Poulhès, A. and P. Mirial, *Dynaski, an Agent-based Model to Simulate Skiers in a Ski area*. Procedia Computer Science, 2017. **109**: p. 84-91.
51. Barras, T., et al., *Queues in Ski Resort Graphs: the Ski-Optim Model*. 2014.
52. Holleczeck, T. and G. Tröster, *Particle-based model for skiing traffic*. Physical Review E, 2012. **85**(5): p. 056101.
53. Ermagun, A. and D. Levinson, *Spatiotemporal traffic forecasting: review and proposed directions*. Transport Reviews, 2018. **38**(6): p. 786-814.
54. Sadek, A., *Artificial intelligence in transportation: information for application*. Transportation Research Board Circular (E-C113), TRB, National Research Council, Washington, DC Available online at: < <http://onlinepubs.trb.org/onlinepubs/circulars/ec113.pdf>, 2007.

55. Neumann, T., et al. *Deriving traffic volumes from probe vehicle data using a fundamental diagram approach*. in *Proceedings 13th World Conference on Transport Research*. 2013.
56. Anuar, K., F. Habtemichael, and M. Cetin. *Estimating traffic flow rate on freeways from probe vehicle data and fundamental diagram*. in *2015 IEEE 18th International Conference on Intelligent Transportation Systems*. 2015. IEEE.
57. Bulteau, E., et al. *Traffic flow estimation using higher-order speed statistics*. in *Transportation Research Board 92nd Annual Meeting*. 2013.
58. Qiu, C.M. and W. Zhang, *Managing long queues for holiday sales shopping*. *Journal of Revenue and Pricing Management*, 2016. **15**(1): p. 52-65.
59. Drew, D.R., *Traffic flow theory and control*. 1968.
60. Vandaele, N., T. Van Woensel, and A. Verbruggen, *A queueing based traffic flow model*. *Transportation Research Part D: Transport and Environment*, 2000. **5**(2): p. 121-135.
61. Neely, M.J., *Stochastic network optimization with application to communication and queueing systems*. *Synthesis Lectures on Communication Networks*, 2010. **3**(1): p. 1-211.
62. Ishak, S. and C. Alecsandru, *Optimizing traffic prediction performance of neural networks under various topological, input, and traffic condition settings*. *Journal of Transportation Engineering*, 2004. **130**(4): p. 452-465.
63. Neuhold, R., et al., *Predicting and Optimizing Traffic Flow at Toll Plazas*. *Transportation Research Procedia*, 2019. **37**: p. 330-337.
64. Mihăiță, A.S., L. Dupont, and M. Camargo, *Multi-objective traffic signal optimization using 3D mesoscopic simulation and evolutionary algorithms*. *Simulation Modelling Practice and Theory*, 2018. **86**: p. 120-138.
65. Fu, Y., et al., *Artificial intelligence to manage network traffic of 5G wireless networks*. *IEEE Network*, 2018. **32**(6): p. 58-64.
66. Papageorgiou, M., et al., *Review of road traffic control strategies*. *Proceedings of the IEEE*, 2003. **91**(12): p. 2043-2067.
67. Castillo, E., et al., *A state-of-the-art review of the sensor location, flow observability, estimation, and prediction problems in traffic networks*. *Journal of Sensors*, 2015. **2015**.
68. Radosavljevic, V., *Gaussian conditional random fields for regression in remote sensing*. 2011: Temple University.

69. Bugarić, U. and D. Petrović, *Modeliranje sistema opsluživanja*. Mašinski fakultet Beograd, Beograd, 2011.
70. Brownlee, K.A., *Statistical theory and methodology in science and engineering*. Vol. 150. 1965: Wiley New York.
71. Lewis, P.W. and G.S. Shedler, *Simulation of nonhomogeneous Poisson processes by thinning*. Naval research logistics quarterly, 1979. **26**(3): p. 403-413.
72. Bishop, C.M., *Pattern recognition and machine learning*. 2006: springer.
73. Connor, J.T., R.D. Martin, and L.E. Atlas, *Recurrent neural networks and robust time series prediction*. IEEE transactions on neural networks, 1994. **5**(2): p. 240-254.
74. Goodfellow, I., Y. Bengio, and A. Courville, *Deep learning*. 2016: MIT press.
75. Mandic, D.P. and J. Chambers, *Recurrent neural networks for prediction: learning algorithms, architectures and stability*. 2001: John Wiley & Sons, Inc.
76. Jozefowicz, R., W. Zaremba, and I. Sutskever. *An empirical exploration of recurrent network architectures*. in *International Conference on Machine Learning*. 2015.
77. Sundermeyer, M., H. Ney, and R. Schlüter, *From feedforward to recurrent LSTM neural networks for language modeling*. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2015. **23**(3): p. 517-529.
78. Ioffe, S. and C. Szegedy, *Batch normalization: Accelerating deep network training by reducing internal covariate shift*. arXiv preprint arXiv:1502.03167, 2015.
79. Jeter, M., *Mathematical programming: an introduction to optimization*. 2018: Routledge.
80. Nelder, J.A. and R. Mead, *A simplex method for function minimization*. The computer journal, 1965. **7**(4): p. 308-313.
81. Bertsekas, D.P., *Nonlinear programming*. Journal of the Operational Research Society, 1997. **48**(3): p. 334-334.
82. Davis, L., *Handbook of genetic algorithms*. 1991.
83. Mirjalili, S., *The ant lion optimizer*. Advances in Engineering Software, 2015. **83**: p. 80-98.
84. Chang, R.-S., J.-S. Chang, and P.-S. Lin, *An ant algorithm for balanced job scheduling in grids*. Future Generation Computer Systems, 2009. **25**(1): p. 20-27.
85. Storn, R. and K. Price, *Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces*. Journal of global optimization, 1997. **11**(4): p. 341-359.

86. Tan, C., et al. *Social action tracking via noise tolerant time-varying factor graphs*. in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2010. ACM.
87. Davis, P.J. and P. Rabinowitz, *Methods of numerical integration*. 2007: Courier Corporation.
88. Jaakkola, T.S. and M.I. Jordan, *Bayesian parameter estimation via variational methods*. *Statistics and Computing*, 2000. **10**(1): p. 25-37.
89. Scheithauer, G., *Jorge Nocedal and Stephen J. Wright: Numerical Optimization, Springer Series in Operations Research, 1999, ISBN 0-387-98793-2* In the preface the authors state... *Our goal in this book is to give a comprehensive description of the most powerful, state-of-the-art, techniques for solving continuous optimization problems. By presenting the motivating ideas for each*. Springer.
90. Facchinei, F., S. Lucidi, and L. Palagi, *A truncated Newton algorithm for large scale box constrained optimization*. *SIAM Journal on Optimization*, 2002. **12**(4): p. 1100-1125.
91. Boggs, P.T. and J.W. Tolle, *Sequential quadratic programming*. *Acta numerica*, 1995. **4**: p. 1-51.
92. Fitriyani, S.R. and H. Murfi. *The K-means with mini batch algorithm for topics detection on online news*. in *2016 4th International Conference on Information and Communication Technology (ICoICT)*. 2016. IEEE.
93. Moon, T.K., *The expectation-maximization algorithm*. *IEEE Signal processing magazine*, 1996. **13**(6): p. 47-60.
94. Taskar, B., et al. *Learning structured prediction models: A large margin approach*. in *Proceedings of the 22nd international conference on Machine learning*. 2005. ACM.
95. Kamarianakis, Y., W. Shen, and L. Wynter, *Real-time road traffic forecasting using regime-switching space-time models and adaptive LASSO*. *Applied stochastic models in business and industry*, 2012. **28**(4): p. 297-315.
96. Yang, S. and S. Qian, *Understanding and predicting travel time with spatio-temporal features of network traffic flow, weather and incidents*. arXiv preprint arXiv:1901.06766, 2019.
97. Yu, D., et al., *Forecasting short-term traffic speed based on multiple attributes of adjacent roads*. *Knowledge-Based Systems*, 2019. **163**: p. 472-484.
98. Mohri, M., A. Rostamizadeh, and A. Talwalkar, *Foundations of machine learning*. 2018: MIT press.

99. Massey Jr, F.J., *The Kolmogorov-Smirnov test for goodness of fit*. Journal of the American statistical Association, 1951. **46**(253): p. 68-78.
100. Elisseeff, A. and J. Weston. *A kernel method for multi-labelled classification*. in *Advances in neural information processing systems*. 2002.
101. Pavlidis, P. and W.N. Grundy, *Combining microarray expression data and phylogenetic profiles to learn gene functional categories using support vector machines*. 2000.
102. Boutell, M.R., et al., *Learning multi-label scene classification*. Pattern recognition, 2004. **37**(9): p. 1757-1771.
103. Trohidis, K., et al. *Multi-label classification of music into emotions*. in *ISMIR*. 2008.
104. Briggs, F., et al. *New methods for acoustic classification of multiple simultaneous bird species in a noisy environment*. in *IEEE International Workshop on Machine Learning for Signal Processing*. 2013.

Прилог 1.

Биографија кандидата

Андрија Петровић рођен је у Београду 23.08.1991. Основну школу „Љуба Ненадовић“ завршио је 2006. године у Београду, као носилац Вукове дипломе и ђак године. Средњу школу „13. београдску гимназију“ завршио је 2010. године у Београду и награђен је за изузетан успех током школовања. Основне студије на Машинском факултету Универзитета у Београду уписао је 2010. године и завршио 2013. године са просечном оценом 9,91. Мастер студије на Машинском факултету Универзитета у Београду, смер Процесна техника и заштита животне средине, уписао је 2013. године и завршио 2015. године са просечном оценом 9,90. Мастер рад је одбранио са оценом 10 на тему „Концептуално решење и економска процена рада система за производњу електричне енергије на бази органског ранкиновог циклуса“. Дипломирао је на Факултету за Пословну економију Универзитета Сингидунум 2014. године, смер финансије и банкарство, са просечном оценом 9,45. Мастер студије на модлулу Пословна економија завршио је 2016. године са просечном оценом 10,00. Мастер рад је одбранио са оценом 10 на тему „Динамика промене вредности емисионих дозвола и рентабилност инвестиције тригенеративног постројења“.

Тренутно је студент треће године докторских студија на Факултету организационих наука Универзитета у Београду, где је положио све испите на докторским студијама и студент је треће године докторских студија на Машинском факултету Универзитета у Београду. Докторске студије на Факултету организационих наука кандидат је уписао школске године 2015/16. Докторске студије на Машинском факултету кандидате је уписао школске године 2015/16.

Награђиван је за изузетан успех током студија на Машинском факултету 2010/11, 2011/12, 2012/13, 2013/14, 2014/15 године. Награђен је као најбољи студент на основним академским студијама генерације уписане 2010/11 године, као најбољи студент на мастер академским студијама генерације уписане 2013/14 године и као најбољи студент генерације уписане 2010/11 године. Добитник награде за најбољи мастер рад на конгресу Процесинг 2016.

Био је стипендиста града Београда за 2011/12 годину, стипендиста Републике Србије за 2013/14 годину, стипендиста Фонда за младе таленте“ за 2012/13 и 2014/15

годину, стипендиста „Термоелектране Никола Тесла“ за годину 2012/13, стипендиста фирме „Ејектоинжењеринг Београд“ за 2012/13, 2013/14 и 2014/15 годину.

Учествовао је у раду летње школе организоване под покровитељством NTNU (Norwegian University of Science and Technology) у оквиру пројекта „Herd energy project - quality improvement of masters programs in sustainable energy and environment“. Био је практикант фирме „Про-инг Београд“ у периоду јун-август 2011 године. Похађао је више курсева организованих од стране фирме „Merit Plan“. У оквиру Erasmus+ програма размене студената похађао је 2. семестар прве године докторских студија школске 2015./16. на Машинском факултету Универзитета Политехника Темишвар. Стипендиран је од компаније Сага д.о.о. од маја 2018 на пројекту развоја алгоритама за машинско учење заједно са професорима са Факултета организационих наука.

Запослен је од 2018. године као асистент на катедри за Индустијско инжењерство на Машинском факултету Универзитета у Београду.

Поседује сертификат FCE– B2 познавање енглеског језика. Служи се руским језиком. Активно користи програмерске софтвере Visual Basic, MatLab, Python, Fortran, Delphi, Eviews, Ansys - Fluent, AMPL, SQL. Зна да употребљава софтверске пакете за обраду текста MS OFFICE (Word, Excell, Power point) и Latex, софтвере за техничко цртање AutoCAD и Solid Works. Поседује возачку дозволу Б категорије. Објавио је до сада 7 радова на SCI листи од тога четири M21a категорије, један M21 категорије и два M22 категорије.

Прилог 2.

Изјава о ауторству

Потписани _____ Андрија А. Петровић _____

Број индекса _____ 5006/2015 _____

Изјављујем

да је докторска дисертација под насловом

Развој интегрисаног модела за предикцију гужви и одређивање

оптималног броја активних канала у модулу

- резултат сопственог истраживачког рада,
- да предложена дисертација у целини ни у деловима није била предложена за добијање било које дипломе према студијским програмима других високошколских установа,
- да су резултати коректно наведени и
- да нисам кршио ауторска права и користио интелектуалну својину других лица.

Потпис докторанда

У Београду, 30.10.2019

Прилог 3.

**Изјава о истоветности штампане и електронске верзије
докторског рада**

Име и презиме аутора Андрија А. Петровић
Број индекса 5006/2015
Студијски програм Квантитативни менаџмент
Наслов рада **Развој интегрисаног модела за предикцију гужви и одређивање
оптималног броја активних канала у модулу**
Ментор проф. др Борис Делибашић;
Коментор проф. др Угљеша Бугарић

Потписани Андрија А. Петровић

Изјављујем да је штампана верзија мог докторског рада истоветна електронској верзији коју сам предао за објављивање на порталу **Дигиталног репозиторијума Универзитета у Београду**.

Дозвољавам да се објаве моји лични подаци везани за добијање академског звања доктора наука, као што су име и презиме, година и место рођења и датум одбране рада. Ови лични подаци могу се објавити на мрежним страницама дигиталне библиотеке, у електронском каталогу и у публикацијама Универзитета у Београду.

Потпис докторанда

У Београду, 30.10.2019

Прилог 4.

Изјава о коришћењу

Овлашћујем Универзитетску библиотеку „Светозар Марковић“ да у Дигитални репозиторијум Универзитета у Београду унесе моју докторску дисертацију под насловом:

**Развој интегрисаног модела за предикцију гужви и одређивање
оптималног броја активних канала у модулу**

која је моје ауторско дело.

Дисертацију са свим прилозима предао сам у електронском формату погодном за трајно архивирање.

Моју докторску дисертацију похрањену у Дигитални репозиторијум Универзитета у Београду могу да користе сви који поштују одредбе садржане у одабраном типу лиценце Креативне заједнице (Creative Commons) за коју сам се одлучио.

①. Ауторство

2. Ауторство - некомерцијално

3. Ауторство – некомерцијално – без прераде

4. Ауторство – некомерцијално – делити под истим условима

5. Ауторство – без прераде

6. Ауторство – делити под истим условима

(Молимо да заокружите само једну од шест понуђених лиценци, кратак опис лиценци дат је на полеђини листа).

Потпис докторанда

У Београду, 30.10.2019
