

Већу за студије при Универзитету

Предмет: Извештај о урађеној докторској дисертацији кандидата Милоша Симића

Одлуком Већа за студије при универзитету бр. 06-4217/IX-1658/4-24 од 8.7.2024. године, именовани смо за чланове Комисије за преглед, оцену и одбрану докторске дисертација кандидата Милоша Симића, под насловом:

Оптимизација алгоритамаских шесћова нормалности заснованих на неуронским мрежама

После прегледа достављене дисертације и других пратећих материјала и разговора с кандидатом, Комисија је сачинила следећи

ИЗВЕШТАЈ

1 Увод

1.1 Хронологија одобравања и израде докторске дисертације

Кандидат Милош Симић уписао је интердисциплинарне докторске студије Универзитета у Београду, смер Интелигентни системи, 2018. године. Током студија је положио све испите и испунио све обавезе у вези са студијским истраживачким радом предвиђене планом и програмом.

Кандидат је пријавио тему докторске дисертације под насловом *Оптимизација алгоритамаских шесћова нормалности заснованих на неуронским мрежама*, а Веће за студије при Универзитету је на седници одржаној 15. маја 2023. године донело одлуку број 06-4594/VII-1658/2-23 о именовању чланова комисије за оцену научне заснованости теме докторске дисертације под насловом *Оптимизација алгоритамаских шесћова нормалности заснованих на неуронским мрежама*. Веће за интердисциплинарне, мултидисциплинаре и трансдисциплинарне студије на седници одржаној 30. августа 2023. године, донело је одлуку број 06-355/III-1658/4-23 којом је усвојен извештај комисије за оцену научне заснованости предложене теме докторске дисертације под насловом *Оптимизација алгоритамаских шесћова нормалности заснованих на неуронским мрежама* и којом су за менторе ове докторске дисертације именовани др Зорица Станимировић, редовни професор, и др Срђан Станковић, професор емеритус.

Веће за студије при Универзитету је на седници одржаној 8.7.2024. године донело одлуку број 06-4217/IX-1658/4-24 да се образује комисија за преглед, оцену и одбрану докторске дисертације под насловом *Оптимизација алгоритамаских шесћова нормалности заснованих на неуронским мрежама* кандидата Милоша Симића, у саставу:

1. Др Бошко Николић, редовни професор, Електротехнички факултет, Универзитет у Београду,

2. Др Вељко Јеремић, редовни професор, Факултет организационих наука, Универзитет у Београду
3. Др Стефан Мишковић, доцент, Математички факултет, Универзитет у Београду

1.2 Научна област дисертације

Докторска дисертација припада научним областима Математика и Рачунарске науке и ужим областима: Вештачка интелигенција, Статистика и Оптимизација, при чему тема има мултидисциплинарни карактер. Мултидисциплинарност дисертације потиче из два извора: методологије и применљивости резултата. Са методолошког аспекта, мултидисциплинарност се огледа у томе што се методе искоришћене у дисертацији налазе у пресеку статистике и оптимизације као грана математике, и вештачке интелигенције као гране рачунарских наука и интердисциплинарног поља уско повезаног са бројним другим научним областима. Са аспекта примене добијених резултата, мултидисциплинарност потиче из тога што се тестови нормалности користе у скоро свим областима науке које се служе статистиком у испитивању научних хипотеза и провери ваљаности математичких модела.

Ментори докторске дисертације су др Зорица Станимировић, редовни професор Математичког факултета Универзитета у Београду, и др Срђан Станковић, професор емеритус Електротехничког факултета Универзитета у Београду. Наведени ментори су аутори великог броја научних радова у истакнутим међународним часописима и испуњавају све формалне и законске услове за менторе ове дисертације. Релевантни радови ментора наведени су приликом пријаве теме докторске дисертације.

1.3 Биографски подаци о кандидату

Милош Симић рођен је 1991. године у Крагујевцу, где је завршио основну школу *Живадинка Дивац* и природно-математички смер Прве крагујевачке гимназије као носилац Вукове дипломе. Основне академске студије информатике завршио је 2014. године на Природно-математичком факултету у Крагујевцу са просечном оценом 10,00 и као студент генерације, са дипломским радом *Бореалис – груписивена мрежа за књижевна сиваралашииво и кришику* под менторством професора доктора Владимира Цвјетковића. На истој установи је 2015. године завршио мастер студије информатике са просеком 10,00, одбранивши мастер рад *Примена оптимизације колонијом ичела у испитивању задовољивости формула вероватности логике $LPPP_2^{ext}$* под менторством професора доктора Небојше Икодиновића. Добитник је стипендије Фонда за младе таленте за најбоље студенте завршних година основних и мастер академских студија, као и Универзитета у Крагујевцу за најбоље студенте завршних година основних академских студија. Од 2018. године је студент докторских студија Интелигентних система при Универзитету у Београду, где се бави комбиновањем статистике и вештачке интелигенције под менторством доктора Срђана Станковића, професора емеритуса Универзитета у Београду, и др Зорице Станимировић, редовног професора Математичког факултета у Београду. Изабран је у научно звање истраживач приправник на Електротехничком факултету у Београду 2020. године. Аутор је два објављена научна рада. Године 2020. је његов рад *How to Control the Error Rates of Binary Classifiers* представљен као један од десет најбољих на догађају *New in ML* организованом у оквиру конференције *NeurIPS 2020*. Његово решење за објашњив модел одобравања кредита постиже четврти најбољи резултат на хакатону *Нове љубљанске банке 2021*. године. Године 2022. је добио награду *Kaggle ML Research Spotlight* компаније *Kaggle*. Од 2021. године ради

као аутор, а од 2022. као уредник техничких чланака из области вештачке интелигенције и рачунарских наука у *Baeldung*-у. Добитник је више награда из области књижевности. Заступљен је у антологијама и књижевној периодици, а 2021. године његову драму *Европеја – њуїї малої народа у Евроїу*, објавила је издавачка кућа *Корацї*.

1.4 Библиографија кандидата

M21

Simić, M. (2021). *Testing for normality with neural networks*. *Neural Computing & Applications* 33, 16279–16313. **IF2021=5.102** <https://doi.org/10.1007/s00521-021-06229-7>

M51

Simić, M. (2017). *Is the Bee Colony Optimisation algorithm suitable for continuous numerical optimisation?* *International Journal of Metaheuristics*, 6(4), 279–308. doi: <https://doi.org/10.1504/ijmheur.2017.086978>

2 Опис дисертације

2.1 Обим, садржај и структура дисертације

Докторска дисертација под насловом *Оптимизација алгоритамских шестова нормалности заснованих на неуронским мрежама* написана је на 103 стране формата А4. Садржи 26 илустрација, 12 табела, шест псеудокодова алгоритама и 112 библиографских референци. Дисертација је написана на српском језику и ћириличном писму.

Текст дисертације организован је у осам поглавља: 1. Увод (3 стране); 2. Тестирање нормалности (11 страна); 3. Примена генетских алгоритама у развоју једносмерних неуронских мрежа за испитивање нормалности (19 страна); 4. Извођење статистичких тестова из бинарних класификатора (17 страна); 5. Тестови нормалности засновани на једносмерним мрежама (10 страна); 6. Омнибус тест нормалности заснован на дугорочној рекурентној мрежи типа LSTM (10 страна); 7. Дискусија (4 стране); и 8. Закључак (1 страна). Поред тога, дисертација садржи: насловну страну на српском и енглеском језику; страну са подацима о менторима и члановима комисије за преглед, оцену и одбрану дисертације; сажетак на српском и енглеском језику, са кључним речима; садржај; библиографију; биографију кандидата; изјаву о ауторству; изјаву о истовестности штампане и електронске верзије докторског рада; изјаву о коришћењу.

2.2 Кратак опис појединачних поглавља

У првом поглављу је дат увод у проблем тестирања нормалности и представљени су изазови са којима се сусрећу стандардни тестови нормалности. Описани су проблеми до сада предложених неуронских мрежа за испитивање нормалности. Дефинисане су хипотезе од којих се пошло у истраживању.

У другом поглављу су представљени класични статистички тестови нормалности. У питању су најмоћнији, најчешће коришћени и највише испитивани омнибус и усмерени тестови нормалности. Поред тога, детаљније су представљене мреже претходно развијене за испитивање нормалности и демонстрирана је њихова осетљивост на варијабилност испитиване расподеле.

У трећем поглављу је детаљно представљен генетски алгоритам (ГА) који тежи да истовремено оптимизује четири групе параметара које је неопходно подесити да би се тренирала неуронска мрежа за испитивање нормалности. У питању су параметри генерисања узорака за тренирање, параметри репрезентације (трансформације генерисаних узорака), параметри који одређују мрежну архитектуру и параметри тренирања мреже на тако генерисаном и трансформисаном скупу узорака. Ови параметри представљени су генима који чине хромозоме (генетски код) јединки у популацији генетског алгоритма. Објашњена је варијабилна дужина и хијерархијска структура хромозома и предложене су варијанте генетских оператора укрштања и мутације дизајниране тако да буду применљиве на варијабилне хромозоме хијерархијске структуре. Детаљно је представљена стохастичка функција прилагођености, која поред класификационе тачности мреже у обзир узима и стабилност тачности на расподелама различите варијабилности, као и монотоност криве тачности на узорцима растућих обима.

У четвртом поглављу су представљене методе које су у области машинског учења предложене за контролу вероватноће грешке бинарних класификатора на циљној класи. Предложене су три нове методе за контролу грешке. Перформансе свих тих метода испитане су опсежном евалуацијом на четири скупа података различите структуре и величине.

У петом поглављу представљени су омнибус и усмерени мрежни тестови нормалности засновани на једносмерним неуронским мрежама добијеним применом варијанте ГА предложене у трећем поглављу. Утврђено је да добијене мреже нису осетљиве на варијансу испитиване расподеле. Оне су претворене у статистичке тестове нормалности контролом вероватноће грешке на класи нормалних расподела помоћу методе која је у евалуацији у четвртом поглављу показала најбоље резултате. Извршено је поређење омнибус и усмерених мрежних тестова са стандардним статистичким омнибус и усмереним тестовима нормалности. На основу анализе добијених резултата, дате су препоруке за примену тестова нормалности у зависности од величине узорка, типа не-нормалне расподеле и изабраног прага статистичке значајности.

Шесто поглавље представља омнибус мрежни тест изведен из рекурентне мреже типа ЛСТМ, која је добијена применом ГА из главе 3. Извршено је поређење тог теста са једносмерним омнибус мрежним тестом из пете главе и стандардним омнибус статистичким тестовима нормалности. Дате су смернице за примену тестова.

Седмо поглавље садржи дискусију о доказаности почетних хипотези и анализу добијених резултата. Анализирани су предности и мане предложеног приступа за развој тестова нормалности заснованих на неуронским мрежама. Идентификовани су изазови са којима се овај приступ сусреће.

У осмом поглављу налази се закључак који даје завршне напомене и назначава могуће правце будућег истраживања у оквиру области којима припада предмет дисертације.

3 Оцена дисертације

3.1 Савременост и оригиналност

Разматрана докторска дисертација представљана оригиналан научни рад који повезује области вештачке интелигенције, статистике и оптимизације. У дисертацији је предложена

прва примена генетског алгоритма за конструкцију и оптимизацију перформанси мрежа које испитују нормалности. При томе, варијанта генетског алгоритма којом је то постигнуто оптимизује и скуп података за тренирање мреже и тежи да минимизује осетљивост мрежа на варијабилност и очекивање дистрибуција одабраних за генерисање узорака за тренирање. Такође, у дисертацији су предложене нове методе за контролу вероватноће грешке бинарних класификатора на датој класи, чиме је омогућено да се из мрежа добијених генетским алгоритмом изведу статистички тестови нормалности којима се може контролисати вероватноћа грешке прве врсте.

Емпиријске студије моћи тестова нормалности чине значајан корпус статистичке литературе и представљају активно поље истраживања. Такође, у савременој литератури се све више пажње посвећује комбиновању модела и метода машинског учења са статистичким моделима и метода, као и примени метахеуристика у проналажењу оптималне архитектуре неуронских мрежа, што је и учињено у дисертацији кандидата. На основу наведеног, може се закључити да разматрана дисертација представља оригиналан и савремен допринос литератури из ужих области којима припада предмет дисертације.

3.2 Осврт на референтну и коришћену литературу

У оквиру дисертације наведено је 112 библиографских референци. Готово све референце могу се сврстати у једну од следеће четири области. У првој области су радови у којима су предложене мреже за испитивање нормалности расподела. Другу област чини литература о тестовима нормалности, у коју спадају како радови у којима су предложени разни омнибус и усмерени тестови нормалности, тако и емпиријске студије перформанси тих тестова на различитим класама не-нормалних расподела. У трећој области су методе које су у оквиру машинског учења развијене за контролу грешке бинарних класификатора на изабраној класи. Четврту област чине радови о примени еволутивних метахеуристика у одређивању архитектуре неуронских мрежа и подешавању њихових параметара.

Референцирани радови представљају научне резултате који су релевантни за предмет разматране дисертације. Међу њима се могу пронаћи радови објављени од средине двадесетог века, као и новија литература са актуелним резултатима. Кроз избор и преглед литературе, кандидат је показао да на адекватном нивоу познаје области истраживања у чије оквире спада ова дисертација.

3.3 Опис и адекватност примењених научних метода

У разматраној дисертацији је објашњено да конструкција неуронских мрежа за испитивање нормалности захтева истовремену оптимизацију параметара из четири групе. У питању су параметри генерисања података, параметри њихове репрезентације (трансформације у облик који мрежа може прихватити на улазу), параметри мрежне архитектуре и параметри тренирања. Дисертација за оптимизацију ових параметара користи варијанту генетског алгоритма (ГА) са посебно дизајнираним генетским операторима и функцијом прилагођености. Генетски код сваке јединке у популацији ГА одговара једном подешавању наведених параметара и представља један нумерички експеримент који тече у четири фазе. У првој фази се генеришу подаци за тренирање, у другој се добијени подаци преводе у форму одређену параметрима репрезентације, у трећој фази се конструише мрежа задате архитектуре, а у четвртој фази се та мрежа тренира на трансформисаним подацима за

тренирање. ГА служи као метода за итеративно извођење експеримената кроз промену подешавања параметара док се не добије мрежа довољно добрих перформанси.

Мреже добијене помоћу ГА претворене су у статистички тест нормалности коме се може задати праг значајности, односно, коме се може контролисати вероватноћа погрешне класификације нормалних узорака. У оквиру машинског учења је развијено више метода које бинарном класификатору могу контролисати вероватноћу грешке на изабраној класи. У овој дисертацији је изведена опсежна евалуација тих метода на четири скупа података различите величине и структуре. На мреже добијене генетским алгоритмом је примењена метода која је показала најбоље перформансе у евалуацији. Нумеричком симулацијом је утврђено да она успешно контролише вероватноћу грешке добијених мрежа на класи нормалних узорака.

Моћ добијених мрежних тестова испитана је на узорцима из неколико група расподела које се у литератури користе за евалуацију тестова нормалности. Извршено је поређење мрежних тестова са најмоћнијим и најчешће коришћеним омнибус и усмереним тестовима нормалности на узорцима са десет до сто елемената. Добијени резултати анализирани су адекватним статистичким методама и праћени квалитативном анализом.

3.4 Применљивост остварених резултата

У анализи података у научном истраживању и индустријским применама, нормалност расподела представља једну од најчешћих претпоставки за коришћење класичних статистичких тестова и верификацију модела добијених на основу података. Мрежни тестови нормалности развијени у оквиру ове дисертације могу пронаћи примену у свим научним дисциплинама које користе статистичке методе за испитивање хипотеза, као и у индустријским истраживањима која се служе статистиком.

Предложене методе за контролу вероватноће грешке бинарних класификатора на изабраној класи не захтевају измену процеса тренирања, тако да се могу применити и на постојеће класификаторе. Те методе највећу примену могу наћи у проблемима бинарне класификације у којима је од пресудног значаја да се на унапред одређену вредност ограничи вероватноћа грешке на критичној класи.

Генетски алгоритам искоришћен је не само за оптимизацију мрежне архитектуре и репрезентације узорака, већ и за оптимизацију параметара тренирања и генерисања нормалних и не-нормалних узорака за тренирање. Та методологија применљива је и у конструкцији других модела за испитивање нормалности, али и за друге проблеме у којима је могуће генерисати податке за тренирање и тестирање. У питању су првенствено други класични статистички проблеми, али и проблем испитивања да ли је дати текстуални или графички садржај креиран помоћу генератора као што су велики језички модели и дубоке мреже.

3.5 Оцена достигнутих способности кандидата за самосталан научни рад

Кандидат Милош Симић је током израде докторске дисертације показао самосталност, систематичност и стручност у сагледавању и решавању истраживачких проблема, као и у критичкој анализи добијених резултата. Кандидат је успешно испитао полазне хипотезе формулисане у предлогу дисертације, показујући да влада знањима везаним за област истраживања и научном методологијом, што указује на његову способност да објективно и у

целини сагледа истраживачки процес и услове потребне за његову реализацију. Комисија сматра да кандидат поседује све квалитете неопходне за самосталан истраживачки рад.

4 Остварени научни допринос

4.1 Приказ научних доприноса

У овој дисертацији остварени су следећи научни доприноси:

- Предложена је методологија за развој неуронских мрежа и других модела машинског учења за тестирање нормалности. Методологија је заснована на примени генетског алгоритма са стохастичком функцијом прилагођености и генетским операторима дизајнираним тако да буду применљиви на хромозоме променљивих дужина и са хијерархијским односом између гена. Предложена варијанта генетског алгоритма служи за истовремену оптимизацију мрежне архитектуре и параметара тренирања, као и параметара репрезентације и генерисања података за тренирање.
- Развијен је омнибус тест нормалности заснован на једносмерним неуронским мрежама. Опсежном емпиријском евалуацијом је утврђено да његове перформансе не зависе од очекивања и варијансе тестиране расподе, да му се може контролисати вероватноћа погрешне класификације нормалног узорка као не-нормалног, као и да надмашује класичне статистичке тестове нормалности на одређеним класама не-нормалних расподела.
- Развијено је више усмерених тестова нормалности за типове не-нормалних расподела који се од класе нормалних дистрибуција разликују по коефицијентима искошености и тежине репова. У емпиријској студији су ти тестови били моћнији од класичних усмерених тестова нормалности или упоредиви по моћи са њима.
- Развијен је омнибус тест нормалности заснован на рекурентној неуронској мрежи типа LSTM.
- Предложена је метода за контролу вероватноће грешке бинарних класификатора на циљној класи, која нумерички предизлаз класификатора третира као излазну статистику и за конкретан објекат који треба класификовати рачуна р-вредност добијеног предизлаза.
- Показано је како се метода бинарне класификације заснована на посредном тестирању расподела растојања до најближих суседа може модификовати тако да коришћењем композитне нулте хипотезе контролише вероватноћу да се погрешно класификује објекат циљне класе.
- Такође, предложена је и тежинска верзија модификације засноване на композитној нултој хипотези, која користи тежинска растојања и адекватна је за случај када су објекти циљне класе ближи суседима из супротне класе.
- Изведена је опсежна емпиријска евалуација метода за контролу грешке које су у литератури и овој дисертацији коришћене за контролу вероватноће грешке бинарних класификатора на циљној класи.

4.2 Критичка анализа резултата истраживања

Сагледавањем добијених резултата у односу на циљеве истраживања и полазне хипотезе, може се закључити да је у разматраној дисертацији примењена адекватна методологија и да су у потпуности испитане полазне хипотезе. Добијени резултати су научно значајни, а методологија примењена и формулисана у овој дисертацији је оригинална и актуелна.

Два највећа доприноса дисертације су: (1) нова методологија, заснована на генетском алгоритму, за развој неуронских мрежа и других модела машинског учења за тестирање нормалности, која је применљива и на сличне проблеме и (2) нове методе за контролу грешке бинарних класификатора и њихова опсежна емпиријска евалуација.

4.3 Објављени радови

Кандидат је самостални аутор следећег рада који је повезан са докторском дисертацијом и објављен у часопису на СЦИ листи:

- **Simić, M.** (2021). *Testing for normality with neural networks*. *Neural Computing & Applications* 33, 16279–16313. *IF2021=5.102 (M21)* <https://doi.org/10.1007/s00521-021-06229-7>

4.4 Провера оригиналности докторске дисертације

Провера оригиналности докторске дисертације обављена је на начин прописан Правилником о поступку провере оригиналности докторских дисертација које се бране на Универзитету у Београду (Гласник Универзитета у Београду, бр. 204/22.6.2018). Увидом у извештај програма *iThenticate* утврђено је да степен подударања текста износи 3%. При томе, подударност ни са једним извором не износи више од 1%. Извори са којима је утврђен највећи степен подударности су верзије радова кандидата постављене на електронску платформу *arXiv*, који су повезани са наведеном дисертацијом. Подударност са осталим изворима последица је коришћења тзв. општих фраза и формулација (на пример: *у директној сујројности са начином на који...*; *... се сасиоји у шоме да*; *Универзитет у Београду*) и уобичајене терминологије и математичке нотације. Стога сматрамо да је докторска дисертација кандидата у потпуности оригинална и да су поштована академска правила цитирања.

5 Закључак и предлог комисије за одбрану докторске дисертације

Докторска дисертација под насловом *Оптимизација алгоритамаских шестова нормалности заснованих на неуронским мрежама*, кандидата Милоша Симића, представља оригинално и самостално научно дело. У питању је методолошки утемељен мултидисциплинаран научни рад из ужих области: Вештачка интелигенција, Статистика и Оптимизација. Добијени резултати и остварен допринос одликују се научном релевантношћу и високом применљивошћу у пракси.

Дисертација је написана према одобреној пријави. У дисертацији су успешно и адекватно испитане хипотезе наведене у пријави теме и остварени циљеви истраживања. Дисертација је у свему израђена у складу са прописаним стандардима и испуњава све прописане услове да би се приступило њеној одбрани.

Комисија, на основу укупне оцене дисертације и горе наведеног, предлаже Већу за студије при универзитету да прихвати овај извештај и докторску дисертацију под насловом *Оптимизација алгоритамаских шестова нормалности заснованих на неуронским мрежама* кандидата Милоша Симића, да дисертацију изложи на увид јавности и да је упуту на коначно усвајање Већу за интердисциплинарне, мултидисциплинарне и трансдисциплинарне студије Универзитета у Београду, након чега би се приступило усменој одбрани дисертације пред комисијом у истом саставу.

У Београду, _____

ЧЛАНОВИ КОМИСИЈЕ

др Бошко Николић, редовни професор
Електротехнички факултет, Универзитет у Београду

др Вељко Јеремић, редовни професор
Факултет организационих наука, Универзитет у Београду

др Стефан Мишковић, доцент
Математички факултет, Универзитет у Београду