

ИЗВЕШТАЈ О ОЦЕНИ ДОКТОРСКЕ ДИСЕРТАЦИЈЕ

I ПОДАЦИ О КОМИСИЈИ		
1. Датум и орган који је именовано комисију: 15.09.2022. Наставно-научно веће Природно-математичког факултета, Универзитета у Новом Саду		
2. Састав комисије у складу са <i>Правилима докторских студија Универзитета у Новом Саду</i> :		
1. др Милош Стојаковић	редовни професор	теоријске основе информатике, 01.4.2016.
презиме и име	звање	ужа научна област и датум избора
Природно-математички факултет, Универзитет у Новом Саду		председник
установа у којој је запослен-а		функција у комисији
2. др Наташа Крејић	редовни професор	нумеричка математика, 14.06.2004.
презиме и име	звање	ужа научна област и датум избора
Природно-математички факултет, Универзитет у Новом Саду		члан
установа у којој је запослен-а		функција у комисији
3. др Милош Ивановић	ванредни професор	меродологије рачунарства, 13.04.2022.
презиме и име	звање	ужа научна област и датум избора
Природно-математички факултет, Универзитет у Крагујевцу		члан
установа у којој је запослен-а		функција у комисији
4. др Душан Јаковетић	ванредни професор	математичко моделирање, 15.11.2020.
презиме и име	звање	ужа научна област и датум избора
Природно-математички факултет, Универзитет у Новом Саду		ментор
установа у којој је запослен-а		функција у комисији

5. др Данијела Боберић Крстићев	ванредни професор	информациони системи, 01.01.2021.
презиме и име	звање	ужа научна област и датум избора
Природно-математички факултет, Универзитет у Новом Саду		ментор
установа у којој је запослен-а		функција у комисији

II ПОДАЦИ О КАНДИДАТУ

- Име, име једног родитеља, презиме:
Лидија, Рудолф, Фодор
- Датум рођења, општина, држава:
25.07.1989. Бачка Топола, Србија
- Назив факултета, назив претходно завршеног нивоа студија и стечени стручни/академски назив:
Природно-математички факултет, Универзитет у Новом Саду, мастер академске студије, мастер информатичар
- Година уписа на докторске студије и назив студијског програма докторских студија:
2019, информатика

III НАСЛОВ ДОКТОРСКЕ ДИСЕРТАЦИЈЕ:

Имплементација и анализа класе алгоритама за дистрибуирану конвексну оптимизацију: Евалуација перформанси и особина на практичним HPC кластерима

(Implementation and analysis of a class of algorithms for distributed convex optimization: Performance evaluation and tradeoffs in practical HPC clusters)

IV ПРЕГЛЕД ДОКТОРСКЕ ДИСЕРТАЦИЈЕ:

Навести кратак садржај са назнаком броја страница, поглавља, слика, схема, графика и сл.

Докторска дисертација је написана на 225 страна на енглеском језику, латиничним писмом. Дисертација садржи 4 поглавља, 19 табела са нумеричким подацима, 52 графичких илустрација у виду слика и дијаграма и 163 цитата. На почетку дисертације, дати су: кључна документацијска информација, извод на енглеском и српском језику, садржај рада, и списак табела и слика.

Истраживања реализована у оквиру ове дисертације, структурирана су у следећа поглавља и потпоглавља:

- Увод
 - Дистрибуирана конвексна оптимизација
 - Методе дистрибуиране оптимизације
 - Рачунарство високих перформанси
 - Мотивација и циљеви
 - Доприноси
 - Доприноси везани за методе примарног типа
 - Доприноси везани за методе дуалног типа
 - Преглед радова из области дисертације
 - Преглед садржаја дисертације
- Методе дистрибуиране оптимизације примарног типа
 - Теоријске основе
 - Оптимизациони и мрежни модели
 - Алгоритамски оквир предложеног скупа метода
 - Анализа конвергенције
 - Имплементација
 - Имплементација алгорита за строго конвексне квадратне функције циља

- 2.2.2. Имплементација алгоритма за функције логистичког губитка
- 2.2.3. Поређење са ADMM имплементацијом
- 2.2.4. Мерење времена извршавања паралелног програма
- 2.3. Експерименти
 - 2.3.1. Инфраструктура
 - 2.3.2. Међурезултати и експерименти
 - 2.3.3. Експериментални резултати за одабран скуп метода
- 2.4. Закључци о предложеној класи примарних метода
- 3. Дуална метода дистрибуиране оптимизације
 - 3.1. Теоријске основе
 - 3.1.1. Модел проблема и предлог паралелног метода
 - 3.2. Имплементација
 - 3.2.1. Улазни подаци
 - 3.2.2. Критеријум заустављања
 - 3.2.3. Паралелна имплементација ADMM-базираног алгоритма за конвексни приступ груписању
 - 3.3. Експерименти
 - 3.3.1. Временска захтевност различитих делова алгоритма
 - 3.3.2. Евалуација тачности решења
 - 3.3.3. Евалуација скалабилности решења
 - 3.3.4. Избор вредности параметра γ
 - 3.3.5. Поређење са другим методама груписања
 - 3.3.6. Тестирање над реалним, индустријским скупом података
 - 3.4. Додатна разматрања о имплементацији
 - 3.5. Поређење MPI и COMPSs паралелних апликација
 - 3.6. Додатни алгоритми машинског учења, који се ослањају на ADMM
 - 3.6.1. ADMM-базирана lasso регресија
 - 3.6.2. ADMM-базирана логистичка регресија
 - 3.7. Закључци о предложеној примени дуалне, ADMM методе
- 4. Закључак
 - 4.1. Резиме достигнућа тезе
 - 4.2. Примене
 - 4.3. Будући рад

На крају тезе, дате су смернице за приступ изворном коду, ради поновљивости резултата. Осим тога, дат је и проширен извод на српском језику, латиничним писмом, као и биографија кандидата.

V ВРЕДНОВАЊЕ ПОЈЕДИНИХ ДЕЛОВА ДОКТОРСКЕ ДИСЕРТАЦИЈЕ:

Дисертација садржи све неопходне елементе прописане важећим правилницима, добро је структурирана и има добар начин излагања.

Комисија сматра да је *наслов* докторске дисертације прецизно формулисан и јасно указује на разматрану проблематику и реализовано истраживање.

Докторску дисертацију чини 4 поглавља, која су логично поређана и добро систематизована.

У уводном, *првом поглављу*, представљена су уводна разматрања, и дати су описи кључних концепата дисертације. Најпре је изнет сажетак прегледа метода дистрибуиране оптимизације и особина рачунарства високих перформанси, као и њиховог споја у циљу решавања временски и ресурсно захтевних проблема. Затим, описани су и главни доприноси дисертације, који се могу поделити на две велике целине: доприноси у области метода примарног и метода дуалног типа. У области метода примарног типа, предложен је скуп метода првог и другог реда, и дат је опис

њихових имплементационих аспеката, кроз развој паралелних алгоритама. Осим тога, теза обухвата описе опсежних емпиријских евалуација датих метода. Акцентат је на евалуацији спарсификације комуникације, уз поређење различитих приступа спарсификовања комуникације међу чворовима, као и на поређењу усмерене и неусмерене комуникације. Из аспекта оптимизационих метода дуалног типа, предложена је метода за конвексно груписање, која се заснива на приступу ADMM (Alternating Directions Method of Multipliers) и изнети су детаљи развоја паралелног алгорита за овај приступ, уз детаљне емпиријске евалуације методе, у смислу тачности и скалабилности. На крају овог поглавља, дат је преглед литературе из области дисертације, који укључује преглед радова везано за примену, теоријске и практичне напретке метода дистрибуиране оптимизације.

Комисија закључује да су уводне одреднице, мотивација и доприноси адекватно дефинисани, као и да је преглед актуелног стања литературе адекватног обима и дат у складу са постојећом литературом и предметом истраживања дисертације.

Кроз **друго поглавље**, описан је предложени скуп метода примарног типа, као први правац истраживања у оквиру тезе. Најпре су описани теоријски концепти (потпоглавље 2.1), потребни за дефинисање и разумевање разматраних метода. Након тога, постављен је јасан алгоритамаски оквир, који прецизно дефинише кораке за решавање посматраног оптимизационог проблема. У оквиру описа имплементације (потпоглавље 2.2), разматра се јединствен приступ, који је лако прилагодљив различитим функцијама циља, а дати су примери имплементације за два случаја функције циља: квадратне функције и функције логистичког губитка. Описани су различити аспекти имплементационих детаља употребом MPI приступа за паралелизацију, као и адаптација имплементације, у циљу побољшања перформанси алгоритама. Посматрају се имплементације метода првог и другог реда, са различитим приступима за спарсификовање комуникације, која може бити усмерена или неусмерена. Детаљном евалуацијом ових метода (у потпоглављу 2.3) на рачунарском кластеру, долази се до значајних закључака. Пре свега, експериментално се утврђује скуп метода, које су од интереса за разматрање, у смислу начина постављања вероватноће и усмерења комуникације. За одабрани скуп метода, дата су поређења времена извршавања за исте улазне податке, приказане су особине скалабилности, убрзања и профили перформанси. На основу описаног скупа емпиријских евалуација, закључује се да методе првог реда у начелу имају боље перформансе, када се ради са великим скуповима података, док методе другог реда имају предност над мањим подацима. Такође, идентификовани су најефикаснији приступи за спарсификацију комуникације и показано је да и принцип усмерене спарсификоване комуникације показује велики потенцијал.

Комисија позитивно оцењује поглавље 2, посвећено методама дистрибуиране оптимизације примарног типа.

Унутар **трећег поглавља**, предложен је приступ, који се заснива на идеји конвексног груписања, а ослања се на методу дуалног типа, ADMM. Најпре су описани неопходни теоријски концепти и модел предложеног метода (потпоглавље 3.1), уз детаљно објашњење алгоритамаских корака паралелне имплементације. Наведени су имплементациони аспекти (потпоглавље 3.2) у контексту COMPSs технологије за паралелизацију. Дата је и детаљна евалуација особина алгоритама на рачунарском кластеру (потпоглавље 3.3). Ово подразумева анализу утрошка времена извршавања на различите сегменте алгоритама, али и евалуацију тачности груписања. Показује се да предложени алгоритама може да постигне тачност груписања, која је упоредива са резултатима груписања других приступа. Приказани су резултати поређења са неколико различитих, прихваћених приступа за груписање. Експерименти показују и да се алгоритама добро скалира, и да постиже задовољавајуће убрзање са повећањем броја чворова. Осим тога, овакав приступ је ефикаснији од већине других алгоритама груписања, с обзиром на његову паралелну природу. Алгоритама је успешно тестиран и над реалним, индустријским подацима, што је такође описано у овом поглављу. Такође, унутар поглавља 3 (у потпоглављу 3.4), описано је и поређење два различита приступа за паралелизацију истог алгоритама: MPI и COMPSs, где су изнесене предности оба приступа, уз анализу перформанси и једноставности развоја паралелног програма.

Додатно, описане су још две додатне имплементације алгоритама машинског учења, које се базирају на ADMM приступу (потпоглавље 3.6). За предложен метод груписања, закључује се да постиже задовољавајући ниво тачности и да се добро скалира. Осим тога, описан приступ је лако прилагодљив за потенцијалне нове примене додатних алгоритама машинског учења.

Комисија позитивно оцењује поглавље 3, посвећено методама дистрибуиране оптимизације дуалног типа.

Четврто поглавље посвећено је закључцима дисертације. Оно обухвата преглед доприноса дисертације, али и описе потенцијалних примена алгоритама. Посебно се наглашава проширивост, која омогућује још распрострањенију примену предложених метода. Осим тога, дате су и смернице за будући рад у смеру обе области, које су у фокусу дисертације.

Комисија позитивно оцењује 4. поглавље дисертације – Закључци.

Комисија позитивно оцењује све делове докторске дисертације кандидаткиње Лидије Фодор, као и начин на који су резултати истраживања приказани и тумачени у овој докторској дисертацији.

VI СПИСАК НАУЧНИХ И СТРУЧНИХ РАДОВА КОЈИ СУ ОБЈАВЉЕНИ ИЛИ ПРИХВАЋЕНИ ЗА ОБЈАВЉИВАЊЕ НА ОСНОВУ РЕЗУЛТАТА ИСТРАЖИВАЊА У ОКВИРУ РАДА НА ДОКТОРСКОЈ ДИСЕРТАЦИЈИ:

Рад у међународном часопису (M23):

1. L. Fodor, D. Jakovetic, N. Krejic, N. K. Jerinkic, and S. Skrbic, "Performance evaluation and analysis of distributed multi-agent optimization algorithms with sparsified directed communication," EURASIP J. Adv. Signal Process, vol. 25, 2021.
<https://doi.org/10.1186/s13634-021-00736-4>

VII ЗАКЉУЧЦИ ОДНОСНО РЕЗУЛТАТИ ИСТРАЖИВАЊА:

Резултати и доприноси ове докторске дисертације, тичу се класе алгоритама дистрибуиране оптимизације, и могу се поделити у две велике целине.

Прву целину чине методе дистрибуиране оптимизације примарног типа. На основу анализе литературе, итврђено је да постоји врло ограничен број научних резултата, који се тичу практичне евалуације оптимизационих метода. У контексту ових метода, дисертација има неколико важних резултата и доприноса:

- Развој паралелне имплементације отвореног кода у MPI технологији, за посматрану класу метода, уз анализу имплементационих аспеката, који утичу на перформансе алгорита. Дата имплементација је погодна за поновну примену, уз једноставно прилагођавање, за произвољне функције циља. Тиме се омогућује приступачан, паралелан и скалабилан начин за решавање датог оптимизационог проблема. На овај начин, доприноси се приближавању теоријских и практичних аспеката у области.
- Анализа и поређење особина класе метода, на реалном, рачунарском кластеру. Иако постоји мноштво теоријских резултата, који се баве напрелима метода дистрибуиране оптимизације, врло је мало практичних резултата евалуације метода. У дисертацији је показано да су предложене методе скалабилне, односно да постижу значајно убрзање са повећањем броја чворова, који учествују у решавању проблема. Осим тога, спроведено је детаљно поређење метода и уочени су сценарији у којима одређене методе имају боље перформансе од других. На тај начин, стиче се увид у утицај особина метода на перформансе: да ли су првог или другог реда, на који начин спарсификују комуникацију и да ли је комуникација усмерена или неусмерена. Примена усмерене спарсификоване комуникације резултује и новом методом, која до сада није разматрана у литератури.
- Евалуација метода, извршена је над разним, јавно доступним скуповима података, и испитан је и утицај природе података на ефикасност појединих метода.

Другу целину представља примена методе дистрибуиране оптимизације дуалног типа, ADMM, за развој новог алгоритма груписања. У литератури се појављују широк опсег предлога метода за конвексно груписање, али је број развијених паралелних имплементација и њихових евалуација врло ограничен. У овом делу дисертације, такође можемо уочити неколико резултата и доприноса:

- Дизајн и имплементација нове методе груписања отвореног кода, која се заснива на ADMM приступу и ослања се на идеје конвексног груписања. Дата паралелна имплементација је развијена у COMPSs технологији, и помоћу предложеног приступа, лако се може прилагодити за друге алгоритме машинског учења. Овим се поново добија једноставан начин за креирање паралелних решења посматраних проблема. У дисертацији је описана имплементација још два додатна алгоритма машинског учења, који се ослањају на ADMM, и имплементирани су помоћу COMPSs технологије.
- Евалуација развијеног алгоритма је показала добре особине скалабилности и висок ниво тачности. Поређења са другим методама груписања, резултирају упоредивим нивоом тачности. Евалуације методе су извршаване над разним скуповима подака, како синтетичким, тако и реалним, јавно доступним. Осим тога, алгоритам је примењен и над реалним, индустријским подацима, што доприноси унапређењу перформанси датих индустријских случајева коришћења, и отвара могућности за сличне, додатне примене.
- Извршено је поређење поменуте две технологије за паралелизацију: MPI и COMPSs. Паралелни алгоритам за груписање је имплементиран у обе технологије и тестиран је над истим случајевима коришћења на кластеру, како би се добио увид у разликама по питању перформанси. Осим дискусије о перформансама, уочене су предности оба приступа, у смислу компромиса између једноставности паралелизације и ефикасности извршавања.

Комисија сматра да укупан рад кандидата, по свом карактеру и обиму, у потпуности одговара дефинисаној теми. Сагласно изнетим ставовима, Комисија позитивно оцењује начин на који су резултати истраживања приказани и тумачени у овој докторској дисертацији.

VIII ОЦЕНА НАЧИНА ПРИКАЗА И ТУМАЧЕЊА РЕЗУЛТАТА ИСТРАЖИВАЊА:

Приказ свих резултата истраживања је ваљано организован по логичким целинама, а формулација резултата је прецизна и илустрована примерима. Тумачење добијених резултата је коректно изведено, уз њихову детаљну анализу.

Текст дисертације је проверен у софтверу за детекцију плагијаризма iThenticate у Библиотеци Департмана за математику и информатику Природно-математичког факултета, са вредношћу резултујућег индекса сличности 21%. На основу резултата провере, Комисија је констатовала да је већина подударности, везана за радове (један објављен и један у процесу ревизије, али доступан као preprint), у којима је кандидаткиња први аутор, а који делимично приказују резултате дисертације. Остатак преклапања се односи на поједине опште коришћене фразе. Стога се закључује да је докторска дисертација оригинално ауторско дело кандидаткиње Лидије Фодор. Са Извештајем о подударности упознати су сви чланови Комисије.

На основу наведеног, комисија је донела позитивну оцену за начин приказа и тумачења резултата, са закључком да је докторска дисертација оригинално ауторско дело кандидаткиње Лидије Фодор.

IX КОНАЧНА ОЦЕНА ДОКТОРСКЕ ДИСЕРТАЦИЈЕ:

1. Да ли је дисертација написана у складу са образложењем наведеним у пријави теме?

Дисертација је у потпуности написана у складу са образложењем наведеним у пријави теме.

2. Да ли дисертација садржи све битне елементе?

Дисертација садржи све битне елементе. Дат је адекватно опширан увод у концепте и теоријске основе, који служе као основа за касније описане резултате. Изложени су сви битни познати резултати на које се дисертација ослања. Обиман списак библиографских референци садржи релевантне радове и сведочи о добром познавању области. Дисертација је прегледна и добро организована.

3. По чему је дисертација оригиналан допринос науци?

Докторска дисертација кандидаткиње Лидије Фодор представља оригиналан научни рад, са теоријским и експерименталним елементима и подразумева неколико оригиналних доприноса. Пре свега, дисертација описује развој и анализу паралелних имплементације класе примарних и дуалних метода, са поређењем одговарајућих метода, њиховом детаљном анализом и практичним потврђивањем теоријских концепата. Описују се и две нове методе: метода примарног типа са усмереном спарсификованом комуникацијом и дуална метода за конвексно груписање, базирана на ADMM приступу. Широко скуп емпиријских евалуација метода, обезбеђује важне увиде у перформансе, скалабилност и примену предложених метода.

Доприноси дисертације се могу поделити у две целине: доприноси везани за методе примарног типа и доприносу везани за методу дуалног типа. Истакнути оригинални доприноси дисертације се могу обухватити на следећи начин:

1. Евалуација стварних бенефита спарсификовања комуникације међу чворовима, који учествују у израчунавању, чиме се потврђују теоријски аспекти посматраних метода примарног типа.
2. Упоредна анализа различитих приступа спарсификовању комуникације, и идентификација најефикаснијих метода примарног типа у задатим сценаријима, као и поређење усмерене и неусмерене комуникације, у комбинацији са различитим приступима спарсификацији. Уз то, појављује се предлог нове методе примарног типа, са усмереном спарсификованом комуникацијом.
3. Развој и евалуација новог, паралелног приступа за груписање у контексту метода дуалног типа, коришћењем особина конвексног груписања, и експериментална демонстрација ралности методе.
4. Евалуација скалабилности развијених имплементација метода, као и могућност поновне употребе развијених паралелних имплементација кроз програмски код отвореног типа, за нове примене, кроз минималне адаптације. Такође, демонстрирана је примењивост предложене паралелне методе за груписање над реалним, индустријским подацима, чиме се доприноси унапређењу индустријских случајева коришћења, и отвара се могућност сличних примена у будућности.
5. Упоредба различитих приступа за развој паралелних алгоритама дистрибуиране оптимизације.

4. Који су недостаци дисертације и какав је њихов утицај на резултат истраживања?

Дисертација нема недостатака.

X ПРЕДЛОГ:
На основу наведеног, комисија предлаже:
а) да се докторска дисертација прихвати, а кандидату одобри одбрана; б) да се докторска дисертација врати кандидату на дораду (да се допуни односно измени); в) да се докторска дисертација одбије.

Нови Сад, 26.09.2022.:

1. др Милош Стојаковић, редовни професор
_____, председник

2. др Наташа Крејић, редовни професор
_____, члан

3. др Милош Ивановић, ванредни професор
_____, члан

4. др Душан Јаковетић, ванредни професор
_____, ментор

5. др Данијела Боберић Крстићев, ванредни професор
_____, ментор

НАПОМЕНА: Члан комисије који не жели да потпише извештај јер се не слаже са мишљењем већине чланова комисије, дужан је да унесе у извештај образложење односно разлоге због којих не жели да потпише извештај и да исти потпише.