

ИЗВЕШТАЈ О ОЦЕНИ ДОКТОРСКЕ ДИСЕРТАЦИЈЕ  
-обавезна садржина- свака рубрика мора бити попуњена

<b>I ПОДАЦИ О КОМИСИЈИ</b>
<p>1. Датум и орган који је именовео комисију <b>06.07.2016. Наставно-научно веће Факултета техничких наука</b></p> <p>2. Састав комисије са назнаком имена и презимена сваког члана, звања, назива уже научне области за коју је изабран у звање, датума избора у звање и назив факултета, установе у којој је члан комисије запослен:</p> <p>1. Др Вељко Милутиновић, редовни професор, УНО Рачунарска техника и информатика, 01.12.1997. Електротехнички факултет, Универзитет у Београду</p> <p>2. Др Војин Шенк, редовни професор, УНО Телекомуникације и обрада сигнала, 18.07.2003. Факултет техничких наука, Универзитет у Новом Саду</p> <p>3. Др Владимир Црнојевић, редовни професор, УНО Рачунарске науке, 15.02.2016. Природно-математички факултет, Универзитет у Новом Саду</p> <p>4. Др Срђан Шкрбић, ванредни професор, УНО Информациони системи, 15.10.2015. Природно-математички факултет, Универзитет у Новом Саду</p> <p>5. Др Татјана Лончар-Турукало, доцент, УНО Телекомуникације и обрада сигнала, 26.04.2012. Факултет техничких наука, Универзитет у Новом Саду</p> <p>6. Др Дејан Вукобратовић, ванредни професор, УНО Телекомуникације и обрада сигнала, 01.04.2014, Факултет техничких наука, Универзитет у Новом Саду</p>
<b>II ПОДАЦИ О КАНДИДАТУ</b>
<p>1. Име, име једног родитеља, презиме: <b>Сања, Стево, Брдар</b></p> <p>2. Датум рођења, општина, држава: <b>19.06.1982. Рума, Србија</b></p> <p>3. Назив факултета, назив студијског програма дипломских академских студија – мастер и стечени стручни назив <b>Факултет техничких наука, Енергетика, електроника и телекомуникације, дипломирани инжењер електротехнике и рачунарства – мастер</b></p> <p>4. Година уписа на докторске студије и назив студијског програма докторских студија <b>2007. Енергетика, електроника и телекомуникације</b></p> <p>5. Назив факултета, назив магистарске тезе, научна област и датум одбране: -</p>

б. Научна област из које је стечено академско звање магистра наука: -
<b>III НАСЛОВ ДОКТОРСКЕ ДИСЕРТАЦИЈЕ:</b> Алгоритми интегративног кластеровања података применом ненегативне факторизације матрице (назив на енглеском: „ <i>Non-negative matrix factorization for integrative clustering</i> “)
<b>IV ПРЕГЛЕД ДОКТОРСКЕ ДИСЕРТАЦИЈЕ:</b>  Докторска дисертација је изложена на 115 страна, обухвата 32 графикона, 18 табела, 130 референци и 3 прилога. Написана је на енглеском језику. На почетку дисертације дата је кључна документација, резиме, индекс графикона, индекс табела и садржај.  Дисертација садржи осам поглавља:  <ol style="list-style-type: none"> <li>1. Увод (<i>Introduction</i>)</li> <li>2. Преглед релевантних области (<i>Related work</i>)</li> <li>3. Негативна факторизација матрице за интегративно кластеровање (<i>NMF for Integrative Clustering</i>)</li> <li>4. Опсежно поређење метода интегративног кластеровања (<i>Comprehensive benchmarking of integrative clusterings</i>)</li> <li>5. Негативна факторизација матрице за интегративно откривање функционално повезаних група гена (<i>NMF for integrative discovery of functionally related genes</i>)</li> <li>6. Регуларизована ненегативна факторизација матрице за интегративно откривање функционално повезаних група гена (<i>Regularized NMF for integrative discovery of functionally related genes</i>)</li> <li>7. Негативна факторизација матрице за стабилно издавајање кластера у микробиолошким узорцима (<i>NMF for Stable Assessment of Clusters in Microbiome Samples</i>)</li> <li>8. Закључак (<i>Conclusion</i>)</li> </ol> У оквиру дисертације се још налазе: Литература и прилози (опис података, табеле са резултатима поређења метода и продужени резиме на српском језику).
<b>V ВРЕДНОВАЊЕ ПОЈЕДИНИХ ДЕЛОВА ДОКТОРСКЕ ДИСЕРТАЦИЈЕ:</b>  У првом поглављу описани су предмет истраживања и мотивација за примену на конкретним проблемима са подручја биоинформатике. Предмет истраживања су алгоритми кластеровања – груписања података и могућности њиховог унапређења. Назначени су савремени приступи у кластеровању података: интегративно, полунадгледано и примењено на великој количини података – <i>Big Data</i> . Мотивација за примену у биоинформатици је образложена експоненцијалним растом количине геномских података и њиховом великом хетерогеношћу. Представљени су разноврсни биолошки подаци у облику секвенци, експресија, интеракција.  Друго поглавље пружа добар увид у релевантне области које су потребне за разумевање наредних поглавља и доприноса дисертације. Описани су алгоритми за кластеровање података и поступци валидације, постојећи добро познати и новији алгоритми интегративног кластеровања као и математичка метода ненегативне факторизације матрице. Истакнут је значај ненегативне факторизације матрице у области машинског учења који произилази из могућности анализе добијених фактора и увида у делове и њихову повезаност унутар целине.  У трећем поглављу предлаже се метода за интеграцију заснована на примени ненегативне факторизације матрице и референцирају се радови на чијим основама је настала предложена метода. Метода се извршава у две фазе. Прва обухвата формирање ансамбла резултата вишеструким извршавањем појединачних кластеровања и представљањем добијених резултата у облику бинарне матрице где

врста означава кластер, а колона објекат. При томе се основни алгоритам кластеровања покреће са различитим параметрима, иницијализацијама, подскуповима обележја, подскуповима узорака и/или се користи више различитих извора података. У другој фази се извршава факторизација матрице и издвајање коначних кластера коришћењем информација садржаних у факторима матрице ансамбла. Пружа се опција да излазни кластери буду преклапајући или непреклапајући, што је од посебне важности за биолошке податке где објекат може припадати у више кластера.

Четврто поглавље представља резултате опсежног поређења предложене методе и 5 алтернативних алгоритама. Евалуација је извршена коришћењем две интерне и две екстерне мере квалитета кластеровања. Експерименти су обухватили три групе података: „UCI Machine Learning“ подаци (20 сетова података), синтетички подаци (20 сетова података) и генске експресије из студија геномике рака (30 сетова података). Додатно су анализирани сетови генских експресија пацијената оболелих од рака у контексту диверзитета/слагања резултата појединачних кластеровања. Проблематика података у облику генских експресија произилази из много већег броја обележја од броја узорака, што значајно смањује поузданост резултата кластеровања и могућност откривања знања. Интеграција је рађена на различитим подскуповима обележја и параметрима кернел функције у „Kernel K-means“ методи кластеровања.

У петом поглављу метода интегративне кластеризације података примењена је на проблем откривања функционално кохерентних група гена/протеина. Експерименти су вршени на три скупа података који су значајни за откривање функционалне повезаности. Интеграција је рађена по различитим изворима података и мерама сличности/растојања (Пирсонова корелација, међусобна информација и Еуклидско растојање). Евалуација методе је извршена над генском онтологијом која представља хијерархијски организовану категоризацију функција гена. Интуитивним графичким приказом представљени су резултати кластеровања индивидуалним методама и добијени резултат након интеграције. Резултати јасно показују да се интеграцијом добијају функционално кохерентније групе гена.

Шесто поглавље разматра проширење методе увођењем регуларизације у поступак интеграције. Регуларизација се врши структуром графа где чворови означавају објекте који се кластерују, а ивице везу између објеката. Мења се алгоритам за факторизацију матрице, а део који се односи на издвајање кластера након факторизације остаје исти. Регуларизована верзија алгоритма је примењена на подацима из претходног поглавља, а граф структура се формира на основу додатних података: сличности између протеинских секвенци или постојања интеракције између протеина. Регуларизација је допринела побољшању резултата.

Седмо поглавље представља резултате примене предложене методе за интегративно кластеровање на још један проблем са подручја биоинформатике. Равојем технологија секвенцирања омогућено је добијање детаљних геномских информација директно из окружења. Актуелна област истраживања је метагеномика која се бави анализом микробиолошких заједница. У тим анализама алгоритми кластеровања имају битну улогу. У дисертацији је рађено на подацима који садрже приближно 69 милиона маркер секвенци - 16S rRNA добијених секвенцирањем 1967 узорака. Диверзитет микробиолошких заједница може се измерити разноврсним мерама али постоји дилема коју меру изабрати. У дисертацији је предложена интеграција резултата кластеровања узорака добијених применом 24 различите мере диверзитета. Интеграцијом се постиже стабилан резултат кластеровања, на нивоу најбољег резултата у ансамблу и резултат је робустан на промене параметара.

У осмом поглављу изложени су закључци и наведени могући правци даљих истраживања у областима машинског учења и биоинформатике.

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

**Sanja Brdar**, Vladimir Crnojević, Blaž Zupan, Integrative Clustering by Nonnegative Matrix Factorization Can Reveal Coherent Functional Groups From Gene Profile Data, IEEE Journal of Biomedical and Health Informatics, 19(2), 698-708. 2015. (M21)

**Sanja Brdar**, Vladimir Crnojević, Ensemble Approaches for Stable Assessment of Clusters in Microbiome Samples, Computational Intelligence Methods for Bioinformatics and Biostatistics, Stirling, Scotland, UK, 2016. (M33)

**Sanja Brdar**, Non-negative Matrix Factorization for Integrative Clustering of Bioinformatics Data, Belgrade Bioinformatics Conference – BelBI, Belgrade, Serbia, 2016. (M34)

**Sanja Brdar**, Vladimir Crnojević, Blaž Zupan, Non-Negative Matrix Factorization as a Framework for Integration of Gene Clusters, Proceedings of the Sixth International Workshop on Machine Learning in Systems Biology, Basel, Switzerland, 2012. (M34)

**Sanja Brdar**, Dubravko Čulibrk, Vladimir Crnojević, Consensus Clustering for Integrated Discovery of Functionally Related Gene Sets, Proceedings of the Eight Conference on Digital speech and image processing, Kovačica, Serbia 2012. (M63)

**Sanja Brdar**, Non-negative Matrix Factorization for Integrative Clustering, Thesis presented at seminar of Knowledge management in bioinformatics group at Humboldt University, Berlin Germany, September 2015.

**VII ЗАКЉУЧЦИ ОДНОСНО РЕЗУЛТАТИ ИСТРАЖИВАЊА**

За анализу огромних количина сложених и хетерогених података који су нам данас доступни и који се константно генеришу потребно је осмислити нове алгоритме за њихову обраду. У дисертацији су предложени интегративни приступи за кластеровање – груписање података - засновани на ненегативној факторизацији матрице са циљем повећања поузданости, робустности на присуство шума и екстремних вредности у подацима, омогућавања фузије података. Интегративним методама решавају се проблеми који су присутни код индивидуалног кластеровања: различите иницијализације, дилеме око избора параметара, мера сличности, скупа обележја, скупа података.

Најважнији резултати који су остварени у оквиру дисертације:

- Развој алгоритама за интегративно кластеровање заснованих на примени ненегативне факторизације матрице: преклапајући, непреклапајући, регуларизовани.
- Примена на проблеме са подручја биоинформатике: откривање функционално кохерентних група гена, груписање пацијента на основу генских експресија и интеграција кластеровања у метагеномици.
- Имплементација софтверског модула који обухвата предложене алгоритме и остале најзначајније алгоритме интегративног кластеровања са којима је вршено поређење.

Детаљне и систематичне евалуације, рађене на великом броју различитих скупова података унапредиле су сазнања о алгоритмима за интегративно кластеровање и показале да предложена метода спада у групу бољих метода. У решавању конкретних проблема са подручја биоинформатике предложена метода дала је значајно боље резултате од осталих метода интеграције. Проширење у виду

<p>регуларизације омогућило је увођење додатних информација и доменског предзнања у поступак кластеровања података. Тиме су резултати додатно побољшани.</p>
<p><b>VIII ОЦЕНА НАЧИНА ПРИКАЗА И ТУМАЧЕЊА РЕЗУЛТАТА ИСТРАЖИВАЊА</b></p>
<p>Кандидаткиња Сања Брдар у потпуности је обавила истраживања која су била предвиђена планом датим у пријави докторске дисертације. Целокупан приказ дисертације је добро и јасно структуриран. Резултати су детаљно и систематично интерпретирани и упоређени са релевантним методама из области. Статистичка обрада података и добијених резултата обухватила је: пермутацијске тестове, подзорковање, тестирање статистичке значајности, АНОВА и непараметарске тестове. Велики број графичких приказа омогућава лакше разумевање резултата представљених у дисертацији. Изведени су закључци утемељени на великом броју експеримената и разноврсним подацима што показује да кандидаткиња влада материјом и поседује висок ниво знања из области ненадгледаног машинског учења и биоинформатике. Дат је обиман преглед релевантне литературе. У складу са наведеним, Комисија позитивно оцењује начин приказа и тумачења резултата истраживања.</p>
<p><b>IX КОНАЧНА ОЦЕНА ДОКТОРСKE ДИСЕРТАЦИЈЕ:</b></p>
<p>1. Да ли је дисертација написана у складу са образложењем наведеним у пријави теме</p> <p>Докторска дисертација у потпуности је написана у складу са образложењем наведеним у пријави теме.</p>
<p>2. Да ли дисертација садржи све битне елементе</p> <p>Дисертација садржи све битне елементе научно-истраживачког рада. Тема, садржај, преглед литературе, методологија, приказ и тумачење резултата задовољавају захтеве нивоа докторске дисертације.</p>
<p>3. По чему је дисертација оригиналан допринос науци</p> <p>Докторска дисертација садржи оригиналне научне доприносе у домену ненадгледаног машинског учења. У дисертацији су предложене методе за интегративно кластеровање које су даље примењене на разноврсним подацима првенствено са подручја биоинформатике, али и на подацима типичним за машинско учење и синтетичким подацима. Из опсежног експерименталног истраживања које је кандидаткиња спровела произашли су резултати применљиви за решавање проблема предикције функције гена/протеина, груписања оболелих од рака на основу измерених генских експресија и интеграције мера диверзитета у метагеномици. Интегративно кластеровање података је веома актуелна тема истраживања. Кандидаткиња је тему детаљно сагледала и критички обрадила. Постављене су хипотезе да интеграција (података, параметара, објеката, обележја и/или узорака) може допринети побољшању резултата кластеризације и умањити ризик од погрешног одабира једне реализације кластеризације. Резултати истраживања потврђују постављене хипотезе.</p> <p>Резултати докторске дисертације објављени су у међународном научном часопису са импакт фактором и саопштени на међународним скуповима, чиме се даље потврђује да докторска дисертација представља оригиналан допринос науци.</p>
<p>4. Недостаци дисертације и њихов утицај на резултат истраживања</p> <p>Дисертација нема недостатке који утичу на резултате истраживања.</p>

<b>X ПРЕДЛОГ:</b>
На основу укупне оцене дисертације, комисија предлаже:
- да се докторска дисертација под називом „Алгоритми интегративног кластеровања података применом ненегативне факторизације матрице“ прихвати, а кандидаткињи Сањи Брдар одобри одбрана.

ПОТПИСИ ЧЛАНОВА КОМИСИЈЕ

---

др Вељко Милутиновић, редовни професор,  
Електротехнички факултет, Београд

---

др Војин Шенк, редовни професор,  
Факултет техничких наука, Нови Сад

---

др Владимир Црнојевић, редовни професор,  
Природно-математички факултет, Нови Сад

---

др Срђан Шкрбић, ванредни професор,  
Природно-математички факултет, Нови Сад

---

др Татјана Лончар-Турукало, доцент,  
Факултет техничких наука, Нови Сад

---

**Ментор:** др Дејан Вукобратовић, ванредни  
професор, Факултет техничких наука, Нови  
Сад

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.