

18.09.2013. године
Београд,

УНИВЕРЗИТЕТ У БЕОГРАДУ
Веће научних области техничких
наука

Београд
Студентски трг бр 1

У прилогу дописа достављамо потребну документацију за давање сагласности на предлог теме докторске дисертације кандидата Милоша Јовановића.

Захтев се доставља у складу са чланом 54 Закона о Универзитету.

С поштовањем,

СЕКРЕТАР ФАКУЛТЕТА

Гордана Алексић Славински, дипл. правник

Факултет организационих наука

УНИВЕРЗИТЕТ У БЕОГРАДУ

04-03-11/150
(Број захтева)

Веће научних области техничких наука
(Назив стручног већа коме се захтев упућује
сходночл 6. и члану 7. ст. 1. овог Правилника)

18.09.2013.
(Датум)

ЗАХТЕВ
за давање сагласности на предлог теме докторске дисертације

Молимо да, сходо члану 57. ст. 3. Закона о универзитету («Службени гласник РС» бр. 21/02), дате сагласност на предлог теме докторске дисертације: (пун назив предложене теме докторске дисертације)
АУТОМАТСКО ГЕНЕРИСАЊЕ АЛГОРИТАМА СТАБАЛА ОДЛУЧИВАЊА ЗА КЛАСИФИКАЦИЈУ

НАУЧНА ОБЛАСТ: ОРГАНИЗАЦИОНЕ НАУКЕ

ПОДАЦИ О КАНДИДАТУ:

1. Име, име једног од родитеља и презиме кандидата:
Милош (Златко) Јовановић
2. Назив и седиште Факултета на коме је стекао високо образовање: ФОН
3. Година дипломирања: 2006. године
4. Назив приступног рада кандидата: Аутоматско генерисање алгоритама дрва одлучивања за класификацију
5. Назив факултета на коме приступни рад одбрањен: ФОН
6. Година одбране приступног рада: 2013. година
7. Назив факултета на коме је кандидат завршио докторске студије: /
Одсек, смер или група: /
Година завршетка докторских студија: _____

Обавештавамо вас да је: НАСТАВНО НАУЧНО ВЕЋЕ ФОН-А
(назив надлежног тела факултета)

на седници одржаној: 18.09.2013. године размотрило предложену тему и закључило да је тема подобна за израду докторске дисертације.

ДЕКАН ФАКУЛТЕТА

Проф. др Милан Мартић

Прилог:

1. Предлог теме докторске дисертације са образложењем.
2. Акт надлежног тела факултета о подобности теме за израду докторске дисертације.

ПОДАЦИ О МЕНТОРУ

за кандидата Милоша Јовановића

Име и презиме ментора: **Милија Сукновић**

Звање: **Редовни професор**

Списак радова који квалификују ментора за вођење докторске дисертације:

- [1] Сукновић, М., Б. Делибашић, М. Јовановић, М. Вукићевић, Д. Бечејски-Вујаклија и З. Обрадовић, *Reusable Components in Decision Tree Induction Algorithms*, Computational Statistics. DOI: <http://dx.doi.org/10.1007/s00180-011-0242-8>, ISSN: 1613-9658, IF=0.5 за 2010.г., 2011.
- [2] Делибашић Б, М. Јовановић, М. Вукићевић, М. Сукновић и З. Обрадовић, *Component-based decision trees for classification*, Intelligent Data Analysis, Vol. 15, No. 5, DOI: <http://dx.doi.org/10.3233/IDA-2011-0489>, ISSN: 1088-467X, (print), 1571-4128 (online), IF=0.4 за 2010.г., 2011.
- [3] Радојевић, Г и М. Сукновић, *Scoring models: towards the more realistic approach*, Computer Science and Information Systems, Vol. 6, No. 1, p. 45-69, www.doiserbia.nb.rs/ft.aspx?id=1820-02140901045R, ISSN: 1820-0214, DOI: 10.2298/CSIS0901045R, IF=0.324 за 2010.г., 2009.
- [4] Вуковић, С., Делибашић, Б., Узелац, А., Сукновић, М., *A case-based reasoning model that uses preference theory functions for credit scoring*, Expert Systems With Applications, DOI: <http://dx.doi.org/10.1016/j.eswa.2012.01.181>, IF=1.926 за 2010.г., 2012.
- [5] Делибашић, Б., Вукићевић, М., Јовановић, М., Kirchen, K., Ruhland, J., Сукновић, М., *An architecture for component-based desing of representative-based clustering algorithms*, Data and Knoweledge Engineering, DOI: <http://dx.doi.org/10.1016/j.datak.2012.03.005>, IF=1.722 за 2010.г., 2012.
- [6] Вукићевић, М., Јовановић, М., Делибашић, Б., Ишљамовић, С., Сукновић, М., *Reusable Component-Based Architecture for Decision Tree Algorithm Design*, International Journal on Artificial Intelligence Tools, DOI: <http://dx.doi.org/10.1142/s0218213012500224>, IF=0.330 за 2010.г., 2012.
- [7] Вујошевић, Д., Ковачевић, И., Сукновић, М., Лалић Н., *A Comparison of the sability of Performing Ad Hoc Querying on Dimensionally Modeled Data Versus Operationally Modeled Data*, Decision Support Systems, DOI: <http://dx.doi.org/10.1016/j.dss.2012.05.004>, IF=2.135 за 2010.г., 2012.
- [8] Вукићевић, М., Kirchner, K., Делибашић, Б., Јовановић, М., Ruhland, J., Сукновић, М., *Finding best algorithmic components for clustering microarray data*, Knowledge and Information Systems, DOI: ,IF=2.225 за 2011.г., 2012., (Dobijena potvrda o prihvatanju rada).

- [9] Јовановић, М., Делибашић, Б., Вукићевић, М., Сукновић, М., Мартић, М., "Evolutionary approach for automated component-based decision tree algorithm design", appear in Volume 18(1), Intelligent Data Analysis - An International Journal, DOI:, IF=0.448 за 2011.г., 2012., (Dobijena potvrda o prihvatanju rada).
- [10] Делибашић, Б., Вукићевић, М., Јовановић, М., Сукновић, М., "White-Box or Black-Box Decision Tree Algorithms: Which to use in education?", IEEE Transactions on Education, DOI:, IF=1.021 за 2011.г., 2012., (Dobijena potvrda o prihvatanju rada).

Заокружити одговарајућу опцију(А,Б,В или Г):

- А) У случају менторства дисертације на докторским студијама у групацији техничко-технолошких, природно-математичких и медицинских наука ментор треба да има најмање три рада са SCI, SSCI, AHCI или SCIE листе.
- Б) У случају менторства дисертације на докторским студијама у групацији друштвено-хуманистичких наука ментор треба да има најмање три рада са релевантне листе научних часописа (Релевантна листа научних часописа обухвата SCI, SSCI, AHCI и SCIE листе, листу часописа које је Министарство за науку класификовало као M24 и додатну листу часописа коју ће, на предлог универзитета, донети Национални савет за високо образовање. Посебно се вреднују и монографије које Министарство науке класификује као M11, M12, M13, M14, M41 и M51.)
- В) У случају израде докторске дисертације према ранијим прописима за кандидате који су стекли академски назив магистра наука ментор треба да има пет радова (рефернци) које га, по оцени Већа научних области, квалификују за ментора односне дисертације.
- Г) У случају да у ужој научној области нема квалификованих наставника, приложити одлуку Већа докторских студија о именовању редовног професора за ментора.

ДЕКАН ФАКУЛТЕТА

Датум, 18.09.2013.г.

М.П.

Поводом предлога Већа студијског програма докторских академских студија, а на основу члана 56. Статута Факултета, на седници Наставно-научног већа одржаној 18.9.2013. године, једногласно је донета следећа

ОДЛУКА

Усваја се Извештај Комисије за оцену научне заснованости пријављене докторске дисертације кандидата Милоша Јовановића под насловом »АУТОМАТСКО ГЕНЕРИСАЊЕ АЛГОРИТАМА СТАБАЛА ОДЛУЧИВАЊА ЗА КЛАСИФИКАЦИЈУ«. За ментора се именује др Милија Сукновић, ред. проф. ФОН-а. Извештај се доставља Већу научних области техничких наука Универзитета у Београду.

Председник Наставно научног већа

Проф. др Милан Мартић

Доставити:

- Ментору
- Стручном сараднику за ПДС
- Секретару органа управљања и стручних органа
- Архиви

Наставно-научном већу
Факултета организационих наука
Универзитета у Београду

Предмет: Подобност теме и кандидата **Милоша Јовановића** за израду докторске дисертације

Одлуком 05-01 бг 3/109-13 бр. од 15.07.2013. године, именовани смо за чланове Комисије за оцену подобности теме и кандидата **Милоша Јовановића** за израду докторске дисертације и научне заснованости теме **Аутоматско генерисање алгоритама стабала одлучивања за класификацију**.

На основу материјала приложеног уз Захтев кандидата, Комисија подноси следећи

ИЗВЕШТАЈ

1. Подаци о кандидату

1.1. Биографски подаци

Милош Јовановић је рођен 28.02.1982. године у Струги, Република Македонија. Од тада живи у Београду, где је завршио основно образовање и средње образовање.

Факултет Организационих Наука, Универзитета у Београду, уписује 2001. године. Дипломира на одсеку за Информационе системе, 2006. године са просечном оценом 8,93. Тема дипломског рада: „Репрезентација знања као мост између Дејта-мајнинга и Експертних система“.

Положио је све испите на докторским студијама са просечном оценом 10, на Факултету организационих наука, Универзитета у Београду, на одсеку за Операциона истраживања.

Током 2009. године завршава и два курса за усавршавање научно-истраживачких знања и вештина, организованих од стране Министарства Науке Републике Србије. Курсеве на тему „Академске вештине“ спровео је др Steve A. Quartie.

Од децембра 2007. године је запослен на Факултету организационих наука, на катедри за Организацију пословних система, Центар а пословно одлучивање, где ради до данас. Предмети на којима ради основним и мастер студијама су: Теорија одлучивања, Информациони системи за подршку одлучивању и Системи за подршку одлучивању, Пословна интелигенција, Дејта мајнинг, Складишта података, Менаџерско одлучивање.

Педагошки рад кандидата се може оценити као изузетно успешан, што потврђују резултати вишегодишњих анкета које Факултет спроводи код студената (просек оцена 4,5, на скали 1-5)

Коаутор је једне стручне књиге „Алгоритми машинског учења за откривање законитости у подацима“ (Делибашић, Сукновић, Јовановић - 2009.), која се користи као уџбеник на предмету Дејта мајнинг, на мастер студијама Факултета организационих наука.

У марту 2012. године је био ангажован као гостујући предавач на предмету „MW 31.1 Business intelligence“, на *Friedrich-Schiller* Универзитету у Јени (Немачка), у оквиру катедре за Информационе системе.

Током свог истраживачког рада је био и рецензент у неколико међународних часописа, међу којима и: International Journal of Computational Intelligence Systems и International Journal of Engineering Education.

У 2012. години постаје лиценцирани експерт за софтвер *RapidMiner*, један од најкоришћенијих софтвера отвореног кода за откривање законитости у подацима и анализу података.

1.2. Стечено научноистраживачко искуство

У наставку ће бити приказани научни резултати кандидата, објављених на конференцијама и часописима, од датума уписа докторских студија, категоризовани у складу са спецификацијом Министарства просвете и науке Србије. Сви радови су у вези са областима Система за подршку одлучивању, Складишта података, Откривања законитости у подацима (енг. *Data mining*).

M21:

1.Vukićević M., Kirchner K., Delibašić B., Jovanović M., Ruhland J., Suknović M.: *Finding best algorithmic components for clustering microarray data*, Knowledge and Information Systems, Vol 35 , No 1, 2012, pp. 111-130.

M22:

2.Delibašić B., Vukićević M., Jovanović M., Kirchner K., Ruhland J., Suknović M.: *An architecture for component-based design of representative-based clustering algorithms*, Data & Knowledge Engineering, Vol 75, May 2012., pp. 78-98.

3.Delibašić B., Vukićević M., Jovanović M., Suknović M.: *White-Box or Black-Box Decision Tree Algorithms: Which to Use in Education?*, IEEE Transactions on Education, Vol 56, No 3, 2011., pp. 287-291.

4.Jovanović M., Vukićević M., Milovanović M., Minović M.: *Using data mining on student behavior and cognitive style data for improving e-learning systems: a case study*, International Journal of Computational Intelligence Systems, Vol 5, No 3, 2012, pp. 597-610.

M23:

5.Delibašić B., Kirchner K., Ruhland J., Jovanović M., Vukićević M.: *Reusable components for partitioning clustering algorithms*, Artificial Intelligence Review, Vol 32, No 1, 2009, pp. 59-75.

6.Suknović M., Delibašić B., Jovanović M., Vukićević M., Bečejski-Vujaklija D., Obradović Z.: *Reusable Components in Decision Tree Induction Algorithms*, Computational Statistics, Vol 27, No 1, 2012, pp. 127-148.

7.Delibašić B., Jovanović M., Vukićević M., Suknović M., Obradović Z.: *Component-based decision trees for classification*, Intelligent Data Analysis, Vol 15, No 5, 2011, pp. 671-693.

8.Vukićević M., Jovanović M., Delibašić B., Išlamović S., Suknović M.: *Reusable component-based architecture for decision tree algorithm design*, International Journal on Artificial Intelligence Tools, Vol 21, No 5, 2012., online.

M33:

9.Vukićević M., Jovanović M., Delibašić B., Suknović M.: *WhiBo - RapidMiner plug-in for component based data mining algorithm design*, Proceedings of the 1st RapidMiner Community Meeting and Conference, Dortmund, Germany, 2010., pp. 30-35.

10. Jovanović M., Delibašić B., Vukićević M., Suknović M.: *Optimizing performance of decision tree component-based algorithms using evolutionary algorithm in RapidMiner*, Proceedings of the 2nd RapidMiner Community Meeting and Conference, Dublin, Ireland, 2011., pp. 135-149.

11. Vukićević M., Delibašić B., Jovanović M., Suknović M., Obradović Z.: *Internal Evaluation Measures as Proxies for External Indices in Clustering Gene Expression Data*, Proceedings of the 2011 IEEE International Conference on Bioinformatics and Biomedicine (BIBM11), Atlanta, Georgia, USA, 2011., pp. 574-577.

12. Vukićević M., Delibašić B., Obradović Z., Jovanović M., Suknović M.: *A Method for Design of Data-tailored Partitioning Algorithms for Optimizing the Number of Clusters in Microarray Analysis*, Proceedings of the 2012 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology, San Diego, CA, USA, 2012., pp. 252-259.

13. Jovanović M., Stojanović J., Vukićević M., Stojanović V., Delibašić B., Suknović M.: *NeurophRM: Integration of the Neuroph framework into RapidMiner*, Proceedings of the 3rd RapidMiner Community Meeting and Conference, Budapest, Hungary, 2012, pp. 39-50.

M34:

14. Bečejski-Vujaklija D., Delibašić B., Jovanović M.: *Business intelligence model for electronic documents archive*, Proceedings of the 23 EURO conference, Bonn, Germany, 2009., pp. 137.
15. Jovanović M., Delibašić B., Vukićević M.: *A "white box" data mining platform for decision support in decision tree induction algorithm design*, Proceedings of the 23 EURO conference, Bonn, Germany, 2009., pp. 137.
16. Delibašić B., Suknović M., Jovanović M.: *Patterns as building blocks in decision-tree algorithms*, Proceedings of the 23 EURO conference, Bonn, Germany, 2009., pp. 269.
17. Delibašić B., Jovanović M., Vukićević M., Suknović M., Kirchner K., Ruhland J., Obradović Z.: *A decision support system architecture for data mining based on reusable components (patterns)*, Proceedings of the EWG-DSS London 2011 Workshop on Decision Support Systems, London, UK, 2011., pp. 35.
18. Jovanović M., Vukićević M., Išlamović S., Suknović M.: *Automatic evolutionary design of decision tree algorithm for prediction of university student success*, Proceedings of the Stochastic Modeling Techniques and Data Analysis International Conference (SMTDA 2012), Chania, Crete, Greece, 2012., pp. 48-49.

M51:

19. Suknović M., Jovanović M., Delibašić B., Vukićević M.: *Business intelligence system development over document meta data in an organization*, Management - časopis za teoriju i praksu menadžmenta, Vol. 15, No 54, 2010., pp. 5-13.

M63:

20. Vukićević M., Suknović M., Delibašić B., Jovanović M.: *Modelovanje sporo menjajućih dimenzija data marta za praćenje parametara nastavnog procesa*, Zbornik radova sa konferencije SYMOPIS, Soko Banja 2008., str. 283-286.
21. Vukićević M., Suknović M., Delibašić B., Jovanović M.: *Projektovanje sistema poslovne inteligencije za potrebe unapređenja nastavnog procesa*, Zbornik radova sa konferencije SYMORG, Beograd 2008., CD format.
22. Jovanović M., Delibašić B., Suknović M., Bečejski-Vujaklija D., Vukićević M.: *Primena poslovne inteligencije na oblast sistema upravljanja dokumentima*, Zbornik radova sa konferencije SYMORG, Beograd 2008., CD format.
23. Delibašić B., Suknović M., Bečejski-Vujaklija D., Jovanović M., Vukićević M.: *Pattern platforma za dejta majning*, Zbornik radova sa konferencije SYMOPIS, Soko Banja 2008., pp. 287 - 290.
24. Vukićević M., Delibašić B., Suknović M., Jovanović M.: *Whibo – generic decision tree environment for effective decision making*, Zbornik radova sa konferencije SYMOPIS, Ivanjica 2009., pp. 705-708.
25. Jovanović M., Suknović M., Vukićević M., Delibašić B.: *A white box approach in modeling phase of the data mining process*, Zbornik radova sa konferencije SYMOPIS, Ivanjica 2009., pp. 709-712.
26. Jovanović M., Delibašić B., Vukićević M., Suknović M.: *An open-source platform for design and testing of data mining algorithms*, Zbornik radova sa konferencije SYMOPIS, Tara 2010., pp. 769-772.
27. Vukićević M., Suknović M., Jovanović M.: *An open source environment for generic partitioning clustering algorithm design*, Zbornik radova sa konferencije SYMORG, Zlatibor 2010., CD format.
28. Vukićević M., Išlamović S., Jovanović M., Delibašić B., Suknović M.: *Primena neuronskih mreža za predviđanje uspeha studenata*, Zbornik radova sa konferencije YU INFO, Kopaonik 2012., pp. 680-684.

M85:

29. Delibašić B., Jovanović M., Vukićević M., Suknović M.: *WhiBo: An open-source data mining framework*, Platforma za razvoj algoritama za otkrivanje zakonitosti u podacima napisana u programskom jeziku Java, besplatno dostupna na Internet adresi: www.whibo.fon.bg.ac.rs, 2009.

Научноистраживачки пројекти у којима је кандидат учествовао:

- Пројекат Министарства науке Републике Србије, 2008/09., број пројекта: TP12013, Тема: *Развој платформе за моделовање компоненти и документовање развијених модела унутар стандардног процеса дејта мајнинга*, руководиоца: др Борис Делибашић.
- Пројекат Министарства науке Републике Србије, 2011-2014., број пројекта: ИИИ 47003, Тема: *Инфраструктура за електронски подржано учење у Србији*, руководиоца: др Владан Девеџић.

- Пројекат Министарства науке Републике Србије, 2011-2014., број пројекта: ИИИ 47008, Тема: *Интеракција етиопатогенетских механизма парадонтопатије и периимплантитиса са системским болестима данашњице*, руководилац др Војислав Лековић.

На докторским студијама кандидат је положио следеће испите (просек оцена 10.0):

- Нелинеарно програмирање, проф. др Вера Ковачевић-Вујчић
- Комбинаторна оптимизација, проф. др Мирјана Чангаловић
- Вишекритеријумска оптимизација, проф. др Милан Мартић
- Метакхеуристике, проф. др Мирјана Чангаловић
- Пословна интелигенција – изабрана поглавља, проф. др Борис Делибашић
- Одлучивање - изабрана поглавља, проф. др Милија Сукновић
- Меко рачунање - одабрана поглавља, проф. др Мирко Вујошевић
- Нови трендови у операционим истраживањима, проф. др Милан Мартић
- Вештачка интелигенција - одабрана поглавља, проф. др Владан Девеџић

На основним студијама Факултета организационих наука, кандидат је ангажован на следећим предметима:

- Теорија одлучивања
- Пословна интелигенција
- Системи за подршку одлучивању

На мастер студијама Факултета организационих наука, кандидат је ангажован на следећим предметима:

- Дејта мајнинг
- Складишта података
- Системи пословне интелигенције
- Изабрана поглавља из Теорије одлучивања

1.3. Оцена подобности кандидата за рад на предложеној теми

Узимајући у обзир:

- резултате и ток досадашњег **образовања**;
- **истраживања** на тему откривања законитости у подацима, стабала одлучивања и алгоритама, објављеним у признатим међународним часописима и конференцијама;
- **држања наставе** из области из које је предложена тема;
- и учешћа на **пројектима**;

закључујемо да је Милош Јовановић одлично припремљен и квалификован да ову тему самостално истражује и начини доприносе у тој области.

2. Предмет и циљ истраживања

Стабла одлучивања представљају модел процеса доношења одлука, који у корацима, кроз „гране“ стабла, описује на који начин се користе информације да би се дошло до неке одлуке, представљене у „листу“ стабла одлучивања.

Стабла одлучивања могу ручно дефинисати експерти из личног искуства, или се могу изградити из података без присуства експерта (такозвана "индуктивна стабла одлучивања"). У овом раду се изучава приступ изградње стабала одлучивања из података, што делом припада области машинског учења и откривања законитости у подацима (енг. *data mining*).

Постоје бројни алгоритми за изградњу стабала одлучивања који су обрађивани у литератури. Ови алгоритми су прављени као монолитни и у облику процедуре чијим извршавањем се добија резултујуће стабло одлучивања. Овај приступ се назива приступ „црних кутија“. Са друге стране, постоји начин описивања оваквих алгоритама као низа компоненти за решавање одређених потпроблема, чиме се отвара могућност комбиновања компоненти (делова) различитих алгоритама стабала одлучивања за класификацију и генерисање мноштва нових алгоритама.

Дакле, **предмет** изучавања овог истраживања су:

- алгоритми за изградњу стабла одлучивања, као модела за класификацију,
- структура ових алгоритама, и простор алгоритама који се добија комбиновањем делова алгоритама (компоненти),
- аутоматско генерисање алгоритама за стабла одлучивања за специфичну примену, тако што се тражи оптимална комбинација компонената.

Алгоритми се оцењују преко своје могућности да креирају стабла која прецизно предвиђају класе објеката од разматрања. Квалитет алгоритама се може егзактно мерити (Pang-Ning et al., 2006), што ће бити коришћено за поређење различитих алгоритама и дискусију о утицају појединих фактора на тај квалитет.

Циљеви овог истраживања су:

- систематизација прегледа литературе о алгоритама за изградњу стабла одлучивања за класификацију;
- истраживање структура ових алгоритама и ограничења приликом састављања алгоритама из компоненти;
- дефинисање простора алгоритама и испитивање оптимизационих техника за претраживање тог простора;
- изучавање могућности за аутоматско генерисање алгоритама за стабла одлучивања, као и параметризацију компоненти;
- анализа конструисаних алгоритама на специфичним применама стабла одлучивања, како би се извршила карактеризација и опис "добрих" компоненти;
- изучавање практичне применљивости овог приступа, као и оправданости у односу на алтернативне приступе;
- изградња софтверског алата који би служио као подршка одлучивању, аналитичарима који примењују алгоритме за стабла одлучивања, што би значајно олакшало и убрзало примену.

У наставку ће бити дат преглед релевантних истраживања на основу којих су дефинисани циљеви и који илуструју актуелност и потребу за оваквом темом.

Изучавање алгоритама за изградњу стабла одлучивања потиче од седамдесетих година предходног века, упоредо са развојем рачунарства. Међу запаженијим алгоритмима су сигурно и ID3 (Quinlan, 1986), C4.5 (Quinlan, 1993), CART (Breiman et al., 1984) и CHAID (Kass, 1980), и они су присутни у готово сваком алату за откривање законитости у подацима (ОЗП). Такође су препознати међу популарнијим алгоритмима за ОЗП уопште, не само у оквиру алгоритама за стабла одлучивања (Wu et al., 2008). Како су се стабла одлучивања показала корисна, развијају се и унапређења основних алгоритама, попут (Mantaras, 1991) или (Loh & Shih, 1997).

Приликом сваког унапређења алгоритама, аутори покушавају да покажу да су унапређени алгоритми објективно бољи од оригиналних. Али многе студије нису могле да потврде супериорност неког од алгоритама, због чега заправо ниједан алгоритам не бива напуштен. Разлози за све то постају јаснији када Волперт (Wolpert, 1996) теоретски доказује да се за сваки специфичан проблем може конструисати алгоритам који је оптималан за тај проблем, иако тај алгоритам на осталим проблемима може бити јако лош. Ово је био јак сигнал да се одустане од потраге за "најбољим" алгоритмом, и да је добро да постоје више алгоритама, где сваки има потенцијал да буде "најбољи" на специфичном проблему.

Пратећи ове сигнале, Сукновић, Делибашић са коауторима (Suknović et al., 2012; Delibašić et al., 2011) покушавају да развију много већи број алгоритама, комбинацијом делова постојећих алгоритама. Ово се испоставило могућим, јер већина алгоритама за изградњу стабла одлучивања има сличну генералну структуру, док се разликују појединачни избори решавање потпроблема унутар алгорита.

Размењујући компоненте алгорита, сада је могуће претраживати читав простор алгоритама. Како се додају нове компоненте у систем, простор могућих алгоритама расте експоненцијално, што значи да врло брзо неће бити могуће састављати алгоритама из компоненти ручно. Са друге стране, по (Wolpert, 1996), јасно је да би то донело боље резултате алгоритама.

Избором алгоритама су се истраживачи и раније бавили, где је семинални рад Рајса (Rice, 1976) описао и основне проблеме при том процесу. Дobar преглед новијег датума се може наћи код Смит-Мајлс (Smith-Miles, 2009). Напори у овим истраживањима се више базирају на томе како исправно мерити успешност алгоритама и како помоћи одлучивање при избору алгоритама за неки задатак. Разлог због чега они не разматрају оптимизационе технике, је што су генерално алгоритми за ОЗП доста различити. Због тога је предмет ове дисертације дефинише у оквиру једне фамилије алгоритама (за стабла одлучивања), где је могуће препознати заједничку структуру тих алгоритама.

Како је компонентни дизајн решења много раније примењен и на друге области, попут кластеровања (Delibašić et al., 2009), дизајна софтвера (Gamma et al., 1995), па чак и архитектуре (Alexander, 1979), постоји оправдано мишљење да би развијени систем над стаблима одлучивања могао да се примени и на друге фамилије алгоритама.

Овакав проблем још није директно решаван у литератури, али постоје слични и аналогни проблеми који јесу били предмет истраживања. Генерални преглед могућности спајања оптимизације (тј. операционих истраживања) и алгоритама ОЗП се може наћи у Олафсон и Ли (Olafsson & Li, 2008). Један од активних приступа у избору алгоритама је и мета-учење, где се применом самих алгоритама за ОЗП покушава предвидети најбољи алгоритама за примену на неком реланом проблему (Smith-Miles, 2009). Овај приступ није у супротности са темом овог доктората, већ је комплементаран, јер се коришћењем резултата мета-учења може знатно смањити простор за претрагу, који би се могао претражити егзактно, уместо апроксимацијом као у приступу мета-учења. Други приступ који је аналоган предложеној теми, је приступ генетског програмирања (Koza, 1991), где се на много нижем нивоу алгоритми представљају појединачним инструкцијама, па се генетским алгоритмом оптимизује низ инструкција, како би се добио алгоритама који решава неки проблем (нпр. сортирање бројева). Проблем код овог приступа је што је на јако ниском нивоу (инструкција), што би за компликоване алгоритме, какви су алгоритми ОЗП, било врло непрактично. Најзад, компоненте алгоритама се могу третирали као хеуристике, па би предложени систем могао да се посматра и као хиперхеуристика, као у Вела са коауторима (Vella et al., 2009).

Референце:

1. Alexander C.: *The timeless way of building*, Oxford University Press, 1979.
2. Breiman L., Friedman J., Stone C.J., Olshen R.A.: *Classification and regression trees*, CRC Press, 1984.
3. Delibašić B., Kirchner K., Ruhland J., Jovanović M., Vukićević M.: *Reusable components for partitioning clustering algorithms*, Artificial Intelligence Review, Vol 32, No 1, 2009, pp. 59-75.
4. Delibašić B., Jovanović M., Vukićević M., Suknović M., Obradović Z.: *Component-based decision trees for classification*, Intelligent Data Analysis, Vol 15, No 5, 2011, pp. 671-693.
5. Gamma E., Helm R., Johnson R., Vlissides J.M.: *Design patterns: elements of reusable object-oriented software*. Addison-Wesley, Reading, MA, 1995.
6. Kass G.V.: *An exploratory technique for investigating large quantities of categorical data*, Applied Statistics, Vol 29, 1980., pp. 119-127.
7. Koza J.: *Concept formation and decision tree induction using the genetic programming paradigm*, Parallel Problem Solving from Nature, Vol 1, 1991., pp. 1-6.
8. Loh W.Y., Shih Y.S.: *Split selection methods for classification trees*, Statistica Sinica, Vol 7, 1997., pp. 815-840.
9. Mantaras R.L.: *A distance-based attribute selection measure for decision tree induction*, Machine Learning, Vol 6, 1991., pp. 81-92.
10. Olafsson S., Li X., Wu S.: *Operations research and data mining*. European Journal of Operational Research, Vol 187, No 3, 2008., pp. 1429-1448.
11. Pang-Ning T., Steinbach M., Kumar V.: *Introduction to data mining*. Addison-Wesley. 2006.
12. Quinlan J.R.: *Induction of decision trees*, Machine Learning, Vol 1, 1986., pp. 81-106.

13. Quinlan J.R.: *C4.5 programs for machine learning*, Morgan Kaufmann, 1993.
14. Rice J.: *The algorithm selection problem*. Advances in Computers, Vol 15, 1976., pp. 65-118.
15. Smith-Miles K.A.: *Cross-disciplinary perspectives on meta-learning for algorithm selection*, ACM Computing Surveys, Vol 41, No 1, 2009., pp. 1-25.
16. Suknović M., Delibašić B., Jovanović M., Vukićević M., Bečejski-Vujaklija D., Obradović Z.: *Reusable Components in Decision Tree Induction Algorithms*, Computational Statistics, Vol 27, No 1, 2012, pp. 127-148.
17. Vella A., Corne D., Murphy C.: *Hyper-heuristic decision tree induction*. Proceedings of the 2009 World Congress on Nature & Biologically Inspired Computing, Coimbatore, India, 2009., pp. 409-414.
18. Wolpert D.H.: *The lack of a priori distinctions between learning algorithms*, Neural computation, Vol 8, 1996., pp. 1341-1390.
19. Wu X., Kumar V., Quinlan J.R., Ghosh J., Yang Q., Motoda H., McLachlan G.J., Ng A., Liu B., Yu P.S., Zhou Z.H., Steinbach M., Hand D.J., Steinberg D.: *Top 10 algorithms in data mining*, Knowledge information systems, Vol 14, 2008., pp. 1-37.

3. Полазне хипотезе

На основу задатих циљева, проучавања предходних истраживања из релевантне литературе и њихових закључака, могу се дефинисати хипотезе, као претпоставке које се могу експериментално верификовати.

Основна хипотеза је да се **простор алгоритама креиран разменом компоненти (делова) алгоритама може ефективно и ефикасно претраживати**. Верификовањем ове хипотезе би се отворио простор за аутоматско генерисање и других алгоритама (осим стабла одлучивања) представљених компонентама.

Као посебне хипотезе јављају се неколико претпоставки које произилазе из основне и које је могуће верификовати. Претпоставља се да **познавање заједничке (генеричке) структуре алгоритама омогућава примену ефикаснијих оптимизационих техника**. Верификовањем ове хипотезе би се створио мотив за детаљнији опис и изучавање особина и структуре алгоритама, ради побољшавања процеса избора алгоритама, што има велике користи и за мета-учење.

Даље, претпоставља се и да **пронађени оптимални (или субоптимални) алгоритми могу бити значајно бољи од случајних (или неких познатих) алгоритама**. Верификовањем ове хипотезе ће бити дат и јак практичан мотив за коришћење овог приступа.

Конечно, верује се да перформансе алгоритма не зависе подједнако од свих делова (компоненти). **Анализом аутоматски конструисаних алгоритама на неком проблему, могуће је открити који делови алгоритма су заслужнији за његове перформансе**.

4. Научне методе истраживања

Наведене хипотезе ће бити верификоване строгом научном методологијом, заснованом на емпиријској (експерименталној) потврди.

Прво, прегледом и **проучавањем литературе, критичким освртом на њу и систематизацијом** треба да се стекне јасна слика о проблемима и потенцијалним решењима за реализацију задатих циљева. Изучавање се алгоритми за стабла одлучивања, структура тих алгоритама и постојећа унапређења, са фокусом на компонентни дизајн за стабла одлучивања. Посебно ће се изучавати оптимизационе технике, као и мета-хеуристички методи за оптимизацију. Поред литературе, проучавање се и доступни алати за изградњу стабала одлучивања и решавања класификационих проблема уопште, као и компарација доступних алата. Такође, биће дат преглед примена алгоритама стабла одлучивања, а посебно и анализа потенцијалне користи аутоматског генерисања алгоритама за изградњу стабала одлучивања у реалним проблемима.

Треба јасно **дефинисати егзактне мере квалитета** алгоритма, на начин који би омогућио **експериментално поређење са другим приступима** у литератури. Процена колико добијено стабло одлучивања добро предвиђа класе ће се вршити искључиво на примерима који нису коришћени приликом изградње стабла, како би се мерила генерализација добијеног модела класификације.

Квалитет алгоритама ће се проверавати и на добро познатим (**бенчмарк**) проблемима, како би се резултати упоредили са постојећим приступима. Ово такође треба да омогући **транспарентност и поновљивост** свих експеримената. Алгоритми ће се проверавати на што већем броју примера, како закључци не би били појединачни, већ генерални.

Експерименти ће се спроводити на истој, јасно дефинисаној, хардверској инфраструктури, како би поређења била **фер и коректна**.

Оптимизационе методе ће се оцењивати са тиме колико је решење близу оптималном, као и колико времена је потребно за извршавање.

Подаци од спроведених експеримената ће бити анализирани и биће извучени **закључци, објашњења**, као и дискутована **нова истраживачка питања** која ће се тиме отворити.

5. Очекивани научни допринос

Предложена тема је истраживачки врло релевантна, базира се на актуелним истраживањима у области и може изродити низ доприноса научној заједници, међу којима вреди истаћи следеће:

- Сачињавање и систематизација прегледа литературе из области стабала одлучивања, уз уочавање сличних структура међу алгоритмима, које се могу искористити за генерализацију.
- Стварање оквира у коме се алгоритми могу аутоматски генерисати за одређени проблем који се решава, а комбиновањем делова алгоритама. Иако тема изучава применљивост на стабла одлучивања, овакав концепт би потенцијално био користан и код многих других фамилија алгоритама.
- Унапређење самих оптимизационих метода, прилагођавањем за специфичности претраживања простора алгоритама.
- Уочавање и описивање важности појединих компонената (делова) алгоритама за изградњу стабла одлучивања, а на основу анализа аутоматски конструисаних алгоритама.
- Емпиријска евалуација перформанси описаног система и поређење са постојећим приступима.

Истраживање на овој теми ће сигурно произвести и низ стручних доприноса, попут:

- Развоја софтверског алата за аутоматско генерисање алгоритама за изградњу стабала одлучивања, и интеграцију тог алата у актуелне алате за откривање законитости у подацима.
- Повећање ефикасности и ефективности примена стабла одлучивања у пракси, аутоматизацијом посла који засада аналитичари обављају ручно.

6. План истраживања и структура рада

Истраживање ће се обавити у следећим фазама:

Фаза 1:

- истраживање и систематизација литературе, које ће покривати области:
 - стабла одлучивања за класификацију, као и њихових модификација;
 - алгоритама базираних на компонентама;
 - аутоматске параметризација алгоритама;
 - оптимизационих метода које се могу применити на простор алгоритама;
 - аналогне приступе попут мета-учења, генетског програмирања, хипер хеуристика, итд.
- систематизација методолошког приступа који су примењивали слични радови из литературе, због лакшег поређења са постојећим резултатима;
- преглед и поређење доступних софтверских алата и практичне употребљивости предложеног приступа, посебно ради проширивања неког од постојећих алата.

Фаза 2:

- прецизирање коначних хипотеза, које се могу експериментално тестирати;
- креирање експеримената за њихово тестирање, на начин који ће омогућити проверу резултата и поређење са досадашњим резултатима доступним у литератури;
- спровођење експеримената на дефинисаној инфраструктури и контролисаним условима.

Фаза 3:

- бележење података из експеримената;
- приказивање и тумачења резултата;
- евентуалне промене и понављање модификованих (унапређених) експеримената, а на основу сазнања из првих резултата;
- дискусија резултата и отворених истраживачких питања.

7. Закључак и предлог

Закључујемо да је предложена тема "Аутоматско генерисање алгоритама стабала одлучивања за класификацију" актуелна и релевантна у датој ужој научној области моделирање пословних система и пословно одлучивање, као и да су очекивани научни доприноси интересантни како научној, тако и стручној заједници.

Кандидат Милош Јовановић, на основу предходног образовања, истраживачког рада, извођења наставе, учешћа у пројектима и другим активностима из области предложене теме доктората, поседује знања, искуство и квалификације да самостално истражује поменуту тему.

Као ментор се предлаже проф. др Милија Сукновић, редовни професор Факултета организационих наука, Универзитета у Београду.

ЧЛАНОВИ КОМИСИЈЕ

.....
Проф. др Милија Сукновић,
редовни професор
Факултет организационих наука
Универзитета у Београду

.....
Проф. др Борис Делибашић
ванредни професор
Факултет организационих наука
Универзитета у Београду

.....
др Драган Радојевић
научни саветник
Институт „Михајло Пупин“