

НАСТАВНО-НАУЧНОМ ВЕЋУ

Предмет: Подобност теме и кандидата Владисава Јелисавчића мастер инжењера електротехнике и рачунарства за израду докторске дисертације

Одлуком Наставно-научног већа Електротехничког факултета бр. 5039/11-1 од 5.10.2016. године, именовани смо за чланове Комисије за оцену подобности теме и кандидата Владисава Јелисавчића мастер инжењера електротехнике и рачунарства за израду докторске дисертације и научне заснованости теме **Структурирано учење над великим подацима**.

На основу материјала приложеног уз Захтев кандидата, Комисија подноси следећи

ИЗВЕШТАЈ

1. Подаци о кандидату

1.1. Биографски подаци

Владисав Јелисавчић је рођен 17.07.1987. године у Обреновцу. Основне студије је завршио на Електротехничком факултету, смер Рачунарска техника и информатика, са просечном оценом 8.70. Дипломирао 2010. године, успешно одбранивши дипломски рад на тему: "Визуални симулатор експертских система" код професора Бошка Николића. Мастер студије је завршио на истом смеру 2011., одбранивши мастер рад: "Евалуација алгоритама за моделовање знања на основу објављених радова" код професора Вељка Милутиновића. Докторске студије је уписао 2011/2012. године на модулу Рачунарска техника и информатика и положио је све испите предвиђене Наставним планом и програмом модула са просечном оценом 9.8. Од 2012. године је запослен као истраживач у Математичком институту Српске академије наука и уметности. Тренутно је ангажован на пројекту Министарства просвете, науке и технолошког развоја, ИИИ44006: *Развој нових информационо-комуникационих технологија коришћењем напредних математичких метода, са применама у медицини, енергетици, е-Управи, телекомуникацијама и заштити националне баштине*.

1.2. Стечено научноистраживачко искуство

На докторским студијама на Електротехничком факултету Универзитета у Београду кандидат је положио следеће испите:

- Архитектуре за подршку софтверским системима (9 ЕСПБ)

оцена 10 (десет)

- Одабрана поглавља из архитектуре рачунарских система (9 ЕСПБ)	оцена 10 (десет)
- Локалне рачунарске мреже (9 ЕСПБ)	оцена 10 (десет)
- Моделовање и мерење рачунарских перформанси (9 ЕСПБ)	оцена 10 (десет)
- Пројектовање рачунара у ВЛСИ техници (9 ЕСПБ)	оцена 10 (десет)
- Класификација и естимација сигнала (9 ЕСПБ)	оцена 8 (осам)
- Технике обраде и препознавања говорног сигнала (9 ЕСПБ)	оцена 10 (десет)
- Моделовање и мерење рачунарских перформанси (9 ЕСПБ)	оцена 10 (десет)
- Вештачка интелигенција и експертски системи (9 ЕСПБ)	оцена 10 (десет)
- Увод у научни рад (6 ЕСПБ)	оцена 10 (десет)

Поред испита, кандидат је обавио и два студијска истраживачка рада и један научно-стручни рад, што је предвиђено планом докторских студија.

Научно-истраживачко искуство је стекао учествовањем у пројектима Министарства за науку и технолошки развој (МНТР) Републике Србије. Пројекти су реализовани на Математичком институту Српске академије наука и уметности, или у Иновационом центру Електротехничког факултета. Пројекти припадају области машинског учења и дата мининга. Био је ангажован на ЕУ ФП7 пројекту BALCON: Boosting EU-Western Balkan Countries research collaboration in the Monitoring and Control area, GA: 288076. Такође био је ангажован као истраживач на иновационом пројекту МПНТР Примена метода за проналажење знања над великом количином података. Такође, поседује и искуство рада на неколико софтверских пројеката отвореног кода везаних за big data. Тренутно се води као један од активних комитера на Apache Ignite пројекту за дистрибуиране софтверске кеш системе.

Године 2014. одлази на стручно усавршавање на Temple University, Philadelphia, Pennsylvania, USA. Током боравка учествовао је на пројекту Prospective Analysis of Large and Complex Partially Observed Temporal Social Networks, Defense Advanced Projects Agency, DARPA-GRAPHS, AFOSR award number FA 9550-12-1-0406, где је стекао искуство из области структурираног учења.

Владисав Јелисавчић је коаутор три рада објављена на међународним конференцијама. Учествовао је у изради три техничка решења, категорије M82 и M83.

Публикације на научним скуповима

1. Stojkovic, I., Jelisavcic, V., Milutinovic, V., Obradovic, Z., "Distance Based Modeling of Interactions in Structured Regression," Proc. 25th International Joint Conference on Artificial Intelligence (IJCAI), New York, NY, July 2016, pp. 2032 – 2038
2. Jelisavčić V., Furlan B., Protić J., Milutinović V., "Knowledge Modeling and Classification of Scientific Papers Based on Topic Modelilng," YUINFO, Kopaonik, Serbia, 2012, pp. 664-669
3. Jelisavčić V., Furlan B., Protić J., Milutinović V., "Topic Models and Advanced Algorithms for Profiling of Knowledge in Scientific Papers," MIPRO, Proceedings of the 35th International Convention, Opatija, Croatia, 2012, pp. 1030 – 1035

Дана 19.10.2016. године одржана је јавна усмена одбрана теме докторске дисертације пред комисијом коју су чинили:

- др Јелица Протић, ванредни професор Електротехничког факултета Универзитета у Београду,

- др Милош Ковачевић, ванредни професор Грађевинског факултета Универзитета у Београду

и

- др Горан Квашчев, доцент Електротехничког факултета Универзитета у Београду.

Комисија је кандидата на јавној усменој одбрани оценила оценом задовољно.

1.3. Оцена подобности кандидата за рад на предложеној теми

Комисија је размотрила све претходно наведене резултате досадашњег рада кандидата. Кандидат је кроз положене испите на докторским студијама, кроз остварено учешће у научним и стручним пројектима, стручном усавршавању у иностранству, успешну одбрану теме докторске дисертације, и објављене научне радове показао да је стекао неопходно искуство за самосталан научно-истраживачки рад. Комисија сматра да је кандидат способан да током израде докторске дисертације оствари резултате који ће бити верификовани од стране међународне научне заједнице. Комисија констатује да кандидат Владисав Јелисавчић испуњава све потребне услове да приступи изради докторске дисертације под наведеном темом. Такође, комисија констатује да је радни назив теме превише уопштен, и стога предлаже да се назив теме промени у **"Структурирано учење над великим подацима засновано на вероватносним графовским моделима"**. Предложени назив тезе на енглеском је „Structured learning from big data based on probabilistic graphical models“.

2. Предмет и циљ истраживања

Структурирано учење је област која се бави учењем сложених објеката из сложених (структурираних) података. Структурирани подаци, у овом контексту, представљају податке који се састоје од више међусобно зависних делова. Код учења над оваквим подацима потребно је искористити информацију која се садржи и у самим деловима, као и у њиховим међусобним односима. Област структурираног учења је важан сегмент машинског учења, предмет је активног истраживања уз бројне актуелне проблеме и покрива учење из текста [1], сигнала (аудио [2], слика [3][4], видео [5]), интеракција у социјалним мрежама [6], биомедицинских (генетских [7][8], болничких [9]) података, и бројне друге домене.

Структурирана предикција је дискриминативна подврста структурираног учења и бави се проналажењем веза између улаза и сложених структурираних излаза. За разлику од регресије или класификације, код којих је циљ предвидети један излаз, објекти о којима се закључује структурираном регресијом, односно структурираном класификацијом, имају сложену структуру.

Један од израженијих изазова структурираног учења представља скалабилност. Са повећањем димензије проблема, значајно се повећава време потребно за обучавање, услед превелике сложености алгоритама за учење. У општем случају сложеност алгоритама за обучавање расте експоненцијално са повећањем димензије проблема. Због тога модели структурираног учења користе структурирану природу проблема, чиме се сложеност своди на полиномијалну. Једна од најчешћих претпоставка код структурираног учења је тзв. Марковљева претпоставка код које се, на основу Хемерсли-Клифорд теореме, структура проблема може представити неусмереним графом. Ова претпоставка представља основ за читаву класу алгоритама за структурирано учење познату као Марковљева случајна поља, која ће бити предмет истраживања у овој тези. Другачија врста претпоставке заснива се на моделовању сличности у простору излаза помоћу кернел функција. Оваква претпоставка представља основ за класу Структурираних кернел метода. Такође, временска комплексност алгоритама за учење зависи и од типа проблема. Уколико је проблем континуалне природе, најчешће су неопходне додатне претпоставке како би проблем био трактабилан. Најчешће претпоставке које се користе се односе на функцију расподеле променљивих. На пример, код Марковљевих случајних поља, коришћење Гаусовске претпоставке омогућава аналитичке изразе за закључивање, чиме се у општем случају добија алгоритама за учење кубне сложености. У случају много-димензионалних проблема који постају све актуелнији, алгоритми са кубном сложеностју постају недопустиво спори. На пример, учење ретког Марковљевог случајног поља са око десет хиљада променљивих са тренутно актуелним алгоритмима на модерном хардверу траје око сат времена [10]. Поређења ради, поједини проблеми у анализи социјалних мрежа [6], климатских промена [9], активности неурона у мозгу [11], генетских података [8], могу имати неколико редова величине више променљивих, као и много гушћу структуру.

Додатан изазов при структурираном учењу представља величина скупа за обучавање. Са повећањем димензије проблема, драстично се повећава број параметара које је неопходно оптимизовати како би се алгоритам успешно обучио, чиме се уједно повећава и простор претраге параметара. Са повећањем простора претраге, неопходан је већи број примера за обучавање који често нису доступни (нпр. за анализу активности неурона у мозгу са FMRI снимка помоћу 1000 воксела, „наивним“ приступом било би неопходно естимирати милион параметара, користећи преко милион снимака). Услед тога, модели често уносе додатне претпоставке, као што су реткост (тзв. sparsity) структуре, или одређени образац као што је то случај у просторно и темпорално уређеним проблемима. Проналажење адекватних претпоставки је предмет активног истраживања, чија ефикасност може зависити и од проблема на који се примењује.

Постоји више различитих типова алгоритама за структурирано учење међу којима су најзаступљенији: Вероватносни графовски модели (Probabilistic Graphical Models) [15], Неуралне мреже (Neural Networks) [16], Кернел методи (Kernel Methods) [17][18][19], и Стабла одлучивања (Decision Trees) [20]. У овој дисертацији, предмет истраживања су Вероватносни графовски модели, и то Марковљева случајна поља (Markov Random Field) и Условна случајна поља (Conditional Random Field).

Вероватносни графовски модели су класа алгоритама за структурирано учење код којих је проблем представљен помоћу случајних променљивих. Марковљева случајна поља су вероватносни графовски модели у којем су променљиве представљене као чворови, док су интеракције између променљивих представљене као гране неусмереног графа [15]. Одсуство гране у графу одговара условној независности између две променљиве, док присуство гране у графу моделује интеракцију између суседних променљивих. Условна случајна поља су дискриминативна верзија Марковљевих случајних поља, која служи за структурирану предикцију. Основни проблеми код структурираног учења су: закључивање, декодовање, и

учење. Закључивање представља процес одређивања највероватнијих вредности променљивих за дате параметре модела. У општем случају, егзактно закључивање није изводљиво и неопходно је усвајање додатних претпоставки или налажење апроксимативног решења. Декодовање представља проналажење вероватноће за дате вредности променљивих. Учење представља проналажење оптималних вредности параметара за дати модел помоћу датих података уз претпоставку да подаци одговарају моделу. Учење се најчешће обавља проналажењем максималне веродостојности (енг. maximum likelihood), помоћу метода конвексне оптимизације.

Ефикасност учења, као и закључивања на основу наученог модела, у великој мери зависи од особина самог модела и претпоставки над којима је заснован, као и прилагођености модела проблему.

Главни допринос ове тезе биће проналажење и анализа нових модела за структурирано учење и алгорита за њихово обучавање.

- [1] Lafferty, J., McCallum, A. & Pereira, F.C.N., „Conditional random fields: Probabilistic models for segmenting and labeling sequence data.,“ ICML '01 Proceedings of the Eighteenth International Conference on Machine Learning 8, 282-289 (2001).
- [2] Fosler-Lussier, E., He, Y., Jyothi, P. & Prabhavalkar, R., „Conditional random fields in speech, audio, and language processing,“ Proceedings of the IEEE 101, 1054-1075 (2013).
- [3] Li, S.Z., „Markov Random Field Modeling in Image Analysis,“ Pattern Recognition Letters 26, 395-398 (Springer: 2009).
- [4] Wang, C., Komodakis, N. & Paragios, N., „Markov Random Field modeling, inference & learning in computer vision & image understanding: A survey,“ Computer Vision and Image Understanding 117, 1610-1627 (2013).
- [5] Bhole, C. & Pal, C., „Fully Automatic Person Segmentation in Unconstrained Video Using Spatio-temporal Conditional Random Fields,“ Image and Vision Computing (2016)
- [6] Xin, X., King, I., Deng, H. & Lyu, M.R., „A social recommendation framework based on multi-scale continuous conditional random fields,“ Proceedings of the 18th ACM conference on Information and knowledge management 1247-1256 (2009).
- [7] DeCaprio, D. Vinson, J. P., Pearson, M. D., Montgomery, P., Doherty, M., & Galagan, J. E, Conrad, „Gene prediction using conditional random fields,“ Genome Research 17, 1389-1398 (2007).
- [8] Dobra, A., Hans, C., Jones, B., Nevins, J.R., Yao, G., West, M., „Sparse graphical models for exploring gene expression data,“ Journal of Multivariate Analysis 90(1), 196–212 (2004)
- [9] Stojkovic, I., Jelisavcic, V., Milutinovic, V., Obradovic, Z., "Distance Based Modeling of Interactions in Structured Regression," Proc. 25th International Joint Conference on Artificial Intelligence (IJCAI), New York, NY, pp. 2032 – 2038, (2016).
- [10] Hsieh, C., „QUIC : Quadratic Approximation for Sparse Inverse Covariance Estimation,“ Journal of Machine Learning Research 15, 2911-2947 (2014).
- [11] Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., et al., „Advances in functional and structural mr image analysis and implementation as fsl,“ Neuroimage 23, S208–S219 (2004).
- [12] Rolfs, B., Rajaratnam, B., Guillot, D., Wong, I. & Maleki, A., „Iterative Thresholding Algorithm for Sparse Inverse Covariance Estimation,“ Advances in Neural Information Processing Systems 25 1583-1591 (2012).
- [13] Friedman, J., Hastie, T. & Tibshirani, R., „Sparse inverse covariance estimation with the graphical lasso,“ Biostatistics 9, 432-441 (2008).
- [14] Johnson, C.C., Jalali, A. & Ravikumar, P., „High-dimensional Sparse Inverse Covariance Estimation using Greedy Methods,“ 15th International Conference on Artificial Intelligence and Statistics, Proceedings of 574-582 (2012).

3. Полазне хипотезе

Постојећи алгоритми за структурирано учење у општем случају нису примењиви на тзв. big data проблеме који су све присутнији. Највећи проблем структурираном учењу много-димензионалних модела представља превелика временска сложеност алгоритама за њихово учење и закључивање. У општем случају, учење Марковљевих случајних поља са континуалним променљивама је тежак проблем услед недостатка аналитичког решења интеграла који се појављује у изразу за закључивање. Услед тога, модели који се баве структурираним учењем у континуалном домену се најчешће заснивају на претпоставци да променљиве имају Гаусову расподелу, што омогућава егзактне изразе за закључивање. Такође, модели за структурирано учење се заснивају на претпоставци да је структура проблема ретка. Оваква претпоставка се најчешће индукује у моделу помоћу додатног регуларизационог члана заснованог на L1 норми [21]. Такође, поред претпоставке да је структура проблема ректа, код неких проблема често је могуће уочити додатне особине, као што је нпр. scale-free својство [25], које могу учинити проблем учења лакшим или ефикаснијим.

Полазна хипотеза ове дисертације је да је одговарајућом репараметризацијом и увођењем додатних претпоставки могуће направити вероватносни графовски модел и алгоритме за његово обучавање и закључивање који дају боље перформансе (у виду повећане брзине учења и/или смањења грешке предикције) у односу на постојеће.

4. Научне методе истраживања

У склопу овог истраживања биће извршен преглед постојећих методолошких приступа и решења структурираног учења. Нови алгоритам који ће бити осмишљен и чија анализа ће бити предмет ове дисертације ће бити имплементиран и примењен на неколико различитих скупова података из више различитих домена. Један скуп експеримената ће се вршити над вештачки генерисаним (тзв. синтетичким) подацима, који ће за циљ имати емпиријску анализу и верификацију под идеалним претпоставкама. Други скуп експеримената ће се вршити над правим подацима из неколико различитих домена, чији ће циљ бити евалуација ефикасности и примењивости у реалним условима. Имплементација ће бити поређена са постојећим решењима који тренутно представљају тзв. state-of-the-art, тј са постојећим научним доприносима из области, као што су [11] [12] [13] [14]. Биће коришћено неколико различитих широко заступљених метрика за поређење и анализу резултата, као што су метрике за квалитет научене структуре (precision, recall, accuracy, F score), метрике за прецизност предикције наученог модела (RMSE, R-squared), као и брзину и скалабилност учења (процесорско време, укупно време по броју процесорских јединица).

Биће извршена теоријска анализа постављеног модела и анализа алгорита за учење. У склопу теоријске анализе, биће испитана конвергентност алгорита учења, брзина учења, временска и просторна комплексност, као и анализа скалабилности у односу на повећање димензије проблема и повећање броја процесорских јединица коришћених за обучавање.

[15] Koller, D. & Friedman, N., "Probabilistic Graphical Models Principles and Techniques," Foundations 2009, 1231 (2009).

[16] Peng, J., Bo, L. & Xu, J. "Conditional Neural Fields," Advances in Neural Information Processing Systems 9, str 1-9, (2009).

[17] Mauricio A. Álvarez, Lorenzo Rosasco and Neil D. Lawrence, "Kernels for Vector-Valued Functions: A Review", Foundations and Trends® in Machine Learning: Vol. 4: No. 3, pp 195-266.

- [18] Joachims, T., Hofmann, T., Yue, Y. & Yu, C.-N., "Predicting structured objects with support vector machines," *Communications of the ACM* 52, 97 (2009).
- [19] Ping, W., Liu, Q. & Ihler, A., "Marginal Structured SVM with Hidden Variables," *International Conference on Machine Learning (ICML)* 32, 12 (2014).
- [20] Celine Vens, Jan Struyf, Leander Schietgat, Sašo Džeroski, and Hendrik Blockeel, "Decision trees for hierarchical multi-label classification," *Machine Learning*, vol 73, str 185-214, 2008.
- [21] Bach, F., Jenatton, R., Mairal, J., Obozinski, G., et al., "Convex optimization with sparsity-inducing norms," *Optimization for Machine Learning* 5 (2011).
- [22] Banerjee, O., El Ghaoui, L., d'Aspremont, A., "Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data," *The Journal of Machine Learning Research* 9, 485–516 (2008).
- [23] Banerjee, O., Ghaoui, L.E., d'Aspremont, A., Natsoulis, G., "Convex optimization techniques for fitting sparse gaussian graphical models," In *Proceedings of the 23rd international conference on Machine learning*, pp. 89–96. ACM (2006).
- [24] Jankova, J. & Geer, S. van de, "Confidence intervals for high-dimensional inverse covariance estimation," *Electronic Journal of Statistics* 9, 17 (2015).
- [25] Barabási, Albert-László. "Scale-free networks: a decade and beyond," *Science* 325.593, pp. 412-413 (1999).

5. Очекивани научни допринос

Главни очекивани научни допринос ове тезе су нови алгоритми за учење Марковљевих случајних поља, са побољшаним карактеристикама у односу на постојеће. Нека од могућих очекиваних побољшања су брзина учења, већа скалабилност, боље моделовање података насталих из дистрибуција са тежим реповима, као и побољшања у структурираној предикцији. Као резултат ове тезе, очекивани научни доприноси су:

- Преглед, систематизација, и анализа постојећих алгоритама за структурирано учење
- Прототип вероватносног графовског модела и алгорита за његово обучавање
- Теоријска и експериментална евалуација предложених вероватносних графовских модела и алгоритама за њихово обучавање, као и упоредна анализа са постојећим решењима
- Имплементација предложеног алгорита на савременим софтверским платформама за обраду великих података
- Анализа примењивости имплементираног алгорита и евалуација на реалном случају коришћења

6. План истраживања и структура рада

Планирано је да се истраживање спроведе у четири фазе:

- Прва фаза обухвата анализу постојећих ретких неусмерених графовских модела и алгоритама за њихово учење. У оквиру ове фазе биће направљен пресек тренутног стања у области структурираног учења, са посебним освртом на вероватносне графовске моделе.
- Друга фаза предвиђа дефинисање и детаљна анализа хипотеза на основу којих се ће бити направљени нови модели и алгоритми за њихово учење. У оквиру ове фазе биће

анализирано неколико проблема структурираног учења, на основу којих ће бити постављен одговарајући теоријски модел.

- Трећа фаза предвиђа имплементацију алгоритама за учење базираних на анализи претпоставки постављених у фази два. У оквиру ове фазе биће најпре имплементиран прототип у неком од језика погодних за матричну обраду података (Octave/Matlab или Python), затим ће бити реализован софтверски пакет заснован на савременим платформама за обраду великих података (нпр. MapReduce, Apache Spark, Apache Flink, ...).
- Четврта фаза предвиђа анализу добијених резултата и доношење закључка. У оквиру ове фазе биће извршено поређење перформанси алгоритама проистеклих из ове тезе са постојећим алгоритмима за структурирано учење.

7. Закључак и предлог

На основу свега изложеног, Комисија констатује да кандидат Владисав Јелисавчић испуњава потребне услове да може приступити изради докторске дисертације, под измењеним радним насловом **"Структурирано учење над великим подацима засновано на вероватносним графовским моделима"**.

Ужа научна област предложене теме је рачунарска техника и информатика. За ментора се предлаже проф. др Вељко Милутиновић, редовни професор у пензији Електротехничког факултета Универзитета у Београду.

Стога комисија предлаже Наставно-научном већу Електротехничког факултета у Београду да одобри Владисаву Јелисавчићу израду докторске дисертације под горе наведеним насловом под руководством ментора проф. др Вељка Милутиновића.

У Београду, 20.10.2016. године

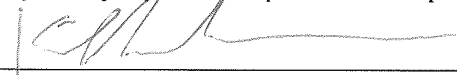
ЧЛАНОВИ КОМИСИЈЕ



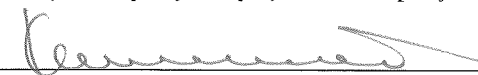
др Вељко Милутиновић, редовни професор у пензији
Универзитет у Београду – Електротехнички факултет



др Јелица Протић, ванредни професор
Универзитет у Београду – Електротехнички факултет



др Милош Ковачевић, ванредни професор
Универзитет у Београду – Грађевински факултет



др Горан Квашчев, доцент
Универзитет у Београду – Електротехнички факултет