

Undergraduate Thesis Defense

Written by Marinka

Saturday, 21 July 2012 19:05 - Last Updated Sunday, 30 March 2014 16:36

Yesterday I defended my undergraduate Thesis entitled **A Matrix Factorization Approach for Inference of Prediction Models from Heterogeneous Data Sources** (slo: Pristop matrične faktorizacije za gradnjo napovednih modelov iz heterogenih podatkovnih virov) at [University of Ljubljana](http://www.uni-lj.si/en/about_university_of_ljubljana.aspx), [Faculty of Computer and Information Science](http://www.fri.uni-lj.si/en/) and [Faculty of Mathematics and Physics](http://www.fmf.uni-lj.si/en/).

With that Thesis I completed a four year interdisciplinary university program consisting of eight semesters of lectures and one year of Diploma thesis work in less than four years. Diploma leads to the degree of University dipl.ing. of Computer Science and Mathematics. I graduated summa cum laude with the average grade being 10.0 out of 10.0.

My brother Slavko took some pictures and shot the defense. The movie and pictures are available at [zitnik.si](http://zitnik.si/wordpress/2012/07/20/marinka-zitnik-sestra-zagovor-diplomske-naloge/) site.

Abstract

Today we are witnessing rapid growth of data both in quantity and variety in all areas of human endeavour. Integrative treatment of these sources of information is a major challenge. We propose a new computation framework for inference of prediction models based on symmetric penalized matrix tri-factorization and intermediate strategy for data integration. Major advantages of the approach are an elegant mathematical formulation of the problem, an integration of any kind of data that can be expressed in matrix form, and high predictive accuracy.

We tested the effectiveness of the proposed framework on predicting gene annotations of social amoebae *D. dictyostelium*. The developed model integrates gene expressions, protein-protein interactions and known gene annotations. The model achieves higher accuracy than standard techniques of early and late integration, which combine inputs and predictions, respectively, and have in the past been favourably reported for their accuracy.

With the proposed approach we have also predicted that there is a set of genes of *D. dictyostelium* that may have a role in bacterial resistance and which were previously not associated with this function. Until now, only a handful of genes were known to participate in related bacterial recognition pathways. Expanding the list of such genes is crucial in the studies of mechanisms for bacterial resistance and can contribute to the research in development of alternative antibacterial therapy. Our predictions were experimentally confirmed in wet-lab experiments at the collaborating institution (Baylor College of Medicine, Houston, USA).

Note: Complete Thesis will be available for download after some aspects of the proposed approach will be presented in the scientific journal.

I am looking forward to starting my PhD studies in the autumn. Until then ... many new interesting projects ...