Notice of the Final Oral Examination
for the Degree of Master of Science

of

LAN ZHENG

MSc (Beijing University of Posts and Telecommunications, 2012)

“Performance Evaluation of Latent Factor Models for Rating Prediction”

Department of Computer Science

Thursday April 16, 2015
2:00 P.M.
Engineering Computer Science Building
468

Supervisory Committee:
Dr. Kui Wu, Department of Computer Science, University of Victoria (Co-Supervisor)
Dr. Alex Thomo, Department of Computer Science, UVic (Co-Supervisor)

External Examiner:
Dr. Venkatesh Srinivasan, Department of Computer Science, UVic

Chair of Oral Examination:
Dr. Ashoka Bhat, Department of Engineering, UVic

Dr. David Capson, Dean, Faculty of Graduate Studies
Abstract

Since the Netflix Prize competition, latent factor models (LFMs) have become the comparison “staples” for many of the recent recommender methods. Meanwhile, it is still unclear to understand the impact of data preprocessing and updating algorithms on LFMs. The performance improvement of LFMs over baseline approaches, however, hovers at only low percentage numbers. Therefore, it is time for a better understanding of their real power beyond the overall root mean square error (RMSE), which as it happens, lies at a very compressed range, without providing too much chance for deeper insight. We introduce an experiment based handbook of LFMs and reveal data preprocessing and updating algorithms’ power. We perform a detailed experimental study regarding the performance of classical staple LFMs on a classical dataset, Movielens 1M, that sheds light on a much more pronounced excellence of LFMs for particular categories of users and items, for RMSE and other measures. In particular, LFMs exhibit surprising and excellent advantages when handling several difficult user and item categories. By comparing the distributions of test ratings and predicted ratings, we show that the performance of LFMs is influenced by rating distribution. We then propose a method to estimate the performance of LFMs for a given rating dataset. Also, we provide a very simple, open-source library that implements staple LFMs achieving a similar performance as some very recent (2013) developments in LFMs, and at the same time being more transparent than some other libraries in wide use.