Program: KDD Cup and Workshop 2007

August 12, 2007, San Jose, California

8:45am – 8:50am Welcome

8:50am – 9:00am ***Netflix Prize Session* I** (Session Chair, Charles Elkan)

8:50am – 9:00am The Netflix Prize
Jim Bennett, Netflix

9:00am – 10:00am Invited Talk I: Restricted Boltzmann Machines for Collaborative Filtering
Andriy Mnih, University of Toronto

10:00am – 10:30am Coffee Break and Poster Set-Up

10:30am – 12:00pm: ***Netflix Prize Session* II** (Session Chair, Padhraic Smyth)

10:30am – 11:00am Short Presentations

Methods for Large Scale SVD with Missing Values
Miklós Kurucz, András A. Benczúr, and Károly Csalogány, Hungarian Academy of Sciences

Collaborative Filtering via Ensembles of Matrix Factorizations
Mingrui Wu, Max Planck Institute for Biological Cybernetics

Improving Regularized Singular Value Decomposition for Collaborative Filtering
Arkadiusz Paterek, Warsaw University

11:00am – 11:30am Variational Bayesian Approach to Movie Rating Prediction
Yew Jin Lim, National University of Singapore
Yee Whye Teh, University College London

11:30am – 12:00pm On the Gravity Recommendation System
Gábor Takács, István Pilászy, Bottyán Nemeth, and Domonkos Tikk, Budapest University of Technology and Economics

12:00 – 1:30pm Lunch

1:30pm – 3:00pm ***Combined Session*** (Session Chairs, Jim Bennett and Bing Liu)

1:30pm – 2:00pm: Invited Talk II: Improving Collaborative Filtering by Learning a Feature Relevance Prior
Su-In Lee, Stanford University.

2:00pm – 2:35pm Improved Neighborhood-Based Collaborative Filtering
Robert M. Bell and Yehuda Koren, AT&T Labs – Research

2:35pm – 2:45pm The KDD cup competition tasks, Bing Liu

2:45pm – 3:05pm KDD Cup Short Presentations

First runner-up of task 1: A Classical Predictive Modeling Approach for Task "Who Rated What?" of the KDD Cup 2007
Jorge Sueiras, Alfonso Salafranca, and Jose Luis Florez, Neometric, Neo Metrics

Second runner-up of task 1: Predicting Who Rated What in Large-Scale Datasets
Yan Liu and Zhenzhen Kou, IBM T. J. Watson Research Center

First runner-up of task 2: A Combination of Approaches to Solve Task "How Many Ratings?" of the KDD Cup 2007
Jorge Sueiras, Daniel Vélez, and José Luis Flórez, Neo Metrics

K-split Based Approach to Predict Movie Rating Frequency
Hariprasad Bommaganti, Yahoo Inc.
Anand Nagarajan, Symbram LLC

3:05pm – 4:00pm Coffee Break and Poster Session

4:00pm – 5:00pm ***KDD Cup Winners Session*** (Chair, Domonkos Tikk)

4:00pm – 4:30pm *First place winner of task* 1: Who Rated What: a Combination of SVD, Correlation and Frequent Sequence Mining
Miklós Kurucz, András A. Benczúr, Tamás Kiss, István Nagy, Adrienn Szabó, and Balázs Torma, Hungarian Academy of Sciences

4:30pm – 5:00pm *First place winner of task* 2: Making the Most of Your Data: KDD Cup 2007 "How Many Ratings" Winner's Report
Saharon Rosset, Claudia Perlich, and Yan Liu, IBM T. J. Watson Research Center

Closing remarks

Abstracts of Invited Talks

Invited talk 1: *Restricted Boltzmann Machines for Collaborative Filtering,* Andriy Mnih, University of Toronto

Abstract: Restricted Boltzmann Machines (RBMs) are a class of two-layer undirected probabilistic graphical models for modelling distributions of high-dimensional datavectors. RBMs are well-suited for modelling large datasets because approximate learning and exact inference can be performed efficiently in these models. I will introduce RBMs, show how they can be used for collaborative filtering, and describe our application of RBMs to the Netflix Prize dataset. I will also outline an approach to making RBMs more powerful by stacking them to obtain multilayer models.

Joint work with Ruslan Salakhutdinov and Geoffrey Hinton.

Invited talk 2: *Learning a Relevance Prior for Collaborative Filtering*, Su-In Lee, Stanford University

Abstract: A collaborative filtering task can be viewed as an ensemble of related prediction tasks, where each task is to predict the user preference of a particular product given the user's preference levels for other products. In this study, we propose a transfer learning scheme – learning on multiple related tasks with shared information – to improve the prediction performance. Assuming that each task is to predict the ratings on a particular movie ('task' movie) using the ratings on other movies as features ('feature' movies), all tasks are 'related' in the sense that they can share the underlying structure regarding which feature movie's ratings give information on which task movie's ratings. The proposed algorithm is based on an informed strategy for feature selection. In this problem, as in many others, we are faced with a huge number of features; therefore, selecting relevant features is essential for achieving good generalization performance. Most feature selection algorithms consider all features to be a priori equally likely to be relevant. In this study, we use transfer learning to construct an informative prior on feature relevance. We assume that features themselves have meta-features that are predictive of their relevance to the prediction task, and model their relevance as a function of the meta-features using hyperparameters (called meta-priors) that are estimated through learning. In our movie rating prediction example, the meta-features of a combination of a task and a feature movie can include whether one is sequel to the other or whether they share a genre, actresses/actors, directors, key words, and so on. For example, the feature movie Star Wars IV is more likely to be relevant to the task movie Star Wars I. We present a convex optimization algorithm for simultaneously learning the meta-priors and feature weights from an ensemble of related prediction tasks that share a similar relevance structure. Our approach transfers the meta-priors among different tasks, allowing it to deal with situations where feature relevance varies over the tasks, a setting that most transfer learning algorithms cannot handle. In our experiment on a subset of the Netflix dataset, we show that our approach improves the prediction performance and our algorithm learns what meta-features are important to determine the prior relevance of a feature movie to a task movie. Our algorithm also applies to a broad range of other problems with multiple related prediction tasks.