April 2, 2012

Re: Jessie W. H. Zou Memorial Award

Dear Award Selection Committee:

My name is Jennifer Nguyen and I am a soon-to-be graduate of the Honours Statistics program here at the University of Waterloo. Over the past year, I have been working on a recommender system project supervised by Professor Mu Zhu of the Department of Statistics and Actuarial Science.

As of late, many businesses and services have integrated the use of recommender systems in their marketing strategies to maintain customer satisfaction and loyalty. Recommender systems are an efficient statistical marketing tool because not only do they help customers find new items, they find new items that suit the customer's taste. For example, the online store *Amazon.ca* recommends to shoppers new products that they may like based on their past purchases. This personal aspect is important because customers are more receptive to marketing that is catered specifically to them.

A popular recommender system approach is collaborative filtering (CF). Users are recommended items that other users with similar preferences have liked in the past. One way to perform CF is to use matrix factorization, where the user-item ratings matrix is factored into two latent matrices. These two matrices are used to predict users' unknown ratings in the original ratings matrix.

While CF has its strengths, a disadvantage of the CF matrix factorization approach is that it suffers from the "cold start" problem. This problem arises when a new user enters the system and has not rated many items. Consequently, it is difficult for the system to recommend items to these users because the system is unsure of the user's preference. Similarly, new items that have not been rated by many users are also difficult to recommend.

Recently, many works have tried to mitigate the cold start problem by introducing "content boosting" methods to augment the ratings matrix with supplemental information. In our work, we incorporate information about the items directly into the matrix factorization algorithm. Intuitively, if two items share a number of common attributes, their feature vectors should be close together in the latent feature space. Hence, we place an additional constraint to force those item feature vectors to be more aligned with one another.

Using this "alignment-biased" algorithm, we are able to demonstrate through repeated trials that not only is recommendation accuracy improved, but the recommendations are also more interpretable. This ability for a recommender system to explain its recommendation is a highly desirable trait. It is because consumers are more willing to trust its recommendations if there is supporting justification.

With the vast number of items available to consumers in today's world, recommender systems are an important statistical tool to help consumers focus their attention. The results of our work have been submitted to the Statistical Analysis and Data Mining Journal for review. I would be grateful to receive this award and hope that the Award Committee considers me a deserving candidate.

Sincerely,

Jennifer Nguyen