R E S E A R C H S T A T E M E N T

Rong Zheng

Stern School of Business, New York University

1. Overview of my research

My research interests lie in the area of developing data mining methods for improving personalized solutions to business intelligence problems. This is increasingly important research area because, as Internet and information technologies have become a central platform for people's daily activities, very large amounts of behavioral data have been generated, and this makes it possible to understand, model and target users at the individual and micro-segment levels. My research has intended to develop better computational methods in the context of target marketing and recommender systems, as well as analyzing writing styles in online settings. The details of my research are described below.

2. Details of my research

1) Improving target marketing via optimal customer segmentation. (with Prof. Alex Tuzhilin)

I am interested in the question of how to segment a customer base in order to achieve best possible performance in the predictions of customer behavior. This is an important research problem because it is directly related to developing better solutions for target marketing. In our current study [9], we formulate the optimal customer segmentation problem as a combinatorial optimization problem and show that it is NP-hard. We also develop different sub-optimal but computationally tractable methods. Our methods group customers into segments in an overlapping fashion, allowing a customer to belong to multiple segments. In our empirical studies we show that our approach dominates previously proposed non-overlapping methods. In addition, our best-performing method achieves better performance as compared to the exchange method, which is used in clusterwise regression analysis [2].

2) Query language for multi-dimensional recommender system. (with Prof. Gedas Adomavicius and Prof. Alex Tuzhilin)

Despite the increasing importance of recommendation technologies popularized by Amazon and Netflix, the current recommender systems have two limitations: (a) passive role of users when recommendations are generated and (b) the restricted nature of the traditional user-item paradigm. My research addresses these two limitations. In particular, we develop a comprehensive query language *REQUEST* that allows users to customize recommendations

by formulating them in the ways satisfying personalized needs of the users [1]. In addition, *REQUEST* is designed particularly for handling multi-dimensional recommendations, thus treating user-item two dimensional applications as a special case. We also defined a set of recommendation algebraic operators based on an OLAP algebra and provide an algorithm that maps *REQUEST* queries into a sequence of algebraic operations. In the current e-commerce applications, implementing *REQUEST* could increase user's satisfaction by providing a way to get users directly involved into the recommender system as active participants. This leads to a series of questions, such as optimizing predictions for multi-dimensional recommendation requests, and modeling users based on their requests, which will be a part of my future research on this subject.

3) Leveraging social networks to improve collaborative filtering. (with Prof. Foster Provost and Dennis Wilkinson)

The recent rise of online social networking provides a new source of data for recommender systems. Collaborative filtering (CF), as the most popular recommendation generating technique, is particularly relevant because it bases its predictions on past behavior of like-minded people. I am currently interested in evaluating the predictive power of user selected social neighbors, as compared to most like-minded people in typical recommender systems [4, 8]. Our results showed that the prediction based on social networks neighbors can rival CF in prediction accuracy. These results are significant for three reasons. First, social network based CF is significantly more computationally efficient than traditional CF and may be implemented in large-scale web commerce and social networking communities. Second, the strong predictive power of social neighbors may indicate social selection and social influence effect that explain evolution of social networks. Third, the result indirectly suggests a non-linear relationship between like-mindness and predictive performance [11], which may improve the quality of recommendations.

4) Statistical analysis of writing styles in online settings. (with Prof. Jason Li)

Writing style has been considered as one of intrinsic properties of people. Internet applications have created numerous sources of data, such as blogs, Web pages, emails, reviews and other user-generated content sources, from which writing styles can be extracted. I am interested in such questions as: (a) Can we automatically identify authors based on the analysis of writing styles? (b) How to measure and quantify writing styles? (c) Does writing style play a role in people's assessments in online reviews, and blogs? In our previous studies, we have shown that writing style features have strong discriminating power in author identification in various online settings [5, 6, 7]. In a recent study [3], we show the high correlation between helpfulness of online reviews and writing style features. I currently study the effects of writing styles in determining the popularity of web blogs [10].

3. Summary and future research plan

In summary, my research focuses on building personalized behavioral models of consumers, such as purchasing, rating and writing in e-Commerce applications. The results of my studies contribute to an improved understanding about consumers and consequently a better alignment between consumers and service providers. My research methodology lies at the intersection of information system, computer science, marketing science, social science, statistics and computational linguistics. This helps me to approach my research goals from multiple research perspectives.

In addition to advancing my current research projects, I am also interest in applying data mining techniques to other marketing research problems, where I focus on building prediction-oriented models. I foresee many exciting and high-impact research opportunities in this interdisciplinary area. Also, I would like to extend my work to the studies of economics of IS. I am interested in applying econometric methodologies when it is appropriate. I have been helping Prof. Anindya Ghose to collect and process data in his various high impact studies. I recently started a project with Prof. Ghose on the economical value of Wishlist in e-commerce applications.

4. References

Published papers

1. Adomavicius, Gedas, Tuzhilin, Alex and Zheng, Rong. REQUEST: A Query Language for Customizing Recommendations. Conditional accepted by Information System Research.

2. Zheng, Rong, and Tuzhilin, Alex. Partitioning Customers Using Overlapping Segmentation Methods. To appear in Proceedings of the 18th Workshop on Information Technologies and Systems (WITS2008), December 2008.

3. Li, Jiexun, and Zheng, Rong. Stylometric Feature Selection for Assessing Review Helpfulness. To appear in Proceedings of the 18th Workshop on Information Technologies and Systems (WITS2008), December 2008.

4. Zheng, Rong, Provost, Foster, and Ghose, Anindya. Social Network Collaborative Filtering: Preliminary Results. Proceedings of the Sixth Workshop on eBusiness (WEB2007), December 2007.

5. Zheng, Rong, Li, Jiexun, Huang, Zan and Chen, Hsinchun. A Framework for Authorship Identification of Online Messages: Writing Style Features and Classification Techniques. Journal of the American Society for Information Science and Technology (JASIST), 57(3):378-393 (2006)

6. Li, Jiexun, Zheng, Rong and Chen, Hsinchun. From Fingerprint to Writeprint: Feature Selection for Authorship Identification. Communication of ACM 49(4): 76-82 (2006)

7. Zheng, Rong, Qin, Yi, Huang, Zan, and Chen, Hsinchun. Authorship Analysis in Cybercrime Investigation. Proceedings of the first IEEE International Conference on Intelligence and Security Informatics, (ISI2003), Tucson, AZ. Published in Lecture Notes in Computer Science (LNCS).

Working papers

8. Zheng, Rong, Wilkinson, Dennis and Provost, Foster. Social Network Collaborative Filtering. Under review at ACM Transactions on the WEB (TWEB).

9. Zheng, Rong and Tuzhilin, Alex. Improving Target Marketing via Overlapping Customer Segmentation. To be submitted to IEEE Transactions on Knowledge and Data Engineering (TKDE).

10. Zheng, Rong and Li, Jiexun. Stylometric Analysis for Popular Web blogs. In preparation for KDD2009.

11. Non-Linear Weighting in Collaborative Filtering. With Foster Provost.