Addressing Challenges in Inferring Preferences from Current-Session Data

Jeffrey Parsons
Faculty of Business Administration
Memorial University of Newfoundland
jeffreyp@mun.ca

Paul Ralph
Sauder School of Business
The University of British Columbia
paulralph@gmail.com

Introduction

Generating effective online recommendations is an important and challenging issue in electronic commerce. The benefits of improving recommender accuracy by even a small amount are substantial. For example, the recent announcement by NetFlix of a $1 million prize for a system that can improve on the accuracy of their current recommender by 10% points clearly to the commercial importance and value of effective recommendations (http://www.netflixprize.com).

Research on the dominant approaches to recommendation focuses on making more effective use of large data sets. For example, collaborative filtering research seeks to improve algorithms that compare user characteristics or behavior to other users and recommending items that similar users have preferred or purchased. Advances in recommendation quality in this area are incremental (Herlocker et al., 2004). In addition, such approaches tend to be obtrusive, requiring considerable effort from users to be effective (e.g., having a user rate many items to generate effective recommendations) and raising privacy and security concerns. Moreover, they do not accommodate additional, different preference-related data that might be relevant in a given situation.

In light of such issues, we have developed DESIRE, a recommender system that requires only a relatively small amount of single-session data to make recommendations that reflect inferred preferences in a specific session or context (Parsons et al., 2004). DESIRE uses ***viewing time*** as an indicator of preference toward an item being viewed. The system decomposes items by attributes, and infers preferences toward values (numeric or categorical) based on time spent viewing a set of items having a reasonable distribution of attribute values. Recommendations are made by calculating a weighted desirability measure for each item in a repository based on the inferred preferences for values of its attributes.

The system has been tested in the laboratory with promising results. During testing, we identified two challenges that merit attention in doing research on recommenders using session data and profile-based recommendations. First, because such systems are primarily intended to infer current preferences from current data, they must function with the very limited data that might be provided during a session (e.g., inferring preferences from a small sample of viewed items). In this context, it is critically important to embed prior knowledge of the relative importance of attributes to a specific preference context. Moreover, the relative importance of attributes is likely to be product and situation specific. Second, in an online setting, preference is only one of many factors that affect viewing time for an item or page. Thus, to make effective recommendations with limited data, it is necessary to statistically control for known effects of other factors on viewing time and isolate the viewing time – preference relationship. In this research, we present initial approaches to addressing each of these two challenges.

Relative Importance of Attributes

DESIRE's recommendation strategy estimates a user's preferences for items based on the items' attributes. This requires several parameters including the relative importance of each attribute (i.e., the degree to which each attribute influences overall preference). For example, when buying a car, fuel efficiency might be more important than trunk space. Collecting or estimating these relative importance weights is difficult for several reasons. First, the relative importance of attributes to preference varies with the type of item (e.g., a camera vs. a lamp), context (e.g., purchase vs. rental) and individual. Second, in an e-commerce setting it may be impractical to obtain personalized information about a user's relative weighting of attribute importance, although such information could be collected if user profiles are used. Thus, aggregate data from a sample or population might be needed. Third, simply asking a representative sample of users to indicate their relative importance weights for the attributes of some item class is unlikely to work because people are generally unable to express the degree to which different beliefs influence their attitudes (Fishbein & Ajzen, 1975). This leaves two valid approaches to generating estimates of the relative importance of attributes. We have had some success with both.

The first is to run the recommendation algorithm in reverse. Normally, the recommendation algorithm computes estimated preference ratings given information about items, attributes and the relative importance of attributes. Instead, suppose that preference ratings are available and relative importance weights are not. This could occur in historical data from an existing e-catalog where purchase history or explicit ratings could provide preference ratings. Given this information, a systematic search for the optimal relative weights is possible via repeatedly running the recommendation algorithm in reverse. Search techniques such as hill-climbing or genetic algorithms can be used. The key limitation of this approach is that time complexity of the solution space search may force the research to settle for a less-than-optimal solution.

The second is to survey users and compute implicit relative weights using a multivariate regression. More specifically, participants can be given a number of items with attribute data and asked to indicate their overall preference for the item on a fine-grained (e.g., 9 or 11 point) scale. Given sufficient ratings and attribute data, the researcher can then run a stepwise, multivariate, ordinal, logistic regression with preference as the dependent variable and the attribute data as independent variables. The resultant coefficients of the regression equation give the relative importance weights. The limitation of this approach lies in its reliance on self-reported preferences, which may not reflect actual preferences as they can suffer from order-effects and other psychological biases.

Isolating the Impact of Viewing Time on Preference

The time a person spends viewing an object is influenced by a variety of factors, including preference for the item, complexity of the item, etc. (see Figure 1 – please note: the dashed arrows indicate other strong correlations). Current understanding of viewing time is based on old psychological studies (see Heinrich 1970 for a summary). As it was not clear whether these findings were generalizable to an e-commerce context up to four decades later, we attempted a replication. Approximately 200 participants were asked to rate 25 items on each of the dimensions hypothesized to affect viewing time. We used stepwise regression to determine the relative impact of each dimension. Because our sample size was large, even very small effects were significant at the $p<.001$ level.

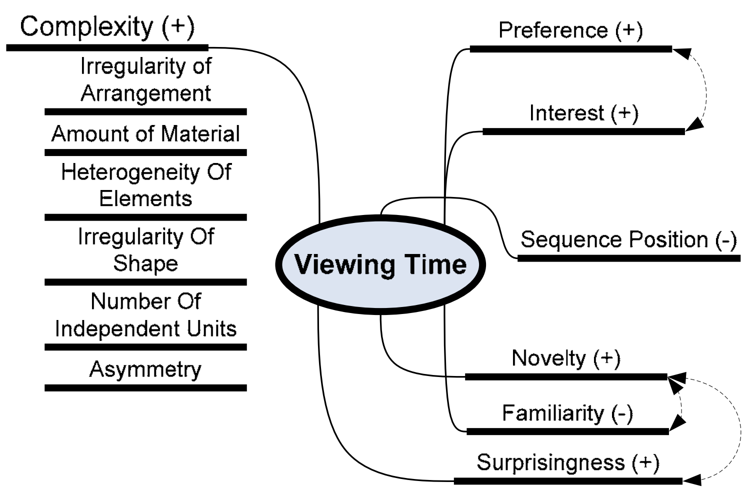

Figure 1: Factors Influencing Viewing Time

We found that most of the hypothesized relationships held in the new context; many factors other than preference affect viewing time. The fact that preference is one of many factors that affects viewing time makes estimating preferences from viewing times a difficult, unsolved problem.

Conclusion

At the conference, we will present approaches taken to the empirical study of the two issues highlighted above, and discuss progress made and remaining challenges.

References

Fishbein, M. & Ajzen, I. (1975), *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research*. Reading, MA: Addison-Wesley.
Heinrich, P. 1970. Free looking time: A method for determining preference. *Psychologie und praxis*, 14 (2), 79-93.
Herlocker, J., Konstan, J., Terveen, L & Riedl, J. (2004) Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, 22 (1): 5-53.
Parsons, J., Ralph, P. & Gallagher K. (2004). Using viewing time to infer user preference in recommender systems. *AAAI Workshop in Semantic Web Personalization*, San Jose, California, July.