Learning Profiles Quickly from User Interactions

Pelin Atahan and Sumit Sarkar
School of Management, The University of Texas at Dallas
pxa041000@utdallas.edu, sumit@utdallas.edu

Research question

Personalization systems play a fundamental role in online business settings. They increase customer retention and loyalty by providing enhanced personalized services, and consequently improve the profitability of firms by deploying targeted marketing strategies and exploiting cross-sell and up-sell opportunities. The performance of these systems depends on the comprehensiveness, accuracy and the reliability of the user profile available to the system, and the matching algorithms that use the profile to identify the best products and services to offer a customer. While a substantial amount of research has been conducted to address the matching aspect, relatively little work has been conducted to determine the appropriate profiles for customers to improve the performance of such systems.

The knowledge needed for the user profile can be implicitly or explicitly acquired from users. Explicit knowledge refers to characteristics or attributes of the user whose values are directly provided by the user herself, whereas implicit knowledge is elicited from the behavior of the user as manifested from the user interactions with a site. Navigation history, purchase history, and time spent on specific web page are some of the sources of implicit knowledge. Acquiring explicit knowledge may not always be possible or desirable as it requires user effort. In addition, many users are reluctant to provide personal information due to privacy concerns. As a result, the majority of recent research on personalization techniques focuses on acquiring implicit knowledge and combining it with available explicit feedback, if any.

This work contributes to the growing body of literature on learning user profiles. Billsus and Pazzani (2000) present a framework for user modeling for adaptive news access that models short term and long term interest profiles. Widyantoro et al. (2001) use a three descriptor representation to learn user interest dynamics. Middleton et al. (2004) introduce an ontological approach to learning interest profiles in order to integrate domain knowledge. The interest profiles consist of topics derived from the research topic ontology along with scores learnt from the documents viewed or rated by the user. These approaches typically use a feature vector representation of user profiles, derived from pages viewed or products rated.

The techniques presented in the existing literature for learning user profiles typically require a long time to build accurate and reliable profiles. This is because these techniques use a passive approach to learning such profiles. They do not consider if some specific user characteristics (profile attributes) may be important for recommendation purposes, and consequently how to expedite the learning process. Of course, the faster a

firm learns the important profile attributes, the sooner it can start creating value to the users and thereby improve customer satisfaction and eventually its profits.

We are studying how user profiles can be learnt as quickly as possible based on a customer's interaction with a firm's site. The profile learning approach we are considering is general and not limited to learning specific attributes.

Approach

When a customer begins interacting with a site, very little is usually known about the customer's attributes that make up the profile. Over time, as the customer continues to interact (e.g., by clicking on links offered, purchasing products, etc.) the site will be better able to determine the values for the profile attributes. Given the general uncertainty associated with the profile of a customer, we consider a probabilistic framework to represent and update the profile. The probabilistic framework fits well with the problem addressed in this study, as it enables a formal approach to revise the beliefs regarding a customer's profile.

We approach the attribute value determination problem as a classification problem. We propose a Bayesian mechanism to determine the likelihood of an individual belonging to a certain class. For illustration purposes, consider a situation where knowing the gender of a visitor could help a firm target recommendations more meaningfully. The site would then benefit from quickly learning the gender of its visitors. As the user navigates through the website, the site can update its beliefs regarding the user's likelihood of belonging to a certain class, male or female in this case, based on the links clicked on by the user and the statistics available for the selected links.

Each link clicked on by the user provides information regarding the user's likelihood of belonging to a certain class. Given the ability to learn a profile from a visitor's clickstream, the next issue to consider is how to determine what links/products the firm should offer so that it can learn the profile quickly. To determine the optimal set of links to provide, we will examine how the firm can determine the informative value of providing an individual link. We will then extend this approach to determine the expected information value when multiple links are provided. Finally, we will develop a methodology to determine the optimal set of links to provide at each stage given the (uncertain) knowledge of a users profile at that stage.

Main findings/expected contributions

The primary objective of this research is to provide a robust and rigorous methodology to learn user profiles as quickly as possible using implicit feedback. The methodology developed will allow a site to rapidly learn profiles of customers, and tailor its services accordingly. The probabilistic framework we are considering will have minimal data requirements making it easy to implement in practice.

Current status of manuscript

The Bayesian update mechanism is mostly developed. We are developing the techniques for determining the informative value of links, and will thereafter study ways to determine the optimal link set.

References

1. Billsus, D. and M. J. Pazzani. 2000. User Modeling for Adaptive News Access. *User Modeling and User-Adapted Interaction* 10(2-3), pp. 147-180.

2. Middleton, S. E., N. R. Shadbolt, and D. C. De Roure. 2004. Ontological User Profiling in Recommender Systems. *ACM Transactions on Information Systems,* Vol. 22, No. 1, pp. 54–88.

3. Widyantoro, D. H., T. R. Ioerger, and J. Yen. 2001. Learning User Interest Dynamics with a Three-Descriptor Representation. *Journal of the American Society of Information Science and Technology (JASIST),* Vol. 52, No. 3, pp. 212-225.