Integrating User Behavior and Collaborative Methods in Recommender Systems

Position Paper

Alexander Tuzhilin
IS Department, Stern School of Business
New York University
atuzhili@stern.nyu.edu

Gediminas Adomavicius
Computer Science Department
New York University
adomavic@cs.nyu.edu

For recommender systems to be successful, they need to achieve a certain level of accuracy in their recommendations that is acceptable to the users. In order to achieve higher levels of accuracy, several researchers advocated the integration of the collaborative and the content-based filtering approaches [Balabanovic & Shoham 1997, Konstan et al. 1998, Pazzani 1999]. In fact, Pazzani [1999] shows that the system that combines the two approaches achieves 71% accuracy of predictions vs. 61% for a pure content-based and vs. 57% to 69% for a pure collaborative approach (depending on parameters used in the collaborative filtering approach).

From the marketing perspective, in order to make accurate recommendations to the user, it is important to understand the behavior of that person. Between the collaborative and content-based approaches, only the latter tries to model on-line behavior of users. However, in this paper we argue that the content-based approach provides only a limited understanding of such behavior and that this approach needs to be extended to capture full on-line behavior of the users based on their transactional histories. More specifically, in the content-based approach the on-line behavior of the users is used for recommending to them "items similar to those a given user has liked in the past" [Balabanovic & Shoham 1997]. However, such description of "behavior" is limited in the following sense. First, this definition focuses on a particular type of behavior where the recommended item is determined exclusively in terms of the attributes (keywords) of the set of the previously considered items and does not take into account other factors determining behavior of the user, such as time. Second, the term "similar" means keyword similarity between items based on information retrieval methods [Salton 1989].

This limited view of behavior prohibits recommender systems from making different types of recommendations depending on specific contexts in which these recommendations are made. For example, consider a system that recommends movie rentals to its customers. Assume that the user prefers to watch action movies on weekdays and European classics on weekends. Most recommender systems would not differentiate between these two situations and would recommend the same type of rental regardless of the day of the week. Similarly, assume that the user tends to rent only one movie at a time on weekdays and that it is Wednesday. Also assume that she has just rented a movie. Then the content-based recommender systems would not know whether to bother the user with any additional recommendations or not.

> *Position Statement 1:* In order to provide more accurate recommendations, it is necessary to base them on a thorough analysis of the *on-line behavior of the user* that is much broader than the behavior captured by current content-based filtering systems.

Capturing On-Line Behavior of Users

The on-line behavior of a user can be learned from the transactional history of that user using various data mining methods, such as the ones described in [Fayyad et al. 1996], and such behavior can be included as a part of that user's profile. In order to describe the structure of these profiles, consider an example of a movie rental database with the following schema:

RENTAL(CustID, MovieID, DateRented, DateReturned, DiscountType, Rating)
CUSTOMER(CustID, Name, Address, Telephone, …)
MOVIE(MovieID, Title, Director, Year, Studio, ….)

where CUSTOMER is a relation containing demographic and other information on customers, MOVIE is a relation containing the information on all the movies available for rent, and RENTAL is a relation containing transactional information on the movies rented by customers, including ratings of the movies by the customers.

A user profile should contain several types of information, including factual information and the rules describing the behavior of that user. The factual information comprises the *factual profile* and can include demographic and psychographic data about the user, the content information about the items that the user considered in the past, and ratings of these items. For example, the factual profile may contain information on an average number of movies rented per month, the types of movies the customer likes (e.g. adventure, action movies, etc.), most favorite actors and directors, the ratings of different movies, etc. In addition to this, the profile may contain rules describing that user's behavior. Examples of such rules are: on weekdays customer X rents action movies; on weekends, whenever customer X rents a classical movie, she also rents an adventure movie; in the past three months, whenever customer X rents an Italian movie, it has Marcello Mastroianni in the leading male role; among the adventure movies, customer X rated the movies with Harrison Ford in the leading male role higher than adventure movies with any other male actor. These rules can be expressed using standard first-order logic formalisms. For instance, the rule "on weekends, whenever customer X rents a classics movie, she also rents an adventure movie" can be expressed for customer XYZ as IF day_of_week = "weekend" AND movie1_type = "classic" THEN movie2_type = "adventure." These rules are combined together into a *behavioral profile*.

Rules comprising the behavioral part of a user profile can be obtained by applying data mining methods [Fayyad et al. 1996] to the transactional data of the user. We have developed a system that discovers relevant behavioral rules that is described in [Adomavicius & Tuzhilin].

Capturing User Behavior: Behavioral Profiles vs. Content-Based Approach

In general, the behavioral profile and the content-based approaches to recommendations are quite different. The content-based approach provides actual recommendations for the items by analyzing the content (descriptions) of the items that have been rated by the user before. In contrast to this, the behavioral profile does not provide any recommendations by itself. It only captures the on-line behavior of the users. However, this behavior can be used in subsequent recommendations if coupled with other techniques, such as collaborative methods. We will address this issue in the next section.

Nevertheless, the two approaches are related because the content-based approach also captures user behavior. However, it does this only in a limited sense by storing in the user profile the content information and the ratings of the items previously selected by the user. In contrast to this, the behavioral profiles can describe much richer types of user behavior than user profiles from the content-based approach. In particular, behavioral rules are not limited to item ratings and to the content information of the items and can also contain various other parameters of user's behavior, such as time, date, and promotional information, including coupons and discounts.

Integration of the Behavioral Profile and Collaborative Approaches

As stated in the previous section, behavioral profiles only capture user behavior and do not provide actual recommendations. Therefore, it is important to augment the behavioral profiling approach with the collaborative approach. This is in line with basic principles of marketing, according to which customer recommendations should be based:

1. on understanding behavior of that customer

2. on what other customers similar to the given one prefer.

Integration of the two approaches should provide for more accurate recommendations than the collaborative approach alone. For example, consider the rule stating that on weekends the customer usually rents adventure movies, i.e., *IF* day_of_week = "weekend" *THEN* movie_type = "adventure." Also assume that it is a weekend now. Then we can restrict our recommendations only to the set of adventure movies. Therefore, our next position statement is:

> *Position Statement 2:* It is important to develop methods for integrating the behavioral profiling with the collaborative filtering approaches to recommendations into one integral approach.

We are working on this problem now. In particular, we are examining the following approaches.

Profiles Drive Collaborative Methods. In this approach, we start with the behavioral profile of the user. We first apply the behavioral rules to reduce the consideration set of items that should be used for the recommendations in the present situation. For example, if it is a weekend now, then the previously mentioned rule would reduce the consideration set of movies only to adventure movies. After we apply all the rules from a user profile to the current "state of the world" and obtain the consideration set of items from which the actual recommendation should be made, the next step is to use collaborative filtering to select one of these items. This can be done using standard collaborative methods, but on a smaller consideration set. We expect this to increase the accuracy of recommendations in comparison to the stand-alone collaborative filtering method.

Profiles Are Used After Collaborative Filtering. In this approach, we first generate a preliminary set of possible recommendations using standard collaborative filtering techniques. We then use behavioral profiles to reorder or even eliminate some of the items that were preliminarily recommended. For example, if collaborative filtering approach recommends three comedies followed by two adventure movies, and it is a weekend now, then the previously described behavioral rule would move adventure movies above comedies in the ranking list.

In this section, we considered integration of behavioral profiles and collaborative methods into one integral approach and maintained that it should outperform the stand-alone collaborative filtering approach. However, it is also important to compare this integrated approach with the combination of content-based and collaborative approaches.

We argued in the previous section that the profiling approach captures a much richer class of behaviors than the content-based approach. However, it is still not clear if profiling and collaborative filtering *consistently* results in better recommendations than the combination of content-based and collaborative filtering approaches. Based on a preliminary analysis of this issue, we believe that this is indeed the case. However, before we provide a definite proof to this, we intend to use *all the three* approaches together: behavioral profiling, content-based, and collaborative approaches. For example, the profiling approach would first restrict the consideration set as described above, and then the combination of the content-based and collaborative filtering approaches would be used on this restricted consideration set.

Conclusions

The literature on recommender systems distinguishes between the collaborative and the content-based approaches and also considers the combination of the two as a more accurate approach. In this paper we argued that it is also important to base user recommendations on the in-depth analysis of user behavior that goes beyond the content-based and collaborative methods. We also described how such behavior could be captured with the user profiles that contain factual and behavioral information. This approach is in line with the basic principles of marketing science, according to which customer recommendations should be based on (1) a thorough understanding of the customer's behavior, and (2) on the preferences of other customers similar to the given one.

We also maintained that it is important to develop methods that combine the behavioral profiles and the collaborative approaches into one integral method. We are currently working on the development of such techniques.

Bibliography

1. Adomavicius, G. and Tuzhilin, A. *Building Customer Profiles in Personalization Applications Using Data Mining Methods*. Submitted for publication.

2. Balabanovic, M. and Shoham, Y. Fab: Content-Based, Collaborative recommendation. *Communications of ACM*, March 1997, vol. 40, no.3, pp. 66-72.

3. Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., and Uthurusamy, R. eds., *Advances in Knowledge Discovery and Data Mining*. AAAI/MIT Press; 1996.

4. Konstan, J. A., Riedl, J., Borchers, A., and Herlocker, J. L. Recommender Systems: A GroupLens Perspective. *Recommender Systems. Papers from 1998 Workshop. Technical Report WS-98-08*, pp. 60-64. AAAI Press, 1998.

5. Pazzani, M. J. A Framework for Collaborative, Content-Based and Demographic Filtering. *Artificial Intelligence Review.* To appear.

6. Salton, G. *Automatic Text Processing*. Addison-Wesley, 1989.