An Intelligent Knowledge Sharing Strategy Featuring Item-Based Collaborative Filtering and Case Based Reasoning

Zeina Chedrawy and Syed Sibte Raza Abidi
Faculty of Computer Science, Dalhousie University, Halifax B3H 1W5, Canada
{chedrawy,sraza}@cs.dal.ca

Abstract

In this paper, we propose a new approach for combining item-based Collaborative Filtering (CF) with Case Based Reasoning (CBR) to pursue personalized information filtering in a knowledge sharing context. Functionally, our personalized information filtering approach allows the use of recommendations by peers with similar interests and domain experts to guide the selection of information deemed relevant to an active user's profile. We apply item-based similarity computation in a CF framework to retrieve N information objects based on the user's interests and recommended by peer. The N information objects are then subjected to a CBR based compositional adaptation method to further select relevant information objects from the N retrieved past cases in order to generate a more fine-grained recommendation.

1. Introduction

Knowledge management (KM) methodologies, methods and applications enable organizations to generate value from their knowledge-based assets. KM methods encompass the capture and sharing of a wide variety of knowledge objects, represented using different knowledge modalities, within an organization. Typically, knowledge sharing permits individuals to share ideas, work in groups, brainstorm and collaborate; all these activities potentially lead to collaborative learning and collective improvement in the performance, intellect and judgment of the participants. However, it is important to understand that 'pushing' information to individuals without considering their interests, needs and intellectual expertise might compromise the optimal use of the information being shared. Rather, knowledge sharing—involving the identification, filtering and delivery of information—should be supported by intelligent solutions to ensure that the right and relevant information is provided to or retrieved by an individual—i.e. personalized information filtering.

We argue that personalized information filtering, in the realm of knowledge sharing, should be guided by three elements: (a) the interests and experiences of the individual seeking information for a particular purpose; (b) the ratings/recommendations of like-minded peers for potential knowledge objects that can be provided to the user; and (c) the past responses and noted experiences—i.e. past information retrieval requests and the corresponding information being recommended—of domain experts to similar information filtering situations. The above tantamount to personalized information filtering [2] characterized at three levels—i.e. personal, community of peers, and expert's experiences. Here, we are suggesting a collaborative information filtering strategy in which personal interests initiate and shape the information retrieval criterion and serve as a precursor to peer based recommendations for pertinent information objects, and finally the peer recommendations are compared against past expert's experiences to streamline and/or enhance the relevance of the information content being filtered. In this manner, knowledge workers are able to leverage and exploit the past experiences and recommendations of not only their peers but also of the domain experts when seeking information about a particular topic. Recent work in recommendation systems along similar lines includes intelligent recommendation aides for selecting web sites [7], news stories [4], and TV listings [6].

Collaborative Filtering (CF) techniques facilitate knowledge sharing based on recommendations, provided by a community of practice or like-minded users [14]. CF systems are employed in an interactive and iterative manner by their users. The main idea is to compare the user-model of an active user, defined in terms of user preferences and characteristics, with the user models of previous users in order to find k similar users. The historical user models are then used to determine the likely preferences of the active user, and the predicted relevant information content, deemed as personalized information, is provided to the active user. Hence, peer within a community of practice, in general, rely on others' experiences when they need to choose an option/action/knowledge item and yet they do not have experience to make an informed judgment.

The AI based reasoning paradigm of Case-based Reasoning (CBR) provides analogy based recommendations based on historical models (or past experiences)—to solve similar new problems [1]. Typically, CBR recommends the entire solution of previous cases as the solution to the new case, despite any inherent dissimilarity between the new and past cases. However, it has been argued that the solution of selected past cases should be appropriately adapted towards the new case so that a more realistic recommendation is possible. Compositional adaptation methods, within CBR, allow item (or attribute) based solution adaptation that leads to more fine-grained tuning of past solutions towards the new problem [3].

In this paper, we present an intelligent information filtering strategy that is a hybrid of item-based collaborative filtering [14] and case based reasoning methods (see figure 1).

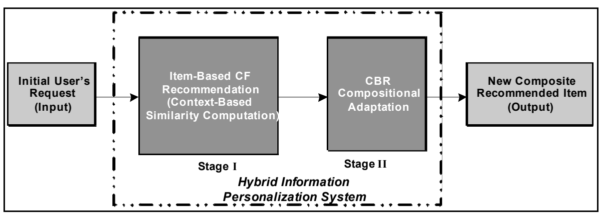

Figure 1. An illustration of our hybrid framework for generating composite recommendations

Our personalized information filtering approach allows the use of recommendations by past users (with similar interests) and domain experts to guide the selection of information deemed relevant to an active user's profile—i.e. interest, goals and needs. In our work, information filtering is implemented in two stages; in stage I, we apply item-based similarity computation in a CF framework to retrieve N information objects based on the user's interests and also recommended by past users with similar interests. The requirement for item-based CF being that the active user should have rated at least one item before he/she can get recommendations based on his/her profile. We have developed a multi-feature item-based CF strategy that allows a more detailed information selection criterion. In stage II, we apply compositional adaptation—an attribute based information selection approach—to selectively collect relevant information objects from the N retrieved past cases to generate a new fine-grained solution/recommendation that is related to the user's initially stated interest either to the similar items' features or to the solutions assigned to similar items in order to generate a new composite, yet well-focused recommendation. We conclude that, in our approach, we are able to leverage the past recommendations of peers (through CF) and domain experts (through CBR) to dynamically generate a new fine-grained solution/recommendation that is highly focused to the individualistic knowledge needs of a specific user.

2. Collaborative Filtering Framework

Functionally, CF builds a database of preferences/ratings done by distinct users on specific items. Based on the work done by Sarwar et al. [14], given a list of m users $U = \{u_1, u_2 \ldots u_m\}$ and a list of n items $I = \{i_1, i_2, \ldots, i_n\}$, each user u_i has a list of items I_{ui}, which he has already rated. Preferences are usually given as rating scores. A CF algorithm finds an item likeliness that can be of two forms:

- *Prediction*, a numerical value, $P_{a,j}$, expressing the predicted likeliness of item $i_j \notin I_{ua}$ for the active user. This predicted value is within the same scale as the preference values provided by u_a.
- *Recommendation*, a list of items, $I_r \subset I$ that the active user will like. The recommended list is on items that have not been already chosen by the active user, $I_r \cap I_{ua} = \varnothing$.

Item-based (Model-based) CF algorithms address the scalability challenge of CF [14] and consider the relation between items rather than users to compute item similarity, recommend a list of items and provide prediction to the active user. First a model for users' ratings is created, it creates a user-item matrix $M(u,i)$ that contains the rating of user u on item i. Before any recommendation or prediction takes place, a similarity metric between items has to be established and corresponding similarities computed. In order to compute the similarity between two items i and j, the set U of all users who have rated both items is first identified, then a similarity technique is applied. There exist numerous similarity techniques that can be used in order to find similarity between items; we have chosen the "adjusted cosine" method proposed by Sarwar et al. in [14]. Given the matrix $M(u,i)$ which denotes the rating of user u on item i, the similarity between items i and j is given by:

$$sim(i,j) = \frac{\sum_{u \in U}(M(u,i) - \overline{M}(u))(M(u,j) - \overline{M}(u))}{\sqrt{\sum_{u \in U}(M(u,i) - \overline{M}(u))^2}\sqrt{\sum_{u \in U}(M(u,j) - \overline{M}(u))^2}} \quad (1)$$

$\overline{M}(u)$ is the average rating of user u on all rated items

2.1. Multi-Feature Items' Similarity Computation

In our work, the CF similarity computation (formula 1) was extended to cover multiple features/ratings, thereby establishing a context based on a set of features. The notion of context in our work allows us to establish similarity based on features that are more relevant to the user as per the current information needs. This context-based similarity determines how close the compared items are with respect to some perspective. Similar approach, applied to the case retrieval stage in CBR systems, is illustrated in [8].

In a CF system, for personalized information retrieval purposes, a user model is used that stipulates a set of preferences and characteristics of the active user. In essence, the user model contains all items rated by the target user along with the ratings required with respect to all features/attributes. For instance, consider an active user u who has rated f number of item-defining attributes/features of an item i. $M_t(u,i)$ is then the rating of the active user u for feature t on a given item i and (formula 1) results in a similarity $sim_t(i,j)$ between two items i and j with respect to feature t.

In addition, we assumed that not all features have equal effects when computing similarity between items; one feature might have a larger contribution to the solution than the other. In fact, if the active user is interested in a feature more than another, he would choose a larger weight to the feature he/she is mostly interested in and smaller weights to the features he/she considers irrelevant. This allows the user to specify a weight value W_t relative to every t^{th} feature. The overall similarity referred to as multi-feature similarity or context-based similarity, is given by:

$$overall\_sim(i,j) = \frac{\sum_{t=1}^{f} W_t * sim_t(i,j)}{\sum_{t=1}^{f} W_t} \qquad (2)$$

2.2. Item-Based CF Case Retrieval Procedure

Item-based CF compares the user model of an active user with the user models of other users so that a set of recommendations is issued that answers the active user needs and interests. The user model structure can be expressed as a vector of attribute/value pairs describing the items rated by the user, and the corresponding ratings' values with respect to all given features. The algorithm for case retrieval works as follows:

- *Step 1* - Preferences of all users are stored in a database and are expressed as rating scores on chosen items. An item is characterized by a set of features and ratings are assigned to every item feature. The active user specifies his interests and needs in terms of preferred item features.

- *Step 2* - Based on the features that interest the active user, item-based CF searches for all items that match with the user needs. The context-based multi-feature similarity metric is used which covers all selected features. One similarity value is computed with respect to each feature, and then similarities with respect to selected features are combined into one overall similarity that reflects the context-based similarity between items (formula 2) and will be used through the remaining part of the algorithm.
- *Step 3* - After computing all similarities between items, a *Top-N* recommendation procedure follows:
  - Find the set R of all items rated by the active user u.
  - For every item in R, find the set of k most similar items.
  - The unions of the sets of k most similar items form the set S.
  - Remove from S all items that have been rated by u.
  - For every item i in S, compute its similarity *Sim* to the set R. This similarity is the sum of the similarities (calculated in *Step 2*) between all rated items by u and item i.
  - Sort the set S by the similarity *Sim* and select the *Top-N* items.

The selected N items would most likely interest the active user because the selected N items are the most similar to the set of items rated by the active user. Every recommended item is a retrieved case that will be subject to the CBR compositional adaptation as a next stage of our hybrid recommendation system.

2.3. Performance of the Multi-Feature Similarity Metric in Recommendation

Recall is a metric commonly used to measure the quality of a recommendation. Recall is the fraction of interesting items that can be located. Let *hits* be the total number of recommended items that were really rated by any of the users but were excluded from the training data to serve as test data; when a set of recommended items is generated for a user, if the rated item in the test set exists in the recommended set, then the number of hits is incremented. Let t be the total number of users in the test set, then recall is given by:

$$\text{Recall} = \frac{hits}{t} \quad (3)$$

In order to evaluate our CF framework, we used the dataset taken from MovieLens web-based recommender system [19]. The original dataset consists of 100,000 ratings from 943 users on 1,682 items. In our experiments, we reduced the dataset to 600 users and 250 items. In order to study the performance of the context-based similarity measure, the current dataset is further modified to contain two more rating attributes for different perspectives. The resulting dataset is divided randomly into training and testing sets. The training set is used to compute the top *N* recommendations while the test set is used to measure the performance of the recommender system. We fixed the number of recommendations *N* to 30. For every neighborhood size value, recall is calculated. During evaluation, the rating values in the test set were ignored and we considered that if a user has rated an item, then he would have interest in it. Hence, our main objective is to match the recommended set of items to the set of items that exist in the test set for an active user.

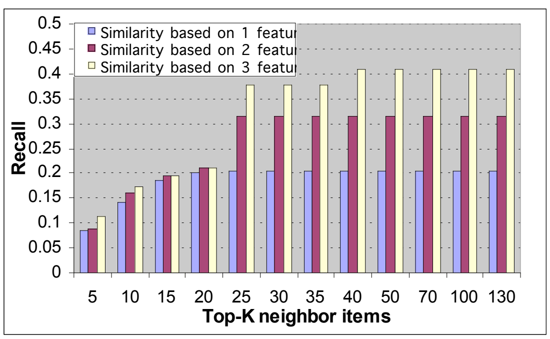

Figure 2. Effect of the multi-feature similarity metric and neighborhood size on the quality of recommendation

Figure 2 shows that the multi-feature similarity measure along with the item neighborhood size improve items' recommendation. In fact, when the similarity is based on only one feature rather than multiple features, the similar item space which includes all items that are similar to the target item is large. As a result, the system will recommend items that are of no interest to the target user. However, when the similarity is based on more features, the similar item space is further reduced; hence the recommendation becomes more focused.

2.4. Prediction of User Preferences

In order to predict the preference of an active user *u* on a specific item *i* corresponding to a feature *k*, the item-based collaborative filtering algorithm looks into the matrix *M*, computes similarity between the target item and items rated by the target user. Prediction on the target item can be generated by taking a weighted average of the target user's ratings on the generated *N* similar items as in formula 3.

$$P_k(u,i) = \frac{\sum_N sim_k(i,N) M_k(u,N)}{\sum_N |\, sim_k(i,N) \,|} \quad (4)$$

Each rating $M_k(u,N)$ with respect to feature *k* is weighted by the corresponding similarity between the target item and the similar item *N*, $sim_k(i,N)$. As per (formula 3), the predicted value is normalized by dividing over the sum of the similarities between the *N* similar items and the target item *i*. $P_k(u,i)$ refers to the predicted rating value for user *u* on target item *i* with respect to feature *k*. Only the top *N* similar items are used for prediction generation.

3. Case Based Reasoning for Information Personalization

The CBR component of our information retrieval strategy using the top *N* items retrieved by the CF method as the solution, and proceeds to apply compositional adaptation to the retrieved solution (i.e. information objects relevant to the user) to compose a more focused and personalized information recommendation. CBR makes associations along generalized relationships between problem descriptors and conclusions [1,16]. Typically, the solution to the new case is determined based on the solutions of the similar past cases; however, in our work we attempt to get a solution that better fits the problem at hand, hence the solutions are dynamically adapted. We apply compositional adaptation method that combines selected components (sub-solutions) from multiple cases to produce a composite solution that is more focused than the original past solutions.

Adaptation is considered as the most difficult task in CBR [17]. We argue that one limitation of traditional information retrieval approaches is that the information object is presented as a whole, provided it matches some gross relevance criterion. In our work, we use the CBR-mediated compositional adaptation approach to select pertinent information 'segments' from multiple past solutions to generate fine-grained personalized information content. Our approach is applicable when the solution consists of multiple independent components [17], which can then be independently selected and combined to form the final solution [13].

The efficacy of the second stage of our information retrieval approach is materialized in the fact that the final information content, recommended to the user, is designed as a 'composite' of individual sub-solutions, where each sub-solution addresses a particular interest of the user [3].

4. A Collaborative Filtering Framework Featuring Case-Based Reasoning: The Algorithm

In this section, we present the compositional adaptation algorithm—involving both case retrieval and adaptation processes—as the final stage of our information retrieval strategy.

As mentioned earlier in this paper, the new framework combines collaborative filtering methods, in particular, item-based collaborative filtering with the CBR methodology in a knowledge sharing context in order to ameliorate recommendations—that makes up the personalized information content, presented to the active user. Item-based similarity computation of the CF framework is applied to retrieve N items similar to the active user's taste. After the N candidate items are identified— through the case retrieval stage of CF— the items are either represented as a series of features or each item is assigned a solution that is an information item comprising a set of features. Compositional adaptation is applied to the multiple cases' solutions to produce a new composite recommendation.

4.1. Solution Adaptation via Compositional Adaptation

After the similar cases are identified— through the case retrieval of CF—their corresponding solutions need to be adapted so that a fine grained personalized solution is derived and exposed to the active user. A new approach to compositional adaptation has been used for proposing suitable solutions to the active user. The basis for our adaptation strategy is defined by two factors: i) the frequency of occurrence of a solution component in the similar cases and ii) the degree of similarity between the user profile and the retrieved case. For instance, if a solution component appears in several similar cases that are retrieved, then there exists a high possibility that this component would be part of the final solution. Below is the procedure for compositional adaptation:

- *Step 1* - For every retrieved case, compute the total distance between the retrieved case and the user request (user preferences/ratings). This distance is the similarity value obtained in Step 3 of Section 3.2—the similarity of the recommended item/case to the set of items rated by the active user. Let $Dist_{Ci}$ be the distance for the similar item/case C_i.

- *Step 2* - Compute the normalized distance for each similar case C_i over the entire set of n retrieved cases as follows:

$$Temp = \sum_{i=1}^{n} 1/Dist_{C_i} \qquad (5)$$

The normalized distance for the retrieved past case C_i is calculated as:

$$NDist_{C_i} = 1/(Dist_{C_i} * Temp) \qquad (6)$$

- *Step 3* - Determine the appropriateness degree of available solution components. Let *Comp* be a component of a solution from a past case, and AD_{Comp} be the appropriateness degree for *Comp*, then

  For k = 1 to n

  If *Comp* exists in the solution of the similar case C_k

$$AD_{Comp} = AD_{Comp} + NDist_{C_k} \qquad (7)$$

In order to compute the appropriateness degree of each component, the normalized distance of similar cases that contain this component are added to one another. If AD_{Comp} is greater than some predefined threshold value, then the component *Comp* would appear in the final solution.

- *Step 4* - After combining the components from multiple cases to form the final solution, the resulting new case is added to the case base if it satisfies the active user needs and preferences.

5. Concluding Remarks and Future Work

In this paper, we have introduced a new intelligent information retrieval that is a hybrid of Collaborative Filtering and Case Based Reasoning schemes in order to ameliorate information personalization in a knowledge sharing context. The suggested item-based similarity CF technique is context-based and is characterized by its capability of using selected features to find the similarity between cases. The process is known as case retrieval stage where the resulting similar cases are then used as input to the CBR framework.

CBR systems are responsible of building solutions based on past experiences when solving a new problem. When solutions of similar cases are derived, their corresponding components are combined efficiently through compositional adaptation so that a well-focused final solution is obtained. In compositional adaptation, only relevant sub-solutions are chosen to be part of the final solution. For instance, the frequency of occurrence of a sub-solution along with the degree of similarity between the retrieved case and the user request determine whether the sub-solution is to belong to the final solution or not. It was shown in [3,5,13] that compositional adaptation has achieved significant results in many applications such as intelligent tutoring systems, adaptive hypermedia for internet portals. The efficacy of CBR based compositional adaptation in our work is under investigation through a series of experiments with the same dataset used for CF experimentation. At an intuitive level the proposal of using compositional adaptation is appealing and provides an opportunity to produce well-focused and accurate item-based information personalization.

Finally, we believe that the suggested personalized information filtering framework will provide an active user fine-grained information content that will satisfy his/her needs. It will be demonstrated that the hybrid of context-based similarity and compositional adaptation techniques will significantly impact the effectiveness of the final information content delivered to the user.

6. References

[1] Aamodt A. and Plaza E., "Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches", In AI communications. IOS Press, 1994.

[2] Abecker A., Bernardi A., Hinkelmann K., Kuhn O. and Sintek M., "Toward a Technology for Organizational Memories", In IEEE Intelligent Systems, 1998.

[3] Abidi S.S.R., "Designing Adaptive Hypermedia for Internet Portals: A Personalization Strategy Featuring Case Base Reasoning With Compositional Adaptation", In Garijo FJ, Riquelme JC & Toro M (Eds.) Lecture Notes in Artificial Intelligence 2527: Advances in Artificial Intelligence (IBERAMIA 2002). Springer Verlag: Berlin, 2002.

[4] Ardissono L., Console L. and Torre I, "An Adaptive System for the Personalized Access to News", In AI Communications, 2001.

[5] Arshadi N. and Badie K., "A Compositional Approach to Solution Adaptation in Case-based Reasoning and its Application to Tutoring Library", In Proceedings of 8th German Workshop on Case-Based Reasoning. Lammerbuckel, 2000.

[6] Cotter P. and Smyth B., "PTV: Intelligent Personalised TV Guides", In AAAI/IAAI 2000.

[7] Eliassi-Rad T., and Shavlik J., "A Theory-Refinement Approach to Information Extraction", In Proceedings of the 18th International Conference on Machine Learning , Williamstown, MA, 2001.

[8] Jurisica I., "Context-Based Similarity Applied to Retrieval of Relevant Cases", Technical Report DKBS-TR-94-5, University of Toronto, Department of Computer Science, Toronto, Ontario, 1994.

[9] Karypis G., "Evaluation of Item-Based Top-N Recommendation Algorithms", Technical Report #00-046, Dept. of Computer Science, University of Minnesota, Minneapolis, MN, 2000.

[10] Moubasher B., Jin X. and Zhou Y., "Semantically Enhanced Collaborative Filtering on the Web", In Proceedings of the European web mining forum, Bettina Berendt et al. (ed.), LNAI 3209, Springer 2004.

[11] Pennock D.M. and Horvitz E., "Collaborative Filtering by Personality Diagnosis: A Hybrid Memory- and Model-Based Approach", In IJCAI Workshop on Machine Learning for Information Filtering, Stockholm, Sweden, August 1999.

[12] Resnick P, Iacovou N., Suchak M., Bergstrom, and Riedl J., "GroupLens: An Open Architecture for Collaborative Filtering of Netnews", In Proceedings of ACM 1994 Conference on Computer Supported Cooperative Work, Chapel Hill, 1994.

[13] Reyhani N., Badie K. and Kharrat M., "A New Approach to Compositional Adaptation Based on Optimizing the Global Distance Function and Its Application in an Intelligent Tutoring System", In IRI 2003.

[14] Sarwar B. and Karypis G., Konstan J. and Riedl, J. "Item-based Collaborative Filtering Recommendation Algorithms", In Proceedings International WWW Conference (10), Hong-Kong, 2001.

[15] Sarwar B., Karypis G., Konstan J. and Riedl J., "Analysis of Recommendation Algorithms for e-commerce", In Proceedings of ACME-Commerce, 2000.

[16] Schmidt R., Vorobieva O. and Gierl L., "Case-Based Adaptation Problems in Medicine", In Reimer U. et al.: Proceedings of WM2003: Professionelles Wissensmanagement – Erfahrungen und Visionen. Kollen-Verlag Bonn, 2003.

[17] Wike W. and Bergmann R., "Techniques and Knowledge Used for Adaptation During Case-Based Problem Solving", In Proceedings of the 11th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems: Tasks and Methods in Applied Artificial Intelligence, June 01-04, 1998.

[18] Wike W., Smyth B. and Cunningham P., "Using Configuration Techniques for Adaptation", In M Lenz, B Bartsch-Sporl, H-D Burkhard, S Wess (eds.): Case-Based Reasoning Technology, From Foundations to Applications, Springer-Verlag, Berlin Heidelberg New York, 1998.

[19] http://www.grouplens.org