Case Based Reasoning for Information Personalization: Using a Context-Sensitive Compositional Case Adaptation Approach

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 present an intelligent information filtering strategy that is a hybrid of item-based Collaborative Filtering (CF) and Case Based Reasoning (CBR) methods. Information filtering is implemented in two phases: in phase I, we have developed a multi-feature item-based CF strategy that allows creating a detailed context for filtering the information and retrieving N information objects based on user's interests and also preferred by similar users with similar tastes. In phase II, we use the N retrieved items as input to the CBR information filtering system and apply CBR-based compositional adaptation technique to selectively collect distinct information components of the N retrieved past items pairs to produce a composite recommendation that better addresses the initial user's interests and needs. We show that the hybrid of context-based similarity and compositional adaptation techniques improves significantly the quality of the recommendations presented to the user in terms of accurate and precise personalized information content.

Keywords
Collaborative Filtering; Information Filtering; Case-Based Reasoning; Information Personalization; Compositional Adaptation.

I. INTRODUCTION

Nowadays, the volume of information over the Internet is increasing at a tremendous rate; therefore, the search for 'relevant' and 'useful' information is becoming proportionally difficult. There is a need and corresponding demand to provide tools to characterize, capture and determine both the relevance and utility of available information items in order to (a) regulate the flow of information to users; and (b) direct users to the right information, based on the user's current specific interests and needs—i.e. to provide personalized information to users.

Recommender systems, used widely over the Internet, act as mediators between information sources and information seekers. Such systems use a variety of methods to filter information in a systematic and transparent manner so that the personalized information being recommended to a user is pertinent to his/her current information needs and interests [16]. Content based approaches facilitate the finding and filtering of information by comparing the user's profile with some meta-description of the information item. Collaborative filtering approaches do not require a content description of the information item, rather they filter information by making use of the experiences and opinions of peers with similar interests—peers critique the information items by rating them along a number of quality and utility dimensions. In this case, the filtering mechanism recommends information that has been measured for its utility in a certain context.

The role of context in information filtering for generating personalized information is paramount as it determines both the relevance and usefulness of the recommended information. Context implies a generalized set of intrinsic relationships between a set of perspectives believed in some way to help make clear and to understand the current information-mediated task, event or discussion and the corresponding information needed [8]. Context, in particular for information personalization, can be characterized by i) a set of discernable feature-value pairs that help describe the information filtering task; and ii) a set of similar past information seeking tasks and corresponding experiences/solutions, preferably the solutions recommended by domain experts. We argue that recommender systems for personalized information should allow both the specification of context through a variety of methods and the subsequent filtering of information based on the context. In our work, we pursue contextual information personalization in the realm of collaborative filtering, whereby context is determined by the peer reviewed rating of information items along a set of perspectives chosen by the user.

In this paper we present a context-sensitive information personalization strategy that is a hybrid of item-based Collaborative Filtering (CF) [17] and Case Based Reasoning (CBR) methods [1]. Information filtering via personalized information is achieved in two phases: In phase I, we apply item-based similarity computation in a CF framework to retrieve N information objects based on the user's interests and also preferred by past users with similar interests. We have developed a multi-feature item-based CF strategy that allows the user to specify a context to guide the information filtering criterion. In phase II, we apply CBR-based compositional adaptation technique [20]—a component-based information selection approach—to selectively collect distinct information components from the overall N retrieved past items/problem-solution pairs—i.e. select only the most relevant and useful information components from relevant solutions (as opposed to the whole solution) in order to generate a new fine-grained

recommendation for a specific user. We conclude that, through our approach, we are not only able to leverage the original past recommendations of peers (through CF) and domain experts (through CBR) but the application of compositional adaptation methods in a CF framework allows the dynamic adaptation of recommendations based on specific contexts, and hence is highly focused to the individualistic information needs of the user.

II. RELATED WORK

Collaborative Filtering systems are considered the most successful recommender systems and have achieved widespread success on the web [17]. The basic idea behind CF is to recommend items/objects (e.g. music [3], movies [6], books [14]) to a user who is seeking for advice from a set of preferences expressed by other users at a previous time [17,15,11]. These preferences are then compared in order to find similarities between users or items and generate personalized information retrieval.

The AI based reasoning paradigm of CBR provides analogy based recommendations based on historical models or past experiences [1]. The problem-solving principles of CBR make them an interesting candidate for integration with similarity based information filtering methods, such as CF methods [10].

Recently, integrating CBR in CF has gained a lot of attention and success [5]. Many hybrid systems were designed that combine both techniques such as the personalized recommendation TV system PTV [6]. PTV operates in the TV listings domain; it combines CBR and CF techniques by interleaving the recommendations, which are the TV programs, in order to produce personalized TV guides for users. PTV uses user profiling and information filtering techniques to generate web-based TV viewing guides that are personalized for the viewing preferences of individual users. PTV examines the potential benefits of both CF and CBR by developing CBR methods that employs CF style ratings profiles directly as cases. CoCoA [3] is a recommendation system for music compilation. It uses a case-based retrieval engine based on the CF approach to create a collection of sound tracks. Other hybrid systems were also developed such as Entrée [5], Tapestry [9], GroupLens [9].

Our work combines item-based CF with the CBR techniques using case adaptation as a final means to generate focused personalized information. The following sections give a detailed illustration of our framework.

III. OUR PERSONALIZED INFORMATION FRAMEWORK

We argue that personalized information is characterized at i) a personal level, i.e. the interests of an individual, and ii) the community of peers level, i.e. the ratings of like-minded peers.

The proposed framework consists of two phases: the CF retrieval phase and CBR adaptation phase (Fig. 1). First, the user who is seeking for advice enters his request by specifying a set of preferences (Input). In phase I, the CF recommender

system retrieves from its knowledge base—containing all other users' preferences, N information items that have been preferred by users with similar interests and at the same time these items fulfill the user's needs. In phase II, we apply compositional adaptation either to the similar items' components or to the solutions assigned to similar items in order to generate a new composite recommendation (Output). This two-phase process is illustrated in more detail in the following sections.
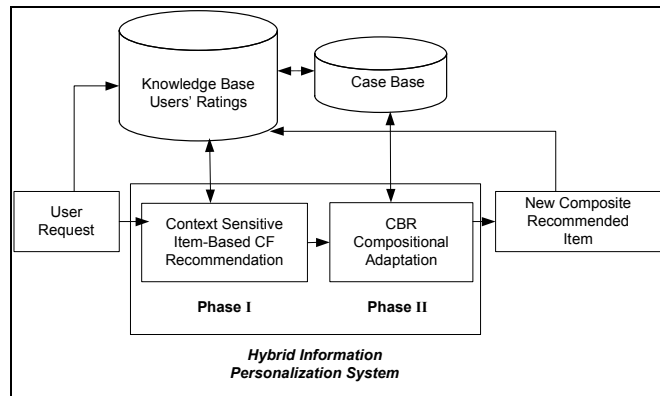

Fig.1. Framework Schematic

IV. PHASE I: CONTEXT-SENSITIVE RECOMMENDATION

The first phase of our framework involves item-based CF in order to select a set of items based on the ratings/recommendations of like-minded peers. A context sensitive CF strategy is presented.

Functionally, CF builds a database of ratings done by distinct users on specific items. Given a list of m users $U = \{u_1, u_2...u_m\}$ and a list of n items $I = \{i_1, i_2, ..., i_n\}$, each user u_i has a list of items I_{ui}, which he has already rated [17]. Typically, CF systems represent user preferences as rating scores showing only that the user has either liked or disliked the rated item. Therefore, one user-item matrix, known as user-model $M(u,i)$, can be created that contains the rating of user u on item i and is used to find the similarity between rated items.

The traditional approach only shows whether the items to be compared are similar along a single perspective without the recourse to a richer context. In our work, we extend the item-based CF proposed by Sarwar et al. [17]; we attempt to create a context for information filtering which involves the use of ratings on items on multiple perspectives along which similarity (or dissimilarity) can be established between two information items. Note that each rating corresponds to one specific criterion; multiple user-item matrices are then generated that designate the ratings with respect to various criteria.

For instance, in the music domain, a compilation (information item) is defined along multiple perspectives: singer performance, song lyric, song rhythm…etc. The original approach considers only one perspective such as 'song lyric' and allows users to express their opinion on the compilation as rating scores along that single perspective. However, in our framework, a context could be defined by one or multiple

perspectives along which music compilations are evaluated and rated. For every perspective, we generate a user-item matrix that contains the ratings for each perspective and a similarity between rated compilations is calculated accordingly. The individual similarities are then combined to compute the similarity vector between compilations over all perspectives.

After a user-model is generated from the user database which contains the information about past users' ratings (Fig. 2), the similarity between two co-rated items i and j is calculated using (1). Similarity between items is determined by calculating correlations of user ratings between items. Items are co-rated if a set of users has rated both items.

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

$M_t(u,i)$ corresponds to the rating of user u on item i with respect to perspective t of context C. $\overline{M_t(u)}$ is the average rating of user u on all rated items. U is the set of all users in the knowledge base.

Since each perspective has its own contribution to the solution and the importance of perspectives changes with the user's interest, a weight value is assigned to every perspective. Let P be the total number of perspectives, W_t the weight assigned to perspective t in context C, $sim_t(i,j)$ the similarity between items i and j with respect to t, the overall similarity $SIM_C(i,j)$ between items i and j is then computed as:

$$SIM_c(i,j) = \frac{\sum_{t=1}^{P} W_t * sim_t(i,j)}{\sum_{t=1}^{P} W_t} \quad (2)$$

This context-based similarity measure is applied before any recommendation takes place. In this regard, we first calculate the individual similarities between i and j for the selected perspectives, then combine them to derive an overall context similarity.

A. Item-Based CF Recommendation

Item-based CF algorithm compares the user model of a user with the user models of other peers so that a set of recommendations is issued that answers the user's needs and interests. An Item-based CF algorithm firstly considers the relationship between items rather than users and then uses these relationships to indirectly generate recommendations. The user model structure is 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 perspectives. Our context-based algorithm for recommending the top N similar items works as follows:

1) Preferences of all users are stored in a database and expressed as rating scores on chosen items. Ratings are assigned to rated/visited items with respect to multiple perspectives of a specified context.
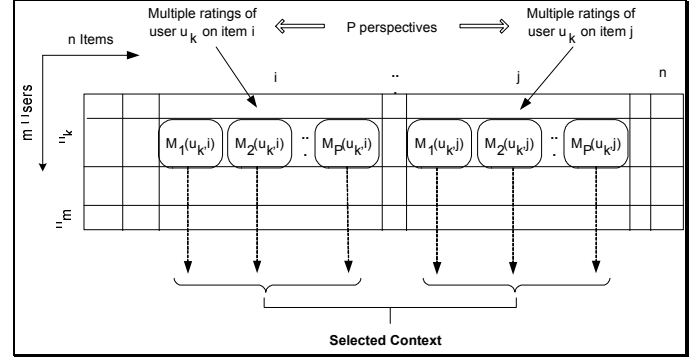

Fig. 2. Detailed representation of multiple user-item ratings

2) Based on the perspectives that interest the user, item-based CF searches for all items that match with the user needs. The similarity metric used to identify similar items is the one introduced in (1) based on the "adjusted cosine" method and covering all selected perspectives. One similarity value is computed with respect to each perspective, and then similarities with respect to selected perspectives are combined into one overall similarity that reflects the context-sensitive similarity between items (2) and will be used through the remaining part of the algorithm.

3) After computing similarities between items, a *Top-N* recommendation procedure follows:
   - Find the set R of all items rated by the user u.
   - For every item in R, find the set of k most similar items.
   - The union of the sets of k most similar items forms 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 overall similarities between all items rated 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 user.

Our item-based CF algorithm employs the context-based similarity metric to compute similarities between visited/rated items. This leads to a recommender system that produces more precise recommendation than that provided by the traditional item-based CF where only one rating is considered in items' similarity computations. We will show how the quality of the recommendation was enhanced by including multiple ratings in calculating items' similarities.

The top N recommended items are subject to adaptation as a next stage of the information retrieval process (Phase II) for generating personalized information content to the user.

B. Performance of the Context-Based Similarity Metric in CF Recommendation

In order to evaluate our framework, we consider the dataset taken from MovieLens web-based recommender system [12] to recommend music compilations. The dataset consists of

100,000 ratings from 943 users on 1,682 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. Then, we divided the resulting dataset 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 10. In addition, we created a dataset that contains each information item and its corresponding components.

For phase I, the quality of our recommender system depends on how accurately the system captures user's preferences as well as its ability to accurately match those preferences with similar users. *Precision* and *Recall* are two metrics commonly used in information retrieval domain. Let *hits* be the total number of recommended items that were really rated by any of the users but were excluded from the training set to be part of the test set; when a set of recommended items is generated for the user, if the rated item in the test set exists in the recommended set, then a hit is recorded, consequently, the number of hits is incremented. Let t be the total number of users in the test set, and N the total number of recommended items. Therefore,

$$precision = \frac{hits}{N} \quad ; \quad recall = \frac{hits}{t} \quad (3)$$

The *F1-metric* [19] is calculated by combining both recall and precision at similar weights is given by:

$$F1 = \frac{2 * recall * precision}{recall + precision} \quad (4)$$

For every neighborhood size, recall and precision were calculated. Fig. 3 shows that by increasing the neighborhood size of an item and using more perspectives in computing similarity between items better recommendation is generated. The abovementioned conclusion is grounded in the observation that when the similarity is based on more perspectives, the item space including all items similar to the target item contains few items that we believe are the ones that mostly interest the individual user.

V. PHASE II: CBR FOR FINE-GRAINED INFORMATION PERSONALIZATION

Functionally, CBR systems involve a case retrieval stage that returns a set of past cases similar to the current one. In our work, we use the set of information items recommended by the CF stage as the set of similar cases (past solutions) retrieved in response to a given user-specified context—i.e. we use the output of the CF stage as input to our case adaptation method, whilst the CF phase serves as the case retrieval stage in a CBR cycle. In phase II, we use CBR as a means to improve the personalized information by applying a compositional adaptation method to further streamline and filter the output of the CF stage. CBR, and in particular compositional adaptation, has been previously applied to personalize information based on a given user profile [2]. Once the top N information items

are retrieved, each item is represented as a case that consists of components characterizing it and compositional adaptation is applied on items' components to selectively collect specific components from multiple items and build the new item/case. This representation is shown in Fig. 4 that also describes the entire CF framework featuring CBR.
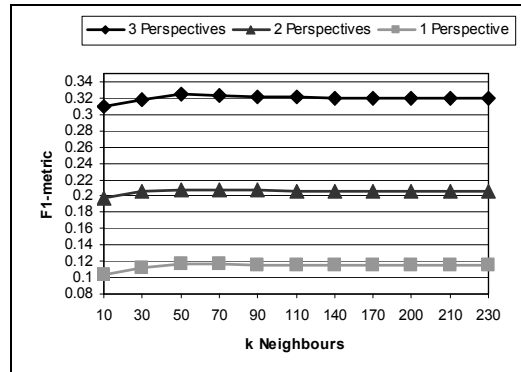

Fig. 3. Effect of context-based similarity and the neighborhood size on the quality of the recommendation

In contrast to other case adaptation methods such as transformational and generative adaptation, compositional adaptation combines solution components from multiple cases to produce a new composite solution; this type is used to fine-tune multiple case adaptation parameters to yield a range of adaptation-driven solution and to modulate similarity assessment parameters to practice varying degrees of inter-case similarities. We argue that one limitation of traditional information filtering approaches is that the information item is presented as a whole, provided it matches some gross relevance criterion.

The rationale for using compositional adaptation is that it allows the selection of parts comprising the whole—those information elements or sub-items that are more related to the user's profile. The efficacy of the compositional adaptation method is inherent in the possibility to select relevant components from multiple similar cases to derive a more focused solution.

A. Solution Adaptation via Compositional Adaptation

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 request and the retrieved case. For instance, if a solution component appears in several similar cases that are retrieved, then there is a high possibility that this component would be part of the final solution. Furthermore, when a solution component of a retrieved case is highly similar to the user's input case, this component will be part of the final solution.
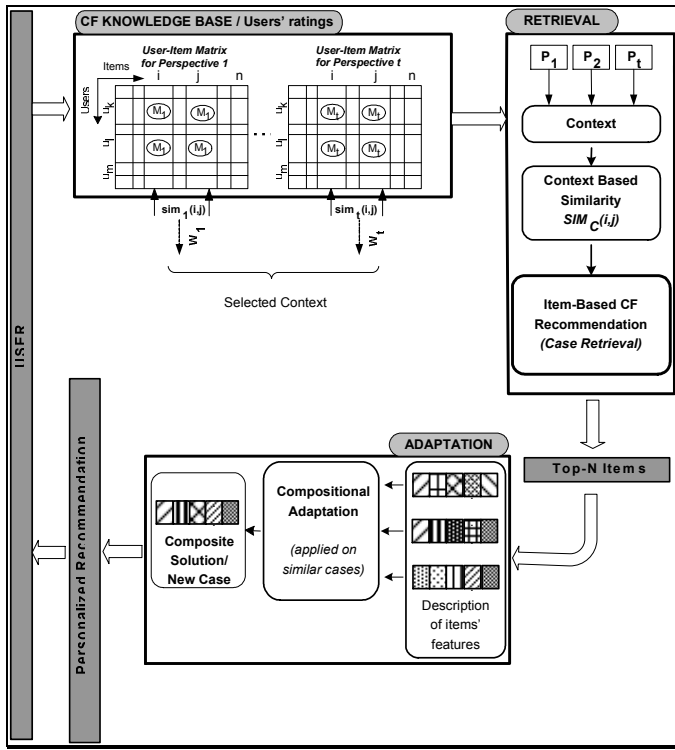
Fig. 4. An illustration of our CF framework with compositional CBR adaptation for personalized recommendation (Components are represented as patterns).

The procedure for our compositional adaptation is described below:

1) Compute the similarity between a retrieved case and a user u, $Sim_i^{R_u}$ computed in (1). Let $S(u,C)$ be the similarity between user u and a similar case C: $S(u,C) = Sim_i^{R_u}$.

2) For every item case C_i, compute the normalized similarity between a retrieved case C_i and the user u over the entire set of RC retrieved cases as follows:

For every user u in the test set,

$$Temp = \sum_{i=1}^{RC} 1 / S(u, C_i) \qquad (5)$$

$$NS(u, C_i) = 1 - (1 / (S(u, C_i) * Temp)) \qquad (6)$$

3) Determine the appropriateness degree of available solution components. Let $Comp_{C_i}$ be a component of a solution from a past case C_i, and $AD_u^{Comp_{C_i}}$ be the appropriateness degree for $Comp_{C_i}$, then

For every user u in the test set,
For i=1 to RC
If $Comp_{C_i}$ exists in the solution of the similar case C_i,

$$AD_u^{Comp_{C_i}} = AD_u^{Comp_{C_i}} + NS(u, C_i) \qquad (7)$$

4) Select the components that are most similar to the initial user's profile. We sort the distinct components by their appropriateness degree, and we select M top components—most similar to user. The M components are amalgamated and one new item is created.

It may be noted that the application of the compositional adaptation method not only takes into account the global similarity between the present and past cases, but is additionally driven by an attribute-level similarity between the current and retrieved previous cases. For information personalization, in a CF framework, this means that the new recommended item derived by combining multiple similar items (Phase II) is an attribute-specific information item that is related to the user's initially provided taste (Phase I).

B. Performance of the Compositional Adaptation Technique

In phase II, the appropriateness degree *(AD)* calculated in (7) is used to measure the efficiency of our compositional adaptation technique. Since *AD* is calculated at a component level, the appropriateness degree of the final recommended item is the average sum of the *AD*s of the M most relevant components. *AD* is calculated for both phases I and II and results are shown in Fig.5-7.
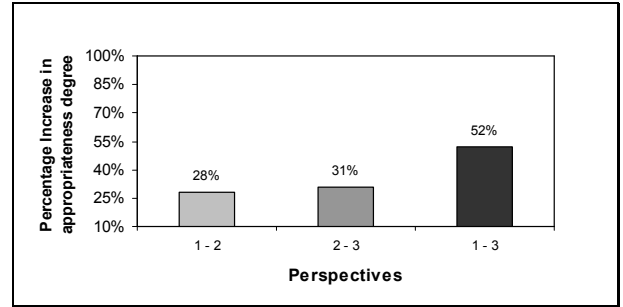

Fig. 5. An illustration of the increase in the appropriateness degree between recommended items and the user from using 1 to 2 perspectives (1-2), 2 to 3 perspectives (2-3) and 1 to 3 perspectives (1-3)
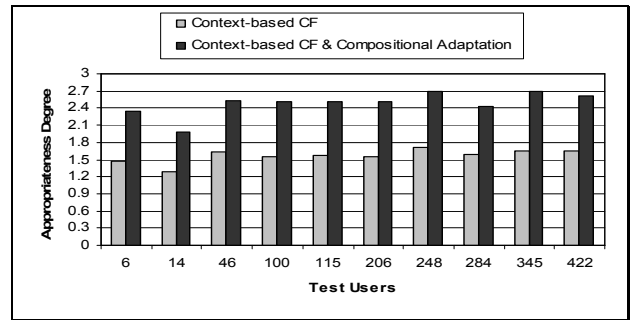

Fig. 6. The appropriateness degree of the composite solution (Phase II) vs. the appropriateness degree of the N recommended items (Phase I). The x-axis refers to the identification numbers of selected users

Our experimental results show that the composite solution is more appropriate and similar to the user's interests—its *AD* is higher than that of the top N recommended items generated in phase I. An average increase of 61% was achieved over all users in the dataset. We conclude that our CBR compositional adaptation has provided the opportunity to pursue personalized

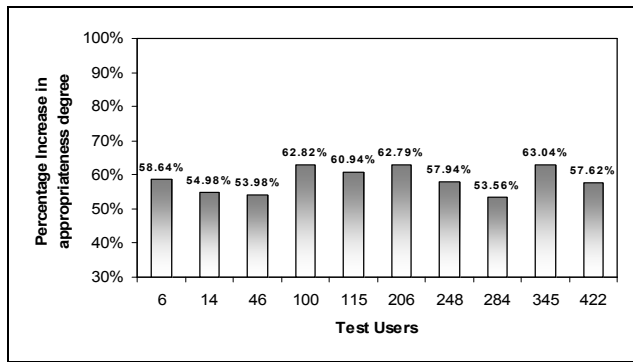information content more focused towards the individualistic needs.

Fig. 7. Impact of the compositional adaptation in generating more focused recommendation. The x-axis shows some selected users' identification numbers, the y-axis denotes the percent increase in the appropriateness degree from Phase I to Phase II.

VI. CONCLUDING REMARKS AND FUTURE WORK

In this paper, we have presented a case for pursuing contextualized CF that is supplemented by the recommendations of both peers and experts/past solutions.

We have presented an intelligent information filtering approach that features a hybrid of item-based CF and CBR methods to achieve more focused and fine-grained personalized information. Our approach could be applied in domains where information items are rated along multiple perspectives and where each item may comprise a set of constituent components; for instance, a book is comprised of chapters, a music compilation of individual songs. We have demonstrated that our framework significantly affected the appropriateness of the final recommendation delivered to the user.

Finally, we believe that our hybrid approach for information personalization has contributed in i) leveraging the CF recommendations by integrating the notion of context and by applying compositional case adaptation, and ii) generating fine-grained personalized information content highly focused to the interests of the user.

We suggest for future work the use of a more systematic approach for quantifying the users' ratings that is based on the Multi-Attribute Utility Theory [21]. For instance, we evaluate initially the overall rating value on every rated item as a weighted addition of its ratings along the multiple perspectives, and finally we compute the similarity between rated items.

REFERENCES

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

[2] 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.

[3] Aguzzoli S., Avesani P., Masssa P., "Compositional CBR via Collaborative Filtering", In *Proceedings of ICCBR'01 Workshop on CBR in electronic commerce*, Vancouver, Canada, 2001.

[4] 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.

[5] Burke R., "A Case-Based Approach to Collaborative Filtering", In *Advances in Case-Based Reasoning*, 5th European Workshop EWCBR 2000, Springer Verlag, New York 2000.

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

[7] Deshpande M., Karypis G., "Item-based Top-N Recommendation Algorithms", In *ACM Transactions on Information Systems (TOIS)*, New York, January 2004.

[8] Dilley R., "*The Problem of Context*", Berghahn Books, New York, 1999.

[9] Geong Y. Y., "An Analysis of Collaborative Filtering Systems", *KMS research paper*, School of Information, University of Texas, Austin, May 2003.

[10] Goker M. H., Smyth B., "Workshop on Case Based Reasoning and Personalization", In *6th European Conference on Case Based Reasoning ECCBR 2002*, Aberdeen, Scotland, September 2002.

[11] Herlocker J. L., Konstan J. A., Terveen L. G., Riedl J. T., "Evaluating Collaborative Filtering Recommender Systems", In *ACM Transactions on Information Systems (TOIS)*, New York, January 2004.

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

[13] 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.

[14] Kangas S., "Collaborative Filtering and Recommendation Systems", *Research Project TTE4-2001-35*, VTT Information Technology, Finland, January 2002.

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

[16] Palme J., "Information Filtering", In *Proceedings of ITS'98 Conference*, May 1998.

[17] 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.

[18] Sarwar B. M., Karypis J., Konstan J. A., Riedl J., "Analysis of Recommendation Algorithms for e-commerce", In *Proceedings of the 2nd ACM conference on Electronic Commerce (EC'00)*, New York, October 2000.

[19] Van Rijsbergen C.J., "*Information Retrieval*",Butterworks, London, 2nd edition, 1979.

[20] Wilke 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.

[21] Winterfeld, D. von and Edwards, W., "Decision Analysis and Behavioral Research", Cambridge, England: Cambridge University Press, 1986.