EXPLOITING AND LEARNING HUMAN TEMPERAMENTS FOR CUSTOMIZED INFORMATION RECOMMENDATION

CHA-HWA LIN AND DENNIS MCLEOD
Computer Science Department
University of Southern California
Los Angeles, CA 90089-0781, USA
{chahwa, mcleod}@usc.edu

ABSTRACT

Human temperaments have been recognized as a predominant factor in determining the activity patterns of human behavior. In our earlier study, a temperament-based filtering method has been proposed to combine concept learning and content-based filtering techniques to incorporate human temperament into the recommendation process of an information system. In this paper, we explain the design of a prototype multiagent system, which is developed, implemented, and experimentally tested by using a group of simulated users generated from the sample users. The notion of human factors, particularly human temperaments, is explored and learned for the representation and segmentation of an information space. Furthermore, the learned temperament concept is employed for the interpretation and measurement of the relevance for classification and recommendation of the information units. The results of our experiments indicate that the accuracy of recommendation using temperament-based filtering exceeds that in content-based filtering. The quality of specific search as well as serendipitous search is enhanced by providing the optimal predictions that are pertinent to not only user interests but also user temperament.

KEY WORDS

Temperament-based information filtering, concept learning, internet search technologies, user modeling, multiagent systems, human factors.

1. Introduction

To cope with the inundating and diverse information space, information customization is increasingly viewed as a critical component of any information access system. Although there are a number of active efforts underway tackling the tasks of adaptive recommendation services for customization, to identify concepts and partition them into appropriate categories have been challenging issues. In the major two techniques of information filtering, the most popular approach is content-based filtering which depends only on the term features of the vector space model without considering any concept or context [1]. The problems encountered in content-based filtering are, for example: items cannot be filtered based on quality, style or point-of-view; information units such as image, audio, video, art or physical items cannot be fully featured by the attached captions; and users may not known their exact interest in order to request relevant information [2]. Although social filtering draws on user evaluations or opinion, no simple conceptual representation is provided to interpret the meaning of the clusters.

Human temperaments have been identified as a predominant factor of the patterns of human behavior by psychologists [3]. In addition, neuroscience research indicates that temperament is an innate property of the brain [4]. The inherent inter-related patterns between user temperaments and user interests may lead to a better understanding of user personality and improve the service of an information system. Consequently, temperament-based filtering was proposed in our earlier paper to incorporate the concept of human temperaments into the filtering process of an information recommendation service [5]. We hypothesized that the accuracy of an information recommendation system can be improved by employing human factors, particularly human temperament, for filtering and customization. However, the potential of the temperament-based filtering method has not been reported. In this follow-up study, a prototype system is developed, implemented and experimentally tested to demonstrate the effectiveness of the proposed temperament-based filtering approach. A user-studies testing is conducted on a web site and a simulation model is employed. Furthermore, varied heuristic selection rules are applied by the filtering agent when it interacts with the simulated user under different task situations and the results outperform content-based filtering.

In what follows, we describe the system architecture, provide the criteria for user simulation, explain the mechanism of the agents for learning, segmentation, clustering, and classification, describe the metrics for performance measurement, present our experimental testbed, and report results from the experiments. Finally, we offer brief concluding remarks and future research directions.

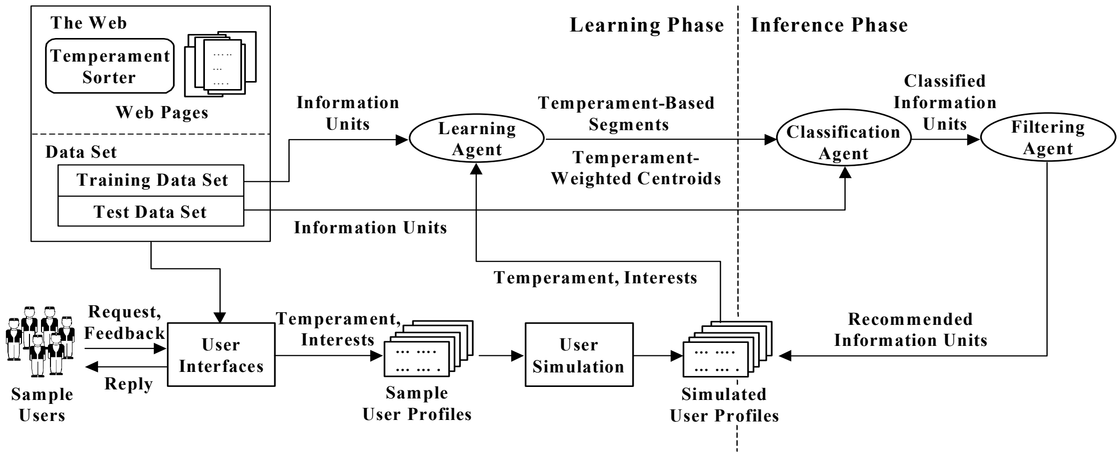


Figure 1. Overview of the system architecture.

2. System Design

The functional architecture of the prototype system is shown in Figure 1. The implementation of the system has two key phases: learning and inference. In the learning phase, a user-studies testing is conducted on a web site. The user interfaces serve as the front end for the end-user to enter ratings (two-level scale), requests (temperament, interest key terms) and feedback (change of ratings) to the system. On the other hand, the interfaces are for the system to monitor and intercept user inputs, and to deliver and present the recommended information units to the user. A simulation model is constructed to generate simulated user profiles based on the user profiles obtained from the sample users. A user profile specifies user temperament and preferences for a particular user. The learning agent learns the temperament concept to segment the information space formed by the training data set after observing the temperament and interest distributions of the simulated users. The information units in a segment are further grouped into clusters by intra-segment similarities and the centroid temperament weights of the clusters are estimated. The learned temperament concept of the temperament-based segments and temperament-weighted centroids is then applied by the classification and filtering agents to classify and recommend information units in the test data set to the simulated users in the inference phase.

2.1 User Simulation

To better measure the effectiveness of the temperament-based filtering mechanism, a huge amount of data could potentially be required. We have therefore chosen to study the behavior of the mechanism with the help of user simulation based on the sample user profiles recorded in the sample survey. To enhance the quality of the behavioral analogy, a simulated user profile of a specific temperament type is built from a set of three randomly selected source user profiles of that temperament type. A simulated user query vector was constructed by adding all the vectors of the information units in a simulated user profile.

2.2 Learning the Temperament Concept

The learning agent learns to estimate and partition an information space by observing the temperament and interest distributions of the simulated users. The knowledge of temperament concept is represented by a set of temperament-based segments in addition to temperament-weighted centroids.

2.2.1 Temperament-based Segments

In the temperament-based filtering model, the information space is partitioned by the intersection of Keirsey's four temperaments. David Keirsey [6] asserted that people can be classified into four temperaments as SJ - sensing and judging, SP - sensing and perceiving, NT - intuiting and thinking, and NF - intuiting and feeling. A statistic report [7] on the percentage distributions of the temperaments in the United States showed that most people are SJs (46.7%) compared with SPs (21.4%), NTs (16.1%), and NFs (15.8%). The segment of an information space is defined as

$$S_n = \bigcap_{t \in T_n} D_t - \sum_{t' \in (T - T_n)} D_{t'}$$

where $n = \{1, 2, \ldots, 2^{|T|}\} = \{1, 2, \ldots, 16\}$, $T = \{SJ, SP, NT, NF\}$, $|T|$ = size of T, $T_n = \{t \mid t$ is a temperament value in segment $S_n\}$, D_t = the set of information units evaluated as "like" by users with temperament t, and $D_{t'}$ = the set of information units evaluated as "like" by users with temperament t'. A sample temperament-segmented information space is shown in Figure 2.

A greater popularity indicates the information unit has a higher probability to be liked among that user population. The popularity of an information unit k within a user population of a particular temperament type t is defined as the conditional probability of interested users

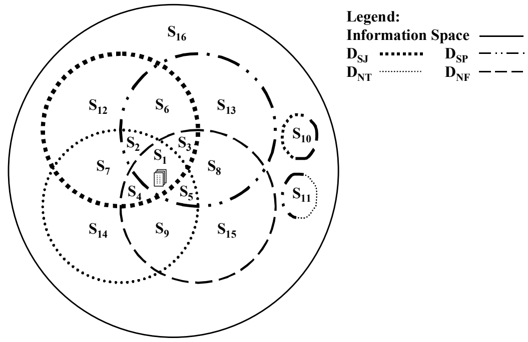


Figure 2. Segments of a sample information space.

given that temperament type, $P(like_k|t)$. In learning the temperament concept to segment the information space, an information unit k of the training data set is not classified into D_t until $P(like_k|t)$ exceeds a predefined threshold θ to maintain a confidence level that the system believes in the learned knowledge. Table 1 shows an example of the statistical results obtained for 5 information units in the training data set rated by 1000 simulated users, where "yes" indicates "like" and "no" indicates "dislike". The popularity of *d3* is zero for all the temperament types, thus *d3* is classified into $S_{16} = D_\phi$ in which an information unit is not evaluated as "like" by any user. On the other hand, three popularity values of *d9* are nonzero: $P(like_{d9}|SJ) = (97/(97+370)) \approx 0.21$, $P(like_{d9}|SP) = (14/(14+200)) \approx 0.07$, and $P(like_{d9}|NT) = (51/(51+110)) \approx 0.32$. Assume that threshold $\theta = 0.10$, then *d9* is considered liked by SJs and NTs but not SPs or NFs, because only $P(like_{d9}|SJ)$ and $P(like_{d9}|NT)$ exceed θ. Hence, *d9* is an element in both D_{SJ} and D_{NT} but not D_{SP} or D_{NF}, and is classified into S_7 where $S_7 = D_{SJ} \cap D_{NT} - D_{SP} - D_{NF}$ by definition. The knowledge of the segment concept is represented by a set of ordered pairs (item id, segment #), such as ($d3$, S_{16}) and ($d9$, S_7).

Item	SJ		SP		NT		NF	
Id	Yes	No	Yes	No	Yes	No	Yes	No
d3	0	467	0	214	0	161	0	158
d9	97	370	14	200	51	110	0	158
d15	66	401	50	164	92	69	57	101
d23	23	444	57	157	46	115	0	158
d36	24	443	59	155	95	66	68	90

Table 1. Example of the statistical results obtained for some information units in the training data set.

2.2.2 Temperament-weighted Centroids

To reduce the size of comparisons when searching within a segment, the information units in the same segment are clustered by content-based approach that adopts conventional vector space model. Both information units and user requests are represented as term vectors in an n-dimensional space. Each term weight TW_i of a term m_i in vector d is computed by TF-IDF (Term Frequency and Inverse Document Frequency) method and is often defined as

$$TW_i = \frac{TF_i \times IDF_i}{\sqrt{\sum_{j=1}^{n} TF_j^2 \times IDF_j^2}}$$

where TF_i = the number of times term m_i appears in information unit d, $IDF_i = \log_2(n/DF_i)$ = the inverse document frequency, DF_i = the number of information units in the collection which contain m_i, and n = the number of information units in the collection. The commonly used cosine similarity of any two vectors V_i and V_j is denoted simply as $Sim(V_i, V_j) = (V_i \bullet V_j) / (|V_i| \times |V_j|)$. A higher cosine value indicates a greater similarity. The basic idea of clustering is to compare a new item with the centroid vector of a cluster and group it into the cluster if the similarity measure is greater than a fixed threshold λ. The centroid vector of a cluster is defined as the mathematical average of the information vectors in the cluster. To keep the number of items neither too large nor too small and without losing representative characteristics in a cluster, λ was set to 0.07. In content-based filtering, with a collection of 2,000 information units in the database, around 200 clusters were generated with an average of about 10 units each. The same λ value was applied in temperament-based filtering when given the user interest key terms, otherwise λ was set to zero when only the segments rather than the clusters were concerned.

The temperament weight w_j of an information unit d_j in segment S_n is set to the fraction of relevant temperaments in S_n multiplied by the the prior probability $P(like_{nj})$ that an information unit d_j in segment S_n is evaluated as "like". Given the probabilities of the temperaments $P(t)$ in the training data, then w_j is just

$$w_j = \frac{|T_n|}{|T|} P(like_{nj}) = \frac{|T_n|}{|T|} \sum_{t \in T_n} P(like_{nj}|t) P(t)$$

where n, T_n, T, and $|T|$ are as defined previously, $|T_n|$ = size of T_n, $P(like_{nj}|t)$ = the conditional probability that an information unit d_j in segment S_n is evaluated as "like", given user temperament t, and $P(t)$ = the percentage distribution of temperament t. The centroid temperament weight e_i of a cluster c_i is the mathematical average of the temperament weights of the information units in the cluster and is given by $e_i = \frac{1}{m} \sum_{j=1}^{m} w_j$ where m = size of cluster c_i and w_j = the temperament weight of an information unit d_j in cluster c_i.

2.3 Inference in the Classification Agent

The assumption underlying temperament-based filtering is that a new information unit having content features similar to that of a widely liked set of information units by a particular group of people is probably liked by that group of people. The classification agent of the temperament-based filtering employs the learned temperament concept to infer the maximum likelihood estimate of the target segment and cluster (S_{target}, C_{target})

of an information vector V_k, for which the popularity similarity $Pop(V_c, V_k) = e_c\, Sim(V_c, V_k)$ between V_k and a centroid vector V_c with temperament weight e_c is maximum. Hence, $(S_{target}, C_{target}) = \arg\max_{s \in S, c \in C_s} Pop(V_c, V_k)$.

2.4 Performance Metrics

One of the main purposes of this work is to study the relative efficiencies of the filtering methods so that proper choices can be made in various practical contexts. We have decided to use percentage accuracy [8] as the performance measure in the experiments. The percentage accuracy is the fraction of positive user feedback in the top ten items recommended by the system. Each filtering method is applied to the training data set and evaluated according to how frequently it recommended the information units taken by the user, particularly for the top 10 items recommended to the user, from the separate test data set.

3. Experimental Environment

The experiments are designed to explore user preferences by accessing the relationships between personality interests for user-customized selection of on-line communication in the document domain.

3.1 Data Collection and Modeling

To populate the database, 2000 hyperlinks to the web pages of the referenced articles were collected from 12 news web sites that contain articles about high technology including cnet (www.cnet.com), Information Week Online (www.iweek.com), PC Week (www.zdnet.com/pcweek), etc. The actual contexts of the referenced articles were obtained and stored in the database by following the hyperlinks. Each article is an information unit. From the sample collection, a fixed dictionary of 16474 terms (word stems) was obtained after Porter algorithm [9] was applied in deleting 566 common words from a standard stop list and removing suffixes. The TF-IDF method was then applied to compute term weights. To avoid over-fitting and reduce memory and communication load, only the 100 highest weighted terms were kept in the vector representations of the information units and user profiles. Recent experiments have shown that using too many words leads to a decrease in performance when classifying web pages using supervised learning methods. The optimum is between 30 and 100 [10]. The produced document vectors formed the sample data set.

In order to obtain statistically significant estimates of filtering performance, the sample data was separated into training data set and test data set in 20 possible ways. Each filtering method was then applied to each training data set and evaluated on the associate test data set. The results of these 20 experiments were averaged. In a pilot test, the sample data was split into 40 groups. The results of the larger groups were similar to the 20 groups, but doubled the cost of storage space and CPU time.

3.2 User Interfaces, Modeling, and Profiles

To evaluate the experimental prototype, user-studies testing on the document domain was conducted between February 9, 2000 and February 16, 2000. The most difficult tasks encountered were to find users with the desired temperament type, especially NT and NF, and the quantity of users willing to test the system. Among the 28 adult users participated in the experiment, there were 18 SJs, 4 SPs, 3 NTs , and 3 NFs. The graphical user interfaces were constructed on a web site (available at http://www.usc.edu/dept/cs/tfiltering/infoworld), for the user to access the document information repository. Each user had a user profile. The user could search the database either by the author name or the article title. In addition, a user could directly browse the article titles as well. Each article title was linked to the original article at a web site. The users were encouraged to pick up at least 200 articles of interest. User preferences are recorded over time as the user picked up more articles. These records constituted the source user profiles learned. A user profile (Figure 3) kept track of the index of each article that was interesting to a particular user. At any point of a search session or any later revisit session, the user could choose to pick up more articles of interest or remove the unwanted articles from the user profile. The user profile was designed to accommodate the evolving structure.

The users who participated in the experiments were adults living in the United States and mostly USC students. Based on the source user profiles, a group of 1000 simulated user profiles were generated by the criteria described in the previous section. By taking into account the statistic results on the percentage distributions of the temperaments in the United States, the simulated users consist of 467 SJs, 214 SPs, 161 NTs and 158 NFs. In an attempt to faithfully reflect the properties in the sample data, a simulated user had 200 articles of interest in the simulated profile as that was the number of articles encouraged to have for a sample user.

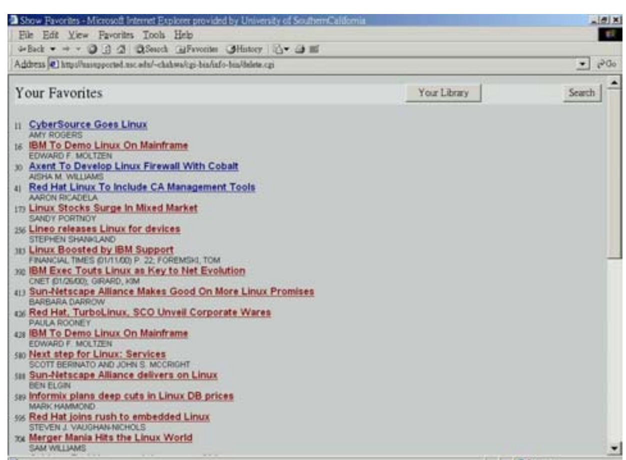
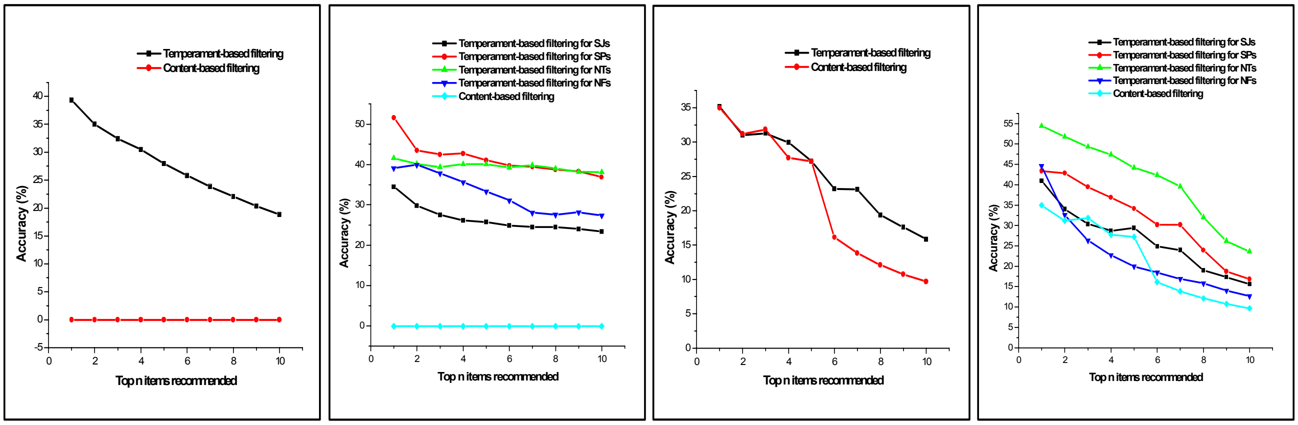


Figure 3. Partial user favorites from a user profile.



(a) **(b)** **(c)** **(d)**

Figure 4. Accuracy of recommendation: temperament-based filtering vs. content-based filtering. (a) Users with unknown temperament and interest. (b) Given the user temperament. (c) Given the user interest key terms. (d) Given the user temperament and the user interest key terms.

4. Performance Evaluation

In order to explore how well the temperament-based filtering method can learn to automatically and adaptively recommend articles to the user, using the user profiles of user temperament and the interest key terms, a series of experiments were conducted. The performance of the method was evaluated in comparison with the content-based filtering method.

4.1 Case 1: For Users with Unknown Temperament and Interest

To help the user find interesting information units in a serendipitous search, temperament-based filtering is viable and productive by offering the information units in segment S_1 to the user (Figure 2). Such filtering strategy is based on the assumption that S_1 contains all those items evaluated as "liked" by everyone in the user population and would probably liked by the new user. In contrast, content-based filtering fails to predict any user behavior as the method requires the user to provide what they want exactly.

The recommendation process consisted of fixing a threshold θ in popularity $P(like_k|t)$ to achieve a certain confidence level for the segmentation. As a higher threshold was applied the accuracy was likely to increase before the over-fitting effect appeared. By ignoring the effect of initial bias, $\theta = 0.10$, instead of 0.05, was used as the base threshold by the temperament filtering method for the remaining analyses of the experiments.

A comparative accuracy of the two methods is shown in Figure 4a. Temperament-based filtering obtains an average accuracy of 39.33% in predicting the top 1 item the user would be interested in, 35.04% in predicting both the top 2 items the user would be interested in, and so on. The gradual downward progression indicates that the accuracy decreases as the difficulty to predict more items increases. In comparison to the content-based filtering which is incapable of operating under this situation, the temperament-based filtering method is robust and suggestive.

4.2 Case 2: Given the User Temperament

In stead of doing a blind navigation, the user is assisted by a ranked list of information units, which matches the temperament of the user by the temperament-based method. The filtering agent recommends the classified testing information units by searching the partial space, which contains that user temperament in its segments. The performance accuracy is estimated for the four human temperaments, SJ, SP, NT, and NF, respectively. However, content-based filtering requires user query to make any further process and no recommendation is observed.

To investigate the heuristic search rule employed by the temperament-based method, the filtering agent made two searches: one tracked only the upper quartile segments in the partial space and the other exhausted all the segments in the partial space. The results showed that the performances of the two searches are almost the same. However, the heuristic rule of searching the upper quartile segments would reduce the operation cost significantly. As illustrated in Figure 4b, temperament-based filtering outperforms content-based filtering when given the user temperament. For users of SP temperament, temperament-based filtering even achieved an average accuracy of 51.64% in predicting the top 1 item the user would be interested.

Interestingly, the percentage accuracy is generally increased when user temperament is provided than when no user profile is involved. For all the top 10 items recommended to the SP or NT users that are relevant to their interests are between 36.90% and 51.64% accuracy, much better than that in case 1, between 18.86% and 39.33. Although the filtering accuracy for SJ users is not that promising, the accuracy for NF users falls between 27.43% and 39.97%, which is improved when compared with that in case 1.

4.3 Case 3: Given the User Interest Key Terms

To evaluate the performance, a simulated user profile that served as the user interest key terms was transformed into a term vector representation and compared with the centroid vectors of the clusters in the information space for both temperament-based and content-based methods. The information units in the cluster of the highest similarity are presented to the user. To avoid overly weighing temperament, cosine similarity measure is used in both methods for processing the relevant search. As illustrated in Figure 4c, temperament-based filtering has a better overall performance that for the top 10 items recommended it maintains a better accuracy ranging between 15.85% and 35.18%, while content-based filtering ranges between 9.68% and 34.95%.

4.4 Case 4: Given the User Temperament and the User Interest Key Terms

In this case, the strategies applied in cases 2 and 3 are combined. The filtering agent of the temperament-based method searches only the upper quartile segments in the partial space, which contains that user temperament. The cosine similarity measure is used in both methods for the relevant search. From the resulting percentage accuracy of the two methods (Figure 4d), the accuracy of temperament-based filtering for SJs, SPs, or NTs in predicting all the top 10 items exceeds that in content-based filtering respectively, while the accuracy for NFs is about the same as that in content-based filtering. This suggests that when adapting information filtering to user temperament as well as interest key terms, more than 85% of the user population would be better satisfied with the recommendations provided by temperament-based method than by content-based method.

5. Conclusion and Research Directions

To be empowered with the knowledge of human temperaments, temperament-based filtering method increases the accuracy of the recommendation service by providing information units that are pertinent to not only user interests but also user temperament. The segmentation and classification mechanism enhances the quality of specific search as well as serendipitous search by providing the adaptive optimal predictions. In addition, the system effectiveness is improved by heuristically searching the partial structurally classified space that matches user temperament. Existing keyword-based document and information retrieval systems (e.g., web search engines) have limitations in their ability to satisfy the user in recognizing user personalities. Temperament-based information filtering presents a general approach to incorporating human temperaments into the optimization of recommendation process. We have developed and experimentally tested the temperament-based filtering mechanism to manage conceptual cohesiveness between the disordered information space presented by a computing system and the user's recognition of the real world. The collaborative agents can learn to classify and filter the information units by utilizing the features of human temperaments and observing user behaviors. The performance improvement reflected in the results of this study demonstrates the feasibility of incorporating human factors into an information recommendation process.

While the experimental results are positive, the filtering algorithms focus on temperament, not with other characteristics of human factors, such as gender, age, education level, experience with system, user demographics, and different cultural features. These problems will be the subject of future research. Currently, we are testing the performance of the temperament-filtering method on an art image domain and the representation styles.

References

[1] G. Salton and M.J. McGill, *Introduction to modern information retrieval*, (New York: McGraw-Hill, 1983).

[2] U. Shardanand and P. Maes, Social information filtering: algorithms for automating 'Word of Mouth', *ACM CHI '95 Proc. on Human Factors in Computing Systems*, Denver, CO, USA, 1995, 210-217.

[3] C. Jung, *Psychological types*, (New York: Harcourt Brace, 1923).

[4] Simonov, P. V. and P. M. Ershov, *Temperament, character, and personality: Biobehavioral concepts in science, art, and social psychology*, (Philadelphia, PA: Gordon and Breach Science Publishers, 1991).

[5] C. Lin and D. McLeod, Temperament-based information filtering: A human factors approach to information recommendation, *Proc. of the IEEE International Conf. on Multimedia & Exposition*, New York, NY, USA, 2000.

[6] D. Keirsey and M. Bates, *Please understand me*, (Del Mar, Ca: Prometheus Nemesis Books, 1978).

[7] A. L. Hammer and W.D. Mitchell, The distribution of MBTI types in the US by gender and ethnic group, *Journal of Psychological Type*, *37*, 1996, 2-15.

[8] L. Chen, and K. Sycara, WebMate: A personal agent for browsing and searching, *ACM AGENTS '98 Proc. of the International Conf. on Autonomous Agents*, Minneapolis, MN, USA, 1998.

[9] M. Porter, An algorithm for suffix stripping, *Program*, 14(3), 130-137.

[10] M. Balabanovic, An adaptive web page recommendation service, *ACM AGENTS '97 Proc. of the First International Conf. on Autonomous Agents*, Marina Del Rey, CA, USA, 1997.