TEMPERAMENT-BASED INFORMATION FILTERING: A HUMAN FACTORS APPROACH TO 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

This paper provides an intelligent multiagent approach to incorporate human temperaments into the filtering process of an information recommendation service. Our approach is to devise a new filtering mechanism, which addresses segmentation, learning, classification, and filtering techniques based on Keirsey's temperament theory, probability theory, the distributions of temperaments, and statistical reasoning. By presenting information units that are consistent with user interests as well as user temperament, the accuracy and precision of the recommendation service may be improved.

1. INTRODUCTION

The rapid proliferation of on-line digital information is changing our everyday life. New techniques are needed to assist people in managing this potentially overwhelming and diverse digital information space. The most popular approach is content-based filtering [2][3][4] which draws on the vector space model of conventional textual information retrieval [12]. However, such keyword-based systems (e.g., web search engines) usually require the user to know what they want exactly and are not suitable for serendipitous search. In addition, multimedia information units cannot be fully featured by key terms or filtered on quality or style [14]. Although social filtering [3][14] addresses some of the limitations of content-based filtering by clustering information by user evaluations or opinion, no predefined concept classes are used to describe or simplify the meaning of the clusters. Moreover, none has taken human temperaments into account in the classification or filtering process.

Psychologists have identified human temperament as a predominant factor in the patterns of human behavior [7][9][11]. Neuroscience research indicates that temperament is an innate property of the brain [15]. Some of the studies have confirmed the relevance of temperament to the public tastes in perceiving or interpreting the information in general, such as art, music, and literature [5][8][10]. The potential for employing human temperament as an effective information filtering technique is strong.

We hypothesize that the accuracy (precision and recall) of an information recommendation system can be significantly improved by employing user temperament for filtering and customization. Temperament-based information filtering model (Figure 1) proposes a solution to characterize the information space by taking human factors, particularly human temperament, into consideration, which may categorize the information space into meaningful classes and improve recommendation service. The model has two key phases: learning and inference. In the learning phase, a novel segmentation framework is introduced to estimate and partition an information space into temperament-based segments by observing the temperament and interest distributions of the sample users. The analysis agent analyzes the data obtained by the user-studies testing. A learning agent is designed to automatically learn the temperament segmentation concepts. The learned knowledge is then applied to infer optimal target segment and cluster values of information units by classification and filtering agents. Furthermore, user profiles are updated by the monitor agent and heuristic selection rules are integrated with the filtering process to facilitate information recommendation when dealing with user requests to the database and user feedback for relevance refinement.

In what follows, we introduce temperament terminology, provide the mechanism for segmentation and learning, and present the design for classification, filtering, and feedback refinement. Finally, future work and conclusions are described.

2. HUMAN TEMPERAMENTS

Carl Jung [7] asserted that people are fundamentally different and can be classified into "psychological types" based on their particular preferences. The four pairs of opposite preferences are Extraverting (E) and Introverting (I), Sensing (S) and iNtuiting (N), Thinking (T) and Feeling (F), and Judging (J) and Perceiving (P). Within each pair of opposite preferences, a person leans toward one or the other in most cases. Later, Katherine Briggs and Isabel Myers [11] adopted Jung's theory and designed the Myers-Briggs Type Indicator (MBTI), a questionnaire for helping people to identify their innate personality types. David Keirsey [9] correlated his theory into the MBTI system and classified the sixteen personality types into four temperaments as SJ, SP, NT, and NF. Each related MBTI type has the two temperament letter codes in it. A statistic report [6] on the percentage distributions of the temperaments in the United States showed that most people are SJs (46.7%), sensing and judging (Table 1).

From the temperament point of view, an information space in the real world can be specified as having a temperament-segmented structure by identifying the interdependent feature of human temperaments and preferences.

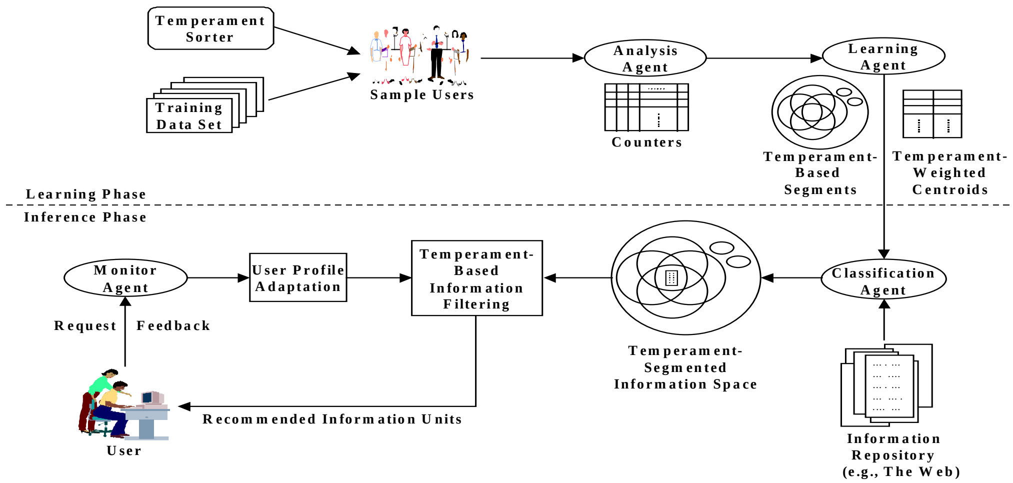

Figure 1. The Architecture of the Temperament-Based Information Filtering Model.

Temperament (%)	MBTI (%)			
SJ 46.7	ESTJ 9.9	ESFJ 9.6	ISTJ 15.6	ISFJ 11.5
SP 21.4	ESTP 4.8	ESFP 5.7	ISTP 6.4	ISFP 4.5
NT 16.1	ENTJ 2.8	ENTP 4.7	INTJ 3.5	INTP 5.2
NF 15.8	ENFJ 2.5	ENFP 6.3	INFJ 2.6	INFP 4.3

Table 1. The Percentage Distributions of the Temperaments in the United States.

3. SEGMENTATION AND LEARNING

The temperament concept is represented by a set of temperament-based segments and centroids of clusters in a sample information space. A temperament-segmented sample information space is used to estimate the temperament-based segmentation of a target information space in the inference phase.

3.1 Segmentation Function

The segmentation function is introduced to model the temperament-based segments of the sample information space and is formulated by Keirsey's temperament theory and probability theory. Keirsey's four temperaments form the four concepts, SJ, SP, NT, and NF. The same terms are used interchangeably to indicate either temperaments or concepts in this paper for convenience. In an information space, each information unit has two possible target values, like and dislike. A concept t of an information space is defined as the set of all the information units evaluated "like" by users with temperament t. An information unit evaluated as "like" or "dislike" by a user with known temperament, will fall into one of the 16 segments produced by all the possible intersections of the four temperaments. For simplicity, consider only the positive examples that information units are evaluated as "like" by users.

The segment of the information space partitioned by the intersection of the four temperaments may be 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'. For example, $S_1 = D_{SJ} \cap D_{SP} \cap D_{NT} \cap D_{NF}$, $S_2 = D_{SJ} \cap D_{SP} \cap D_{NT} - D_{NF}$, ..., and $S_{16} = D_\phi$, where D_ϕ = the set of information units not evaluated as "like" by any user. The relevant temperaments of these segments are $T_1 = \{SJ, SP, NT, NF\}$, $T_2 = \{SJ, SP, NT\}$, ..., and $T_{16} = \phi$.

To eliminate bias and false alarms in the training data set, an information unit is learned to be in one of the segments except S_{16} only when sufficient evidence has been observed to pass a preset threshold; otherwise, the information unit is retained in S_{16}. In addition, to search systematically, segments are sorted into descending order by their accumulated percentage distributions of temperaments, $\sum_{t \in T_n} P(t)$. Thus, a segment with a lower index will contain information units having a higher accumulated probability (or popularity).

3.2 Temperament Weight

It is obvious that an information unit in a segment with a higher prior probability indicates the likelihood of a larger interested population. However, consider the situation that a segment consists of more temperament types than the other segments especially when the information units in these segments have the same prior probability. The information units in the segment consisting of more temperament types exhibit more human diversity in the interested population and thus may have heavier weights. To cope with such observations, temperament weight is introduced to quantify and estimate the relative influence of various temperaments on the popularity measure of an information unit. The temperament weight of an information unit d_j in segment S_n can be set to the fraction of relevant temperaments in S_n multiplied by the prior probability $P(like_{nj})$ as

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

where n, T_n, T, and $|T|$ are as defined previously, $|T_n|$ = size of T_n, and $P(like_{nj})$ = the prior probability that an information unit d_j in segment S_n is evaluated as "like". Since temperaments are mutually exclusive with $\sum_{t \in T} P(t)$ = 1, by the theorem of total probability, the above equation can be rewritten as

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

where $P(like_{nj}|t)$ = 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.

Within a segment, the information units are grouped into clusters by conventional cosine similarity measure and the centroid vector of a cluster is defined as the mathematical average of the information unit vectors in the cluster. The centroid temperament weight e_i of a cluster c_i may be defined in the same manner as the mathematical average of the temperament weights of the information units in the cluster and thus 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. Thus, a centroid temperament weight is an estimate of the popularity within the user population for the information units in a cluster.

The learned estimate of the centroid temperament weights as well as cosine similarity measure provides a set of probabilities and numerical measures in the following classification process to predict into which cluster of which segment, a new information unit is likely to fall.

4. CLASSIFICATION

The assumption underlying the classification process 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. Also, by observing the fact that a higher centroid temperament weight indicates the information units in that cluster have a greater popularity among the user population, the centroid temperament weight is considered as an important factor when formulating the measure strategy to classify a new information unit.

4.1 Popularity Similarity Measure

The popularity similarity measure is to estimate the level of importance of both popularity and similarity between a new (target) information unit vector V_k and a centroid vector V_c of a temperament-classified cluster c. Thus the measure relies not only on the traditional cosine similarity $Sim(V_c, V_k)$, but also on the temperament weight e_c of V_c. This lead to the definition of the popularity similarity between V_c and V_k be expressed by

$$Pop(V_c, V_k) = e_c Sim(V_c, V_k)$$

4.2 Heuristic Function

The proposed approach to classifying a new information unit into a temperament-based information space is to assign the optimal pair of segment and cluster based on the popularity similarity measure. The heuristic function of the optimal pair is the maximum likelihood estimate of the target segment and cluster, (S_{target}, C_{target}), for which the popularity similarity $Pop(V_c, V_k)$ is maximum, as follows:

$$(S_{target}, C_{target}) = \underset{s \in S, c \in C_s}{\arg\max} Pop(V_c, V_k)$$

where S = {s | s is a segment of a temperament-based information space} and C_s = {c | c is a cluster of segment s}.

The location of the new information unit can be adapted by observing user feedback for that unit. The classification agent also maintains a set of cluster indicators to be used in the filtering process for heuristic selection. A cluster indicator is the nearest neighbor of the centroid vector of that cluster. In contrast with that a centroid is an artificial vector, a cluster indicator is a real information unit.

5. FILTERING

In the temperament-based filtering, user temperament information is used to restrict the search space and the centroid temperament weights are added to the similarity measures to guarantee conceptually meaningful solutions. The filtering agent is to recommend the user a ranked list of information units, which best match not only user interest key terms, but also user temperament. If a user did not give enough descriptions about user temperament or interest key terms, the system will utilize the user profile. A user profile is a record of user behaviors, which include temperament, interest key terms, and feedback of a particular user.

The heuristic selection rules employed by the filtering agent are:

- When given the user temperament, the agent searches only the partial space, which contains that user temperament in its segments.
- When given the user interest key terms, the agent selects items from the optimal target cluster by applying the same heuristic function used in the classification process.
- For a user with unknown interest, the agent collects and returns the most popular items or the cluster indicators of the segments to the user.
- For a user with unknown temperament and interest, the agent returns the information units in the segment with the lowest index where the information units are considered interested by any temperament of people.
- The agent searches only the upper (higher popularity) quartile segments in the information space at quick search mode.

These selection strategies may eliminate redundant similarity measures and simplify the filtering process.

6. FEEDBACK REFINEMENT

The task of feedback refinement addresses the issue of adapting the recommendation to user responses in order to provide a satisfactory recommendation service. For this, the Ide Dec-Hi method [1][13] is adopted and modified. This method adds all relevant document term weights directly to query terms but subtracts only the top-most non-relevant document terms. Although the conventional relevance feedback techniques, Standard Rocchio, Ide Regular, and Ide Dec-Hi, are considered having similar retrieval quality, the Ide Dec-Hi is computationally very efficient and outperforms the others in some cases [13]. However, Aalbersberg [1] pointed out that negative relevant feedback can be omitted without significantly detracting from the retrieval quality. Hence, the modified Ide Dec-Hi method considers only the relevant information units and the modified interest vector is reformulated as

$$V_{k+1} = V_k + \sum_{V_r^k \in D_r^k} V_r^k$$

where V_k = the original interest vector, D_r^k = the set of relevant information units retrieved for interest vector V_k , and V_r^k = a relevant information unit in D_r^k.

The agent utilizes the modified Ide Dec-Hi method to refine the search results when given feedback by the user.

7. FUTURE WORK AND CONCLUSIONS

In this ongoing research, we are experimenting with a prototype system to evaluate the proposed temperament-based filtering approach in comparison with other filtering methods, such as content-based filtering. An experimental user-studies testing procedure will demonstrate the improved effectiveness, by utilizing a combination of simulation and user-studies testing. The prototype information recommendation system we have been developing will be applied to and demonstrated in at least two application domains - an art image collection and a document collection.

We expect this study to provide a new general approach to incorporating human factors, particularly human temperament, into the adaptive information recommendation process. This temperament-based filtering method presents an effective framework for the analysis of the inherent interrelated patterns between user temperaments and user interests, and the modeling of the internal representations for partitioning the diverse information space into meaningful segments. The mechanism for the inference of classifying new information units into the temperament-based information space is also proposed. This classification is built upon known concepts of human factors to simplify system configuration and implementation. In addition, the technique for the optimization of exploring information units by considering the critical dimension of human temperament provides a basis for further exploiting predefined concepts of other human factors or environment properties in the services of the computing systems. The system effectiveness is improved by searching only the part of the structurally classified information space that matches user temperament and the quality of the recommendation service is enhanced by returning information units that are relevant to user interests as well as user temperament. The accuracy of the information system may be increased and the goal to better satisfy the user may be achieved.

8. ACKNOWLEDGEMENTS

This research has been funded in part by the Integrated Media Systems Center, a National Science Foundation Engineering Research Center with additional support from the Annenberg Center for Communication at the University of Southern California and the California Trade and Commerce Agency.

9. REFERENCES

[1] Aalbersberg, I. J., "Incremental Relevance Feedback," in Proc. ACM SIGIR Conference, pages 11-22, 1992.

[2] Armstrong, R., D. Freitag, T. Joachims and T. Mitchell, "WebWatcher: A Learning Apprentice for the World Wide Web," Proc. AAAI Spring Symp. on Information Gathering from Heterogeneous, Distributed Resources, Stanford, CA, 1995.

[3] Balabanovic, Marko, "An Adaptive Web Page Recommendation Service," in ACM AGENTS '97, Proceedings of the First International Conference on Autonomous Agents, Marina Del Rey, CA, 1997.

[4] Chen, Liren, and Katia Sycara, "WebMate: A Personal Agent for Browsing and Searching," in ACM AGENTS '98, Proceedings of the International Conference on Autonomous Agents, Minneapolis, MN, 1998.

[5] Evans, Joan, "Taste and Temperament: A Brief Study of Psychological Types in their Relation to the Visual Arts," Hyperion Press, reprint edition 1979.

[6] Hammer, Allen L. and Wayne D. Mitchell, "The Distribution of MBTI Types in the US by Gender and Ethnic Group," Journal of Psychological Type, Volume 37, pp.2-15, 1996.

[7] Jung, Carl, "Psychological Types," New York: Harcourt Brace, 1923.

[8] Jung, C. G., translated by R. F. C. Hull, "The Sprit in Man, Art, and Literature," Princeton University Press, 1971.

[9] Keirsey, David, and Marilyn Bates, "Please Understand Me," Del Mar, Ca: Prometheus Nemesis Books, 1978.

[10] Kemp, Anthony E., "The Musical Temperament: Psychology and Personality of Musicians," Oxford University Press, New York, 1996.

[11] Myers, Isabel, "Manual: The Myers-Briggs Type Indicator," Palo Alto, California: Consulting Psychologists Press, 1962.

[12] Salton, G. and M. J. McGill, "Introduction to Modern Information Retrieval," McGraw-Hill, 1983.

[13] Salton, G. and C. Buckley, "Improving Retrieval Performance by Relevance Feedback," Journal of the American Society for Information Science, 41(4):288-297, 1990.

[14] Shardanand, Upendra and Pattie Maes, "Social Information Filtering: Algorithms for Automating 'Word of Mouth'," in ACM CHI '95, Conference Proceedings on Human Factors in Computing Systems, Denver, CO, 1995, pp. 210-217.

[15] Simonov, P. V. and P. M. Ershov, "Temperament, Character, and Personality: Biobehavioral Concepts in Science, Art, and Social Psychology," Gordon and Breach Science Publishers, 1991.