Adaptive User Modeling for Personalization of Web Contents*

Alberto Díaz[1] and Pablo Gervás[2]

[1] CES Felipe II – Universidad Complutense de Madrid
c/ Capitán 39, 28300 Aranjuez, Madrid
adiaz@cesfelipesegundo.com

[2] Departamento de Sistemas Informáticos y Programación
Facultad de Informática – Universidad Complutense de Madrid
c/ Juan del Rosal, 8, Madrid 28040
pgervas@sip.ucm.es

Abstract. This paper presents a system for personalization of web contents based on a user model that stores long term and short term interests. Long term interests are modeled through the selection of specific and general categories, and keywords for which the user needs information. However, user needs change over time as a result of his interaction with received information. For this reason, the user model must be capable of adapting to those shifts in interest. In our case, this adaptation of the user model is performed by a short term model obtained from user provided feedback. The evaluation performed with 100 users during 15 days has determined that the combined use of long and short term models performs best when specific and general categories and keywords are used together for the long term model.

1 Introduction

Web content appears in many forms over different domains of application, but in most cases the form of presentation is the same for all users. The contents are static in the sense that they are not adapted to each user. Content personalization is a technique that tries to avoid information overload through the adaptation of web contents to each type of user.

A personalization system is based on 3 main functionalities: content selection, user model adaptation, and content generation. For these functionalities to be carried out in a personalized manner, they must be based on information related to the user that must be reflected in his user profile or user model [8].

Content selection refers to the choice of the particular subset of all available documents that will be more relevant for a given user, as represented in his user profile or model. In order to effect this choice one must have a representation of the documents, a representation of the user profile, and a similarity function that computes the level of adequacy of one to the other.

---

* This research has been partially funded by the Spanish Ministerio de Ciencia y Tecnología (TIC2002-01961).

W. Nejdl and P. De Bra (Eds.): AH 2004, LNCS 3137, pp. 65–74, 2004.

User model adaptation is necessary because user needs change over time as a result of his interaction with information [2]. For this reason the user model must be capable of adapting to those interest changes, it must be dynamic. This adaptation is built upon the interaction of the user with the system, which provides the feedback information used to evolve the profile.

In our case, content generation involves generating a new result web document that contains, for each selected document, its title, its relevance as computed by the system, a summary, and a link to the full document.

In this paper we focus on user model adaptation and the various possible combinations of modeling alternatives for this process. The aim is to identify which is the best way of carrying out the user model adaptation process to improve content selection.

2 Available Methods and Techniques

Existing literature provides different techniques for defining user interests: keywords, stereotypes, semantic networks, neural networks, etc. A particular set of proposals [2; 11] model users by combining long term and short term interests: the short term model represents the most recent user preferences and the long term model represents those expressed over a longer period of time.

Various classification algorithms are available for carrying out content selection depending on the particular representation chosen for user models and documents: cosine formula, rules associated to stereotypes, neural networks, nearest neighbour classifier, naive Bayes classifier, etc.

The feedback techniques needed to achieve a dynamic modeling of the user are based on feedback given by the user with respect to the information elements selected according to his profile. The information obtained in this way can be used to update accordingly the user models in representation had been chosen: term weights, semantic networks, rules associated to stereotypes, etc.

The representation of the text content of the documents by means of techniques based on term weight vectors [9] allows a number of classification algorithms and feedback techniques and constitutes a good option in which to test and compare the relative efficiency of different approaches. The vector associated with a document can be obtained by eliminating the words contained in a stop list and extracting the stems of the remaining words by means of a stemmer. Weights are usually calculated by means of the $tf \cdot idf$ formula, based on frequency of occurrence of terms [9].

Text categorization, the assignment of subject labels to text items, is one of the most prominent text analysis and access task nowadays [10]. In particular, some authors have used Yahoo! categories to semantically characterize the content of documents [6]. They suggest that the best result occurs when using the very brief descriptions of the Yahoo! categories entries.

3 Our Proposal

We propose a browsable user model or user profile that represents user interests from three different points of view [1]. The user model stores three types of information:

personal information, information concerning the format in which information is to be received, and specific information about user interests according to various reference systems that are used to carry out the personalization.

The proposal is similar to [4] but adding more information to the long term model. This additional information comes from the use of the first level categories of Yahoo! Spain as an extra reference framework for describing information needs, that allows users a richer language in which to define their user model.

Long term user interests are modelled with respect to two reference frameworks: the first one based on a domain specific system of classification, and the second one based on the content of the documents.

A basic reference system is the classification system specific to the particular domain under consideration - for instance, in a digital newspaper, this system will be based on the set of sections used by the newspaper -. This system is composed of a set of first level categories that represent different types of information - for instance, examples of sections of digital newspapers would be: national, international, sport, etc. Each web document belongs to a category of that classification system. Information concerning these specific categories is stored as a matrix where rows correspond to specific categories and columns correspond to users (C_{cu}). Users may assign a weight to each specific category to indicate their interest in them.

The reference system based on the content of documents is subdivided in two reference systems, a fine grain model based in keywords and a coarse grain model based on domain independent general categories.

The user can enter a number of keywords to characterise his fine grain model. The appearance of these keywords in the documents will be taken to indicate that the document may be interesting to the user. For each keyword the user introduces a weight that indicates its importance to him. These keywords are stored, for each user u, as a term weight vector (k_u).

To define the coarse grain model the user must choose the general categories in which he is interested. Information concerning these general categories is stored as a matrix where rows correspond to general categories and columns correspond to users (G_{gu}). Users may assign a weight to each general category to indicate their interest in them. The general categories are the first level of categories of Yahoo! Spain and they are represented as term weight vectors (g) obtained from the very brief descriptions of the first level of Yahoo! categories entries.

Short term interests are represented by means of feedback terms. These terms are obtained from user provided feedback over the documents he receives. That is, the user provides positive or negative feedback over the documents he receives, and a set of representative terms is extracted from them. This information is handled by the user model adaptation process, which returns a term weight vector (t_u) for each user. This term weight vector is taken to represent the current short term interests of that user. Short terms interests tend to correspond to temporary information needs whose interest to the user wanes after a short period of time. Therefore their weight must be progressively decreased over time.

Documents are downloaded from the web as HTML documents. For each document, title, category, URL and text are extracted and stored for subsequent processing. Term weight vector representations (d_d) are obtained by application of stop lists, stemmer, and the tf · idf formula for computing actual weights [9].

The only restrictions that must be fulfilled by a domain for the proposed model to be applicable are that there exist textual information associated with web documents and that a domain specific classification exists to classify the documents.

4 Content Selection

Content selection refers to the choice of those among the available documents that are particularly relevant for a user, according to his profile. Once particular representations have been fixed for documents and user model, it becomes feasible to establish which documents are more adequate for each user.

Since we have different reference frameworks in the user model we will indicate how content selection is performed with respect to each one of them, and later we will explore different possible combinations of the resulting selections. Combinations will be based on the relevance obtained for each document within each particular reference framework, and the relative weight used for each reference framework in a particular combination. For all combinations, the final result is a ranking of the set of documents according to the computed overall relevance.

4.1 Selection with Respect to the Long Term Model

As each web document has a preassigned specific category, selection with respect to this reference framework is immediate. Each document is assigned the weight associated with the corresponding specific category in the particular user model. The relevance between a document d, belonging to a specific category c, and a user model u is directly the value assigned to the specific category c by user u:

$$r_{du}^{c} = C_{cu} \quad . \tag{1}$$

The relevance between a document d and a general category g is computed using the cosine formula for similarity within the vector space model [9]:

$$r_{dg} = sim(d_d, g) \quad . \tag{2}$$

The relevance between a document d and the general categories of a user model is computed using the next formula:

$$r_{du}^{g} = \frac{\sum_{i=1}^{14} G_{iu} r_{dg_i}}{\sum_{i=1}^{14} G_{iu}} \quad . \tag{3}$$

The relevance between a document d and the keywords of a user model is computed using the cosine formula for similarity within the vector space model [9]:

$$r_{du}^{k} = sim(d_d, k_u) \tag{4}$$

When all documents have been ordered with respect to the various reference frameworks, the results are integrated using a particular combination of reference frameworks. Therefore, the total relevance between a document d and a user model u is computed with the following formula:

$$r_{du}^{l} = \frac{\alpha r_{du}^{c} + \beta r_{du}^{g} + \chi r_{du}^{k}}{\alpha + \beta + \chi} \tag{5}$$

where Greek letters α, β and χ represent the importance assigned to each reference framework (α, for specific categories, β, for general categories and, χ, for keywords). These are parameters to allow easy configuration of the modes of operation of the system. For this combination to be significant, relevance obtained for each framework must be normalised with respect to the best results for the document collection under consideration.

5 User Model Adaptation

Adaptation of the user model involves obtaining / updating a short term model of the user from the feedback information provided by the user. This model can be used to improve the process of selection in the personalization system.

5.1 Obtaining the Short Term Model

The short term model is obtained as a result of the process of adaptation of the user model. The user receives a web document that contains an automatically generated summary [5] for each of the 10 web documents that the system has found more relevant according to his user profile. With respect to this information the user may interact with the system by giving positive or negative feedback - refraining from providing feedback is interpreted as a contribution as well, taken to imply indifference - for each of the information elements that he has received. The feedback terms of the short term model are obtained from the news items for which either positive or negative feedback has been provided.

Because these terms represent an interest of the user over a short period of time, an algorithm is used to decrement their value over time: each day the starting value of the new weights is obtained by subtracting 0.1 from the previous day's value. Terms that reach a weight less or equal to 0 are eliminated from the model.

To select / update the new feedback terms all documents are preprocessed in the same way as was done for the selection process: stop list and stemmer are applied. The starting point for the adaptation process are the terms of the representation of the documents, with their associated frequency (tf).

The algorithm in [4] is then applied to obtain the feedback terms. The final result of this process is a set of terms ordered according to their new interest value. A subset of them is selected - the 20 most relevant ones - to obtain / update the feedback terms of the short term model.

5.2 Selection with Respect to the Short Term Model

Relevance between a document d and a short term user model u is computed in the same way used for the keywords of the long term model, but using the term weight vector obtained in the process of adaptation of the user model:

$$r_{du}^{s} = r_{du}^{t} = sim(d_d, t_u) \quad (6)$$

5.3 Selection with Respect to the Combined Long Term – Short Term Model

When all documents have been ordered with respect to the different sources of relevance, the results are integrated using a particular combination of reference frameworks. Therefore, the total relevance between a document d and a user model u is computed with the following formula:

$$r_{du} = \frac{\delta r_{du}^{c} + \varepsilon r_{du}^{g} + \phi r_{du}^{k} + \gamma r_{du}^{t}}{\delta + \varepsilon + \phi + \gamma} \quad (7)$$

where Greek letters δ, ε, ϕ, and γ represent the importance assigned to each of the reference frameworks -δ, for specific categories, ε, for general categories, ϕ, for keywords, γ, for feedback terms. Again, for this combination to be significant, the relevance obtained from each reference framework must be normalised with respect to the best results over the document collection being used.

6 Evaluation

As an example of web documents for experimentation we have chosen the web pages of the digital edition of a Spanish newspaper[1]. Experiments are evaluated over data collected for 106 users and the news items corresponding to three weeks – the 14 working days - of the digital edition of the ABC Spanish newspaper. These days correspond to the period 1st-19th Dec 2003. The average of news items per day is 78.5.

To carry out the evaluation, judgements from the user are required as to which news items are relevant or not for each of the days of the experiment. To obtain these judgements users were requested to check the complete set of news items for each day, stating for each one whether it was considered interesting (positive feedback), not interesting (negative feedback) or indifferent (no feedback). Users were explicitly asked not to confine their judgements on interest to relevance with respect to the initial user profiles they had constructed on first accessing the system, but rather to include any news items that they found interesting on discovery, regardless of their similarity with respect to their initial description of their interest. It is hoped that enough information to cover these rogue items will be captured automatically and progressively by the system through the feedback adaptation process.

The three possibilities in each judgment are used to the adaptation process to obtain / update the short term model, but also are used as the relevance judgment of the inter-

[1] This provides a consistent format, which simplifies systematic processing.

est for the user for a news item. However, we consider only two possibilities: interesting, for positive feedback, and not interesting, for negative or no feedback.

Because the evaluation is based on these judgments, significant results can only be obtained for those users that have provided feedback over and above a minimum threshold in terms of number of judgements per day. As the evaluation process involved an effort for the users, only 37.4 users per day actually provided judgments. Additionally, some users only perform feedback for less than 10 news items per day. These users have been eliminated for the evaluation in order to obtain more significant results. This restriction does not constitute a limitation in as much as it is not a particular system that is being evaluated, but rather the relative efficiency of various combinations of methods for specifying information needs. The final collection employed for evaluation presented, on average, 28.6 user per day.

6.1 Metrics

Since our experimental set up combines a binary relevance judgement from the users and a ranking of news items provided by the system, it was decided to use normalised precision [7; 9] as our evaluation metric. In addition, with respect to equal relevance values for consecutive positions of the ranking, the average ranking of the whole set of conflicting positions has been taken as ranking for each and all of them. This adjustment avoids the problem of ordering items at random within the ranking when they have equal relevance.

6.2 Statistical Significance

Data are considered statistically significant if they pass the *sign-test*, with paired samples, at a level of significance of 5% ($p \leq 0.05$). This decision is based on the fact that no specific assumption is made concerning the distribution of data, and that due to the different normalisation processes carried out, it is more convenient to consider relative values instead of absolute values [9].

6.3 Experiment 1

The following experiments have been carried out to check the validity of the proposed model. Each experiment combines different possibilities for long term modeling - only specific categories, L(C), only general categories, L(G), specific and general categories, L(CG), specific categories and keywords, L(CK), general categories and keywords, L(GK), specific and general categories and keywords, L(CGK) - either acting on their own or in combination with the short term model, S. This implies giving different values to the parameters δ, ε, ϕ and γ of formula (7).

For example, L(CG)S is the combination of long and short term model when the long term model includes specific and general categories (δ=1, ε=1, ϕ=0, γ=1).

The results in Table 1 can be interpreted as follows. The first line shows the comparison for alternatives in which the long term model uses only specific categories, L(C). The first two columns indicate that the combination is 16.6 % better than the long term model on its own; the middle columns indicate that the combination is 20.9 % better than the short term model; and the last two columns show that the long term model is 5.2 % better than the short term model. Other lines present results for different configurations of the long term model as described above.

Table 1. Relative increments (% Pr) in normalised precision between different combinations of long term model and short term model

combination > long	% Pr	combination > short	% Pr	long > short	% Pr
L(C)S > L(C)	16.6	L(C)S > S	20.9	L(C) > S	5.2
L(G)S > L(G)	4.4	L(G)S > S	18.4	L(G) > S	14.7
L(K)S > L(K)	26.9	L(K)S > S	11.4	S > L(K)	21.2
L(CG)S > L(CG)	1.7	L(CG)S > S	27.8	L(CG) > S	26.5
L(CK)S > L(CK)	11.8	L(CK)S > S	25.7	L(CK) > S	15.8
L(GK)S > L(GK)	5.6	L(GK)S > S	22.2	L(GK) > S	17.5
L(CGK)S > L(CGK)	2.9	L(CGK)S > S	29.8	L(CGK) > S	27.7

All the results in Table 1 are statistically significant. This leads to the conclusion that the combination of the long term model with the short term model is always better than using each model separately. Also the long term model alone is better than the short term model alone except if the long term model is only constituted for the keywords, in this case, is better the short term model. This means that the worst characterization for the long term model are clearly the keywords.

6.4 Experiment 2

This experiment compares the best performing combinations of previous experiments - long and short term models used together - when the long term model is built using: only specific categories L(C)S (δ=1, ε=0, ϕ=0, γ=1), only general categories L(G)S (δ=0, ε=1, ϕ=0, γ=1), only keywords L(K)S (δ=0, ε=0, ϕ=1, γ=1), specific and general categories L(CG)S (δ=1, ε=1, ϕ=0, γ=1), specific categories and keywords L(CK)S (δ=1, ε=0, ϕ=1, γ=1), general categories and keywords L(GK)S (δ=0, ε=1, ϕ=1, γ=1) and all of the possibilities L(CGK)S (δ=1, ε=1, ϕ=1, γ=1).

Table 2. Relative increments (% Pr) in normalised precision (Pr) between the use of all the reference systems and combinations of L and S together

	L(CGK)S	L(CG)S	L(CK)S	L(GK)S	L(G)S	L(C)S	L(K)S	S
% Pr		2.8	5.5	9.9	11.2	14.0	20.8	29.8
Pr	0.606	0.589	0.572	0.546	0.538	0.521	0.480	0.425

All results are statistically significant (Table 2). This means that the long term / short term combination that uses specific and general categories, and keywords in the long term model, is better than whatever combination of each of them for the long term model.

The second best combination is specific and general categories, and the third is specific categories and keywords, moreover significantly. It can also be observed that the specific categories, when combined, offer the best selection method for the users.

Between L(GK)S, L(G)S y L(C)S there are no significant differences, therefore they performed in a similar way. Finally, the worst result is for the short model alone followed by the combination with the keywords.

6.5 Comparison with Previous Work

The results reported here can be compared with existing evaluations of combinations of methods over similar data. With respect to [3], the set of possible combinations has been extended with the inclusion of short term modeling via relevance feedback. Additionally, normalised precision is used as a metric instead of precision and recall to better take into account the kind of relevance judgements being considered. With respect to [4], the set of possible combinations has been extended with the inclusion of general categories. Also, a new set of data has been collected, with a much higher number of users and evaluation days than in any of the previous collections, to ensure significance of the data.

7 Conclusions

This paper presents the improvement in personalisation achieved by the inclusion of a process of user model adaptation, due to the fact that the selection that is obtained by combining the long term and short term profiles performs better than the one obtained by using the long term model on its own.

The results show that using a combination of a long term model based on specific and general categories and keywords, together with a short term model, improves the adaptation to the user because values of normalised precision increase. These results are obtained for a set of experiments covering various combinations of interest specification methods, and tested over a collection of data for over a hundred users during 14 days. Over this set of data, the conclusion is proved to be statistically significant.

References

1. Amato, G. & Straccia, U., 1999. “User Profile Modeling and Applications to Digital Libraries”. Third European Conference on Research and Advanced Technology for Digital Libraries (ECDL'99), LNCS 1696, pp. 184-197.
2. Billsus, D. & Pazzani. M.J., 2000. “User Modeling for Adaptive News Access”, User Modeling and User-Adapted Interaction Journal 10(2-3), pp. 147-180.
3. Díaz, A., Gervás, P., García, A., 2001. “Sections, Categories and Keywords as Interest Specification Tools for Personalised News Services”. Online Information Review, Volume 25 No 3 (2001), MCB University Press

4. Díaz, A. & Gervás, P., 2004. "Dynamic user modeling in a system for personalization of web contents". Proceedings of CAEPIA 2003, LNAI 3040.
5. Díaz, A. & Gervás, P., 2004. "Item Summarization in Personalisation of News Delivery Systems". Proceedings of TSD 2004, LNAI. In press.
6. Labrou, Y. & Finin, T., 2000. "Yahoo! As an Ontology: Using Yahoo! Categories to Describe Documents". Proceedings of the 8th International Conference on Information and Knowledge Management (CIKM-99), pp. 180-187, ACM Press, November 2-6 2000.
7. Mizzaro, S., 2001. "A New Measure Of Retrieval Effectiveness (or: What's Wrong With Precision And Recall)". International Workshop on Information Retrieval (IR'2001), Infotech Oulu, pp. 43-52.
8. Mizarro, S. & Tasso, C., 2002. "Ephemeral and Persistent Personalization in Adaptive Information Access to Scholarly Publications on the Web". 2nd International Conference on Adaptive Hypermedia and Adaptive Web Based Systems, Málaga, España, 2002.
9. Salton, G., 1989. Automatic Text Processing: The Transformation, Analysis and Retrieval of Information by Computer, Addison-Wesley Publishing, Reading, Massachussets, 1989.
10. Sebastiani, 2002. "Machine learning in automated text categorization". ACM Computing Surveys (CSUR), 34(1), pp. 1-47, ACM Press, 2002.
11. Widyantoro, D.H., Ioerger, T.R., Yen, J., 2001. "Learning User Interest Dynamics with a Three-Descriptor Representation". Journal of the American Society for Information Science and Technology 52(3), pp. 212-225.