IDEA GROUP PUBLISHING
701 E. Chocolate Avenue, Hershey PA 17033-1117, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com
ITP4140

A Knowledge-Based Recommender System for Customized Online Shopping

Fiona Y. Chan and William K. Cheung
Department of Computer Science, Hong Kong Baptist University, Kowloon Ton, Hong Kong
Tel: +852-23395965, Fax: +852-23397892, {fiona, william}@comp.hkbu.edu.hk

ABSTRACT

The concept of personalization has long been advocated to be one of the edges to improve the stickiness of on-line stores. By enabling an on-line store with adequate knowledge about the preference characteristics of different customers, it is possible to provide customized services to further raise the customer satisfaction level. In this paper, we describe in details how to implement a knowledge-based recommender system for supporting such an adaptive store. Our proposed conceptual framework is characterized by a user profiling and product characterization module, a matching engine, an intelligent gift finder, and a backend subsystem for content management. A prototype of an on-line furnishing company has been built for idea illustration. Limitations and future extensions of the proposed system are also discussed.

INTRODUCTION

The development of Web technologies has brought a lot of advantages to merchants for moving their business on line. Within the past few years, a large variety of on-line stores has been started in the cyberspace. However, the survival rate is just around 50%, where some recognized dom-com like Boo.com, Kozmo.com, MVP.com are included [1]. We believe that one important factor determining the success of on-line stores is whether the on-line shopping experience can be enhanced to such an extent that some customers choose to and continue to shop on-line. Along this direction, the concept of personalization has long been advocated as one of the edges to improve the stickiness of on-line stores. A survey, recently conducted by Cyber Dialogue, reveals that customers are more likely to purchase from a site that allows personalization, and register at a site that allows personalization or content customization [2]. To achieve that, an on-line store needs to be enabled with adequate knowledge about customers' preference characteristics and use it effectively to provide personalized services with high precision. A typical example of personalized services is the use of recommender systems.

Recommender systems have been implemented by many big Web retailers, such as Amazon.com and CDNow.com. Typically, they use an intelligent engine to mine the customer's rating records and then create predictive user models for product recommendation. Software products of recommender systems are now available from various companies like NetPerception, Andromedia, Manna, etc. Based on the underlying technology, recommender systems can be broadly categorized as:

- **Knowledge-based** [3] where user models are created explicitly via a knowledge acquisition process.
- **Content-based** [4] where user models are created implicitly by applying machine learning or information retrieval techniques to user preference ratings and features extracted from product description, and
- **Collaborative** [5] where user models are created solely by utilizing overlap of user preference ratings.

In the literature, there exist a lot of works on content-based and collaborative recommender systems. One of their common characteristics is that a substantial amount of good user preference ratings is required before precise recommendations can be provided. However, if a company is lacking such ratings information or it has new items arrived constantly, these two approaches will fail.

Here we argue that before such ratings information can be collected, the knowledge-based approach should provide a good complementary solution. With a similar rationale, Ardissono *et al.* [6] proposed a knowledge-based system using for tailoring the interaction users using a shell called SETA for adaptive Web stores, where stereographical information is also used for user modeling. Sen *et al.* [7] proposed an intelligent buyer agent which aims to educate the user to be a more informed customer by understanding the user query and providing alternatives using a pre-built domain-specific knowledge base, which is based on propositional logic representation. For automatic rule generation, Kim *et al.* [8] have built a prototype system where the decision tree induction algorithm is applied to personalize advertisements.

As there is always a trade-off between personalization and privacy, what kind knowledge needed to be acquired for exchanging personalized services is definitely an important concern of on-line customers. So, the question becomes: "how can the user information requirement be minimized while an acceptable level of recommendation service can still be provided?". In this paper, we restrict the user information needed to only demographic information and describe in details how a related knowledge-based system can be built to support an adaptive on-line store in providing customized recommendation services. Our proposed conceptual framework is characterized by a user profiling and product characterization module, a matching engine, an intelligent gift finder, and a backend management system. A prototype of an on-line furnishing company has been built and is used throughout the paper for idea illustration. The limitations and future extensions of the proposed framework will also be discussed.

SYSTEM OVERVIEW

Knowledge-based systems are characterized by the fact that its two important components, namely the knowledge base and the inference engine (sometimes also called the shell in expert systems) are separated. A typical example is the rule-based system where the knowledge base is represented in the form of a set of if-then rules and forward-chaining reasoning is used in the inference engine. The knowledge engineer can keep on expanding the knowledge base by acquiring more domain knowledge with the inference engine being unchanged at all.

In this project, instead of using the rule-based syntax, a feature vector-based representation is adopted. Also, we assume a conventional 2-tier architecture, where domain knowledge is stored in a relational database and all the functional modules of the inference engine are run on the web server. The knowledge required to be acquired and stored in the database for driving this customized on-line store include:

- **Generic products information,** e.g., product name, price, manufacturing country, etc.
- **Product characteristics,** e.g., degrees of reliability, design style, etc.
- **User demographic information,** e.g., sex, age, occupation, and
- **User preference profiles,** e.g. preferences on reliability, dressing style, etc.

The inference engine contains the following functional modules:

- **User profiling module** which acquires the user demographic information via a simple questionnaire during membership registration and transform the information to create a preference profile for supporting the subsequent matching.
- **Matching engine** which computes the similarity score between user preference profiles and product characteristics to support personalized product ranking shown in the catalog or as special product recommendations.
- **Intelligent gift finder** which can assist the customer via a wizard interface to identify possible gifts for a particular recipient.
- **Back-end management system** for managing the contents for supporting the above modules, which is important as adding adaptability to an on-line store greatly increases its complexity and the store can easily become unmanageable.

To provide personalized product recommendations to customers based on their preferences, one needs to first create the representations for user preferences and product characteristics, and then define a measure for computing the similarity between them (see Section 4).

PRODUCT CHARACTERIZATION & USER PROFILING

Generic Representation

A set of discriminative features $\mathbf{\Phi} := \{\phi_1, \phi_2, \ldots, \phi_N\}$ has first to be identified based on domain knowledge. Then, the user preference can be represented as a vector of preference values on those feature representation $\mathbf{u}=\{u_1, u_2, u_3, \ldots, u_N\}$ $\forall u_i \in [U_{\min}, U_{\max}]$ and the product can be characterized as a vector of values revealing the extent to which it possesses those features, denoted as $\mathbf{p}=\{p_1, p_2, \ldots, p_N\}$ $\forall p_i \in [P_{\min}, P_{\max}]$.

Product Characterization

Based on a chosen set of features, product characteristic vectors **p** have to be created for all the products. Unless for the cases where each product comes with a detailed product description so that some information extraction techniques can be applied, human effort for the creation of **p** is inevitable.

User Profiling

For acquiring user preference profiles **u**, it can be achieved by filling in a questionnaire during the registration process. However, in practice, requiring the users to provide preference values for a long list of features is infeasible as the required effort may simply scare them from continuing to shop in your store. So, the questionnaire for a newly registered user has to be reasonable short and the questions should be easy enough for the user to provide answers. Typical examples are the demographic data like gender, age and occupation, here denote as $\mathbf{d}=\{d_1, d_2, \ldots, d_M\}$ $\forall d_i \in \mathbf{\Delta}_i$, where $\mathbf{\Delta}_i$ is a set of possible stereotypical categories for d_i.[1] However, such a simple representation contradicts the requirement for a discriminative set of features. One solution is using domain knowledge to transform the demographic information **d** user into a preference profile representation **u** containing a rich set of features via a transformation $\boldsymbol{f}_u(\mathbf{d})$: $\mathbf{\Delta} \longrightarrow [U_{min}, U_{max}]^N$ where $\mathbf{\Delta} = \mathbf{\Delta}_1 \times \mathbf{\Delta}_2 \times \ldots \times \mathbf{\Delta}_M$. The precision of the preference profile thus highly relies on that of the transformation.

Another issue related to user profile representation is about the importance of each individual feature. Under the aforementioned feature vector representation, user preferences on the features are assumed to be equally important. However, this is not the case in practice. Some users may consider "color" to be a more important feature than "durability" while some may find it the other way round. The situation can be even worse as this kind of information is usually unconscious for users and hard to be provided precisely. In our system, we model the relative importance of the feature with a weighting vector $\mathbf{w}=\{w_1, w_2, \ldots, w_N\}$ $\forall w_i \in [0,1]$ and $\sum_i w_i = 1$. Also, we introduce one more transformation $\boldsymbol{f}_w(\mathbf{d})$: $\mathbf{\Delta} \longrightarrow [0,1]^N$. This transformation can be interpreted as the relative importance of the features for different combinations of demographic categories. It is hoped that this can free up the user from providing subjective weighting values. Again, the precision of the transformation is crucial to the success of weighting application.

Obtaining the transformations that can effectively reflect the interests of the different demographic categories is by no means straightforward. Some possible objective means include conducting marketing surveys or analyzing past transaction records. Regarding their implementations, the input dimensions of the two transformations are equal to $\prod_{i=1}^{M} card(\mathbf{\Delta})$. Creating them directly may result in large storage requirement as well as tedious work in creating and managing them. By assuming the effect of each elements in **d** on the overall transformation to be independent, the transformation for preference can be decomposed into a set of transformations $\{ \boldsymbol{f}_{uj}(\boldsymbol{d}_j) : \mathbf{\Delta}_j \longrightarrow [U_{\min}, U_{\max}]^N \ \forall_j = 1..M \}$, each corresponds to a particular element in **d**. The storage requirement can then be reduced from $\prod_{i=1}^{M} card(\mathbf{\Delta})$ to $\sum_{i=1}^{M} card(\mathbf{\Delta})$. With the decomposition, the preference profile is then computed as

$$\mathbf{u} = [ui = {}^{1}/M \textstyle\sum_{j=1}^{M} f^{i}_{uj}(d_j)] \ \forall i = 1 \ldots N$$

where f^{i} denotes the i^{th} element of f's output. The range of value for each element in **u** remains to be $[U_{\min}, U_{\max}]$. Similarly, the transformation for weighting can be decomposed as $\{ \boldsymbol{f}_{wj}(d_j) : \mathbf{\Delta}_j \longrightarrow [0,1]^N \}$. More details about the use of the weighting vector are described in Section 4.

An On-line Furnishing Company Prototype

To provide a concrete example for explaining the representation issue, we have built an on-line furnishing company prototype for idea illustration.[2] The furniture items include tables, sofa, beds, quilts, etc. For user profiling and product characterization, the set of features $\mathbf{\Phi}$ we have used is shown in Figure 1 and the range of value for each element in both representations is set to be $U_{min}=P_{min}=-1$ and $U_{max}=P_{max}=+1$. Products with softness="-1" means that they are extremely hard whereas those with softness="1" means that the product is very soft. For demographic information **d**, 3 attributes — gender, year-of-birth and occupation are adopted (i.e., M=3). For the creation of the transformation functions $\boldsymbol{f}_u(d_j)$ and $\boldsymbol{f}_w(d_j)$ (see Figure 2 and Figure 3) as well as the product feature vectors **p,** it is done manually based on domain knowledge.

After a user registers with our system, his/her basic personal demographic information will automatically be stored. If he/she logs onto the system again, a personal preference profile will be created based on the methodology previously described. Recommendation services can thus be provided.

MATCHING ENGINE

Given the user preference profile **u**, the product characteristics **p** and the range of preference values, a similarity measure can then be

Figure 1: Examples of product feature vectors, **p**

Product ID	…	Colorful	Essential	Exotic	Easy to clean	Durable	Safe	Soft	Modern
10002	…	-0.4	0.8	0.4	-0.1	0.5	0.6	0.4	0.7
10023	…	0.6	-0.4	0.3	-0.2	0.4	0.3	0.1	0.6
10045	…	0.3	0.4	0.8	-0.2	-0.1	0.4	0.8	0.1

Figure 2: Examples of preference transformation, $\boldsymbol{f_u}(d_j)$

Category	Colorful	Essential	Exotic	Easy to clean	Durable	Safe	Soft	Modern
35-45 yr old	-0.3	0.7	0.1	0.8	0.8	0.3	0.1	-0.4
Female	0.1	-0.2	0.2	0.5	-0.2	0.5	0.2	-0.1
Housewife	-0.5	0.8	-0.3	0.9	-0.4	0.1	-0.2	-0.1

Figure 3: Examples of weighting transformation, $\boldsymbol{f_w}(d_j)$

Category	Colorful	Essential	Exotic	Easy to clean	Durable	Safe	Soft	Modern
35-45 yr old	0.03	0.18	0.06	0.33	0.24	0.08	0.04	0.04
Female	0.2	0.08	0.11	0.15	0.08	0.09	0.09	0.2
Housewife	0.04	0.17	0.08	0.31	0.22	0.1	0.04	0.04

defined. In our prototype, as the preference value range is [-1,1], one obvious measure is the dot product between **u** and **p** weighted by **w,** given as

$$\mathrm{sim}(\mathbf{u},\mathbf{p},\mathbf{w}) = \sum_{i=1}^{N} u_i p_i w_i$$

with the output equal [-1,1].[3] Based on the similarity scores computed, personalized product ranking can be achieved. It can also be used to customize the catalog for browsing with the hope that the user can identify their intended products with less mouse clicks. The keyword search engine can also benefit by ranking the search results based on the scores so as to improve the chance that the intended items are put on the first few pages of the search results. Besides, when there is a list of new products, personalized recommendation services can be provided to further improve the quality of customer services.

INTELLIGENT GIFT FINDER

Profiling Gift Recipients

For on-line shopping customization, we used to focus on how to acquire the interest of the individual customers so as to provide just-in-time customized services. However, other than buying things for themselves, customers often buy product items to be presented to their friends as gifts. Most of the on-line stores try to fulfill the need by providing some advanced search functions, with the hope that the customers can manage to specify the preference of the gift recipients in the form of some complicated searching criteria. However, this does not conform to our usual shopping habit. Instead, we used to have dialogs with the salesperson in the store describing some basic characteristics of the gift recipient with the hope that the salesperson can effectively provide some relevant recommendations to us for reference. With that idea in mind, our proposed system contains a module called *intelligent gift finder*. It operates like a typical wizard that asks users a sequence of dynamically generated questions for profiling the gift recipient. In our system, the questions are extracted from a predefined question bank[4] with a tree structure where the edges of the tree determine the next question to ask based on the answers of the previous question. Within the question bank, there are two sets of questions. One set of questions, denoted as Θ_H, is to help capturing hard constraints to prune down the product search space (e.g., "what is your budget?"). The other set of questions, denoted as Θ_S, is to help capturing the *preference profile* of the gift recipient (as described in Section 3) for ranking the products in the reduced search space in a customized manner (e.g., "what is the gender of the gift recipient?").

Suppose a user has clicked on the wizard and answered K questions. Denote θ_i as the i[th] question, α_i as his/her corresponding answer. Also, denote $\Gamma(\theta,\alpha)$ as the set of feasible solutions corresponding to the question-answer pair (θ,α). The current set of feasible solutions should then be given as

$$\boldsymbol{\Gamma}_{\mathrm{K}} = \bigcap_{i=1:K;\theta_i \in \Theta_H} \Gamma(\theta_i = \alpha_i).$$

For ranking products in $\boldsymbol{\Gamma}_{\mathrm{K}}$, we need to associate (θ_l, α_l) with some similarity measure. For questions capturing the recipient's demographic information, the associated preference and weighting transformations can be used, and the similarity score can be computed as

$$\sum_{i=1:K;\theta_i \in \Theta_S} sim(\boldsymbol{f}_{u_{\theta_i}}(\alpha_i), \mathbf{p}, \boldsymbol{f}_{w_{\theta_i}}(\alpha_i)).$$

A concept similar to that of our gift finder has been used in the Decision Guide which is developed by Personallogic and currently used in AOL.com.

Implementation and Related Management Tools

It is obvious that embedding the questions and answers into the program code greatly reduces the system's maintainability and extensibility. So, putting them into the database as part of the knowledge is a natural solution. As the question bank adopts a tree-like structure, the linking relationship between the questions has to be stored as well. To ease the effort for maintaining the question bank with tree-like structure, an associated management tool has been developed accordingly. With the help of the tool, internal staff of the store can easily create, update and delete questions and answers. Also, they can easily specify how each question-answer pair is associated with the corresponding conditions to be used in searching the database, though prior knowledge on the database schema is still inevitable for the staff.

FUTURE EXTENSIONS

Personal Adaptation

One major limitation of the proposed framework is that it assumes that user preferences can solely be determined based on their demographic information. In fact, two customers with identical demographic information can only be considered to have the same preferences up to a certain extent on the average. If more precise personalized recommendation service is to be provided, a deeper level of personalized adaptation will be needed. For example, one can further adapt the preference transformations and weighting transformations acquired after the first registration to suit the specific characteristics of the customer. One possible direction is to use relevance feedback, i.e., to modify the transformations based on the characteristics of the customers' highly rated products in a weighted sum manner [9]. Currently, we are studying how to cast the problem under a Bayesian framework. Other open issues include:

- what kind of information should be acquired from the customer to support the personalized adaptation (e.g., customers' ratings, click-streams, etc.),
- in what manner should they be acquired (explicitly or implicitly), and
- how the acquired information should be analyzed to represent the customer preference and combined with the existing transformation tables in an incremental and disciplined manner.

Integration With Collaborative Methods

The collaborative filtering technique is known to be an effective method for identifying like-minded customers solely based on customer ratings and has been used by a number of recommender systems. It will be interesting to see how the knowledge-based approach can take the advantage of collaborative filtering to shorten it time in providing highly precise recommendations. One possibility that has been proposed in the literature is to compute *ratings for features* by aggregating *ratings for products* [10]. Then, the predicted preferences on different features, which will change as more ratings are provided, can be used as relevance feedback or additional evidence with the hope to further increase the precision of the recommender system. Other works along this direction have been reported in the literature, including [11, 12].

CONCLUSION

In this paper, we have demonstrated how customers' stereotypical information can be used to provide customized recommendations using a knowledge-based approach and discussed several ways to further enhance its precision. Performance evaluation of the proposed system has not been conducted due to the lack of empirical data. We believe that the best way of evaluation is to apply it to the real market and we are currently identifying interested industrial parties for the evaluation study.

ENDNOTES

1 Whether a user is willing to provide their demographic information is related to the privacy issue, which is out of the scope of this paper. In general, the user has to sacrifice a certain degree of privacy in order to gain customized services.

2 There are many related on-line companies, e.g., IKEA.com, Maxwellfurniture.com, etc.

3 If the range of preference value is changed, the Pearson correlation or the cosine value between the vector **p** and the vector **u** weighted by **w** can be used.

4 For the creation of the question bank, we believe there exist different cognitive or psychological theories governing how the questions should be set to achieve objective user profile acquisition. However, related considerations are out of the scope of this paper.

REFERENCES

[1] Helft M., Jul 16, 2001, The E-commerce survivors, the Industry Standard Magazine, URL=http://www.thestandard.com/article/0,1902,27593,00.html

[2] May 9, 2001, Personalization Consortium News, URL=http://www.personalization.org/ pr050901.html

[3] Towle B. and Quinn C., Knowledge-based recommender systems using explicit user models, In Knowledge-based Electronic Markets: Paper from AAAI, Technical Report WS-00-04, pp. 74-77, 2000

[4] Mooney. R.J. and Roy L., Content-based book recommending using learning for text categorization, In Proceedings of SIGIR'99 Workshop on Recommender Systems: Algorithms and Evaluation, Berkeley, CA, August, 1999

[5] Resnick P., Iacovou N., Suchak M., Bergstorm P. and Riedl J., GroupLens: An open architecture for collaborative filtering of netnews. In Proceedings of ACM 1994 Conference on Computer Supported Cooperative Work, pp. 175-186, Chapel Hill, NC, 1994

[6] Ardissono L., Goy A., Petrone G., Segnan M. and Torasso P., Tailoring the interaction with users in electronic shops, In Proceedings of the 7th International Conference on User Modeling, Banff, MA, 1999

[7] Sen S., Dutta P.S. and Mukherjee R., Agents that represent buyer's interests in E-commerce, In Knowledge-based Electronic Markets: Paper from AAAI, Technical Report WS-00-04, pp. 63-69, 2000

[8] Kim J.W., Lee B.H., Shaw M.J. Chang H.L. and Nelson M., Application of decision-tree induction techniques to personalized advertisement on internet storefronts, in International Journal of Electronic Commerce, Vol. 5, No. 3, pp. 45-62, Spring, 2001

[9] Rocchio J.J., Relevance feedback in information retrieval, In The SMART Retrieval System, Gerard Salton (ed.), Prentice-Hall, Englewood Cliffs, New Jersey, 1971

[10] Pazzani M.J., A framework for collaborative, content-based and demographic filtering, In Artificial Intelligence Review, Vol. 13, Issue 5/6, pp. 393-408, Dec., 1999

[11] Burke R., Integrating knowledge-based and collaborative-filtering recommender systems, In Artificial Intelligence for Electronic Commerce: Paper from AAAI, Technical Report WS-99-01, pp.69-72, 1999

[12] Tran T. and Cohen R., Hybrid recommender systems for electronic commerce, In Knowledge-based Electronic Markets: Paper from AAAI, Technical Report WS-00-04, pp. 78-84, 2000.

Related Content

Steel Surface Defect Detection Based on SSAM-YOLO
Tianle Yangand Jinghui Li (2023). *International Journal of Information Technologies and Systems Approach (pp. 1-13).*
www.irma-international.org/article/steel-surface-defect-detection-based-on-ssam-yolo/328091

Theory Development in Information Systems Research Using Structural Equation Modeling: Evaluation and Recommendations
Nicholas Robertsand Varun Grover (2009). *Handbook of Research on Contemporary Theoretical Models in Information Systems (pp. 77-94).*
www.irma-international.org/chapter/theory-development-information-systems-research/35825

Addressing Team Dynamics in Virtual Teams: The Role of Soft Systems
Frank Stowelland Shavindrie Cooray (2016). *International Journal of Information Technologies and Systems Approach (pp. 32-53).*
www.irma-international.org/article/addressing-team-dynamics-in-virtual-teams/144306

Two Rough Set-based Software Tools for Analyzing Non-Deterministic Data
Mao Wu, Michinori Nakataand Hiroshi Sakai (2014). *International Journal of Rough Sets and Data Analysis (pp. 32-47).*
www.irma-international.org/article/two-rough-set-based-software-tools-for-analyzing-non-deterministic-data/111311

A Comparative Analysis of a Novel Anomaly Detection Algorithm with Neural Networks
Srijan Das, Arpita Dutta, Saurav Sharmaand Sangharatna Godboley (2017). *International Journal of Rough Sets and Data Analysis (pp. 1-16).*
www.irma-international.org/article/a-comparative-analysis-of-a-novel-anomaly-detection-algorithm-with-neural-networks/186855