Scientific
Research

Combining Personal Ontology and Collaborative Filtering to Design a Document Recommendation System

Deng-Neng CHEN, Yao-Chun CHIANG

Department of Management Information Systems, National Pingtung University of Science and Technology, Taiwan, China.
Email: dnchen@mail.npust.edu.tw

ABSTRACT

With the advance of information technology, people could retrieve and manage their information more easily. However, the information users are still confused of information overloading problem. The recommendation system is designed based on personal preferences. It can recommend the fittest information to users, and it would help users to obtain information more conveniently and quickly. In our research, we design a recommendation system based on personal ontology and collaborative filtering technologies. Personal ontology is constructed by Formal Concept Analysis (FCA) algorithm and the collaborative filtering is design based on ontology similarity comparison among users. In order to evaluate the performance of our recommendation system, we have conducted an experiment to estimate the users' satisfaction of our experiment system. The results show that, combining collaborative filtering technology with FCA in a recommendation system can get better users' satisfaction.

1. Introduction

With the internet technology has been widely used in human life, huge amounts of websites have been built and updated every day. This phenomenon usually makes the internet users at a loss in such a huge amount of information, and this problem is known as "information overloading". Furthermore, the information that hides in the databases is beyond the search engines' reach. In this case, although many internet search engines are available, it is still useless to information users to find what they want. Therefore, many websites, such as Yahoo! news and Amazon online bookstores, launch their own recommendation service on their platforms. They hope their systems could recommend products or information to users automatically and help users to find what they are searching for more quickly. In advance, the recommendation systems could even assist in answering to the potential information in which the users are interested.

Collaborative filtering technology is considered to be an effective way to solve the information overloading problem [1]. This technology mainly emphasizes on the cooperation between people. The system first collects the information of the users and then calculates the similari-

ties among the users. Through this way, the system could learn the preferences of every user and those preferences in common which could be recommended to the users. It will not only present the information that the users are interested in, but also some potential information that may surprise the users. Currently, some famous websites such as Amazon have adopted this technology. This shows that among these recommendation systems, collaborative filtering technology is relatively successful and most commonly used, as well as an excellent system used in electronic commerce [2–4].

Apart from helping finding the demanded information, the recommendation system aims to help the users to search with a faster speed and accuracy by constructing the shared documents and common preferences. It also makes the resources and services on the internet easier to access and share [5]. In this research, we integrate ontology and collaborative filtering to design a system to provide information recommendation service. We adopt Formal Concept Analysis (FCA) to construct a personal ontology to show the conceptual structure of personal preferences. FCA technology has been proved to be helpful in the development of ontology [6–10].

This research is not only engaged in constructing a recommendation system which combines ontology with the collaborative filtering technology, but also compare the users' satisfaction with the system without that technology. We have developed a prototype system and conducted a laboratory experiment to evaluate the users' satisfaction on different recommendation mechanism. The remainder of the paper is organized as the following. Section 2 reviews major literature concerning recommendation systems, ontology and FCA. System architecture and experiment design are shown in Section 3, and data analysis results are discussed in Section 4. Finally, implications and conclusions are described in Section 5.

2. Literature Review

2.1 Recommendation Systems

At the present time, recommendation systems hold more extensive definitions. It can be used to describe personal recommendations from any system or direct the users to find interested or useful targets from multiple possible choices. In this information overloading era, the design and development of recommendation systems is virtually more attractive than search for information depending on individuals, because it could help people make decisions from the complicated information. Currently, recommendation systems are already included in some electronic commerce websites such as Amazon [3].

The earliest recommendation system, developed by Goldberg *et al*. [11] is called Tapestry. It filters the useful information by collaborative filtering system. Collaborative recommendation system is the most famous and commonly used one. The system analyses the behaviors or preferences from the set of users within the system. It finds out the set of users with similar characteristics and takes this relevance as an evidence to induct the potential preferences of the users. Therefore, besides recommend the interested information to the users, this research is expected to recommend the information that may arouse the users' potential demands. In our recommendation system, it will first collect the users' information and calculate similarities of every user. From this way, the system could learn the preferences and the ones in common and find out the users who hold the similar preferences.

2.2 Ontology

Ontology could be defined from many aspects. Schreiber *et al*. [12] defined it as ontology provides a clearly description and conceptualization to express the knowledge in knowledge base from the aspect of knowledge base construction. In addition, Bernaras *et al*. [13] agreed that ontology provides a clear description to conceptualize knowledge in knowledge base. William and Austin [14] also proposes that ontology is a set describing or ex-

pressing concepts or terms of a certain field and can be used to organize the higher level of conceptual knowledge in knowledge base or describe the knowledge of a certain field. The process of its development leads to different definitions of ontology, but one point in common is ontology could help describe knowledge and the conceptual structure. In addition, the importance of ontology is that it matters the expression of knowledge structure and the analysis through ontology so as to present a clear knowledge structure. In one certain field, ontology is the core of expressing knowledge system and would help effectively express through analysis of ontology.

Therefore, the utmost task is to develop terms and relations that could effectively express knowledge so that the certain field or category would be analyzed efficiently. Moreover, the development of ontology would help share the knowledge. Knowledge base could be constructed according to different circumstances due to the share of ontology. For example, different manufacturers could use common terms and grammars to construct and describe the catalog indexes of some product, and then they share and use these indexes in automatic data exchanging systems. This kind of sharing could greatly increase the chances of knowledge reusing [15]. Now that ontology could familiarize the users with knowledge in specific field, users could utilize the conceptual correspondence of ontology to avoid the confusion of conceptions and rapidly find conceptual category in individual ontology. This could make browsing websites and searching information more efficient and convenient [5,16].

2.3 Formal Concept Analysis and Ontology

Formal Concept Analysis (FCA), proposed by Rudolf Wille in 1982, is a data analysis theory to disclose conceptual structures from data set [17]. The characteristic is that structures of data set could produce the graphical visualization, especially the quantitative analysis that the social sciences cannot be fully captured. Ganter and Wille [18] considered that FCA could mainly be used on data analysis such as investigate and process definite data. This data is based on Formal Abstractions of Concepts which is prominent and understandable. Wille [17] combined the target, property and relevance (each target possesses a property) together to present these relations by mathematical definitions of Formal Context and define Formal Concept [19].

The goal of both ontology and FCA is to build conceptual models of knowledge domain. FCA can be viewed as a technology of ontology construction to obtain structured data by concept lattices; it can be used as foundation of developing ontology manually and automatically by extracting concepts from the data set; it can also be used to present the visualization of ontology and
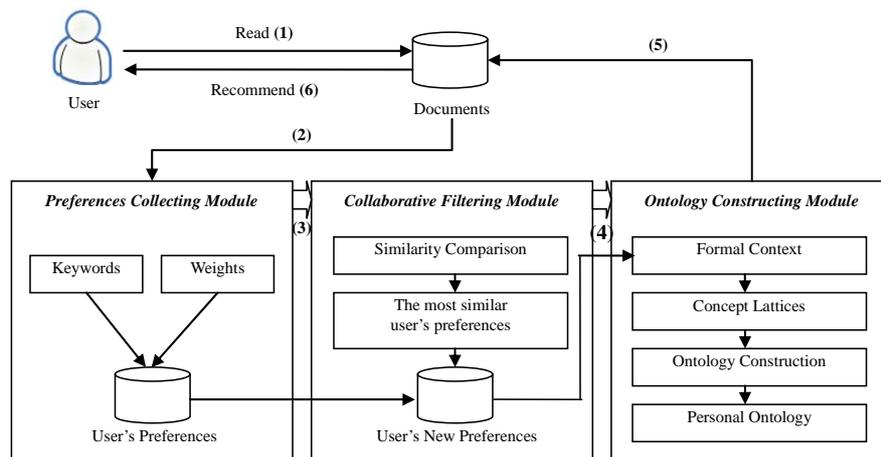
Figure 1. System architecture

help browse and analyze tasks. Among the theories combining FCA and ontology, the most prominent application is to identify the concept of ontology through formal concept [20]. Moreover, Hsu [7] proposes to automatically construct ontology based on FCA theory. It firstly extracts terms that stands for document concepts from term extraction system. Then integrate the binary matrix of document and terms to express independent, interlaced and inherited relations among different concepts and form the diagram of relations of concepts of ontology. The above documents all consider the property of FCA as the concept of ontology and the other relevant concepts as properties. Based on this view ontology is constructed or combined. The researches mentioned above prove that FCA and the concept of ontology could effectively help construct ontology. This research will use the ontology construction by FCA in recommendation systems.

3. System Architecture and Experiment Design

In this research, we aims to develop a recommendation system based on the combination of collaborative filtering technology and ontology. It will not only construct personal ontology with the FCA, but also calculate the users' familiarity to the keywords of all the documents. The users will give scores on those they read and are interested in while browsing them. These scores could show the users' preferences and work as a weights standard in the construction of ontology.

3.1 System Architecture

Figure 1 shows the system architecture of our recommendation system. In the step 1, the users enter the system, and the system assigns 20 documents randomly to users. The users browse and choose the top five docu-

ments they prefer to and give scores from 1 to 5 on the familiarity of the keywords of the 5 documents. In the step 2, the system analyze the collection of keywords and scores in the preference documents and make weights computing in users' preference collection module to prepare for the ontology construction and similarity comparison. In the step 3, with the weights computed in the previous module, the collaborative filtering module will compare the keywords and weights of preference with others. For the sake of time and efficiency, the system will only compare the first 100 users in the database and find the users with the highest similarity. The preference keywords and weights of these couple users will be sent to users' ontology construction module to prepare for the ontology construction. In the step 4, the system intermix the keywords and weights of the user with the highest similar one's. The sum of the keywords and weights will be used to construct the users' preference ontology by ontology construction module based on FCA technology. In the step 5, the system will send the new personal preferences back, and then the system will calculate the weights of each document. Finally, in the step 6, after calculating weighs of each document in the database, the system recommend the top five documents with the highest weights, and measure the user's satisfaction by online questionnaires.

The major modules in the system architecture are shown as follows.

1) Document database

The experimental system recommends documents to the users to read. In the document database, there are 210 mater dissertations focus on electronic commerce selected from Electronic Theses and Dissertations System in Taiwan[1]. The data schema of documents database is composed of eleven fields, including serial number, author's name, year, paper's title, affiliation, abstract, and

five keywords.

2) Preferences collection module

For constructing personal preferences ontology by FCA, we need to collect user's preferences of keywords of documents. We believe that choosing their preference documents of the users cannot fully reflect the degree of their preferences. Therefore, we propose the scoring mechanism of the keywords to modify the weights between the concepts in the process of constructing ontology. In this module, user should select 5 preferred documents and score from 1 to 5 for each keyword in the documents to show their preference degree.

3) Collaborative filtering module

For the collaborative filtering mechanism, our system should have some users' preferences first. Therefore, when a user enters our system, the system can select the fittest user from the database and finish the collaborative filtering. In our experiment, we collect 105 participants' preferences in the database before collaborative filtering mechanism is running.

To find the fittest user from the database, we need a function to calculate the similarities between the users. We define Sims as the degree of the similarities of two users' preferences, and its function is shown as follows.

$$Sims = \frac{\sum_{t=1}^{r} KtWt}{\sum_{i=1}^{m} KiWi + \sum_{j=1}^{n} KjWj}$$

$\sum_{i=1}^{m} KiWi$: the sum of weights of user's preferred concepts

$\sum_{j=1}^{n} KjWj$: the sum of weights of the other user's preferred concepts

$\sum_{t=1}^{r} KtWt$: the sum of weights of the two users' conjunctive preferred concepts

4) Ontology construction module

This module mainly focuses on the weights of keywords collection and constructs the personal ontology. We adopt FCA [17] construct ontology. The steps are as follows.

Step 1: produce the formal contexts of the documents and keywords.

We first extract the collection of the keywords of the chosen documents from the document database. Then we match all the documents with the keywords collection. If the document includes certain keyword it will be marked

as "1". In this way form the formal contexts of the documents and keywords. Because of the scoring mechanism in this research, the keywords collection will be sequenced according to the weights of the users. In the later part the preference discussion will be transformed into the section of tree framework and make the concepts with high weights as higher hierarchy. According to the definition of FCA, this research defines the definition of formal contexts as K, the document collection on e-commerce as E, the keywords collection as T and the binary of the document collection and keywords collection as R. Then their relation can be put into $K := \langle E, T, R \rangle$.

Step 2：Produce all the concepts C

Define A as the subset of E and B as the subset of T, that is, $A \subseteq E$, $B \subseteq T$. If a certain concept is $A \subseteq B$, then it is marked as concept c (A, B). For a concept c (A,B), if all the relations R between A and B can form a biggest matrix, then all the collection of concept c is marked as C.

Step 3: produce the concept lattices between all the concepts

If the collection of all the documents with the keyword B1 is included in the collection of all the documents with the keyword B2, the keyword B1 is marked as the sub-concept of the keyword B2. That is, for all the concepts C, if $B_1 \subseteq B_2$, then $c_1(A_1, B_1)$ is the sub-concept of $c_2(A_2, B_2)$ and expressed as $(A_1, B_1) \leq (A_2, B_2)$. The sign $\leq$ stands for hierarchy of concepts.

Step 4: transform into tree diagram of ontology

While transforming the concept lattices diagram into tree framework of ontology by using breadth-first search, the relations of nodes may be fairly complicated and make the system spend too much time computing. This would lead to the inefficiency of recommendation and failure to promptly recommend documents to users. In order to avoid this, while constructing concept lattices, this research does not take the interlaced relations into account and make the concepts with high weights higher hierarchy. Then the relation contains only the concepts of higher hierarchy and the lower hierarchy. Then by breath-first search transform the relevance of formal contexts into tree framework which is the users' preference ontology.

3.2 Experiment Design

This experiment aims to recommend the users documents through two different recommendation systems and test their satisfaction. First, to be the experiment group, this system constructs ontology with the FCA theory, the scoring system and collaborative filtering technology. The other one, to be the control group, this system constructs ontology with the FCA theory and the scoring system without collaborative filtering. We will introduce

Table 1. User's satisfaction measurement

1. Do the recommendation documents meet your demands?
2. Are the recommendation methods accurate?
3. Are the recommendation methods satisfying?
4. Do you understand the recommendation methods?
5. Do you think the recommendation methods practical?
6. Do you think the recommendation methods reliable?
7. Do you think the recommendation methods clear?
8. Do you think the way of recommending understandable?

the recommendation steps of the first system as follows:

Step 1: Enter into the system: the users first read the introduction of the first page to learn the purpose and contents of the experiment.

Step 2: Assign documents randomly: the system extracts 20 documents randomly from the 210 ones for the users to read.

Step 3: Choose the documents the users prefer to and give scores: the users click the 20 ones to further read the contents and give scores on five interested ones. The system will store the keywords collection and preference scores of the five documents to prepare for the computing or collaborative filtering of the preferences.

Step 4: Ontology constructing for the users and recommends 5 documents to users based on ontology.

Step 5: After reading the recommendation documents, the users could fill in the questionnaires. The satisfaction refers to the satisfaction with information quality. The users should answer eight questions with Likert's five point scale from very dissatisfying to very satisfying. The experiment finishes after the users answer these questions.

4. Data Analysis

To evaluate the user's satisfaction on our experiment system, we have conducted a laboratory experiment research. The system combining personal ontology and collaborative filtering is served as the experiment group, and the system that has only personal ontology recommendation without collaborative filtering is served ad the control group. There are totally 250 qualified participants have been invited to the experiment. By randomly dispatched by the system, 145 samples are assigned for experiment group and 105 for control group. User's satisfaction is measured by questionnaires online. The questionnaire is designed based on DeLone and Mclean's IS (information systems) success model [21,22]. This model proposes a comprehensive perspective to measure the success of an information system and has been widely used to appraise the quality of information systems. In a nutshell, a suc-

cessful information system should have qualified information quality and system quality to satisfy the users. In our research, due to both the experiment group and control group are conducted in the same platform, the system quality are the same in certain. We only adopt the measurements for information quality in our questionnaires. Table 1 shows the user's information quality satisfaction measurements and Likert's five point scale, from very disagree to very agree, is applied.

Factor analysis is applied to evaluate the validity of our measurements. The KMO value of this construct is 0.856. It shows that these measurements are feasible to factor analysis. Extract the dimensions whose eigenvalue is larger than 1 by using principal component analysis and orthogonal rotation through VARIMAX. After factor analysis, we divide the eight questions into two factor components. Question 2, 1, 6, 3 make up the first factor component, and this construct is named as *satisfaction with recommendation results*. Question 8, 4 and 7 make up the second one, and is named as *satisfaction with recommendation process*. Question 5 has the similar factor loading in both the two components (both are more than 0.5). We would delete question 5 after factor analysis.

Table 2 shows the descriptive statistics results of our experiment. The experiment group always gets higher satisfaction both on recommendation results and process.

To verify the experiment group really gets higher users' satisfaction than the control one in statistics, the independent-samples T test is applied. The results are shown in Table 3. No matter on recommendation results or process, users get higher satisfaction significantly. That is to say, the recommendation system based on the combination of ontology and collaborative filtering system is more satisfying than the one based only on personal ontology.

The higher satisfaction of experiment group might cause by the extension capability of combining collaborative filtering results with personal preferences. Due to the original personal ontology is built based on only five

Table 2. Descriptive statistics results of research constructs

Users' Satisfaction	Group	N	Mean	Std. Dev.
Recommendation Results	Experiment	145	3.641	0.591
	Control	105	3.441	0.580
	Total	250	3.557	0.594
Recommendation Process	Experiment	145	3.740	0.658
	Control	105	3.578	0.526
	Total	250	3.672	0.610

Table 3. Independent-samples t test results

Users' Satisfaction	Levene's Test for Equality of Variances		T-test for Equality of Means	
	F	Sig.	t	Sig.
Recommendation Results	0.011	0.916	2.673	0.008**
Recommendation Process	1.841	0.176	2.092	0.037*

interested documents, it will not include all the user's preferences certainly. Collaborative filtering would help to capture user's other preferences that have not been defined in the original personal ontology. In the other words, the expanded personal ontology, combining collaborative filtering results, might cover the potential preferences that have not been discovered. Therefore, the system recommends documents to users based on the expanded personal ontology would cause higher satisfaction.

5. Conclusions

This research is expected to take advantages of collaborative filtering and personal ontology to design an effective recommendation system. Therefore, we have first discussed how to construct personal ontology based on one self's and others' preferences. The personal ontology is built up by FCA method, in advance, we used scoring mechanism to intensify the weights of users' preferences. Then, we elaborated on how to make use of this method to provide personal recommendation service in an electronic documents repository system. We have implemented a prototype system and conducted a laboratory experiment to evaluate the system's performance. The research results show that the users have higher satisfaction with the recommendation system that combined collaborative filtering and ontology technology.

In practice, this research applies collaborative filtering and ontology to provide personal recommendation service on an electronic documents website. This personal recommendation method can be used widely in different online websites such as electronic news website, or e-retail website to recommend news/products to customers.

However, in our experiment, the recommendation service is built based only 210 master theses. Due to the FCA method should calculate the relations among each document, it might cause performance problem when it were used in real repository system that usually has more than ten or hundred thousands of documents. The FCA method should be improved in calculation efficiency when it is used in the larger scale system. In the other, how to extract the proper keywords from documents would be another important and interesting issue. In our system, the recommendation documents database is composed of master theses. As usual, the maser theses have accurate keywords that are defined by the author. However, in some other documents repository system, such as news website, there is no well-defined keyword in the system. How to extract proper keywords from this kind of system would be anther critical problem when our recommendation system is implemented.

REFERENCES

[1] B. Sarwar, J. Konstan, A. Borchers, J. Herlocker, B. Miller, and J. Riedl, "Using Filtering Agents to Improve Prediction Quality in the GroupLens Research Collaborative Filtering System," Proceedings of the 1998 Conference on Computer Supported Cooperative Work, pp. 345–354, 1998.

[2] M. Balabanovic and Y. Shoham, "Fab: Content-Based, Collaborative Recommendation," Communications of the ACM, Vol. 40, pp. 66–72, 1997.

[3] J. B. Schafer, J. Konstan, and J. Riedl, "Recommender System in E-Commerce," Proceedings of the first ACM

Conference on Electronic Commerce, pp. 158–166, 1999.

[4]   S. Staab, and H. Werthner, "Intelligent Systems for Tourism," IEEE Intelligent Systems, Vol. 17, pp. 53–66, 2002.

[5]   L. Khan, D. McLeod, and E. Hovy, "Retrieval Effectiveness of an Ontology-based Model for Information Selection," Very Large Data Bases, Vol. 13, pp. 71–85, 2004.

[6]   H. M. Haav, "A semi-automatic method to ontology design by using FCA," Proceedings of the 2nd International CLA Workshop, Ostrava, pp. 13–25, 2004.

[7]   C. H. Hsu, "Automatically Constructing Ontology on Semantic Web," M.S. thesis, Fu Jen Catholic University, Taiwan, 2004.

[8]   M. Obitko, V. Snasel, J. Smid, and V. Snasel, "Ontology Design with Formal Concept Analysis," In V. Snasel, and R. Belohlavek, (eds.) Concept Lattices and their Applications, Ostrava: Czech Republic, pp. 111–119, 2004.

[9]   G. Stumme, and A. Maedche, "FCA-MERGE: bottom-up merging of ontologies," Proceedings of the Seventeenth International Conference on Artificial Intelligence, pp. 225–234, 2001.

[10] Y. Zhao, X. Wang, and W. Halang, "Ontology Mapping based on Rough Formal Concept Analysis," Proceedings of the Advanced International Conference on Telecommunications and International Conference on Internet and Web Applications and Services, pp. 180, 2006.

[11] B. Ganter, and R. Wille, "Applied lattice theory: Formal Concept Analysis," In G. Grater, (edition), General Lattice Theory, Springer, Berlin, 1997.

[12] G. Schreiber, B. Wielinga, and W. Jansweijer, "The kactus view on the "o" word," Workshop on basic ontological issues in knowledge sharing: international joint conference on artificial intelligence, pp. 159–168, 1995.

[13] A. Bernaras, I. Laresgoiti, and J. Corera, "Building and reusing ontologies for electrical network applications," Proceedings of European Conference on Artificial Intelligence, pp. 298–302, 1996.

[14] S. William and T. Austin, "Ontologies," IEEE Intelligent systems, Vol. 14, pp. 18–19, 1999.

[15] B. Chandrasekaran, J. R. Josephson, and V. R. Benjamins, "What are ontologies, and why do we need them?" IEEE Intelligent systems, Vol. 14, pp. 20–26, 1999.

[16] J. Chaffee and S. Gauch, "Personal Ontologies for Web Navigation," Proceedings of Conference on Information Knowledge Management, pp. 227–234, 2000.

[17] R. Wille, "Restructuring lattice theory: An approach based on hierarchies of concepts," In I. Rival, (edition) Ordered Sets, Reidel, Boston-Dordrecht, pp. 445–470, 1982.

[18] A. Formica, "Ontology-based concept similarity in Formal Concept Analysis," Information Sciences, Vol. 176, pp. 2624–2641, 2006.

[19] K. E. Wolff, "A first course in Formal Concept Analysis - how to understand line diagrams," Advances in Statistical Software, Vol. 4, pp. 429–438, 1994.

[20] P. Cimiano, A. Hotho, G. Stumme, and J. Tane, "Concept Knowledge Processing with Formal Concept Analysis and Ontologies," Proceedings of Second International Conference on Formal Concept Analysis, pp. 189–207, 2004.

[21] W. H. Delone and E. R. Mclean, "Information System Success: The Quest for the Dependent Variable," Information Systems Research, Vol. 3, pp. 60–95, 1992.

[22] H. D. William and R. M. Ephraim, "Information Systems Success: The quest for the dependent variable," Information Systems Research, Vol. 3, pp. 60–95, 1992.