

Applying Adaptive Hybrid Recommendation Technology for Searching Algorithm Learning Articles

Shu-Chen Cheng, Shih-Che Huang

Department of Computer Science and Information Engineering, Southern Taiwan

University of Science and Technology, Taiwan

kittyc@mail.stust.edu.tw

Abstract : In this generation, technology is developed in tremendous speed. The Information on the Internet is increasing at a high speed each day. People get plenty of information via the search engine and spend much time to filter out insignificant information at the same time. Therefore, this research system can filter all kinds of articles to exclude advertisement, news from network bookstore, or insignificant information related to keyword. After filtering, this system will gather up all useful articles and provide users to review. This study proposes the hybrid recommended system with multi-adaptive recommendation to learners. Hybrid recommendation is divided into two ways to recommend “Content-Based Recommendation” and “Collaborative Filtering Recommendation” articles. First, content-based recommendation is based on Term Frequency- Inverse Document Frequency to estimate the characteristic values of articles. Then, we set the weight of difficulty of keywords. After that, people can decide the level of article in the beginning and use it for reference. Besides, collaborative filtering recommendation is applied based on user abilities estimated by IQ tests, quizzes, online tests, ability certificates, and other exams. When the result of two users is similar and one of them thinks it’s useful, the article will be automatically forwarded to the other.

Keyword: Content-Based Recommendation System, Filtering Recommendation, Term Frequency-Inverse Document Frequency.

1. Introduction

1.1 Background information

When searching for articles tend to obstruct because of too much information on the Internet. As shown in figure 1, users need to spend much time filtering information when they read it. In the search result, it also accompanies with irrelevant sites of learning such as: advertisement, information

of bookstore...etc. Owing to this reason, this study uses the hybrid recommendation to assist users to filter the articles they don't need and learn efficiently.



Figure 1. search result by Google

1.2 Purpose of Research

This study will focus on hybrid methods recommended to build a system for recommended articles. The hybrid recommended system is combined with “Content-Based Recommendation” and Collaborative Filtering Recommendation.” Content-Based Recommendation is to analyze the content of articles keyword. And another one Collaborative Filtering Recommendation selects the recommended articles by the ability of learners and peer recommendation.

2 Literature Review

The chapter is to explore how to use the hybrid recommended system. Then introduce the advantage and disadvantage when users use this system and give the advice in detail to improve it. [1]

2.1 Classification

The most important part of Content-Based Recommendation is the classification. We need to analyze the content of article, set keywords, and article categories.

2.1.1 N-Gram word segmentation

After selecting articles by Crawler on the Internet, we need to deal with the content of these articles first. How to analyze these articles? We use each word of the articles to do the cutting so that the article presents the smallest unit of word. Then by smallest word unit of different lengths to make up words, it can form words and sentences. N-Gram word segmentation is used at the beginning[2][3] of the data processing. (Table1)

Table1. length of word segmentation (example: string of Algorithm by Chinese characters)

N-Gram	results
2-Gram	演、算、法、演算、算法
3-Gram	演、算、法、演算、算法、演算法

2.1.2 Term Frequency-Inverse Document Frequency

After using the N-gram word segmentation, the system starts to determine the importance of each string in the article. Take Knapsack algorithm for example, “Knapsack” and “Algorithm” will be the most important word. When the strings of the importance in the article to be found out, these strings are used for the characteristic value of this article.

The characteristic value is determined by Term Frequency-Inverse Document Frequency (TF-IDF). [4][5]

2.2 Recommendation System

2.2.1 Content-Based Recommendation System

On the basis of Content-Based Recommendation System is to analyze the content rather than the user for evaluation. This system will calculate the ability of user’s interest and then calculate the characteristic values of user. Therefore it can find the item user needs. In this study, the most important thing is to recommended articles through keywords and characteristic values. To achieve this recommended system, we need to use characteristic values to make up keyword and given the different scores by different rights then the result of recommendation will be come out.

2.2.2 Collaborative Filtering Recommendation System

CF system is abbreviated form of Collaborative Filtering Recommendation System. This method was successfully applied in a wide variety of fields, such as YouTube, Amazon, etc., are very famous examples. This system is based on the gathering user preferences, personal information, gender, browsing habits and other personal information [6], to make recommendation. Therefore, while user browsing merchandise or other things, CF system will recommend users the section plus the recommended information, not only can make commodities accelerated been viewed but also speed

up information dissemination.

2.2.3 hybrid Recommendation System

Hybrid recommendation System recommended by Content-Based Recommendation System and Collaborative Filtering Recommendation system. We can use these two different systems which have different strength and shortages to improve recommendation system better. Therefore, hybrid will compensate for the shortcomings of their systems to each other. The advantage and disadvantage of two recommendation systems are shown in Table 2. [7][8]

Table 2. Comparisons between Content-Based Recommendation and Filtering Recommendation

	Content-Based Recommendation	Filtering Recommendation
reference value	contents	Information of user
advantage	Quick Recommendation	Find out the preference of user High-automated
shortage	Can't find out the user's preference Much time to do processing information Can't be recommended automatically	New Item Sparsity Cold-start Scalability New user

Both of Content-Based Recommendation System and Collaborative Filtering Recommendation System all have drawbacks, so we combine[9] these two ways to be hybrid recommendation. (Table3)

Table3. hybrid recommendation way

method	hybrid recommendation way
Method of weighting	Take two different recommendation methods to calculate, and in accordance with the result of calculate to given different weighting.
Method of selection	If there is users' information, use Collaborative Filtering Recommendation and Content-Based Recommendation. If not, use Content-Based Recommendation only.
Method of mixing	When both content have recommendation systems, use the two recommended ways to calculate, and then generate a recommendation.

3 Research Method

3.1 System Framework

This study is based on the self-developed Algorithm platform to design a hybrid recommendation system to give users search teaching articles. (Figure.2) We use the Algorithm database to do the recommendation. The hybrid recommendation is divided into two modules which are “Content-Based Recommendation module ”and” collaborative filtering recommendation module”. In the Content-Based Recommendation module, the article needs to through preprocessing module first and then use N-Gram for word segmentation. After that, the characteristic value of the article will be selected which will be the keyword to analyze the degree of the article difficulties and make recommendation. The Collaborative Filtering Recommendation use the calculated module to calculate the users’ ability. Then in accordance with the users’ ability to recommend or accordance with the recommendation module to allows users with the same ability to recommend the article and share with them. Therefore, it will have many different recommended articles to let users read.

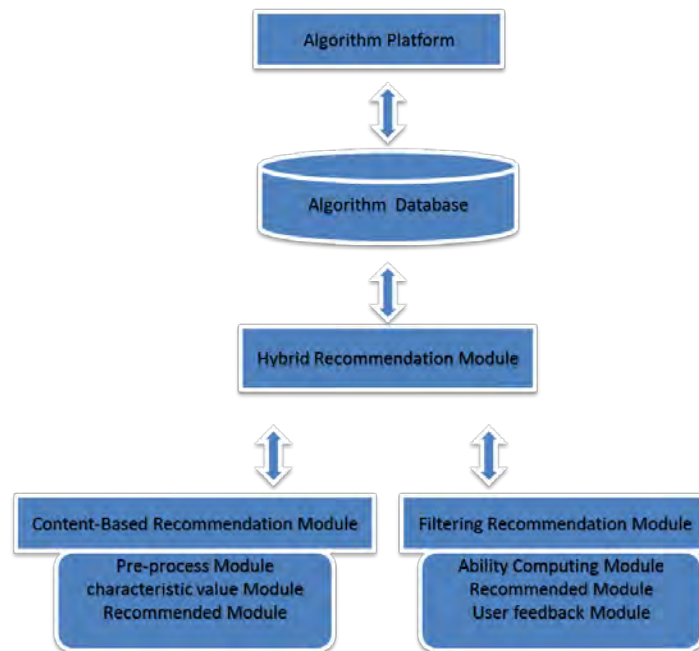


Figure.2 System Framework

3.2 Content-Based Articles Recommendation Method

When the analysis of the article that will be processed by the keyword of characteristic value and then in accordance with the value of calculation to recommend users. The weighting of keyword is the most important part of this method which is in accordance with given to those area of expertise; conducting analysis to improve ease of articles and recommended stability.

3.2.1 Articles Pre-process

Articles are pre-processed by N-Gram segmentation to handle the database articles. There are many Stop Words that needs to be removed. Chinese Stop word List is shown in Table 4.

Table4. Examples of Chinese Stop Words List

Stop Words
嗎 (麼)、吧、呢、罷、呀、啊、啦、哇、哪、罷了、而已、我、你、您、我們、他們、是、全部、所有、不知道

3.2.2 Initial Item Bank

This system is to determine whether the article for Algorithm teaching articles and set up the degree of difficulties to achieve recommendation. It needs to set up positive, negative item bank, and weighting. The positive item bank put in the words represents Algorithm; however, the negative item bank represents non-Algorithm.

First of all, the initial item bank set up by the artificial, the positive item bak select the keyword that can represent Algorithm from the book of Algorithm. The more abstruse terms set up higher weighting that represents the more difficult article. On the contrary. The negative key words comes out from the advertisement of on-line bookstore. Therefore, the system can exclude those of words to choose Algorithm article. (Figure.3)

Forward keyword	Weights	Negative keyword	Weights
Hoffmann	1	The authors introduce	1
greedy	1	Library	1
Binary	1	Purchasing	1
merge	1	Books	1
Bubble	2		
Prim	2		
Kruskal	2		

Figure.3 Positive and Negative Keywords and Weightings

3.2.3 Recommendation Module

When the item bank set up, then the keywords in the item bank will be compared and do the weighting accumulated. If the article in the positive item bank has higher weighting, then it can be used as recommended priorities. This module can be used while users searching the article in accordance with the recommended system.

3.3 Collaborative Filtering Recommendation Methods

The system, another recommendation System, is Collaborative Filtering Recommendation. Since the Algorithm belongs to professional in the field, not like English article which contains preference, and article style such as exercise, movies, and music. Therefore, in the study we propose a Peer Recommendation, which is based on user's ability and preference from groups. They can find appropriate article by sharing and watching the recommended article from the same ability users.

3.3.1 Recommendation by Users' Abilities

Although Algorithm is in the same article; however it can be different by users, articles and the degree of the articles. Sometimes it's too difficult to let beginning user or not good at logical thinking user to understand. On the other hand, if it's too easy to understand, it can't help user who has higher ability learning. This study will be in accordance with the user's ability to recommend the article (Recommend Articles), initially to collect the user's test scores, and recorded data repository, such as intelligence tests, proficiency examinations test program, license, Algorithm online quizzes, etc. After that, set the initial capacity and in accordance with article difficulty keyword weighting to classify and control user ability to give a different degree of difficulty of the article. [11][12]

3.3.2 Group Preferences Recommendation Method

When one of users think the article is helpful, the system will record it. Next time, if there is a user of same ability appears, the system will recommend the article to him. Therefore, user can search appropriate article easily.

3.4 Hybrid Recommendation Method

When user enters the keyword in the Content-Based Recommendation, it will search out the article. After that, the collaborative Filtering Recommendation will offer recommended article to the user. Because Algorithm articles are not much, when users all recommended an article, the article will be exposed highly. Because of this reason, this type of the article order will increase and be found easily. If this article isn't related to the keyword, it will give the system in return. The system will reduce the article exposure to avoid system erroneous judgment that causing users to research on the system distress to make the system more perfect. If the feedback has found that the article isn't related to the

mathematics article or the article can't be connected, the system will reduce the recommendation. When it comes to an certain amount, the system will not recommend to the user and remove it.

4 Conclusions

In this study, we design a hybrid recommendation system which offers user to find appropriate articles according to their abilities, keywords, or groups recommendation articles so that it can achieve hybrid recommendation effect. Users can follow their own abilities or similar users' to read the recommended articles in accordance with collaborative filtering recommendation system. By the way, the feedback from user can increase the article exposure and improve the content-based recommendation. All of these systems can help users find appropriate algorithm articles and enhance learning abilities.

Acknowledgement

This study is supported in part by the Science Council of the Taiwan under Contract No. NSC 100-2511-S-218 -008 -MY3

References

- [1]Adomavicius G. and Tuzhilin A. (2005), Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions, IEEE Transactions on Knowledge and Data Engineering, Vol. 17, No. 6, pp.734-749
- [2]Jerome R. Bellegarda(2000),Exploiting Latent Semantic Information in Statistical Language Modeling,Proceedings of the IEEE,Vol.88,No.8,pp.1279-1296.
- [3]Qiang Yang and Zhang,H.H.(2003),Web-Log Mining for Predictive Web Caching,IEEE Transactions on Knowledge and Data Engineering,Vol.15,No.4,pp.1050-1053.
- [4]Wu HC,Luk RWP,Wong KF,Kwok KL(2008),Interpreting tf-idf term weights as making relevance decisions,ACM Transactions on Information Systems,Vol.26,No.3,pp.13:1-13:36.
- [5]Salton G,Buckley C(1988),Term-weighting approaches in automatic text retrieval,Information Processing & Management, Vol.24,No.5,1988, pp.513-523.
- [6]Rokach L.and Kisilevich S.(2012),Initial Profile Generation in Recommender Systems Using Pairwise Comparison,IEEE Transactions on Systems,Man,and Cybernetics,Part C:Applications and Reviews,Vol.42,No.6,pp.1854-1859.
- [7]Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989).User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. Manag. Sci., vol. 35, no. 8, pp. 982–1003
- [8]Ahmad A. Kardan and Mahnaz Ebrahimi (2013), A novel approach to hybrid recommendation systems based

on association rules mining for content recommendation in asynchronous discussion groups, *Information Sciences*, Vol. 219, pp.93-110.

- [9]Robin Burke(2002),Hybrid Recommender Systems Survey and Experiments.,User Modeling and User-Adapted Interaction,Vol.12,No.4,pp.331-370.
- [10]Wikipedia(n.d.),"Stop words",https://en.wikipedia.org/wiki/Stop_words.
- [11]Srinivasan, A. (1985). Alternative measures of system effectiveness: Associations and implications. *MIS quarterly*, 9, 243-253.
- [12]Fishbein, M., & Ajzen, I. (1975). Belief, attitude, intention and behavior: An introduction to theory and research.