

Study on Collaborative Filtering Algorithm of Fusion Context-Aware Calculation

Yuanyuan Yang and Guiyun Zhang

Abstract—In this paper, the project of context awareness information is introduced into the traditional collaborative filtering algorithm based on item, modeling by context awareness, put forward a kind of combination of context-aware computing collaborative filtering algorithm. The algorithm of the project context similarity calculation of similar fusion to the project, and then preference prediction and recommend using collaborative filtering algorithm, test the indicators of algorithm optimization, efficiency and accuracy of the proposed algorithm.

Index Terms—Collaborative filtering algorithm, context-aware computing, similarity degree.

I. INTRODUCTION

With the development of information technology and the Internet, people gradually enter the era of information overload from the age of information shortage [1]. In order to solve the problem of information overload, there are many solutions, and the recommendation system is the excellent representative. Recommendation algorithm does not need to provide clear requirements, he is by learning the user's interests, or the characteristics of the information. The problem of information overload can be solved by directly recommending new information that users may be interested in. In music gradually become an indispensable part of people's life and music populations have exploded, the choice of music has become an obsession, so music recommendation is also becoming more and more important. But the traditional recommendation is not ideal, and we can't just recommend hot music for users. So we need to consider various factors to meet the user's personalized demand for music. The traditional recommendation algorithm deals with the binary relationship between user and music. Only considering the similarity between users or music, and does not take into account the user's context-aware information and music context-aware information, such as time, location, weather, mood, and users of social information, and so on. And this information will greatly influence users' preferences and choices in the recommendation algorithm. Combining music context-aware of data information, such as the influence of the location information to the user, while driving, user preferences choose move feeling, excited, happy music class, but alone at home or study, the user will choose quiet, relaxing music. In this way, the context-aware

information of music data is combined into the recommendation algorithm to make recommendations more consistent with the user's interest. This recommendation algorithm found and effective utilization of context-aware information to calculate the model called context-aware computing, and the recommended method has been widely applied in various areas [2], [3].

II. THEORETICAL BASIS

A. Traditional Item-Based Collaborative Filtering Algorithm

The item based collaborative filtering algorithm is mainly divided into two steps. One is to calculate the similarity between items. The second is to generate a list of recommendations based on the similarity of the item and the user's historical behavior [4]. The measure methods of item similarity include three kinds: cosine similarity, Pearson correlation coefficient and cosine similarity of correction. Suppose item i and item j , using $N(i)$, $N(j)$ to express the set of users who like item i , j , and the similarity between the items are represented by $\text{sim}(N(i), N(j))$. The similarity degree of the item is presented as follows:

$$\text{sim}(N(i), N(j)) = \cos(N(i), N(j)) = \frac{|N(i) \cap N(j)|}{\sqrt{|N(i)| |N(j)|}} \quad (1)$$

Equation (1) shows the similarity between item i and item j . $|N(i)|$, $|N(j)|$ respectively, the number of users who like item i and j (i.e., the number of users who have acted on the item). And the numerator, $|N(i) \cap N(j)|$ indicated that the number of users who like item i and item j . On this basis, we predict the interest degree of user u for unknown item j . Using P_{uj} to indicate the user's interest in item j , using R_{ui} to represent the interest of the known user u to item i , using W_{ij} to express the similarity between item i and j . With $P_{uj} = R_{ui} * W_{ij}$, the similarity between item i and j is multiplied by the user's known interest in item i , and obtains the interest of user u on unknown item j . The specific calculation formula is as follows:

$$P_{uj} = \sum_{i \in N(u) \cap S(j, K)} W_{ij} * R_{ui} \quad (2)$$

Here, $N(u)$ is a collection of users' favorite items, and $S(j, K)$ is a collection of items j similar to item i . And the number of items j is K .

B. Context and Context-Aware Calculations

Context is a very broad concept that has different definitions in different areas of research. Dey proposed context definition is widely used in the recommendation

Manuscript received February 2, 2018; revised April 28, 2018. This work was financially supported by the National natural science fund project (61572358) and Tianjin natural science fund project (16JCYBJC23600).

Yuanyuan Yang and Guiyun Zhang are with the Department of Computer and Information Engineering, Tianjin Normal University, China (e-mail: echo34yy@163.com, dyxy1999@126.com).

algorithm: context is any information used to describe the state of the entity. In this case, an entity can be a person, place, or an object that is interrelated to the user and application (including the user and the application itself) [5]. There are four types of context-aware information, including the physical context of time, place, weather and temperature. The social context of the user's identity and social objects, the state context of users' mood, behavioral goals, experiences, cognitive abilities, etc. The Interactive media context of recommended information types (text, images, video, and so on) [6]. The context-aware information corresponds to the recommendation algorithm, which refers to the user and the item, which corresponds to music data, which is user and music. In music data users and music's context is also varied, including music's mood, style, theme, age, musical instrument and so on.

If there is a valid context property in the proposed algorithm, the context set of the data in this algorithm can be defined as:

$$\text{Context} = \{C_1, C_2, \dots, C_i, \dots, C_k\} \quad (3)$$

Each of the context attribute C_i values is discrete. the item with different Context property, defined as $C_i, C_i \in \text{Context}$, and different Context properties have different values, the attribute values are discrete, C_i value set for $\{c_{i1}, c_{i2}, \dots, c_{ij}, \dots\}$, $c_{ij} \in C_i, C_{ij}$ describes a specific value for the context attribute C_i that comes with the item itself [7]. For example, in the music recommendation algorithm, we define the context of music data : $\text{Context}\{C_1, C_2\}$, C_1 is the "theme" attribute of music data, C_2 is the "genre" attribute of music data, C_1 has different values {love songs, red songs, children's songs}, C_2 also has different values {classical, hip hop, rock}.

Context-aware computing is the use of context-aware information in the recommendation algorithm to provide users with more accurate recommendation services. There are three ways to integrate context-aware computing into the recommendation algorithm: context prefiltering, context post-filtering, and context model [8], [9]. (1) Context prefiltering refers to the recommended mode, recommended the results in the generation before, using the current context-aware information filter out irrelevant user preferences data, so as to build current context-aware information relevant recommendation data set; Then, the traditional recommendation method is used to process the selected data to predict and generate the recommendation results that meet the current context conditions. (2) The context post-filtering refers to ignore context-aware information at the beginning, recommend the use of traditional algorithm is recommended, to get initial suggestion list, and then filter the list of recommendations based on the current context of the target user. (3) In this way of context model, the information is integrated directly with the traditional recommendation algorithm. Although the calculation is more complex, it can better exploit the user interest and improve the recommendation performance [7].

In this paper, context modeling is applied to improve the accuracy of the proposed algorithm by combining the context similarity of the project with the similarity of the project to

improve the accuracy of the project similarity calculation.

III. COLLABORATIVE FILTERING ALGORITHM OF FUSION CONTEXT-AWARE CALCULATION

A. Introduce Context Similarity in Item Similarity

In the traditional collaborative filtering algorithm, we only calculated the similarity of the item based on the user's behavior of the item. But because of too little user ratings, or the user is not on this item had a behavior, such as the new item, the resulting data sparseness problem and cold start problem. And if in the original project based on similarity, add similar context-aware of the item, to a certain extent can improve data sparseness problem due to less user rating data, for the new item, combined with the item information on current context-aware, to a certain extent can also improve data cold start problem.

B. Calculation of Context Similarity

Similarity in the context of the introduction of context-aware computing recommendation algorithm, we think that the similarity is not only related to the user rating of the item, and relevant to item its own context-aware information, item their own context similarity is higher, the degree of similarity between items should also be increased accordingly. For example, in the music recommendation, we calculate the similarity of song a, b, and c. Assuming that all three of these songs get the same user scores, we get $\text{sim}(a, b) = \text{sim}(b, c) = \text{sim}(a, c)$, But if you can introduce the context-aware information of song a, b, and c, let's say that song a, c's "theme" attribute value is "love song", and b's "theme" attribute value is "red song". So we can get $\text{sim}(a, c) > \text{sim}(a, b) = \text{sim}(b, c)$.

In this paper, the context similarity calculation method is proposed to determine the context similarity of the item by using the common item context instance between the item i and j . If $C(i)$ is the set of context instances owned by item i and the number of context instances that $C(j)$ has for item j , the context similarity of item i and j is calculated as follows:

$$\text{sim}(C(i), C(j)) = \cos(C(i), C(j)) = \frac{|C(i) \cap C(j)|}{\sqrt{|C(i)| |C(j)|}} \quad (4)$$

Equation (4) shows the context similarity between item i and item j . $|C(i)|, |C(j)|$ is the number of context instances owned by item i , and the molecular $|C(i) \cap C(j)|$ is the number of context instances jointly owned by item i and item j .

In the process of calculating the similarity degree of item similarity, we introduced the weighting factor $\alpha, 0 < \alpha < 1$, and the similarity calculation formula of item i and item j is as follows:

$$\text{sim}(i, j) = (\alpha \text{sim}(N(i), N(j)) + (1 - \alpha) \text{sim}(C(i), C(j))) / 2 \quad (5)$$

C. Design of Collaborative Filtering Algorithm for Context-Aware Computing

Context-aware collaborative filtering algorithms are mainly divided into the following steps.

- 1) First, the user - project inversion list is established, as shown in Table I, such as the user has acted on the items (51, 52, 53, 54, 55). We will call (51, 51), (51,

52), (51, 53), (51, 54).....to 1. Continue to populate the co-occurrence matrix of the item. After that, we use the formula (1) cosine similarity calculation method to normalize the matrix and remember it as W_n .

- 2) Set up item context-aware information - inversion lists, convenient statistical different items have the common context-aware of information, and thus fill the item context similarity matrix, using the formula (4) the cosine similarity matrix for normalized processing, for W_c .
- 3) Using formula (5) for the two similarity matrices obtained. The similarity matrix that combines the context similarity is recorded as W_{ij} . $W_{ij} = (\alpha W_n + (1 - \alpha)W_c)/2$
- 4) Set up the user's interest matrix for the item, according to the user group, such as according to user 1 to 51,52,53...weight of interest, established [(51, 13883), (52, 11690), (53, 11351), (54, 10300), (55, 8983)...].Such an interest matrix, the user does not have an item weight of the behavior is 0, the user interest matrix is R_{ui} .

According to formula (2), multiply the third and fourth steps of the item similarity matrix W_{ij} and the user interest matrix R_{ui} . Get the user's interest in the unknown items and rank them according to the weight, and get the final recommendation.

IV. EXPERIMENT AND RESULT ANALYSIS OF COLLABORATIVE FILTERING ALGORITHM BASED ON CONTEXT PERCEPTION CALCULATION

A. Experimental Data Set

The data set used in this article is a set of user activity records from Last.fm online music systems, including user social information, music tags, listening information, and more. The data set includes more than 2,000 listening records for more than 18,000 songs, and the last.fm data set contains information about music. The experiment uses the user's behavioral data of music as a test set and a training set, a ratio of training: test = 3: 1. Table I is part of the user - project inversion list. Table II is part of the music data. Table III is part of the information that the music data contains for context-aware information.

TABLE I: USER ITEM TABLE

User ID	Item ID, weight
2	[(51, 13883), (52, 11690), (53, 11351), (54, 10300), (55, 8983).....]
66	[(51, 910), (59, 367), (61, 603), (68, 459), (72, 1303).....]
4	[(51, 228), (53, 686), (64, 420).....]

TABLE II: MUSIC ITEM COMPARISON CHART

User ID	Item ID, weight
Item ID	name
1	'MALICE MIZER'
2	'Diary of Dreams'

TABLE III: MUSIC ITEM CONTEXT-AWARE INFORMATION

Item ID	Name	Context-aware
52	Morcheeba	chillout
		downtempo
		electronic
		trip-hop
	

B. Evaluation Index

In this paper, the improved algorithm is proposed to recommend K items of user u as $R(u)$, and the set of items that user u likes on test set is $T(u)$. Then evaluating the accuracy of the algorithm by accuracy rate, and the formula as follows:

$$\text{Precision} = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} |R(u)|} \quad (6)$$

C. Result Analysis

According to the improved algorithm, the recommendation result is generated by fusing the data context-aware information: There are two parameters in the algorithm that need to be adjusted according to the actual situation. They are recommended algorithms that are similar to the recommended users in the calculation process. In the similar item set K take different values, and the different weight factor α , the fusion of context similarity recommendation algorithm to get the recommended result.

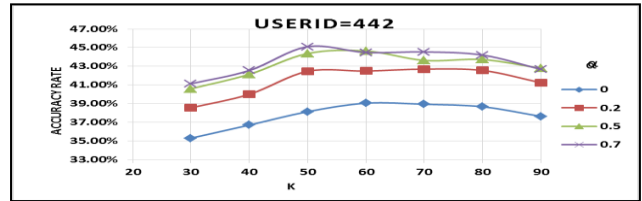


Fig. 1. User ID = 946 at different α and K under the recommended accuracy.

According to Fig. 1, the recommended results of user422, we can see that the recommended results reach the peak value with the accuracy of $\alpha = 0.5 \sim 0.7$ and $K = 50 \sim 70$, and the increase of α increases the accuracy of the recommended results. Which validates the idea that the fusion context awareness information proposed in this paper will improve the accuracy of the recommendation, but the accuracy of a fluctuates between (0.5, 0.7) for α , which also verifies that idealism can't be achieved by considering only the item's contextual relationship and considering only the user-likeness of the item the result of.

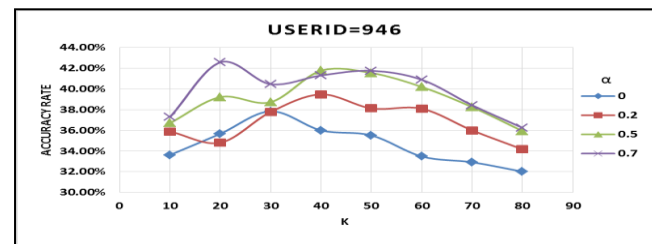


Fig. 2. User ID = 946 at different α and K under the recommended accuracy.

According to Fig. 2, the recommended results of user946, we can see that the recommended results reach the peak value with the accuracy of $\alpha = 0.5 \sim 0.7$ and $K = 20 \sim 50$, while the

increase of α increases the accuracy of the recommended results. The value of α is 0.5 and 0.7, the accuracy of the curve of some cross, the data overlap traces for the user946 recommended results of the degree of discrimination is small, indicating that when the context-aware information and the similarity of the item when the fusion, the excessive dependence on context-aware information cannot be obtained very good result.

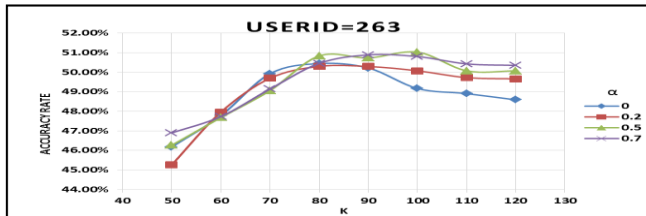


Fig. 3. User ID = 263 at different α and K under the recommended accuracy.

According to Fig. 3, the recommended results of user263, we can see that the recommended results reach the peak value at the accuracy of $\alpha = 0.5 \sim 0.7$ and $K = 80 \sim 100$, while the increase of α increases the accuracy of the recommended results slightly. Under different α values, The rate of the curve has a significant intersection, the data overlap traces, indicating different values of α , the user263 recommended results for the small degree of differentiation, but as the value of α increases, With the increasing proportion of context-aware information, the accuracy rate increased slightly.

Through the analysis and comparison of the recommended results of different users, it is found that the parameter α achieves good recommendation results at $0.5 \sim 0.7$. However, due to the influence of the actual data, the value of K fluctuates more, but the value of K will also get a stable range with the value of α . According to the trend of recommendation results of different users, with the increase of α , α is recommended in the range of 0.5 to 0.7, which is based on context-aware information, and the accuracy of the recommended results is improved. The proposed fusion context-aware information Improve the accuracy of the recommended results of this idea.

V. CONCLUDING REMARKS

Based on the traditional item-based collaborative filtering recommendation algorithm, this paper combines the item context-aware information to optimize the accuracy of the algorithm. It is very important to select the context-sensitive information. This article focuses on the genre characteristics of the music data. After that, it can also explore the context-aware information and the user's context-sensitive information.

ACKNOWLEDGMENT

This work was financially supported by the National

natural science fund project (61572358) and Tianjin natural science fund project (16JCYBJC23600).

REFERENCES

- [1] Information overload. [Online]. Available: http://en.wikipedia.org/wiki/information_overload
- [2] R. Li, "A survey of context-aware computing and its system infrastructure," *Journal of Computer Research & Development*, vol. 44, no. 44, pp. 269-276, 2007.
- [3] Y. L. Zhang, M. M. Ma, and S. P. Wang, "Research of User-Based Collaborative Filtering Recommendation Algorithm Based on Hadoop," in *Proc. International Conference on Computer Information Systems & Industrial Applications*, 2015.
- [4] J. Ye, "Cosine similarity measures for intuitionistic fuzzy sets and their applications," *Elsevier Science Publishers B. V.*, 2011.
- [5] A. K. Dey, "Understanding and using context personal and ubiquitous computing journal," *Personal & Ubiquitous Computing*, vol. 5, no. 1, pp. 4-7, 2001.
- [6] S. W. Lai *et al.*, "Hybrid recommendation models for binary user preference prediction problem," *Kdd Cup*, 2011.
- [7] Y. Ying *et al.*, "Personalization recommender system based on cloud-computing technology," *Computer Engineering & Applications*, 2015.
- [8] G. Adomavicius and A. Tuzhilin, *Context-Aware Recommender Systems. Recommender Systems Handbook*, Springer US, 2015, PP. 2175-2178.
- [9] M. Braunhofer *et al.*, *Context Dependent Preference Acquisition with Personality-Based Active Learning in Mobile Recommender Systems. Learning and Collaboration Technologies. Technology-rich Environments for Learning and Collaboration*, Springer International Publishing, 2014, pp. 105-116.



technology.

Yuanyuan Yang was born in Shanxi and received her bachelor's degree from Tianjin University of Commerce in 2015. She went to Chaoyang University of science and technology to study and communicate as an exchange student in 2017. Currently, She is studying for a master's degree in Tianjin Normal University, majoring in computer application technology, and her tutor is professor Guiyun Zhang. Her research direction is data mining, big data

technology. She has published a paper at the Eleventh International Conference on information. The title is *Analysis of Product Association of E-commerce Website*.



Guiyun Zhang was born in January 1965. In 2002, She completed her bachelor's degree and master's degree in the Department of Mathematics, Tianjin Normal University. In 2006, she received her Ph.D. from the Department of Information Technology, Beijing University of Science and Technology. From 2016 to 2010, she worked as a postdoctoral researcher in the National Key Laboratory of Cognitive Science and Learning at Beijing Normal University. Her research area includes artificial intelligence, data mining, big data, and education technology and so on.

She is director of the Institute of Artificial Intelligence in China, the Chinese Society of Artificial Intelligence, deputy director of the Intelligent Computing Discrete Branch, director of China Society of Basic Computer Education in colleges and universities, China Computer Society YOCSEF Tianjin Subcommittee academic committee (AC), the cognizance expert of Tianjin high and new enterprise. She has published more than 40 papers in EI and magazines. She has presided over and participated in more than 20 projects, including national natural science foundation project, Tianjin natural science foundation project, key project of science and technology support program, more than 20 published textbooks, etc.

Prof. Zhang has won the first prize of the city youth teacher basic skills contest, the advanced individual of Tianjin labor competition demonstration, and the famous teacher of Tianjin normal university.