Application of Time-Weighted User Behavior Analysis Collaborative Filtering Algorithm in Short Video

Zhiqiang Li1, Yang Liu1,* and Difei Yu1

1School of Computer and Communication Engineering, Liaoning Shihua University, Liaoning, China

*862720689@qq.com

Abstract. In the big data environment, the information overload is serious. The traditional search engine does not satisfy the user's demand for information. The user wants to obtain the user's acquired information conveniently and quickly, and the information producer wants to transmit the information to the target user. The intermediary that acts as both is the recommendation system. Recommendation system is an important way to deal with massive information. This technology brings great convenience to viewers in the field of short video. It is very difficult to extract features from missing short video data and from content information of short video. About this question. This paper proposes a time-weighted user behavior analysis collaborative filtering algorithm. This paper starts with the behavior of short video users. In order to compensate for date sparseness, the explicit and implicit behaviors of users are deeply analyzed and a user-video scoring model is established. The user's interest in video will change with time, so we use time weighting to adjust the time variable. Finally, the K most similar neighbors are extracted and recommended to the user in order from high to low. Experiments show that the proposed algorithm alleviates sparsity of date and improves the accuracy of recommendation.

1. Introduction

In recent years, with the rapid development of mobile internet, new media technology is supported by technology. With the support of related technologies such as data storage and video compression, short video has become the mainstream of network information dissemination from the previous mass information dissemination. Short videos are usually short, ranging from a few seconds to a few minutes, and have some innovative features. There are currently dozens of short video apps on the market. However, with the increasing number of short video uploaders and uploaded videos, the quality of short videos has become a huge problem. Nowadays, it has become the biggest problem for users to get interesting short videos from a large number of videos.

With the development of the recommendation system, many scholars and experts have proposed many different recommendation algorithms. The earliest recommendation algorithm for video is VideoRech's online video recommendation system, which is a real-time recommendation system proposed by Mei [1]. Personalize short videos based on the video the user is watching. But this system only considers some superficial information of the video itself. For example: title, author, release date. It does not build interest models for video users. However, short videos contain little or no textual information, and this algorithm does not work for short videos. It is not realistic for the tens of thousands of video data resources. Xing-yao Yang [2]. Add different attributes to the recommended project through artificial annotation. Further look for the relationship between user-attribute and attribute-video. Since...
most of the short video uploaders are ordinary Internet users and uploaded videos are short. The attribute of the video is blurred, and the accuracy of the recommendation is reduced. Jian-wei Niu [3] puts forward the idea of video recommendation based on emotional analysis. Recommend appropriate videos by capturing the current user's emotions. It is suggested that video recommendation is dynamic, but the user's emotion changes with time, mood and weather. This change is one that is extremely complex and unmeasurable. Wei-wei Song [4] proposed a time-weighted label algorithm, and he considered that time variation would have an impact on the recommended effect. But this is mostly recommended for movies and TV shows. Movies and TV shows contain more tags, such as actors, directors, movie types, 3D, and more. Short videos have very few tags, and most of the videos on short video apps are spontaneously uploaded by users. Most of these users do not have professional analysis of video content, and the content and features of the video may be unclear or even erroneous. It is very difficult to extract features from short video content. In addition, the user interest of the video may also change over time. The user may have different evaluations of the same video at different times. This paper uses a Time-Weighted User Behavior Analysis Collaborative Filtering Algorithm (TWUBCF) for these problems.

2. Definition of problems and basic methods

2.1. User behavior information analysis
The information of video user's behavior is stored in the form of logs on the mobile side. The definition of user behavior in the video recommendation system [5] usually includes five aspects: the behavior object, the user who generated the behavior, the content of the behavior and the content of the behavior, and the type of behavior. The behavior information of short video users mainly includes browsing time, browsing records, sharing friends, comments and scoring. These behaviors exist on the computer in the form of logs. According to the user's behavior information, there are two types: explicit behavior and implicit behavior. Explicit behavior requires the user to express his or her own interest in the video. Implicit behavior refers to the inability to directly see the like and dislike of the content.

2.2. Modeling of user behavior information
In order to find the intrinsic relationship between users and videos, this paper adopts the user behavior information modeling method. This paper establishes a model that can reflect user's interest in video content. Explicit behavior is used to judge whether the user is related to the video. In the implicit user behavior [6], we can set a reference value for the viewing time, the number of views, and the number of sharing to determine whether the user is related to the video content. Next, we use an algorithmic flow chart to verify the relevance of user and content. As shown in figure 1.

![Figure 1. User-Video relationship judgment flow figure.](image-url)
From Figure 1, we can see that the user likes the degree is 0, or is 1. 0 means that the user has no relationship with the video, and 1 means that the user has a relationship with the video. Based on the above, we create a scoring matrix \( M \) about user-video. As shown in Table 1, the matrix represents the user's rating of video, and it represents the user's preference for video.

### Table 1. User-Video scoring matrix table.

| User | Video | Item1 | Item2 | \( \ldots \) | Itemn |
|------|-------|-------|-------|-------------|-------|
| U_1  |     | R_{1,1} | R_{1,2} | \( \ldots \) | R_{1,n} |
| \( \vdots \) | \( \vdots \) | \( \vdots \) | \( \vdots \) | \( \vdots \) | \( \vdots \) |
| U_s  |     | R_{s,1} | R_{s,2} | \( \ldots \) | R_{s,n} |

In addition, in order to solve the problem of large data cold start, we let new users register an attribute table before using short video APP, which includes a series of information such as age, gender, region, occupation type, etc.

#### 2.3. Calculating similarity of users

In this paper, the modified cosine similarity [7] is used. It treats user ratings as vectors in space. The modified cosine similarity formula can be expressed as equation (1).

\[
Sim(U_a, U_b) = \frac{\sum_{i \in I'} (R_{a,i} - \bar{R}_a) \times (R_{b,i} - \bar{R}_b)}{\left(\sum_{i \in I'} (R_{a,i} - \bar{R}_a)^2\right)^{1/2} \times \left(\sum_{i \in I'} (R_{b,i} - \bar{R}_b)^2\right)^{1/2}}
\]

(1)

In the above equation (1), \( R_{a,i} \) is the rating of the short video by the user \( U_a \), \( R_{b,i} \) is the rating of the short video by the user \( U_b \), and \( \bar{R}_a \) and \( \bar{R}_b \) are the average of the ratings of the user \( U_a \) and the user \( U_b \). Because everyone has different hobbies and experiences, different user ratings are also different. In order to cope with the above problem, the modified cosine similarity method first calculates the average value of the user for all videos, and then subtracts this average to weaken the difference in the user's scoring scale. The final result is that the value of \( Sim(U_a, U_b) \) is within the interval \([0,1]\).

#### 2.4. Improving the computation of similarity measure

Computing the similarity between short video users mainly depends on the short video they scored together. Assuming that there are fewer videos scored by video users together, there is some uncertainty in this similarity. Herlocker [8] et al. proposed adding a parameter to improve the accuracy of similarity. Based on this theory, we define \( I' \) as a video set that short video user \( U_a \) and short video user \( U_b \) score together, \( I' = I_a \cap I_b \). A threshold \( \gamma \) is set artificially and compared with \( |I'| \). The improved similarity is shown in equation (2).

\[
Sim'(U_a, U_b) = \frac{\min([I'], \gamma) \times Sim(U_a, U_b)}{\gamma}
\]

(2)

From the algorithm of equation (2), we can get that the value range of the improved similarity quantity is still in \([0, 1]\). If the short video user \( U_a \) and the short video user \( U_b \) have a larger number of common scoring videos, according to \( |I'| \geq \gamma \), we can get \( Sim'(U_a, U_b) = Sim(U_a, U_b) \). If the short video user A and the short video user B have a small number of common scoring videos, according to \( |I'| < \gamma \), we can get \( Sim'(U_a, U_b) < Sim(U_a, U_b) \). Through the above methods, we can improve the accuracy of user similarity.
3. Time-weighted user behavior analysis collaborative filtering algorithm

3.1. Calculating time weights
As time goes by, the user's interest will also change [9]. In a short period of time, the user's interest in watching video is basically stable, so the recently added user behavior has a great impact on recommendation prediction. Based on this problem, this paper proposes to add the factor of time to the user behavior analysis collaborative filtering algorithm. This is Time-weighted user behavior analysis collaborative filtering algorithm.

\[
\text{logistic}(t_i) = \frac{1}{1 + e^{-t_i}}
\]

In the above equation (3) \( t \in [-1,1] \), \( \text{logistic}(t_i) \in (-1,1) \). \( \text{logistic}(t_i) \) is a monotonically increasing function. With the increase of time \( t \), \( \text{logistic}(t_i) \) also increases and remains in the range of \((0,1)\). In this paper, we will map the variation range of time \( t \) to \([0,1]\) using a standardized method. The value of \( \text{logistic}(t_i) \) changes almost linearly with time [10]. Based on the characteristics of this function, we will take this function as the weight of our time. Because short video has strong timeliness, in order to accurately calculate the user's nearest neighbor. On the basis of combining equation (2), we add the weight of time. The formula for adding time weights is given by equation (4).

\[
\text{Sim}(U_a, U_b) = \frac{\sum_{i \in I_j}(R_{a,i} \times \text{logistic}(t_i)) - \bar{R}_a \times \sum_{i \in I_j}(R_{b,i} \times \text{logistic}(t_i)) - \bar{R}_b)}{\sqrt{\sum_{i \in I_j}(R_{a,i} \times \text{logistic}(t_i))^2 \times \sum_{i \in I_j}(R_{b,i} \times \text{logistic}(t_i))^2}}
\]

3.2. Recommended results of k NN collaborative filtering
The k NN collaborative filtering recommendation algorithm is the most mature algorithm in data mining. The k NN approach relies primarily on a limited sample of users' proximity, rather than relying on short video types. For cross-over or overlapping sample sets such as video, the k NN method is more suitable than other methods.

By calculating the similarity between users, Time-weighted user behavior analysis collaborative filtering algorithm finds \( k \) neighbor for an unknown \( I_j \) and defines \( S(U_a) \) and \( |S(U_a)| = k \).

\[
R_{a,j} = \bar{R}_a + \frac{\sum_{U \in S(U_a)} \text{Sim}'(U_a, U_b) \times (\bar{R}_{b,j} - \bar{R}_b)}{\sum_{U \in S(U_a)} \text{Sim}'(U_a, U_b)}
\]

In the above equation (5), \( R_{a,j} \) represents the predicted score of the user on the video B. The video is sorted in descending order according to the score value, and the top \( k \) items are selected as the recommendations.

3.3. Process of time-weighted user behavior analysis collaborative filtering algorithm
The first step: Modelling user behaviour. (Obtain the behaviour data of video users watching video and convert the behaviour data into user ratings)

The second step: Calculating user similarity by scoring matrix.
The third step: Weighting Time with Function \( \text{logistic} \).

The fourth step: Calculating the validity of time.
The fifth step: Choosing the appropriate \( k \) value.
The sixth step: Calculating \( k \) video sets of unknown score video after collaborative filtering.
4. Experimental results and analysis

4.1. Experimental data

The test data for the experiment is a data set given by Tubular Labs Research Centre, an American short video data organization. The data set contains approximately 1 million pieces of data. This date included 680 users watching 9343 short videos, and each user watched 17 short videos. The score matrix density of user-video is 15.7%, we can conclude that this data set is sparse.

4.2. Test the quality of the recommendation

In order to verify the validity of the time-weighted user behaviour analysis recommendation algorithm proposed in this paper. This paper uses \textit{Precision} and \textit{Recall} method to test it.

| Table 2. Recommendation system evaluation rules table |
|-----------------------------------------------------|
| User favorite video | User dislike video |
| Method recommended video | TP | FP |
| Method do not recommended video | FN | TN |

From the above Table 2, we can get some of the following equations.

\begin{align}
\text{Precision} &= \frac{TP}{TP + FP} \tag{6} \\
\text{Recall} &= \frac{TP}{TP + FN} \tag{7}
\end{align}

4.3. Experimental result

This paper analyses the data of sixty experimenters watching videos and compares Time-Weighted User Behaviour Analysis Collaborative Filtering Algorithm (TWUBCF), Time-Weighted Tab Collaborative Filtering Recommendation Algorithm (TBCF), Uncertain Neighbors’ Collaborative Filtering Recommendation Algorithm (UNCF), and Item-Based Collaborative Filtering Recommendation Algorithm (IBCF).

\begin{figure}[h]
\centering
\includegraphics[width=0.4\textwidth]{precision.png}
\caption{Comparing precision.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=0.4\textwidth]{recall.png}
\caption{Comparing recall.}
\end{figure}

From the above Figure 2 and Figure 3, it is obvious that the TWUBCF proposed in this paper is far more comprehensive than the other three. The algorithm in this paper compares with TBCF, the former focuses on the relationship between users, while the latter focuses on the relationship between products. For short videos, it is difficult to find the relationship between videos, because short video information...
is less, quality is different, and it is difficult to distinguish. It’s much easier to analyze the relationships between users, just to find similar user groups. UNCF is one of the classic algorithms. It balances the relationship between the item and the user. However, the score matrix of short video is sparse, and it does not fill the matrix in other ways, resulting in poor recommendation quality. TBCF A and TWUBCF are aware of the influence of time factors on the recommendation results, and weighted the time to improve the quality of recommendation. TBCF pays attention to artificially adding tags to videos, which is useful for traditional movies. Movies take a long time, are shot regularly, and are tagged by professionals. Make the tag fully express the characteristics of the movie itself.

5. Conclusion
Based on the characteristics of short video, TWUBCF is proposed in this paper. User behavior is deeply analyzed from the aspect of users. In order to avoid data sparsity, this paper not only analyzes the explicit behavior of users, but also the invisible behavior of users. By improving the similarity, the relationship between users can be more accurately calculated. Time weighting can avoid the decline of recommendation quality caused by the change of user interest over time.

Reference
[1] Mei T, Yang B and Hua X. VideoReach: An Online Video Recommendation System. // International ACM Sigir Conference on Research & Development in Information Retrieval. ACM, 2007.
[2] Yang X, Jiong Y U and Ibrahim T. Collaborative Filtering Recommendation Model Based on Trust Model Filling. J. Computer Engineering, 2015.
[3] Niu J, Zhao X and Zhu L. Affivir: An affect-based Internet video recommendation system. J. Neurocomputing, 2013, 120(10): 422-433.
[4] Song Weiwei, Yang Degang and Zheng Min. Research on Collaborative Filtering Recommendation Algorithm Based on time-weighted labels. J. Journal of Chongqing Normal University (Natural Science Edition), 2016 (5): 113-120.
[5] Jian D, Wang Y and Qi W. Football video recommendation system with automatic rating based on user behavior. // International Congress on Image & Signal Processing. 2017.
[6] Quan S, Lu X and Shi Z. Analysis and Research of the Campus Network User's Behavior Based on k-Means Clustering Algorithm. // International Conference on Digital Manufacturing & Automation. 2013.
[7] Hoppe A, Roxin A and Nicolle C. Dynamic, Behavior-Based User Profiling Using Semantic Web Technologies in a Big Data Context. // Otm Confederated International Conferences on the Move To Meaningful Internet Systems. 2013.
[8] Mclaughlin M R and Herlocker J L. A collaborative filtering algorithm and evaluation metric that accurately model the user experience. // 2004.
[9] He L and Wu F. A Time-context-Based Collaborative Filtering Algorithm. // IEEE International Conference on Granular Computing. IEEE, 2009.
[10] Zhang J, Lin Z and Bo X. An optimized item-based collaborative filtering recommendation algorithm. // IEEE International Conference on Network Infrastructure & Digital Content. 2009.