Journal of Multimedia Information System: ISSN 2383-7632(Online)
VOL.1, NO.1, September 2014, pp. 67-75,

Video Ranking Model: a Data-Mining Solution with the Understood User Engagement

Author: Yongyu Chen[1], Jianxin Chen[1,*], Liang Zhou[1], Ying Yan[2], Ruochen Huang[3], Wei Zhang[1]

Abstract

Nowadays as video services grow rapidly, it is important for the service providers to provide customized services. Video ranking plays a key role for the service providers to attract the subscribers. In this paper we propose a weekly video ranking mechanism based on the quantified user engagement. The traditional QoE ranking mechanism is relatively subjective and usually is accomplished by grading, while QoS is relatively objective and is accomplished by analyzing the quality metrics. The goal of this paper is to establish a ranking mechanism which combines the both advantages of QoS and QoE according to the third-party data collection platform. We use data mining method to classify and analyze the collected data. In order to apply into the actual situation, we first group the videos and then use the regression tree and the decision tree (CART) to narrow down the number of them to a reasonable scale. After that we introduce the analytic hierarchy process (AHP) model and use Elo rating system to improve the fairness of our system. Questionnaire results verify that the proposed solution not only simplifies the computation but also increases the credibility of the system.

Key Words: Data-Mining, AHP, Elo rating system, QoE.

I. INTRODUCTION

Video-based services are growing rapidly, such as YouTube, Youku, and TuDou, etc., which results in the rapid growing of video traffic[1]. Herein, it is important for the service providers to provide customized services to satisfy the increasing requirements of different types of video services. To solve this problem, it is necessary to rank videos according to the subscribes' requests.

Previous studies on the video ranking mainly have three aspects:1) video quality; 2) user specificity; and 3) diversity[2], e.g. YouTube. These ranking mechanisms are used widely and proved to be useful. However they did not take the potency of engagement increasing into considerations and the computation complexity was rather high. On the other side, there are some researchers focusing on the QoS and QoE about video services[3].However QoS and QoE refer to either objective or subjective parts of the video quality respectively. They could not provide a relatively fair evaluation. To solve this problem, in this paper, we propose a new ranking mechanism based on both subjective and objective factors which combines QoS with QoE according to a third-party data collection platform. To improve the reliability we build a test bed to obtain the experiment data about the viewing of the online videos. After carefully analyzing, we find that:

1. Over half of the video sessions are viewed less than 50% of their total lengths. This is due to that the subscriber opened several video sessions at the same time and finally he chose the video that interested him most.
2. Popular videos usually have lower viewing ratio and the recommendation system usually recommends the popular videos.
3. Most video resolutions are 360p. It is due to that the websites they are viewing always choose the quality according to the current network link quality.
4. The same video appears in the same router repetitively, it is probably because the users in the same router often share their user experience with each other which would lead to repetitive viewing.

In order to find the appropriate videos for further ranking, we analyze the features and find out that the

Manuscript Received September 9, 2014; Revised September 15, 2014; Accepted September 30, 2014. (ID No. JMIS-2014-0006)
Corresponding Author (*): Jianxin Chen, Mailbox 214, Nanjing University of Posts and Telecommunications, No.66, Road Xinmofan, Nanjing210003, China. +86 18951893820, chenjx@njupt.edu.cn.
[1] Key Lab of Broadband Wireless Communication and Sensor Network Technology of Ministry of Education,[2]Library, Nanjing University of Posts and Telecommunications, Nanjing, China.
13010634@njupt.edu.cn,chenjx@njupt.edu.cn,liang.zhou@njupt.edu.cn,yany@njupt.edu.cn

regression tree and the decision tree (CART)[8, 9] are helpful to narrow down the video number to a reasonable scale. To guarantee the traditional QoS[8] we propose the AHP model[10] to rank the “candidates”. To provide the traditional QoE we introduce the Elo rating system[11]for the video rating system. With this process we not only simplify the computation but also combine the QoS with QoE successfully, which finally increases the system credibility. We adjust the result gained from the AHP model to fit further ranking in our Elo rating system. We make the two different models work together successfully. Then we verify the ranking system by comparing the results with that of the questionnaires. The proposed mechanism is also helpful to find the potential popular videos and can be used in other areas such as product quality classification and the athlete ranking.

II. PROBLEM FORMULATION

The major goal of our study is to find out how the features affect the user engagement and then rank the videos. We first quantify the user engagement, then introduce the influential features including the application QoS metrics under study. We do not take QoS metrics into account because there are only 20% of total videos that have the rebuffing phenomenon, and most of them were eliminated by the filtering process.

To describe a typical video session, we define a viewer as a specific user who watches video on the website. Constrained by the data, we only monitor the traffic from several famous websites such as YouKu, TuDou and iQiYi. We define a view as a whole video viewing process that a viewer watches the video. Generally, a whole video session consists of *startup delay state*, *playing state*, *rebuffing state* and the *end*.

When a session begins, a viewer sends a request to the server and the browser establishes a connection between them. Then the video session enters the *startup delay state* where the browser begins to buffer the video as well as the advertisement. When the buffer is enough to play the advertisement, the play will begin. In order to reduce the number of the states we have, here the *startup delay state* also includes the *advertisement playing state*. It is due to that we can only obtain the video *begin-to-play time* and the *video session start time* in the data log. After finishing the advertisement playing, the video session enters the *playing state*. If the network link is stable, the session will stay in this state. Otherwise, it will enter the *re-buffer state*, in which the video pauses till the buffer fills enough to play again. Re-buffer often causes earlier quitter phenomenon. According to the data, if a video session enters *re-buffer state* for more than three times, the viewers have the possibility of 70% to quit the video session. In our experiments those data were eliminated so the influence is very small. During the session viewers may also have some interactions such as pausing and fast-forwarding, they all contribute to the final analysis of our data. Additionally, there are two kinds of manners to end a video session:

1. Quit: The viewer quits the video session in *re-buffer state* or *playing state*.

2. Complete: The view session ends when the video completes.

III. Data Analysis

1. Main features analysis

1.1. QoS metrics

Existing QoS studies focus on the potential QoS metrics that may affect the user engagement. The traditional QoS metrics usually consist of application QoS metrics such as *startup delay* and network QoS metrics, e.g. *signal strength*. In this paper we consider only application QoS metrics, as well as relevant external features such as *video popularity* and other positive interactions with the server.

The QoS metrics[3, 8] are the metrics that directly associate with the video service which can be observed by the viewers directly. In this paper we have three metrics:

1. *Startup delay time*. We define *startup delay time* as the total time of the *startup delay state* and denote it as t. $t = t_1 + t_2 + t_3$,where t_1 refers to the ad buffering time; t_2 refers to the *ad playing time*; andt_3 refers to the *video content buffering time*

2. *Total rebuffing time*. When the buffer is completely working out, the video session will try to download the video content till the buffer is able to play again. The whole download time is defined as total rebuffing time.

3.*Display resolution*. The service providers are often able to change the video resolution according to the current network link state. Therefore the *display resolution* is an indicator for the current bit rate and it reflects the video quality. We can not acquire bit rate directly. But we can use third-party software to parse the HTTP stream then acquire the video resolution.

1.2. Other metrics

Besides QoS metrics, external features also affect the QoE levels[8].

1. Viewers demography. User demography includes age, nationality and education，etc. . Our samples cover the the senior undergraduate students, 1st year and 2nd year postgraduates.

2. Video length. The total video duration of the video are sampled.

3. Video viewing ratio. The video viewing ratio is defined as the total viewing time dividing *the video length*. We consider the viewing ratio as the main t feature of the user engagement.

4. *Video popularity*. The total number of views.

5. *Fast-forwarding ratio*. The fast-forwarding ratio is defined as the fast forwarding duration dividing *the video length*. It is used to measure the patience of the viewer.

6. *Video repeating times*. Users often share good experience with others, which lead to repetitive viewing. It is an important reference index.

1.3. Metrics improvement

Existing QoE studies mainly focused on quantifying the user engagement as *the viewing ratio*, and refer to the objective QoS metrics or subjective features. Besides those features, in this paper we present a video interaction comparison system by introducing the Elo rating system. We present two main features into our study[11, 12]:

1. *Viewers interaction*. Traditional Elo rating system compares the result of a match that two 'players' take part. In this paper we record the viewers' interactions to the server and consider them as the match. To make this adaptive to the system, we have assumed three conditions:

1) win (when user K did some interaction such as clicking support on video A while had no interaction with video B, video A wins video B)

2) lose (when user K did some interaction such as clicking support on video B while had no interaction with video A, video A loses to video B)

3) tie (when user K did some interaction such as clicking support on video B and had the same interaction on video A, video A and video B encounters tie)

2. *Viewers replying*. It is defined the same as viewers interaction. We record the viewers replying and consider them as the match. To make this be adaptive to the system, we assume three conditions:

1) win (When user K did some replies on video A while had no reply on video B, A wins video B)

2) lose (When user K did some replies on video B while had no reply on video A, video A loses to video B)

3) tie (When user K did some replies on video B and had the same interaction on video A, video A and video B encounters tie)

2. Data collection platform setup

To enhance the reliability we set up a third-party experimental data collection platform in our campus. We have many routers deployed around our campus, mainly distributing in the offices and dormitories only 3 of them are deployed in the classrooms. Most of the students are connecting to the router through Wi-Fi network, so our clients are consists of PC users as well as the mobile users. To protect the privacy, everyone who connects to the network will receive a web-based notification to ensure that they have knowledge to our experiment. At the same time, each router has a unique person in charge and we will let the participants to sign contracts to ensure that they agree to our data collection regulation. From those routers we can monitor about 1000 video sessions everyday and in this paper we only use the data within one week. Finally we have 7850 video sessions. Constrained by the data processing, all of the video sessions were from Youku,Tudou or iQiYi. Traditional data collection platforms monitor the traffic from the server or client. Most of the server-based collection mechanisms may intrude user privacy to different degrees and they can only collect data from their own website. As for the client data collection mechanisms are mainly based on questionnaires. It is feasible but defective. Apart from the troublesome deployment process it will also remind the user that they were under monitoring, which inversely affects the survey result. In our experiment we deploy off-the-shelf MERCURY MW5430R 750Mbps Dual Band Wi-Fi Wireless Gigabit Router with the operating system overwrite to Openwrt [4](a operating system based on Linux Kernel) on the client side. We use TCPDUMP[5] to capture the uplink and downlink packets via the routers.

The storage space of a single router is 16MB, and a 30s full-speed download will exhaust it. At the same time we monitor more than 10GB data from a single router every day. So it is impossible to collect data directly from the routers. To solve this problem we build a storage array on IBM V7000[6]. We write script on each router and mount them to the storage array with Network File System(NFS)[13]for every booting. So all data are stored on the storage array directly and this will have little impact on the network link state.

To analyze the data, we use TSTAT[7], an open-source traffic statistics and analysis tool. The original TSTAT can only analyze the data from YouTube. So we first rewrite the traffic paring code to make it able to parse the data we collected. By parsing the HTTP packets TSTAT will acquire video session parameters such as *video special code*, *total video length*, *resolution*, viewer operation such as *fast-forwarding*, *pause*, *clicking support*, *making comment* and the corresponding timestamp relevant to each parameter. After the parsing we use the video code and write procedure based on Python to acquire the *video popularity* on the website.

3. Traffic data composition analysis

As it is mentioned above, we have 7850 video sessions from our test bed within one week. About 60% of them are from YouKu, 14% from Sina video, 8% from iQiYi, 7% from TuDou and the rest are coming from other websites, which is depicted in Fig.1.

After analyzing, 50.6% videos lasted less than 10min. Only 3.8% of the videos lasted longer than 100min, which shown in Figure 2. Figure 3 shows that 40% videos were recorded during the weekend. It is reasonable since the main user group were students and they have time at weekends.

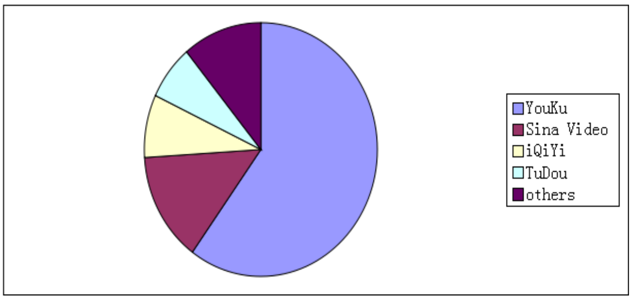

Fig. 1. Video composition

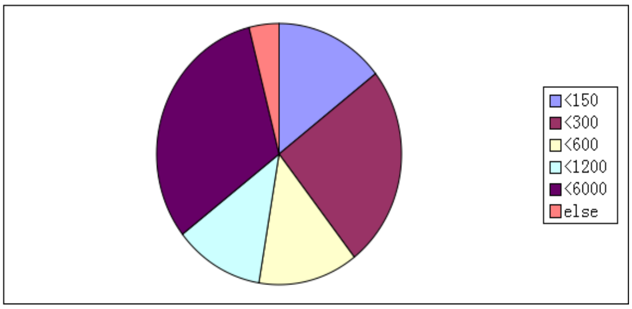

Fig. 2. Length composition

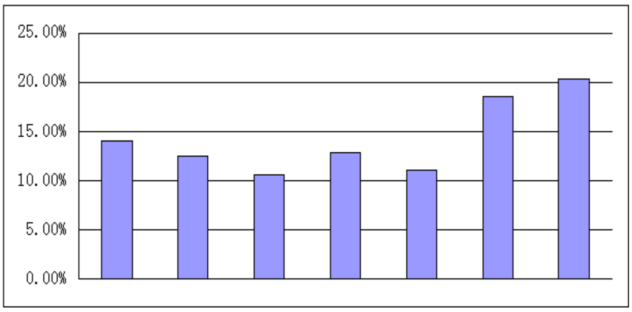

Fig. 3. Time distribution

IV. RANKING MODELS AND CORRESPONDING DATASET ANALYSIS

In this section we first introduce our data filtering model and then analyze its rationality[14, 15]. According to the results from the data filtering model, we use the AHP model to do the basic ranking, and then we introduce the Elo rating system and adjust it to fit our application.

1. Data filtering model

In this part we first analyze the features according to the data filtering model. Suppose that our goal is to find out the top N videos, and the fact is that we have thousands of "candidates". Therefore it is reasonable to narrow down the range of the "candidates" to a reasonable scale.

1.1. Corresponding dataset analysis

Previous researches focused on quantifying the user engagement as the *viewing ratio*. Therein it is reasonable to filter the "candidates" and then rank the videos using the features that are relevant to the *viewing ratio* most. It is not proper to do the filtering or ranking all depending on the mathematical analysis. To avoid dogmatism, we should also take the actual situation such as the advertisement playing rule into consideration.

In this paper we analyze the dataset according to Spearman correlation coefficient analysis on SPSS. The Spearman correlation coefficient has the value among [-1, +1]. Here '-1' means the independent variable can be represented as a monotonically decreasing function of the dependent variable, and '1' means the independent variable can be represented as a monotonically increasing function of the dependent variable. The more closer to zero the absolute value of the number the more independent the two variables are. The analyzing results are shown on Table.1.

Table 1. Correlation Coefficient (CC)

Features	**Spearman(*Viewing ratio*)**	
	CC	Sig
Startup delay time	-0.642	0.000
Video length	-0.736	0.000
Video popularity	-.0.241	0.024
Fast-forward time(all)	-0.221	0.040
Fast-forward time(validate)	-0.452	0.000
Total re-buffering time(all)	0.094	0.385
Total re-buffering time(validate)	0.053	0.840
Page view	-.0385	0.128

If correlation coefficient (CC) ranks high and sig ranks low, the corresponding features are more likely to be correlated. Previous results are mainly quantifying the user engagement as the *viewing ratio*. Our goal is to rank the videos according to the user engagement. Therein it is reasonable to use other features that are correlated to the viewing ratio to do further filtering or ranking.

Noting that the correlation coefficient of the *startup delay time* and *the video length* to *the viewing ratio* are '-0.736' and '-0.642' respectively. It means the longer the *startup delay time* (*video length*) the smaller *theviewing ratio* will be. This is satisfactory with the practical situation because people are often not willing to wait too long for a single video. But this does not mean that we can use these features to filter the "candidates". If so, the longer videos will be wiped out, which is not in accordance with the common sense. After analyzing we find *startup delay* and *video length* have some special relationship with each other, and the correlation coefficient of them is 0.867. In this paper we divide *startup delay time* into 3 parts:

1. *Ad buffering time*.
2. *Ad playing time*.
3. *Video content buffering time*.

The *ad playing time* is strongly associated with *the video length*, and it often takes most part of the *startup delay time*. After the investigation on the videos from YouKu, TuDou and iQiYi, we find they that have two major correlations:

1. The *ad playing time* and *video length* follow a piecewise function, and the relationship is listed in Table.2.
2. Assume that we have divided *the video length* into 6 levels as in table.2. The *ad playing time* may sometimes reduce to the level below the current level when the user has just viewed another video. And the exact relationship is not clear at this time.

Thus we cannot use the *startup delay* and *total video length* to do the filtering, and it is reasonable to group the 'candidates' first.

When a viewer feels tired about the video being watched, he may drag the video progress bar. When he drags the video, the fast-forward action happens. Not all videos have the fast-forward behavior. According to the statistics, there are only 30% of the videos that have the fast-forward action. So we may not use this to filter the 'candidate'. It is worth mentioning that if we consider the 30% videos alone, the correlation between the *fast-forward ratio* and the *viewing ratio* will be larger, so we believe this is some relationship that needs to be finding out, but in this paper we do not consider it.

As it is shown on Table.2, the correlation coefficient between *total re-buffering time* and *the viewing ratio* is low. Therein this will not be used as a feature when filtering the 'candidate'. *Page view is* not correlated to the viewing ratio. Due to the constraint from the sample quantity, we will not use this feature to do filtering.

Although some of the features such as the *total video length* and *startup delay* will not be used in the data filtering process directly, the correlation coefficient analyzing results still work when they come to our ranking model.

Table 2. Relationship between ad playing time and video length.

Video length(second)	**Ad playing time(second)**
<150	5
<300	15
<600	30
<1200	45
<6000	60
others	90

1.2. Data filtering

In this part data filtering model is mainly designed according to the ideology of CART. We use this model to process data and the model can be divided into three major phases[14]:

Phase1: Data grouping. The *total re-buffering time* is strongly associated with *the video length*. If we want to use this feature to do the filtering we have to wipe out the effect of the *ad playing time*. As it is analyzed above, the total *video length* can be divided into 6 levels. And it is reasonable to divide them into 6 groups according to the video length dividing. The number of each group is listed in Table.3. This grouping process is used to keep the length diversity of the final ranking result. Generally, if a video is diverse in video length it may be diverse in video content type as well.

Phase2: Primary filtering. In this process we introduce two new features called *number of replies* and *number of clicks*. We are not going to use the *video repeat times* as a feature in our model because it is small in value, and we have the feature of *video popularity* as the alternative. At the same time, according to the ideology of recursively dividing the independent feature's spaces, we want to introduce the Elo rating system into our model. First we need to consider the number of the replies and clicks in order for further comparing. Those videos which do not have the corresponding behaviors can be wiped out form the "candidates" list. After the first filtering step there is only 89 "candidates" left. The number is obtained after the combination process, which combines the same video together and counts the number of each feature we are going to use. The distribution of the "candidates" left is shown in Table.3. Then we use the mean value of the *viewing rate, the video length* and *processed total re-buffering time* one by one to narrow down the range of the “candidates” in each single group to a reasonable scale and the result is shown in Table.3 as well.

Table 3. Video number change.

Video length(second)	Total number	Number after P2 S1	Number after P2 S2
<150	1091	14	3
<300	1889	20	4
<600	994	11	2
<1200	894	11	3
<6000	2385	27	4
others	297	6	2

Phase3: Limited filtering. According to the ideology of cutting down the number of the data according to the validation features. This phase is designed for further ranking, in which we have taken the grouping problem into consideration. So we first regroup the “candidates” together and then filter them according to the features we are going to use in the ranking model. Considering the remaining “candidates”, we set the goal number down to 5, which should be adjusted in real situation. We choose the top 5 “candidates” according to *viewing ratio*, *totalvideo length*, *fast-forward time*, *startup delay time* and *video popularity* respectively and then collect them together to form the final “candidates” list. For *viewing ratio* and *video popularity* we select the biggest five while for the other features we select the smallest 5. The reason to do so will be explained in next section. It is reasonable that if there is no feature ranking top 5(Suppose our goal is to acquire the top 5 video), the candidate have no reason to stay in the “candidates” list. After this step we are down to 11 “candidates”.

2. AHP model

We want to analyze video from five main aspects: basic engagement level, tolerance level, popularity level and content attraction level. The content attraction level consists of: the video name attraction level and the real content attraction level. Each of the candidate has five main features[10] and each of them refers to an single aspect that we want to consider into:

1. *Viewing ratio*: It represents the basic user engagement level. If this feature ranks high, the user engagement is supposed to rank high.
2. *Total video length*: It represents the user tolerance level. As it is analyzed above, if this feature ranks low, the user engagement is supposed to rank high.
3. *Fast-forward ratio*: It represents the real content attractive level. If this feature ranks low the user engagement is supposed to rank high.
4. *Startup delay*: It represents the video name attractive level. If this feature ranks low, the user engagement is supposed to rank high.
5. *Video popularity*: Total watching number of the video at present. The videos the subscribers are watching are not the latest. So In this paper we can use the *current video totally viewing number* to represent the *video popularity*. In realistic situation, we can use the *video popularity* of last week to analyze.

Our AHP model aims at generating the top five videos according to these five features: *viewing ratio, total video length, fast-forward ratio, startup delay* and *video popularity*. The structure of our video ranking AHP model consists of 3 different layers: goal, criteria and alternatives. The only difference is that the “candidate” here replaces traditional plans and takes up the alternative layer. The relationship is shown in Figure 4.

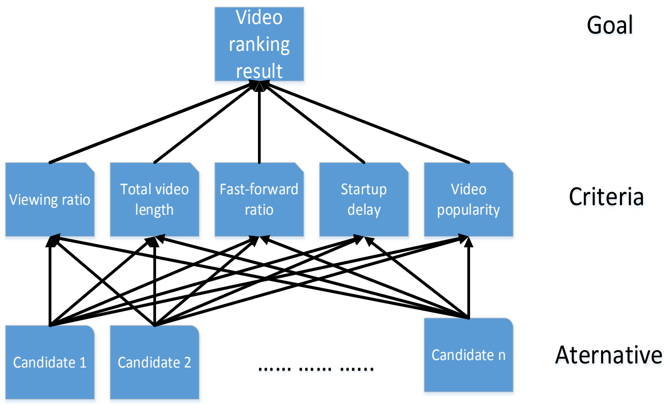

Fig. 4. AHP model hierarchy

The criteria is achieved by the joint consideration of the collected data as well as the goal. The alternative refers to the “candidates” list achieved by the data filtering model.T he relative strength of rating scale ‘a_{ij}’ is achieved by the analysis of the five features listed on the criteria. The rating scale is shown on Table.4, in which C_i refers to the absolute strength. From the AHP model we got the preliminary relative strength ‘R_{ij}’.

For practical applications, our goal is to offer the weekly recommendation of the latest videos. Although the result from the model is acceptable, this plan is still unfair to some later uploaded videos. Assuming that a video is uploaded in the second half of the week, some features will encounter the “time line horizon” problem. For example, the feature *video popularity* of some later uploaded videos may be smaller than that of the videos uploaded before, which will definitely affect the final result. In realistic situation the weekly recommendation result often shows on Monday and we cannot wait until the online time of the former uploaded videos to reach that of others. In order to keep the ranking result fair while reducing the computation complexity we introduce the Elo rating system into our model. Thus we not only take the potency of the use engagement growth into consideration but also solve the "time line horizon" problem successfully.

After the processing of AHP model, we obtain a fundamental ranking result with the eigenvalues that represent each video's relative strengthen. By adjusting these eigenvalues we finally adapt them to further ranking and successfully combine the AHP model with the Elo Rating system. Then we use the Elo Rating system to adjust the result according to the current ranking.

Table 4.Rating scale.

scale a_{ij}	Meaning
1	$C_i = C_j$
3	$C_i > C_j$ and R_{ij} is tiny
5	$C_i > C_j$ and R_{ij} is small
7	$C_i > C_j$ and R_{ij} is moderate
9	$C_i > C_j$ and R_{ij} is big
one of 2,4,6,8	$C_i > C_j$ and R_{ij} is between the adjacent R that mentioned above
1,1/2,...,1/9	$C_i < C_j$ and R_{ij} is the reciprocal of the R that mentioned above

3. Elo rating system

Elo rating system was created by Arpad Elo for the purpose of ranking players[11]. Arpad Elo first used this system to rank the chess players. For its rationality, some multi-player competition began to use it as the integral system. The system gives each player an identical initial score, so the final score depends entirely on their performance during the matching against others. It is reasonable and fair, but we can not use the system directly to characterize the video according to the QoE level. We can not give the identical start level to videos either. Therefore we combine the system with the AHP model and use the result of AHP model to formulate the initial score of each candidate for further ranking.

3.1 System introduction

Assume video i and j represent two different videos respectively. $i = j$ means they are the same. $i \neq j$ means that they are different. E_i and E_j represent the expected scores for video i and video j. There are three possible results for a 'match': win, lose, or tie. Each of them have its number. Assuming that each video's base score is R_i, we change the base value according to the AHP ranking to get the initial rating R_i of video I and then we use equations to get the final R_i. The order is given all depends on the final score that a video got.

To increase the fairness of the system generally, we give the bigger K-factor to the 'weaker' (low in the score) and the smaller K-factor to the 'stronger'. Through this when a 'weaker' wins a 'stronger', it will gain more scores than expectation. With the introduction of K-factor, we not only consider the competition information but also consider the potentiality of a video. We compare the number of replying and the clicking support action of two different videos and limit the range from 1 to 5. Table.5. shows the five levels according to the strength.

Table 5.Rating scale.

Difference of actions	Sacle video ij	Meaning
[-5,-4)	1	Preceding large stronger
[-4,-3)	3	Preceding moderate stronger
[-3,-2)	5	Preceding small stronger
[-2,-1)	7	Preceding tiny stronger
[-1,0)	9	Equal
(0,1]	-1	Equal
(1,2]	-3	Behind tiny stronger
(2,3]	-5	Behind small stronger
(3,4]	-7	Behind moderate stronger
(4,5]	-9	Behind large stronger

The expected score E_i of video i is defined according to Eq.1 :

$$E_i = 1 \Big/ \left(1 + 10^{(R_j - R_i)/40}\right) \tag{1}$$

where R_j is the rating of video j. Through using this equation we successfully enhanced the variation of the score. Through this process when a video 'over-performed' or 'underperformed' than its expectation the score it will obtain will be more truthfulness than other simple integration methods. The equation has been proved to be useful in many fields, and we consider it will help we to rank the videos too.

The maximum possible value per game is called the K-factor. K = 16 denotes the 'stronger' video and K = 32 denotes the 'weaker' video. K-factor may reflect the potentiality of a video. Eq.2 define the updating rating, this equation illustrated the change of the scores before and after a 'match':

$$R_i' = R_i + K(S_i - E_i) \tag{2}$$

Supposing video $i$is expected to score E_i points but actually scored $S_i$points, the equation will give our its final score after the 'match'. In the equation the introduction of the *K*-factor will protect the 'weaker' and finally increase the fairness of our ranking system.

3.2 Final result and its analysis

After filtering, only 11 "candidates" left. We use the AHP model to rank these "candidates" according to the above features. Noting that most of the 11 "candidates" are from YouKu, and the others are from other websites. So we transform the video code to YouKu code. For example, the video code 'ZpTKEqK-ufQ' on TuDou is translated to 'XNzI4NTgwNDky' on YouKu. The ranking result is listed on Table. 6.

Table 6.AHP ranking results

Number	**Video code**	**Weight(AHP)**
1	XNzI4MzE4ODgw	0.1633
2	XNzI3ODY4NDMy	0.1632
3	XNzI4NTgwNDky	0.1245
4	XNzI3NjA1MTYw	0.1099
5	XNzI3NTY5NDY0	0.1061
6	XMzQ0MjkxNDM2	0.0939
7	XNzI3NTU5MDgw	0.0651
8	XNzI4NzI4OTc2	0.0635
9	XNzI2MjMzNTc2	0.0478
10	XNjEyNDg0NTc2	0.0384
11	XNzIxNzc0NTUy	0.0242

Our goal is to recommend videos that have higher quantified user engagement level. The top ranker are mainly trailers of upcoming movies or some short entertaining videos, while other long videos which have high *video popularity* usually rank low. We find that the subscribers are willing to interact with some of the lower rankers. It means their engagement may be high. As we discuss above the result is not fair to the later uploaded videos. Therefore the ranking result is not reasonable.

Then we use the results of the AHP model and adjust the final weight to adapt to the Elo rating system in order to do further ranking. From the improved Elo rating system, the result is improved as those in Table.7.

Most of the ranking changes little, it is due to that the gap between their initial score is big and the viewers made little interaction with the server during their watching. Noticing that, although the initial score of the video 'XNzI3NTU5MDgw'is low but it finally ranks the third. After investigation we think it is normal, we find this video is a funny 'dota film' that combine with a popular song which boys are willing to watch and interact. Although it has low clicks as well as high startup delay it may still attracts many interactions.

Table 7.Elo rating system output

Number	**Video code**	Final Ri
1	XNzI4NTgwNDky	2445
2	XNzI3ODY4NDMy	2432
3	XNzI3NTU5MDgw	2351
4	XNzI4MzE4ODgw	2233
5	XNzI3NTY5NDY0	2071
6	XMzQ0MjkxNDM2	2039
7	XNzI3NjA1MTYw	2006
8	XNzI4NzI4OTc2	1835
9	XNzI2MjMzNTc2	1478
10	XNjEyNDg0NTc2	1384
11	XNzIxNzc0NTUy	1242

To improve the credibility of our ranking system we have done questionnaires from the subscribers after finishing the experiments. Some of the subscribes were asked to score the 11 videos according to their own feeling. Limited by the number of students we only gathered 30 questionnaires. Each of them have the scoring results of 11 videos ranking form 1-5. Thanks to the small number of videos, we convince the subscribers to do the experiment. After we have gained all of the questionnaires we calculated the average score of each video. The ranking order compared with the questionnaire results is listed in Table.8

Table 8.Order comparison

Number	**Video code**	**AHP to Questionnaire**	**Elo to Questionnaire**
1	XNzI3NTU5MDgw	7	2
2	XNzI4NTgwNDky	3	3
3	XNzI3ODY4NDMy	2	1
4	XNzI3NjA1MTYw	4	7
5	XNzIxNzc0NTUy	6	6
6	XNzI3NTY5NDY0	10	10
7	XNzI4MzE4ODgw	1	4
8	XNzI2MjMzNTc2	9	9
9	XNzI4NzI4OTc2	8	8
10	XMzQ0MjkxNDM2	11	11
11	XNjEyNDg0NTc2	5	5

Column 3 and 4 listed the comparison order of our model against the result gained from the questionnaires. By analyzing we find that the AHP plus Elo rating system is better than the AHP model alone. So through this we proved our combination of the two model successfully and reasonable.

V. CONCLUSION

Traffic of websites (YouTube, YouKu, .etc.) and mobile video services are growing rapidly. Various types of video services such as online VoD services[16] are appearing. With the ever-increasing streaming service and video traffic, to develop a ranking system based on the understanding of user engagement is very important. It will help the service providers to maintain their audiences and increase the profits.

In this paper, we combine AHP model with the Elo rating system. According to final results we find that our system is more reliable and objective than the traditional methods. The proposed system is not only unaffected by the "time line horizon" problem, but also has low computational complexity. With the development of computer technology and data processing techniques, our ranking system will be applied into other fields such as the qualitative comparison and the athlete ranking. In the future we will focus on the eliminated data in order to acquire more useful information.

Acknowledgement

This work is partly supported by the State Key DevelopmentProgram of Basic Research of China (2013CB329005), National Natural Science Foundation of China (GrantNo. 61201165 and No. 61271240), Priority Academic ProgramDevelopment of Jiangsu Higher Education Institutions,Wireless Broad Communication and Wireless Sensor networkDepartment of Education Key lab Open Program (GrantNo. NYKL201306), Education Natural Science Foundationof Jiangsu Province (13KJB510026), and Nanjing Universityof Posts and Telecommunications Foundation (GrantNo. NY211032).

References

[1] V. Valancius, N. Laoutaris, L. Massoulié, C. Diot, and P. Rodriguez, "Greening the internet with nano data centers," in *Proceedings of the 5th international conference on Emerging networking experiments and technologies*, pp. 37-48, 2009.

[2] J. Davidson, B. Liebald, J. Liu, P. Nandy, T. Van Vleet, U. Gargi, *et al.*, "The YouTube video recommendation system," in *Proceedings of the fourth ACM conference on Recommender systems*, pp. 293-296, 2010.

[3] M. Z. Shafiq, J. Erman, L. Ji, A. X. Liu, J. Pang, and J. Wang, "Understanding the Impact of Network Dynamics on Mobile Video User Engagement,", 2014.

[4] *Openwrt*. Available: https://openwrt.org/

[5] *TCPDUMP*. Available: http://www.tcpdump.org/.

[6] J. Tate, A. Baer, N. Iron, T. Jahn, P. Jenkin, J. Quintal, *et al.*, *Implementing the IBM Storwize V7000 Unified*: IBM Redbooks, 2013.

[7] *TSTAT*. Available: http://tstat.tlc.polito.it/overview.shtml

[8] A. Balachandran, V. Sekar, A. Akella, S. Seshan, I. Stoica, and H. Zhang, "Developing a predictive model of quality of experience for internet video," in *Proceedings of the ACM SIGCOMM 2013 conference on SIGCOMM*, pp. 339-350, 2013.

[9] W. Y. Loh, "Classification and regression trees," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 1, pp. 14-23, *2011*.

[10] T. L. Saaty, *Fundamentals of decision making and priority theory with the analytic hierarchy process* vol. 6: Rws Publications, 2000.

[11] A. Elo. (2014). *Elo rating system*. Available: http://en.wikipedia.org/wiki/Elo_rating_system

[12] K. Wauters, P. Desmet, and W. Van Den Noortgate, "Monitoring Learners' Proficiency: Weight Adaptation in the Elo Rating System," in *EDM*, pp. 247-252, 2011.

[13] S. Shepler, M. Eisler, D. Robinson, B. Callaghan, R. Thurlow, D. Noveck, *et al.*, "Network file system (NFS) version 4 protocol," *Network*, 2003.

[14] M. Bramer, *Principles of data mining*: Springer, 2013.

[15] A. Halevy, P. Norvig, and F. Pereira, "The unreasonable effectiveness of data," *Intelligent Systems, IEEE*, vol. 24, pp. 8-12, *2009*.

[16] L. Chen, Y. Zhou, and D. M. Chiu, "Video Browsing-A Study of User Behavior in Online VoD Services," in *Computer Communications and Networks (ICCCN), 2013 22nd International Conference on*, pp. 1-7, 2013.

Authors

Yongyu Chen is currently a grad student in Nanjing University of Posts and Telecommunications (NJUPT) under the supervision of Prof. Liang Zhou. He earned an B.Eng in Radio and Television Engineering form the NJUPT in 2012. His research interests include Multi-Media Communication, QoE and Data Mining.