User Behavior Analysis and Prediction on Personal Live Streaming Platform

Xing Lan, Zhao Gao, Yi Sun
University of Chinese Academy of Sciences, Institute of Computing Technology, Chinese Academy, Beijing, China

1. Introduction

The popularity of mobile terminals and the enhancement of network performance make it possible to transmit video in real time. This has led to the rapid development of personal live broadcasting in recent years which occupies an increasing proportion of people's entertainment activities.

User behavior is an important research issue to understand users' activities on the personal living streaming system. However, prior work on this topic is rare. There is almost no research on analyzing users' behavior, nor does it take into account the influence of various factors on users' behavior.

This paper's theme is user behavior analysis and prediction based on a large dataset collected from a big personal live streaming platform in China. First, we measured some important aspects of user behavior and analysed the result to discover the laws of user activities. Then we focused on how long the users watched videos and proposed a new scheme to predict the user's accumulated viewing time in a day(Ada-DT). Finally, we predicted which anchors a user would watch based on knowledge graph(APKG).

2. Measurement

When we researched personal broadcast, a new new video transmission method, we collected real data set from a personal live platform in China, which including total user logs from January 1st to January 9th, 2018.

We extracted useful information from user logs and measured the user behavior in various aspects. Through the analysis of the measurement results, we found the following important conclusions:

1. There was a clear day-to-cycle regularity in both anchors' and viewers' behavior. In general, the peak of user activity was at night in local time, while the valley was in the morning.The viewers' activity would start to drop one hour after the anchors' falling.

2. The average duration of a single viewing was only tens of seconds and was inversely proportional to the total number of views in the current period.

3. For both anchors and viewers,20% users had made 80% contribution to the personal live streaming platform.

4. Social relationships were highly concentrated, 10% of the anchors occupied 90% of the fans on the platform, and 40% of the anchors occupied almost all the fans on the platform.

5. 80% video streams duration were less than 20 minutes, while only 5% were more than 1 hour, 30% video streams occupied 80% of the total video stream length.

6. Most video viewing delays were within 3 seconds, about half of the cases could play the first frame in one seconds.

7. The users had poor tolerance for buffer, most viewers would exit the live room within 4 seconds after buffer. For every 1% increase of buffer ratio, the average viewing time would be reduced by 30%.

3. User viewing time prediction

Due to the generally short duration of one-time viewing by viewers, the accumulated viewing time in one day makes more sense. According to the previous conclusion, 20% active users had made 80% contribution platform, in order to ensure the continuity and stability of the data, we selected users who watched videos more than one hour per day as experimental data sets, 80% of the data set is used as training set and the rest 20% is used as test set.

The accumulated viewing time in a day is a continuous value, so we abstract this problem into a regression problem in machine learning. Because of the periodicity of user behavior, we use user's accumulated viewing time on the previous 8 days and other important features to predict the user's accumulated viewing time on the ninth day.

Taking into account the differences in the behavior of different user groups, we proposed a new scheme which used an Aaboost ensemble model based on CART Decision Tree (Ada-DT) to improve the overall fitness of the model.

Simulation proved our model could avoid the problems encountered during the training of other models, such as the slowness of multiple linear regression training, underfitting of CART regression on partial samples and overfitting of GBDT regression on training set.

As for results, Figure.1 had shown the average prediction error rate of this model was 20% lower than other algorithms and Figure.2 had shown the root mean square error was reduced by 30% over the entire test set over other algorithms.

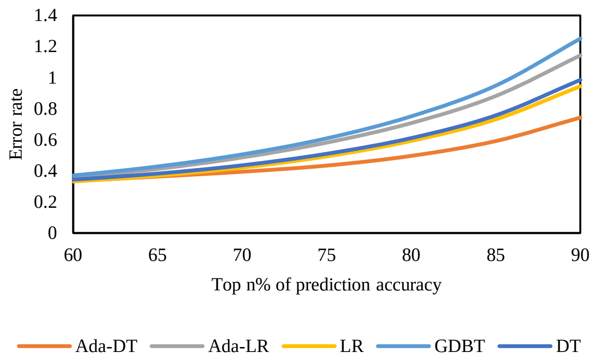

Figure.1 Error Rate Curves of Different Algorithms

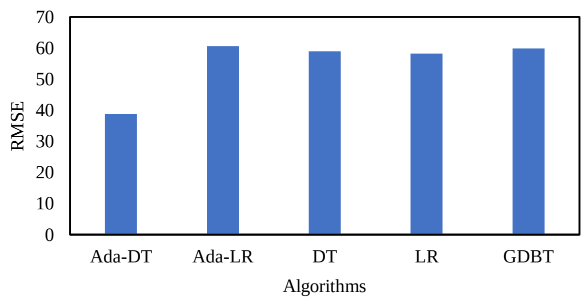

Figure.2 Root Mean Square Error of Different Algorithms

4. Anchors watched prediction

Due to the lack of user category labels, interactive records and attention lists in our data set, it was difficult to obtain users' interests and predict which anchor they would watch.

In response to this problem, we proposed an Anchors Prediction model based on Knowledge Graph (APKG), the experimental data set is same to the previous chapter, this model would constructed the knowledge graph using historical data from the past eight days and then according to the information in the graph to predict which anchors the user would watch on the ninth day.

In addition, we found the influence to users' future behavior decreased over time, so we has used the exponential decay function to update the weight of the interest relationship in knowledge graph based on time series, which meet the human memory curve.

Our model had functions such as user's interest mining, user clustering, hot anchor discovery, and relationship exploration. Simulation results showed that APKG obtained better performance than that of history-based predicition model, for about 10% improvment on precision precision(Figure.4) and recall rate(Figure.3), and about 20% improvement on the TOP-N precision(Figure.5).

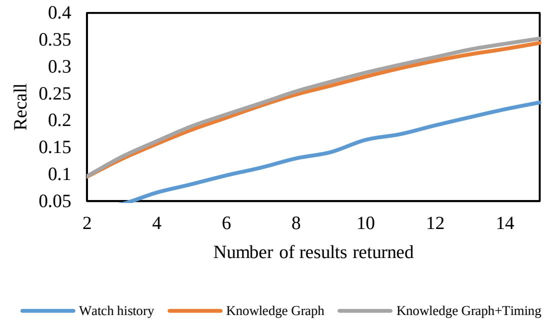

Figure.3 Recall of Different Algorithms

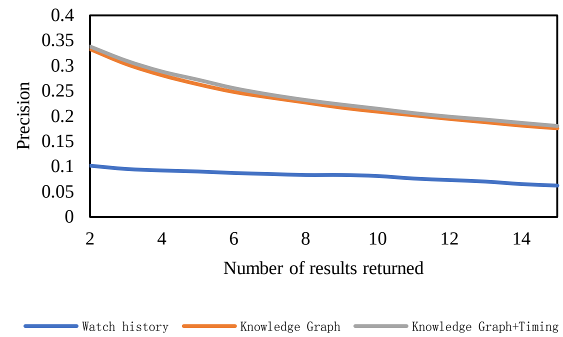

Figure.4 Precision of Different Algorithms

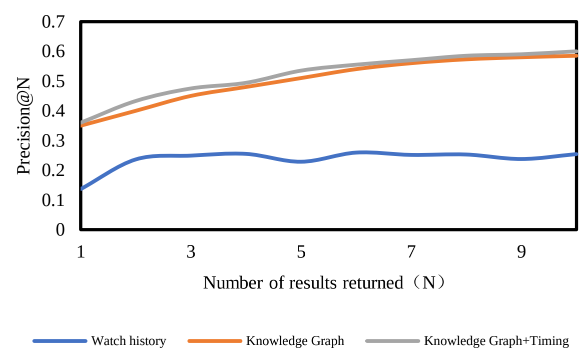

Figure.5 TOP-N precision of Different Algorithms

5. Related Work& References

Zhao.B et al.[1] crawled data analysised Periscope platform overview and video transmission delay. Huan Yan et al.[2] specifically studied the migratory behavior between different content providers and Ming Ma et al.[3] studied some user behavior in MPL based on datase from INKE.

[1] Zhao B Y, Zhao B Y, Zhao B Y, et al. Anatomy of a Personalized Livestreaming System[C]// ACM on Internet Measurement Conference. ACM, 2016:485-498.

[2] Yan H, Lin T H, Wang G, et al. On Migratory Behavior in Video Consumption[C]// ACM, 2017:1109-1118.

[3] Ma M, Zhang L, Liu J, et al. Characterizing User Behaviors in Mobile Personal Livecast[C]// The, Workshop. 2017:43-48.