PREDICTING THE TOP-N POPULAR VIDEOS VIA A CROSS-DOMAIN HYBRID MODEL

[1]KUNCHALA PADMA, [2]D.SAI KRISHNA
[1]MCA Student, [2] Assistant Professor
Department Of MCA
Sree Chaitanya College of Engineering, Karimnagar

ABSTRACT

Predicting the top-N popular videos and their future views for a large batch of newly uploaded videos is of great commercial value to online video services (OVSs). Although many attempts have been made on video popularity prediction, the existing models has a much lower performance in predicting the top-N popular videos than that of the entire video set. The reason for this phenomenon is that most videos in an OVS system are unpopular, so models preferentially learn the popularity trends of unpopular videos to improve their performance on the entire video set. However, in most cases, it is critical to predict the performance on the top-N popular videos which is the focus of this study. The challenge for the task are as follows. First, popular and unpopular videos may have similar early view patterns. Second, prediction models that are overly dependent on early view patterns limit the effects of other features. To address these challenges, we propose a novel multifactor differential influence (MFDI) prediction model based on multivariate linear regression (MLR). The model is designed to improve the discovery of popular videos and their popularity trends are learnt by enhancing the discriminative power of early patterns for different popularity trends and by optimizing the utilization of multi-source data. We evaluate the proposed model using real-world YouTube data, and extensive experiments have demonstrated the effectiveness of our model.

1. INTRODUCTION

Since the popularity of online videos has been proven to be predictable through the statistical analysis of large-scale YouTube data , numerous related studies have been conducted. Szabo and Huberman (S-H) proposed a content-scaling (CS) model based on log-transformed relations between a video's long-term popularity and its early popularity . Their conclusion is one of the most important foundations of popularity prediction research and has been succeed by many related works. All of the approaches cited above achieved initial success, but their shortcomings has been uncovered by subsequent research. Popularity prediction of online videos, especially the prediction of the top-N popular videos is of great importance to support the development of online video services (OVSs). From the perspective of better user experience, the ability to identify the top-N popular videos is beneficial to video services, such as caching and recommendation. From the perspective of commercialization, identifying the top-N popular videos helps the video service providers to maximize their profits, as advertisers are more likely to pay more for popular videos. Although many attempts have been made on popularity prediction of online videos , because most of the videos in an OVS system are unpopular; consequently, models preferentially learn the popularity trends of these unpopular videos to achieve better performance on the video set as a whole. Prediction of the top-N popular videos remains a challenging problem for the following reasons. First, popular and unpopular videos may have similar early view patterns, and this similarity limits the performance benefit of video classification based on early view patterns. We evaluate the proposed model using real-world YouTube data, and extensive experiments have demonstrated the effectiveness of our model.

Popularity prediction of online videos, especially the prediction of the top-N popular videos is of great importance to support the development of online video services (OVSs). From the perspective of better user experience, the ability to identify the top-N popular videos is beneficial to video services, such as caching and recommendation. From the perspective of commercialization, identifying the top-N popular videos helps the video service providers to maximize their profits, as advertisers are more likely to pay more for popular videos. Although many attempts have been made on popularity prediction of online videos [18][14][1][17][4], because most of the videos in an OVS system are unpopular; consequently, models preferentially learn the popularity trends of these unpopular videos to achieve better performance on the video set as a whole. Prediction of the top-N popular videos remains a challenging problem for the following reasons. First, popular and unpopular videos may have similar early view patterns, and this similarity limits the performance benefit of video classification based on early view patterns [6]. Second, existing studies show that the strong correlation between early views and long-term popularity dominates the training of the prediction models. This overdependence on early view patterns prevents models from finding popular videos based on multisource data [8][13].

To address the above problem, we present a novel popularity prediction model named multi-factor differential influence (MFDI) based on multivariate linear regression (MLR). We first enhance the ability of early view patterns to identify different popularity trends. We conduct a large-scale analysis of statistical data of early viewers' attitude-related behavior and the long-term popularity of videos. We find that the increase in the future popularity of videos follows an approximate Rayleigh distribution with respect to the degree of contradiction between early viewers with different attitudes. Based on this discovery, by combining early views with knowledge of early viewers' attitudes, we construct early rating patterns that offer better discriminative power for identifying popularity trend and use these rating patterns to replace early view patterns as the input to the proposed model. Furthermore, we incorporate the popularity of the videos' content on a social network to help the proposed model to discover popular videos and to learn their popularity trends. To overcome the restrictions on multi-source data utilization, we propose a timeaware trade-off mechanism to control the model's relative dependence on enhanced early patterns and social network data. The time-aware trade-off applies higher decay to earlier enhanced patterns and correspondingly increases the degree of denpendence of the model on social network data over time.

2. LITERATURE SURVEY

A peek into the future: predicting the evolution of popularity in user generated content

Content popularity prediction finds application in many areas, including media advertising, content caching, movie revenue estimation, traffic management and macro-economic trends forecasting, to name a few. However, predicting this popularity is difficult due to, among others, the effects of external phenomena, the influence of context such as locality and relevance to users,and the difficulty of forecasting information cascades.

In this paper we identify patterns of temporal evolution that are generalisable to distinct types of data, and show that we can (1) accurately classify content based on the evolution of its popularity over time and (2) predict the value of the content's future popularity. We verify the generality of our method by testing it on YouTube, Digg and Vimeo data sets and find our results to outperform the K-Means baseline when classifying the behaviour of content and the linear regression baseline when predicting its popularity.

Modeling and Predicting Popularity Dynamics via an Influence-based Self-Excited Hawkes Process

Modeling and predicting the popularity dynamics of individual user generated items on online social networks has important implications in a wide range of areas. The challenge of this problem comes from the inequality of the popularity of content and the numerous complex factors. Existing works mainly focus on exploring relevant factors for prediction and fitting the time series of popularity dynamics into certain class of functions, while ignoring the underlying arrival process of attentions. Also, the exogenous effect of user activity variation on the platform has been neglected. In this paper, we propose a probabilistic model using an influence-based self-excited Hawkes process (ISEHP) to characterize the process through which individual microblogs gain their popularity. This model explicitly captures three ingredients: the intrinsic attractiveness of a microblog with exponential time decay, the user-specific triggering effect of each forwardings based on the endogenous influence among users, and the exogenous effect from the platform. We validate the ISEHP model by applying it on Sina Weibo, the most popular microblogging network in China. Experimental results demonstrate that our proposed model consistently outperforms existing prediction models.

Analyzing the video popularity characteristics of large-scale user generated content systems

User generated content (UGC), now with millions of video producers and consumers, is re-shaping the way people watch video and TV. In particular, UGC sites are creating new viewing patterns and social interactions, empowering users to be more creative, and generating new business opportunities. Compared to traditional video-on-demand (VoD) systems, UGC services allow users to request videos from a potentially unlimited selection in an asynchronous fashion. To better understand the impact of UGC services, we have analyzed the world's largest UGC VoD system, YouTube, and a popular similar system in Korea, Daum Videos. In this paper, we first empirically show how UGC services are fundamentally different from traditional VoD services. We then analyze the intrinsic statistical properties of UGC popularity distributions and discuss opportunities to leverage the latent demand for niche videos (or the so-called "the Long Tail" potential), which is not reached today due to information filtering or other system scarcity distortions. Based on traces collected across multiple days, we study the popularity lifetime of UGC videos and the relationship between requests and video age. Finally, we measure the level of content aliasing and illegal content in the system and show the problems aliasing creates in ranking the video popularity accurately. The results presented in this paper are crucial to understanding UGC VoD systems and may have major commercial and technical implications for site administrators and content owners.

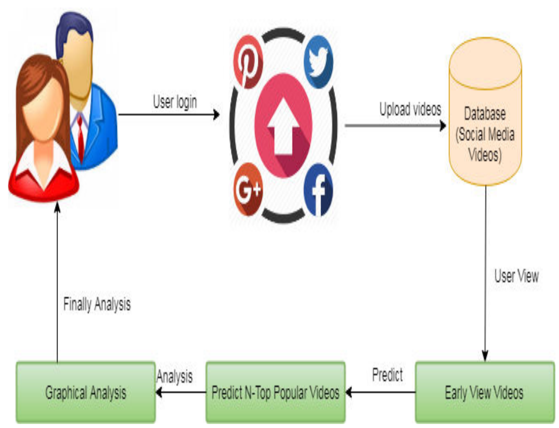
3. EXISTING SYSTEM:

Popularity prediction of online videos, especially the prediction of the top-N popular videos is of great importance to support the development of online video services (OVSs). From the perspective of better user experience, the ability to identify the top-N popular videos is beneficial to video services, such as caching and recommendation. From the perspective of commercialization, identifying the top-N popular videos helps the video service providers to maximize their profits, as advertisers are more likely to pay more for popular videos. Although many attempts have been made on popularity prediction of online videos , because most of the videos in an OVS system are unpopular; consequently, models preferentially learn the popularity trends of these unpopular videos to achieve better performance on the video set as a whole. Prediction of the top-N popular videos remains a challenging problem for the following reasons. First, popular and unpopular videos may have similar early view patterns, and this similarity limits the performance benefit of video classification based on early view patterns. Second, existing studies show that the strong correlation between early views and long-term popularity dominates the training of the prediction models. This overdependence on early view patterns prevents models from finding popular videos based on multisource data.

4. PROPOSED SYSTEM:

We evaluate the proposed model using real-world data consisting of videos from YouTube and social network data from Twitter. Our experimental results show that the proposed model outperforms state-of-the-art models, thereby confirming the benefits of our efforts to improve the prediction performance for the top-N popular videos. The main contributions of this paper can be summarized as follows ,We propose a model for predicting the top-N popular videos. By enhancing the ability of early patterns to distinguish among popularity trends and optimizing the model's utilization of multi-source data, we develop a model that achieves the promised performance. By using the tags of videos as indicators of their content and jointly training a multi-layer perceptron (MLP) network on the popularity data of videos and their related social content, we estimate the contribution of the popularity of a video's content on a social network to the long-term popularity of the video. Predicting the top-N popular videos and their future views for a large batch of newly uploaded videos is of great commercial value to online video services (OVSs). Although many attempts have been made on video popularity prediction, the existing models has a much lower performance in predicting the top-N popular videos than that of the entire video set. The reason for this phenomenon is that most videos in an OVS system are unpopular, so models preferentially learn the popularity trends of unpopular videos to improve their performance on the entire video set. However, in most cases, it is critical to predict the performance on the top-N popular videos which is the focus of this study. The challenge for the task are as follows. First, popular and unpopular videos may have similar early view patterns. Second, prediction models that are overly dependent on early view patterns limit the effects of other features.

5. SYSTEM DESIGN ARCHITECTURE DIAGRAM



6. IMPLEMENETATION MODULES DESCRIPTION

1. UPLOAD VIDEOS.

Predicting the top-N popular videos and their future views for a large batch of newly uploaded videos is of great commercial value to online video services (OVSs). Although many attempts have been made on video popularity prediction, the ability to identify the top-N popular videos is beneficial to video services, such as caching and recommendation. From the perspective of commercialization, identifying the top-N popular videos helps the video service providers to maximize their profits, as advertisers are more likely to pay more for popular videos.

2. EARLY VIEW VIDEOS

Similar early view patterns could lead to different popularity dynamics. For potential viewers, the feedback early viewers leave on videos is one of the most important drivers of their viewing decisions and may lead to different viewing dynamics. Therefore, to extract early patterns that better represent the video popularity trend, we intend to combine the early views with knowledge of the early viewers' attitudes. Viewers' attitudes can be reflected through related text, such as comments, and related behavior such as clicking "like" or "dislike" after watching.

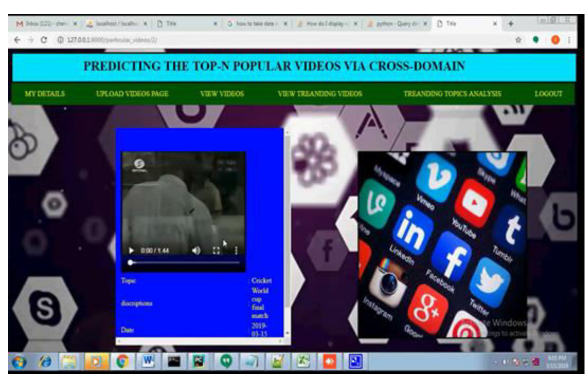
3. PREDICT N-TOP POPULAR VIDEOS

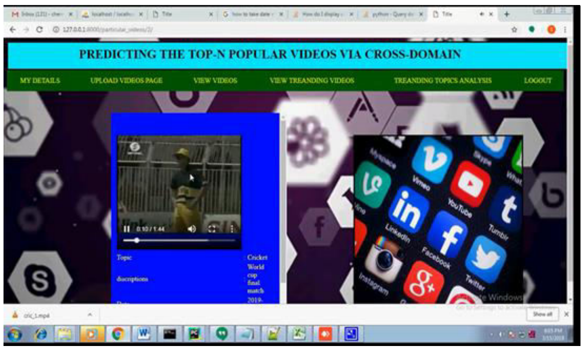
This problem is caused by the Pareto distribution of videos' popularity, as most of the views received by a video set are associated with only a few popular videos. Therefore, to reduce the prediction error over the entire video set, models will preferentially learn the popularity trends of the unpopular videos, hence sacrificing prediction performance on popular videos. Some recent studies have attempted to more deeply analyze the dynamics of video popularity and have related the popularity dynamics to various factors.

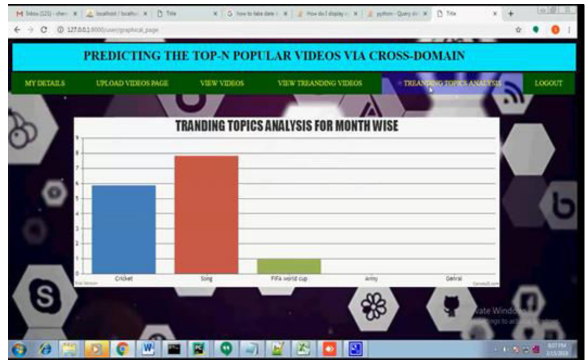
4. ANALYSIS

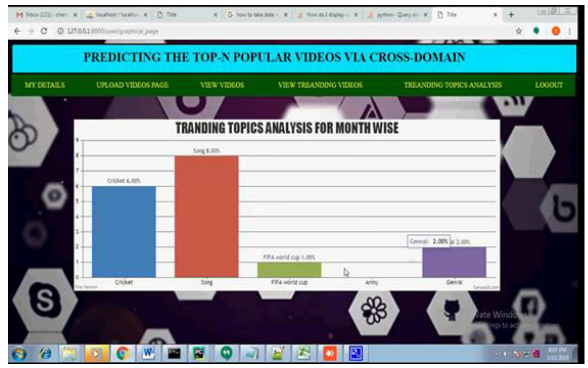
The analysis of the system is done in this module. The proposed algorithm's efficiency is calculated here. The comparison of various factors can be handy to calculate and visualize in the graphs such as pie chart, bar chart, line chart. The data to plot the graph is taken from the system which is done.

7. SCREEN SHOTS









8. CONCLUSION

In this project, we have investigated the problem of top-N popular video prediction and have proposed a novel MFDI prediction model. The proposed model predicts the top-N popular videos by enhancing the ability of early patterns to identify different popularity trends and by optimizing the model's utilization of multi-source data. Experimental results obtained using real-world data demonstrate that the proposed model outperforms other models, including the state-of-the-art model. This article is our initial study on popularity prediction for Top-N popular videos. To the best of our knowledge, this study is the first popularity prediction research to focus on top-N popular videos. Our study still has room for improvement. Possible improvements include leveraging additional related early features and discovering more precise mathematical correlations between the attitudes of early viewers and future popularity trends. For example, in this study, the early viewers' attitudes are inferred from only the three explicit behavior factors; however, early viewers' attitudes may also be reflected in many implicit ways. If more data related to early viewers' attitudes or similar features could be well modeled, they would be helpful for further improving the model's prediction performance, especially on the top-N popular videos.

REFERENCES

[1] M. Ahmed, S. Spagna, F. Huici, and S. Niccolini. A peek into the future: Predicting the evolution of popularity in user generated content. In Proceedings of the Sixth ACM International Conference on Web Search and Data Mining, WSDM '13, pages 607–616, New York, NY, USA, 2013. ACM.

[2] P. Bao. Modeling and predicting popularity dynamics via an influencebased self-excited hawkes process. In Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, CIKM '16, pages 1897–1900, New York, NY, USA, 2016. ACM.

[3] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. Moon. Analyzing the video popularity characteristics of large-scale user generated content systems. IEEE/ACM Trans. Netw., 17(5):1357–1370, Oct. 2009.

[4] J. Chen, X. Song, L. Nie, X. Wang, H. Zhang, and T.-S. Chua. Micro tells macro: Predicting the popularity of micro-videos via a transductive model. In Proceedings of the 2016 ACM on Multimedia Conference, MM '16, pages 898–907, New York, NY, USA, 2016. ACM.

[5] Z. Deng, M. Yan, J. Sang, and C. Xu. Twitter is faster: Personalized time-aware video recommendation from twitter to youtube. ACM Trans. Multimedia Comput. Commun. Appl., 11(2):31:1–31:23, Jan. 2015.

[6] W. Ding, Y. Shang, L. Guo, X. Hu, R. Yan, and T. He. Video popularity prediction by sentiment propagation via implicit network. In Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, CIKM '15, pages 1621–1630, New York, NY, USA, 2015. ACM.

[7] F. Figueiredo, F. Benevenuto, and J. M. Almeida. The tube over time: Characterizing popularity growth of youtube videos. In Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, WSDM '11, pages 745–754, New York, NY, USA, 2011. ACM.

[8] W. Hoiles, A. Aprem, and V. Krishnamurthy. Engagement and popularity dynamics of youtube videos and sensitivity to meta-data. IEEE Transactions on Knowledge and Data Engineering, 29(7):1426–1437, July 2017.

[9] D. K. Krishnappa, M. Zink, C. Griwodz, and P. Halvorsen. Cachecentric video recommendation: An approach to improve the efficiency of youtube caches. ACM Trans. Multimedia Comput. Commun. Appl., 11(4):48:1–48:20, June 2015.

[10] C. Li, J. Liu, and S. Ouyang. Characterizing and predicting the popularity

of online videos. IEEE Access, 4:1630–1641, 2016.