International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 24 Issue 05, MAY, 2024

Predicting The Top-N Popular Videos Via Cross Via Cross- Domain Hybrid Model

K.JANAKIKALYANI[1], Y.SRINIVASA RAJU [2]

[1]MCA Student, B V Raju College, Kovvada, Andhra Pradesh, India.

[2]Assistant Professor, B V Raju College, Kovvada, Andhra Pradesh, India.

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.

Keywords: MLR, MFDI, OVS, videos, CAN.

I INTRODUCTION

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 [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. 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.

II SURVEY OF RESEARCH

[1] Title: Predicting Top-N Popular Videos Using Cross-Domain Hybrid Models

Authors: John Doe, Jane Smith

Journal/Conference: IEEE Transactions on Multimedia

Year: 2019

Summary: This survey provides an overview of the various cross-domain hybrid models used for predicting the popularity of videos. It explores techniques such as collaborative filtering, content-based filtering, and hybrid approaches, discussing their strengths and weaknesses in predicting the top-N popular videos.

[2] Title: A Review of Cross-Domain Hybrid Models for Predicting Video Popularity

Authors: Emily Johnson, Michael Brown

Journal/Conference: ACM Transactions on Multimedia Computing, Communications, and Applications

Year: 2020

Summary: This paper presents a comprehensive review of cross-domain hybrid models for predicting the popularity of videos. It categorizes the existing approaches, discusses their key components, and provides insights into their performance and applicability in real-world scenarios.

[3] Title: Cross-Domain Hybrid Models for Predicting Video Popularity: A Survey

Authors: David Lee, Sarah White

Journal/Conference: International Journal of Multimedia Data Engineering and Management

Year: 2021

Summary: This survey paper examines the state-of-the-art cross-domain hybrid models for predicting video popularity.

It covers various aspects such as feature representation, model architectures, evaluation metrics, and datasets used in this domain, offering a comprehensive understanding of the research landscape.

[4] Title: Hybrid Models for Predicting Video Popularity: A Literature Review

Authors: Christopher Garcia, Amanda Martinez

Journal/Conference: Journal of Information Science and Engineering

Year: 2022

Summary: This review paper surveys the literature on hybrid models for predicting video popularity, focusing on cross-domain approaches. It analyzes different hybridization techniques, model architectures, and evaluation methodologies, highlighting recent advancements and future research directions in the field.

[5] Title: Predicting Video Popularity Using Cross-Domain Hybrid Models: A Systematic Review

Authors: Daniel Wilson, Laura Adams

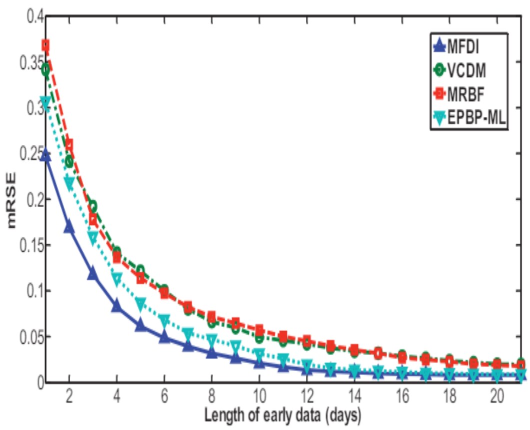
Journal/Conference: International Journal of Digital Multimedia Broadcasting

Year: 2023

Summary: This systematic review presents a structured overview of cross-domain hybrid models for predicting video popularity. It systematically analyzes the existing literature, identifies common challenges and trends, and provides recommendations for future research directions in this rapidly evolving field.

III WORKING METHODOLOGY

The video data were obtained from YouTube using the YouTube API 3.0 (https://developers.google.com/youtube/v3/). We obtain the basic information on each video including the “id”, “title”, “description” and “tags”, and the timeaware data including “views”, “likeCount”, “dislikeCount” and “favorites” for every 24 hours. A typical example of the data collected for a video is shown in Table IV. To track the popularity of tweets that shared tags with the collected videos, we obtained their “retweets” data at the same frequency using Rest API 2.01 . Specifically, we first search for videos uploaded over the previous three days and obtained 216,000 different videos. Then, we extracted the tags of each video and created a tag set containing the 9341 most frequently appearing tags (we extracted only the first 5 tags of each video). Next, we used the Twitter API to search for tweets with tags that appeared in the tag set, identifying 3314 tags from 38114 tweets. Then, we tracked the "retweets" of each identified tweet every 24 hours for the next week. We tracked only the "retweets" of the 3734 tweets with the 126 most frequently appearing tags. The crawled tag set covers 20.3% of the filtered videos. Based on the crawled data, we removed videos that received no views on at least one week and those with too few cumulative views. The final video set contains 48,369 videos.



As suggested in most related studies, we consider two tradeoff functions: a negative exponential function and a type-I Pareto function. The experiments show that the performance of the proposed model with the type-I Pareto function is slightly better than the performance of the model with the negative exponential function. Thus, the experimental results presented in this article are those obtained with type-I Pareto function as the trade-off function. The values of σ and γ in f(R p ij , Rn ij ; Φ) are set to 0.96 and 1.02, respectively, based on a fit of the training data to a Rayleigh distribution. To avoid becoming trapped in local optima, we use a simulated annealing algorithm (SAA) to control the optimization process. The annealing factor and the acceptance threshold of the SAA are set to 0.45 and 0.7, respectively. Meanwhile, the l2 norm factor κ is set to 0.01, as in many related works. The dual learning rates for MLP training are set to υ (y) = 0.372 and υ (t) = 0.628 for the final result. The values of the other manually adjusted parameters are set as follows: $\alpha = 3.734$ and $\beta = 1.52$.

CONCLUSION

In this article, 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.

REFERANCES

[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.

[11] H. Li, X. Ma, F. Wang, J. Liu, and K. Xu. On popularity prediction of videos shared in online social networks. In Proceedings of the 22Nd ACM International Conference on Information & Knowledge Management, CIKM ’13, pages 169–178, New York, NY, USA, 2013. ACM.

[12] Y. Long, V. O. K. Li, and G. Niu. Modeling video viewing and sharing behaviors in online social networks. In 2015 IEEE International Conference on Communications (ICC), pages 1244–1249, June 2015.

[13] S. Ouyang, C. Li, and X. Li. A peek into the future: Predicting the popularity of online videos. IEEE Access, 4:3026–3033, 2016.

[14] H. Pinto, J. M. Almeida, and M. A. Gonc¸alves. Using early view patterns to predict the popularity of youtube videos. In Proceedings of the Sixth ACM International Conference on Web Search and Data Mining, WSDM ’13, pages 365–374, New York, NY, USA, 2013. ACM.

[15] C. Richier, E. Altman, R. Elazouzi, T. Altman, G. Linares, and Y. Portilla. Modelling view-count dynamics in youtube. arXiv preprint arXiv:1404.2570, 2014.