IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 6, June 2021

DOI 10.17148/IJARCCE.2021.10616

A Framework for Prefetching of Social Media Content for Analysing Twitter Data

Saichandana S[1], Kavitha Sooda[2]

Computer Science and Engineering, B.M.S. College of Engineering, Bengaluru, India[1]

Associate Professor, Computer Science and Engineering, B.M.S. College of Engineering, Bengaluru, India[2]

Abstract: In the recent days, online social networks (OSNs) are widely known as one of the largest platforms for the source of information and content that can be shared among people. As a socially oriented learning-based model has been proposed with a view to provide quality of experience (QoE) for social media services. This is for media content prefetching in order to improve the access delay reduction and to enhance OSN user's satisfaction. With a wide spread of data-driven analysis during a period of fourteen months, over the Twitter traces on real-life from over 2,800 users, it is revealed that the social relationship has a wide impact on user's social media behaviour. In order to represent this scenario, a social relationship of clusters for a large group of friends has been conducted, and further to expand a cluster-based machine learning model for socially-oriented learning based prefetch prediction. And then to predict users influence on the social media app, a usage-adaptive prefetch scheduling mechanism has been used by considering different users which might possess heterogeneous user's app usage way. A framework has been proposed that can be evaluated using a trace-driven analysis on the social media. The data can be dumped onto a cloud for further analysis by using different machine learning approaches.

Keywords: Online Social Networks, Prefetching, Quality of Experience, Spice, Twidere, Usage Adaptive prefetching scheduling.

I. INTRODUCTION

Apart from being a platform for social relationship, social networks are emerging as a prominent platform for all the information and of content sharing. Social media is been widely used among everyone. It has become a ubiquitous usage of the media, can be accessed anywhere and at any time. Social media becoming a part of our daily lives, where one gets daily news, entertainment, or something related to technology, etc. Around 47% and 52% of the social media users gets news from Facebook and Twitter, respectively [1]. A shared media contains files such as images and videos, which contains larger data than the text content in the users' posts. To ensure quality of experience (QoE) support for the devices, a promising approach calls for a user-friendly pattern.

A key factor of degrading the mobile OSN user's satisfaction is the access delay in consuming the social media content in OSN. There is a kind of limited network bandwidth, long round-trip time, and wireless connection establishment latency, which would impair user's social media app usage. Considering this issue, a prominent approach is to influence prefetching. That is downloading the media content prior to user's influence on the social media [2]. To know about the advantage of prefetching, is to predict the user's media behaviour in an accurate manner.

By achieving precised content prediction could help to prefetch the most appropriate content, so that this would be useful for the user. Also this is helpful for the access delay reduction and furthermore saves the data traffic and energy consumption, avoiding prefetch of large amount of social media content.

In this paper we have a tendency to propose an approach that uses the distinctive users' behaviours of social media relationship among them for media content prefetching. Finally, socially oriented data mining tools and the cluster-based machine learning algorithms has been leveraged, to deduce a user's social media consumption on the user's app usage way and social relationship preference. Furthermore, we combine such an abstract thought, the app usage way and network combining to execute the entire prefetching mechanism. Precisely, all the twitter tweet traces from Twitter, has more than 1,000,000,000 downloads on Playstore [3] and over 300 million users, and thus growing monthly. This allows us to conduct a data-driven and trace-drive analysis. Over the large-scale of data analysis, social relationship has been revealed which has an analytical impact on the user's tweet behaviour. With this, a social media clustering is been made to classify different user friends into different groups. Also to design a cluster-based model to analyse the user's app usage. Thus allowing users to dump their machine learning procedures to cloud, furthermore to reduce the network traffic and energy consumption and the processing latency.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 6, June 2021

DOI 10.17148/IJARCCE.2021.10616

Some of the analysis made in the paper [4] are as follows:

- To provide a Quality of Experience service for the social media, they proposed a framework analyze the prediction of prefetching the media content called Spice Framework.
- Collecting a large number of Twitter traces for over a period of 5 months around 17,000 users data being collected every month to reveal the user's influence on the social media.
- Such a data-driven analysis is being made on the user's social network. They have developed Latent Bias Model for the prediction of prefetching.
- A social friends clustering for a part of user's friends is been conducted, further to develop a cluster-based mechanism for socially oriented prefetch prediction accordingly. Trace-driven emulation for this traces is been conducted over a group of user's friends to achieve a better performance.
- Then a usage based prefetch scheduling scheme for a set of different users' social media usage pattern is developed.
- They partitioned the Twitter data analysis into several zones. They have partitioned zones into 4 parts, as morning time, afternoon, evening and night schedule.
- They have also classified as a part of users' behavior as busy, active, occasional and break period.
- Then the performance of the framework is been evaluated by a trace-driven evaluation.
- The performance results show that the framework improves the access delay reduction to around 84.5% accuracy, which is much a better improvement over the benchmark procedures.
- To increase the execution process, we can dump the machine learning procedures to the cloud.

The rest of this paper is organised as follows. Section II presents an overview of the Literature Survey, which includes all the related work from the various authors. Section III generalizes the proposed framework for this paper, which briefly describes about the system architecture. Finally, section VI includes the conclusion and the future work of the paper.

II. LITERATURE SURVEY

In most of the social media, there exists a tie between users which is a significant role in the user's social media influence. One among a way this social tie could occur is through the similar user's behaviour or the user friends who act in a same way on the social media. In the social media, the user social influence on different modes of access or new technologies can vary through network as a wide range. So by recognizing the user's influence is a critical task. In this paper [5] they study this problem efficiently. They fairly defined corresponding models that recreate the above-mentioned correlation of users' sources. Further they have proposed 2 simple tests which help to predict users influence as a part of social correlation. They have achieved by giving a theoretical conclusion by providing a high probability by one such test. This is successful in bringing out user's influence in a general model of the correlation.

They have simulated all their test cases with a numerous examples by designing randomly generated influence of users on real-time social media (Flickr). This paper has a limitation which says the techniques only provides a qualitative justification of the existence of users influence but could not provide a quantitative indication. Also no formal verification of results is discussed from their analysis in this paper.

In this paper [6], they accomplished social ties of users in social media in order to improve cooperative device-to-device (D2D) communications. Clearly, as handheld mobile devices being carried by humans, to promote for better and efficient cooperation among the devices, they enhance 2 main social mechanisms that are trust and reciprocity on social media. With this thought, they have developed a mechanism with network assisted relay selection to carry through the game theoretic solution and show that the framework is proof against group and individual rational, truthful, and is computationally efficient. They have evaluated this mechanism's performance by using real-time social media traces. They have also achieved significant performance over the case from the simulation results without the D2D cooperation.

In this paper [7], with a large set of Twitter data, they have made comparison among three influential measures: retweets, indegree and mentions. With these measures, they explored the actions of users influence across diverse themes and time periods using these metrics. They have also made various interesting observations. Firstly, the users with high popularity are said to have high indegree who need not influence on social media by retweeting or mentions.
Secondly, the high influenced users who probably have a great influence against various topics. Thirdly, users influence on the medial is not achieved accidentally or spontaneously but with their cooperative effort by limiting certain tweets to one topic. They have also believed that these observations give some new thoughts for viral activity and recommends that the influential measures such as the high indegree can alone reveal very less user's influence.

In this paper [8], IMP that represents a straightforward interface which detects and conceals the complexity of prefetching. IMP employs a cost-effective analysis for deciding when to prefetch the media content. It also uses a purposive adaptation to trying to minimize response time of an application by achieving costs for energy and data. IMP

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 6, June 2021

DOI 10.17148/IJARCCE.2021.10616

makes use of the existing network to make sure the prefetching does not degrade network performance for foreground activity. It also keeps track for the previous prefetches of the highest rates for the data traffic. These experimental results with the emails and with some other social media applications illustrates that IMP gives an accurate usage of limited resources, by decreasing application response time. Their evaluation result shows that IMP meets the goal with targeted energy and data consumption. Also, they have shown that in most of the cases, IMP performs better than those mechanisms on an interactive prefetch time.

Haewoon Kwak et al. [9] have crawled the complete Twitter app and obtained the results as 41.7 million profiles, 1.47 billion friends, 262 most trending topics, and around106 million tweets. With the friendship analysis, they marked a unimportant followers, which makes them different from the other users. To identify the users influence on the Twitter, they ranked each user with the total number of friendship using PageRank algorithm. A quick glance at retweets shows if any retweets are supposed to target an average number of 1000 tweets by not considering what number of friends is for the first tweet. When a tweet is retweeted, it is retweeted immediately, indicating a wide spread of content after first retweet. They made an analysis that the most trending topics are based on the temporary user's influence. They also observed the trend topics by the user's activity on Twitter. Thus, they concluded that the trending topics are over 85% headlines or some news.

In this paper [10], the existing studies are about data mining on Smartphones. Their observations for the various studies are gathered based on two research analysis; firstly, prediction of user app usage and applications recommendation along with other studies. And also, they have presented many challenges in the data mining of smartphone's app usage analysis.

In order to utilize the mobile connectivity, they have identified an individual's to carryout similar paths daily. Their strategy [11], Breadcrumbs marks all the path of user's location and designs a prognosticative model for that individual user. Based on the previous observations of wireless networks, they have combined this for the BreadCrumbs to generate connectivity models. They have also designed a BreadCrumbs framework and illustrated its activity with the analysis of weeks of usage of a device. Their evaluations shows that this analysis has got sufficiently accurate results, even with a one week of analysis, they could even provide better performance with less power usage for most of the apps. BreadCrumbs even make analysis of the bandwidth quality of applications and also the latency of the cellular network. Their model also predicted the upstream and downstream bandwidth for a particular period of time.

In this paper [12], they have made an app prefetch prediction practically on mobile phones. Their contributions measures are of two-fold. Firstly, they design an app prediction algorithm, APPM, that need not be trained in prior, which adapts to the users app usage dynamics and predicts the probability of not only which next application would be used but also when it will be used, and it also provides a high accuracy without the need of any additional sensor content. Secondly, they have performed a parallel prefetching on the screen unlock, which is a mechanism that leverages the benefits of prefetch prediction. They have conducted their experiments for long-term data traces, and live deployments on the Android Play Market and shows that they outperform previous approaches to prefetch predicting app usage, by also providing some practical ways to prefetching application content on mobile devices.

This paper [13] mainly concentrates on understanding the users' social media influence. They have focused on using Twitter as a micro blog service. From the Twitter dataset, which is collected for this study is being observed that over 72.4% of the Twitter users follows more than 80% of their followers. Around 80.5% users have 80% users following back them. On this basis, TwitterRank algorithm, a supplement of PageRank algorithm has been used to measure users' influential on Twitter. The experiment results show that TwitterRank algorithm performs better than the present Twitter uses, namely PageRank and Topic-sensitive PageRank algorithms.

In this paper [14], from the F-tree construction algorithm and the mining phases which maintains persons information in a tree and processes to get set friends of a strong group. The main advantage of maintaining all such person information with the F-tree is to allow us to get extra knowledge about the social relationship among users. They have easily identified the association of the social relationship among friends groups. All these information may provide us additional knowledge with different aspects which include studying social relationship association and the system administration. Therefore, discovering all the strong friend group association is their main contribution of the paper. They have proposed an algorithm which consists of two steps, that is, construction and mining of friend –tree (F-tree). Their experimental results show that mining with F-trees reduces the false positives number.

A. Social Media Prefetching

This part describes the socially-oriented network analysis and media content prefetching. In paper[4], they have used a cluster-based model to analyse a group of social friendship, which had been significant part in achieving a better prefetching.For the media content prefetching, [8] represents IMP model which is used as a prefetching mechanism to

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 6, June 2021

DOI 10.17148/IJARCCE.2021.10616

reduce energy and data consumption. In [10], they include a model for when to invoke prefetching mechanisms for different categories including network and apps energy usage. They have shown only a qualitative measure but not a quantitative analysis. Among many literature survey [8], [15] and [12]) which aims at designing prefetching approaches for generic purpose.

B. Socially-Oriented Network Analysis

For the socially-oriented network analysis, [5] and [9] observes users' social ties by social graphical structure which uses Flickr as a social media network [9], [7] which detects the social users influence using PageRanking algorithm with number of friendship, retweets or mentions, and similarly [16] uses all the user friendships to describe the users' influence on the social media.

III. PROPOSED FRAMEWORK

The framework illustrates how system architecture works. Spice works as a user-centric as described in paper [4]. A social media app such as Twitter, Instagram, Facebook or WeChat [17] etc, can be used to rank media content based on the usage way of user. The idea here is to know when and how the prefetching mechanism to be invoked, and what amount of data to be prefetched. Followed by this, applies a learning based machine learning model to decide when to prefetch the media contents. The below is the system architecture for prefetching of social media content in Online Social Networks. It works an a user-centric manner as illustrated in the figure 1, and collects all the user twitter traces on the user's Twitter feed while accessing it [3]. All the Tweet traces are fetched from the Twitter REST API present in Twitter Wrapper, which is in turn controlled by Scheduler. We can collect Twitter traces as much as 1000 users data per day. The Task Scheduler queries for new Twitter traces to fetch from their newsfeed. All these data is being sent to the Aggregator.

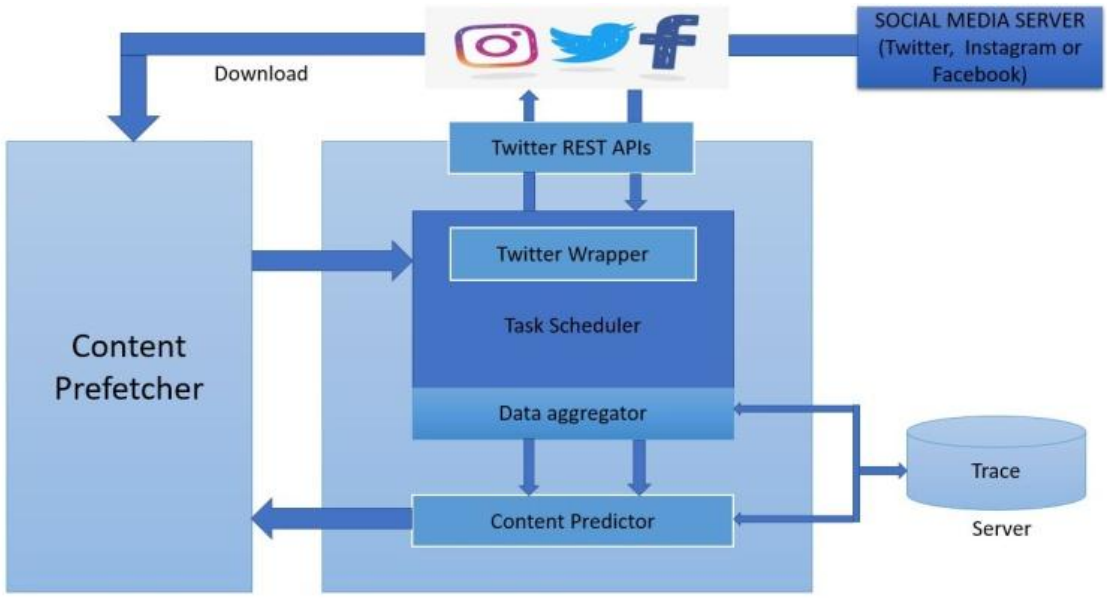

Fig. 1 A proposed framework

All the prefetched data can further filter the duplicity of the users or the existing user. To analyze the prefetching, a popularity of a user could be found out and trace-driven evaluations could be done to it. To prefetch the related media data easily, we can dump the data to a cloud for later use. Further the data is been passed to Content Predictor to which a learning based content prediction model is applied to predict user's app usage. Few other machine-learning algorithms can used to train the model in the Content Predictor Component to get better prefetching accuracy. Specifically, all the new tweets are fetched as an input and trained using a machine learning algorithms to predict most relevant features of media files. These contents are prefetched by a Content Predictor. Later all the media files are then stored on the cloud server for easy access. Once the Twitter data is being fetched by the Twitter API [18], must be preprocessed before applying any statistical methods can be applied. Using some word vector process, tweets are to be cleaned and rescaled. Identify all the unique features of the users' tweets and analyze according to the popularity of the user. Then train the model with some machine learning algorithms to get a best performace results. These machine learning techniques to be trained to each of topics of the event. And further applying some correlation process to get the power of standard value to find the best performance of accuracy.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 6, June 2021

DOI 10.17148/IJARCCE.2021.10616

IV. CONCLUSION

Focusing to develop a prefetching mechanism to fetch media content in Online Social Networks, this paper firstly focuses on identifying the unique behavior of a user in Online Social Networks. Further to implement a cluster analysis on a group of user social friends to predict the users' behavior. Based on the usage of social media apps, an adaptive prefetch scheduling mechanism is been proposed to determine the usage pattern of a user. The performance analysis is made to be more accurate than that of the recent studies made according to the proposed system. To further improve the accuracy, we could offload the machine learning algorithms to the cloud server for later use by the user. In spite of that, the proposed model can also be applied to any other social networks. Furthermore, we can use images and videos as an extension of this paper for the analysis.

REFERENCES

[1]. [Online].Available:http://www.journalism.org/2013/11/14/news-useacross-social-media-platforms/
[2]. Wang, Yichuan, Xin Liu, David Chu, and Yunxin Liu. "Earlybird: Mobile prefetching of social network feeds via content preference mining and usage pattern analysis." In *Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pp. 67-76. 2015.
[3]. "Twidere for twitter. http://twidere.mariotaku.org/."
[4]. Wu, Chao, Xu Chen, Wenwu Zhu, and Yaoxue Zhang. "Socially-driven learning-based prefetching in mobile online social networks." *IEEE/ACM Transactions on Networking* 25, no. 4 (2017): 2320-2333.
[5]. Anagnostopoulos, Aris, Ravi Kumar, and Mohammad Mahdian. "Influence and correlation in social networks." In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 7-15. 2008.
[6]. Chen, Xu, Brian Proulx, Xiaowen Gong, and Junshan Zhang. "Exploiting social ties for cooperative D2D communications: A mobile social networking case." *IEEE/ACM Transactions on Networking* 23, no. 5 (2014): 1471-1484.
[7]. Cha, Meeyoung, Hamed Haddadi, Fabricio Benevenuto, and Krishna Gummadi. "Measuring user influence in twitter: The million follower fallacy." In *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 4, no. 1. 2010.
[8]. Higgins, Brett D., Jason Flinn, Thomas J. Giuli, Brian Noble, Christopher Peplin, and David Watson. "Informed mobile prefetching." In *Proceedings of the 10th international conference on Mobile systems, applications, and services*, pp. 155-168. 2012.
[9]. Kwak, Haewoon, Changhyun Lee, Hosung Park, and Sue Moon. "What is Twitter, a social network or a news media?." In *Proceedings of the 19th international conference on World wide web*, pp. 591-600. 2010.
[10]. Cao, Hong, and Miao Lin. "Mining smartphone data for app usage prediction and recommendations: A survey." *Pervasive and Mobile Computing* 37 (2017): 1-22.
[11]. Nicholson, Anthony J., and Brian D. Noble. "Breadcrumbs: forecasting mobile connectivity." In *Proceedings of the 14th ACM international conference on Mobile computing and networking*, pp. 46-57. 2008.
[12]. Parate, Abhinav, Matthias Böhmer, David Chu, Deepak Ganesan, and Benjamin M. Marlin. "Practical prediction and prefetch for faster access to applications on mobile phones." In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, pp. 275-284. 2013.
[13]. Weng, Jianshu, Ee-Peng Lim, Jing Jiang, and Qi He. "Twitterrank: finding topic-sensitive influential twitterers." In *Proceedings of the third ACM international conference on Web search and data mining*, pp. 261-270. 2010.
[14]. Cameron, Juan J., Carson Kai-Sang Leung, and Syed K. Tanbeer. "Finding strong groups of friends among friends in social networks." In *2011 IEEE Ninth International Conference on Dependable, Autonomic and Secure Computing*, pp. 824-831. IEEE, 2011.
[15]. Ravindranath, Lenin, Sharad Agarwal, Jitendra Padhye, and Christopher Riederer. "Give in to procrastination and stop prefetching." In *Proceedings of the Twelfth ACM Workshop on Hot Topics in Networks*, pp. 1-7. 2013.
[16]. Bakshy, Eytan, Jake M. Hofman, Winter A. Mason, and Duncan J. Watts. "Everyone's an influencer: quantifying influence on twitter." In *Proceedings of the fourth ACM international conference on Web search and data mining*, pp. 65-74. 2011.
[17]. "Wechat. http://en.wikipedia.org/wiki/wechat."
[18]. "Twitter rest apis. https://dev.twitter.com/."

BIOGRAPHIES

Saichandana S pursuing M.Tech in Computer Science and Engineering at B.M.S. College of Engineering, Bengaluru. She has completed her Bachelors of Engineering in the area of Computer Science and Engineering at C Byregowda Institute of Technology, Kolar in the year 2019. Her interest includes studies in the area of Cyber Security and Computer Graphics.

Kavitha Sooda holds Ph.D in Computer Science and Engineering. She has nineteen years of teaching experience and completed her Post-Doctoral work on Higher Education System from IISc, Bengaluru in the year 2018. Her interest includes routing techniques, QoS application, Cognitive Networks, Evolutionary Algorithms and Higher Education System. Currently she works as Associate Professor at B.M.S. College of Engineering, Bengaluru.