Data Mining and Cluster Based Machine Learning in Mobile Online Social Networks using Spice

Miss. Sushma Sri Enapanuri[1], Mrs. Jayashree S Patil[2]

[1]M.Tech(cse), G. Narayanamma Institute of Technology and Science, Hyderabad
[2]Associate Professor, Department of CSE, G. Narayanamma Institute of Technology and Science, Hyderabad

---***---

Abstract - *To distribute the content and information among the people, mobile Online Social Networks (OSNs) are emanating as the most popular platform. With the universal distribution and demand of wireless communication, more and more users are using online social networking resources in wireless telephone devices. Hence, to reduce the access delay and improve quality of experience support to attain mobile user's fulfillment in the media content prefetching, a socially-driven learning based framework called Spice has been proposed.*

The Spice framework is estimated to attain high-performance with high access delay reduction at the low cost of cellular data and energy consumption. A speed-up over the local data training execution on smartphones is also expected by enabling the users to offload their machine learning procedures to a cloud server.

Key Words: *Mobile computing, online social network, multimedia applications, quality of experience.*

1. INTRODUCTION

The Online Networking stage where every single unique individual can associate with others through their cell phones and various contraptions is known as portable online informal organizations (OSN's). The social correspondence plots, for example, Facebook and Twitter have made portable applications to give constant access to web from wherever.   So OSN's, for example, facebook and twitter are filling in as the stage for social collaboration. They are emerging as the exceptionally popular terrace so as to share the substance and data among the individuals. As the versatile web has step by step created from exclusive portable advances and systems to Internet versatile access, the distinction has changed to two kinds: 1. Web – based social networks: It extended the mobile access through mobile browsers and smartphone apps. 2. Native mobile social networks: It focuses on mobile usage such as mobile communication; location based services and augmented reality.

It was accounted for that these days 68% of the OSN administration utilizations happen on cell phones and on a normal, a client burns through 2hours 25minutes every day utilizing OSN administrations. Generally, 52% and 47% of the client's gets news from twitter and facebook and so forth, subsequently, OSN's are showing up as the standard channel for data and substance sharing.

With the expanding notoriety and use of OSN's, there are a portion of the key components for the breakdown of versatile customer's satisfaction in devouring media content.

1. Access delay.

2. High wireless connection establishment latency.

3. Long roundtrip time of data transmission.

Some other factors include,

- Time varying network quality and
- Sporadic network availability, which leads to fluctuating connection.

To beat this, an emphatically engaging and promising methodology called Prefetching i.e., to download the media content sooner than client's utilization at whatever point there is a likelihood for the greatest bit of leeway. This can diminish the access delay, saving both energy and information traffic utilization.

For the intention of improving the forecast exactness of media content prefetching in OSNs on cell phones, a structure of EarlyBird was proposed. The real want is to coordinate tweet preparing properties into a model called direct relapse for accomplishing expectation by mining the client's OSN utilization design. The crucial responses of the earlybird model are that, it has not contributed sufficient examination for social effect among the clients, which assumes a basic job in media content utilization in OSNs.

II RELATED WORK

Prefetching-Based Mobile Dictionary:

To explain the prefetching effect, we describe each working procedure is described in two different cases: In the first case, the prefetch technique is not used and in the second case the prefetching technique is used. During the first case assuming that a user wants to know the meaning of the word w1 in the dictionary. If the user inputs w, into the mobile, then the server will return the result r1 for w1 to the mobile. Next, if the user requests the meaning of w2 in the same way, the server will return the result r2 for w2 to the mobile. Likewise, if the prefetching is not used, the mobile device needs to access the dictionary server for every word, and it causes frequent roundtrips.

The rapidly expanding demand for digital mobile communication services, in conjunction with the recent advancement in mobile technology, has led to the development efforts for future mobile systems directed towards Mobile Computing - a new dimension and requirements for future communication and computing networks. The mobile applications can take advantage of contextual information (such as user location, time of day, nearby people and devices, and user activity) [1]. There are many popular value-added applications that specifically target the mobile client's location as context information such as traffic condition, tourist, weather information, location-dependent advertising and disaster management systems [2].

III METHODOLOY
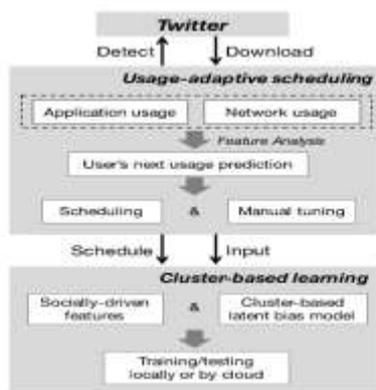
Figure 1: Logical workflow of the Spice mobile media prefetching system

The client follows are gathered from Twidere, an Android Twitter application which has near 500,000 downloads on Google Play and more than 17,000 clients consented to report use information. This aides in leading an information driven investigation and plan, and a sensible follow driven execution assessment.

Through huge scale information investigation, it was uncovered that the social cooperation among clients has extraordinary effect on client's tweet click conduct. Contingent upon this, social companionship bunching is directed to characterize a client's social companions dependent on significance. Next a group based Latent Bias Model is created to appraise the probability of media substance clicking.

Afterward, a versatile planning plan was intended to direct the media tweet prefetching in energy and cell information effective way which empowers the Spice clients to dump their information preparing errands for AI to a cloud server, in order to battle the high vitality utilization and long handling inertness.

In Spice, the prefetching errand characterized ought not exclusively be programmed yet additionally use versatile, which prompts basic cell information stream and battery

productivity prerequisites on a starter that media documents stacking postponement can be ensured.

It is critical to choose when to prefetch as indicated by client's profile, arrange utilization inclination, application use liveliness, and the manual tuning factors. As Spice is a socially-determined usage for versatile media content prefetching. It is critical to deal with the social kinship impact, setting inclination, and OSN media properties.

Towards this objective, a socially-determined learning-based calculation was presented which would be affected by the social fellowship and setting highlights.

IV EXPERIMENTS

Data Prefetching is the process of retrieving and caching data before the user has requested it. It is a technique proposed to reduce latency to data and ensure data access in case of disconnection. It can speed up the user's experience with his mobile app.

Modern mobile operating systems such as Android and IoS enable applications to cache data on the device. On Android Settings –> Storage, one can easily see the amount of cache storage used by each application. This section looks at algorithms that each application can use to determine what should be cached; the next section looks at how the operating system should allocate cache spaces among the applications. Caching is an extensively studied subject in system research, from models of data access, to universal algorithms such as LRU or LFU, to specialized algorithms for file systems or databases. Unfortunately, most of the research results do not apply to caching in mobile applications for one simple reason: the source of the data access is not a program, it's a human!

For example, the locality model, which states that a recently used item is likely to be used again, rarely applies to caching on mobile devices. In fact, in the case of entertainment applications, items consumed (e.g. a video that has been watched, an article that has been read) are less likely to be consumed again.

Caching in mobile applications needs non-traditional algorithms. There are three distinct goals of caching on mobile devices:

• reduce latency of mobile web access and improve user experience of mobile applications;

• reduce data cost;

• improve availability of mobile applications during periods of significant degradation in cellular data services, such as crowded events.

Each of the above goals needs its own specialized algorithms.

Caching for Latency Reduction:

For mobile web, caching on the mobile device is implemented by the browser, and using browser cache to reduce web latency is generally well-understood [8]. All CSS files and java scripts on the critical rendering path should be marked cacheable by the web site through explicit Expires or Cache-Control headers. Unfortunately, browser implementations offer no guarantee of caching any file. HTML5 local storage [12] gives web sites much better control on what are cached and what are not, though the size cap on local storage limits its utility. There are commercial solutions that transform web sites for better browser caching on mobile [9].

Many mobile applications also use caching to reduce data fetches and improve user experience, though the impact of caching on user-interaction latency varies by application.

Caching to Reduce Data Cost:

Traditional algorithms such as LRU can be applicable in certain types of applications. If the common usage pattern of the application is such that a user might concentrate on a small set of objects for some period of time, then LRU can be used. For example, a document reader application might want to cache recently opened documents, since the user is likely to work on them again. However, for entertainment applications and social applications, LRU is often the wrong algorithm to use. Instead, a "predicted probability of reuse" function, similar to the predicted probability of consumption function, should be used. Many applications have good models to predict reuse of the data objects; for example, video applications might know that music videos tend to be re-watched more often than non-music videos, social applications might know that pictures with lots of comments tend to be viewed and commented again, etc. Assuming that the function exists, the caching algorithm should simply cache objects in order of descending probability of reuse. Free diagnostic tools also exist to help mobile application developers identify what objects should be cached, for example, the Application Resource Optimizer from AT&T [7].

Caching to Improve Availability:

When the goal of caching is to improve application ability under little or no mobile wireless connectivity, the value of an object isn't just the likelihood of a user needing it, but also the likelihood of the user needing it at a time of low wireless connectivity. Since low wireless connectivity is typically a property of the geo area, a mobile application can monitor the wireless performance of its data fetches as the device roams, and keep track of the bandwidth experienced at each geo location. The application cache can then favour objects that are needed at low-connectivity locations, either through semantic connection (e.g. in the case of map application), or through observations of past user behaviour (e.g. for users with daily routines, objects needed at a particular geo location follow a diurnal pattern).

V RESULTS

HOME SCREEN

This is the home page where initially user dataset is uploaded which contains the information related to user and their friends list.

UPLOAD DATASET

Dataset is uploaded

AFTER SUCCESSFUL UPLOADING OF DATA SET

Successful uploading of the dataset

GENERATE CLUSTER

Generation of three different clusters

CONTENT PREDICTION

An alert message is displayed.

VIEW PREDICTION RESULT
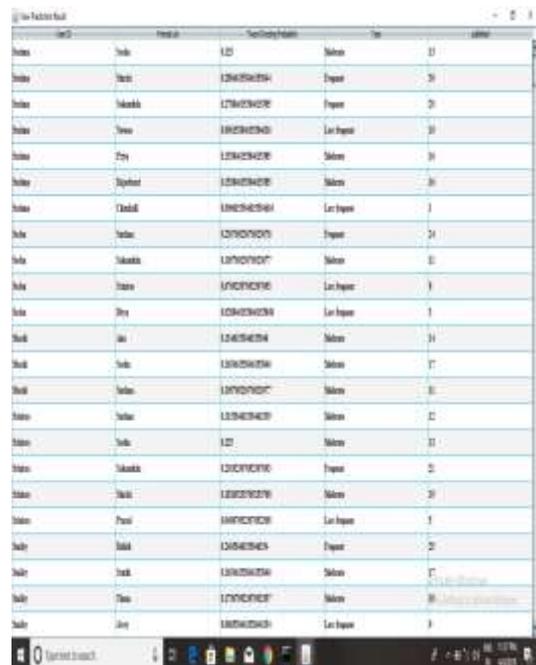
The results of tweet checking probability, friend type, number of tweets published.

CLUSTER GRAPH
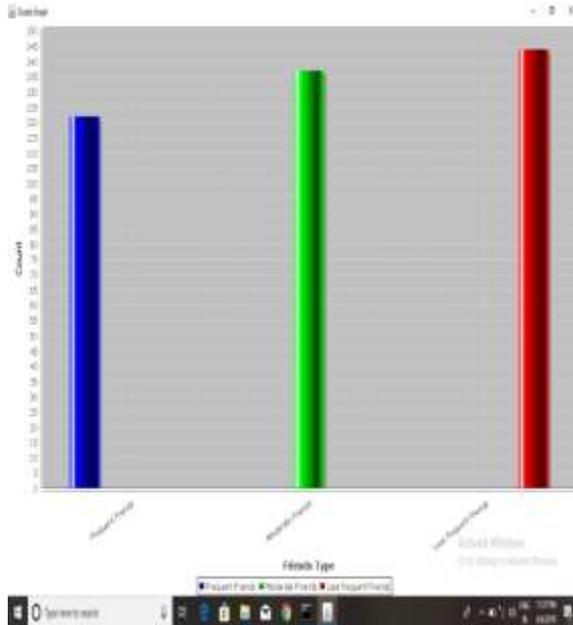
Generation of cluster graph based on the friend type and count values.

VI CONCLUSION

Due to fundamental challenges in scaling wireless data networks, wireless data remains a non-negligible cost item for many users in the world. Intelligent caching and preloading of data in mobile applications can significantly improve user's mobile experience, and is an area that needs active research. To facilitate the research community's work in this area, the following data need to be collected, suitably anonymized, and made available for researcher's use.

REFERNCES

[1] C. Wu et al., "Spice: Socially-driven learning-based mobile media prefetching," in Proc. 35th Annu. IEEE Int. Conf. Compute. Commun. (IEEE INFOCOM), Apr. 2016, pp. 1–9.

[2] A. Lella, A. Lipsman, and K. Dreyer. U.S. Digital Future in Focus, accessed on Apr. 2014. [Online].http://www.comscore.com/Insights/Presentations-andWhitepapers/2014/2014-US-Digital-Future-in-Focus

[3] S. Kemp. Digital, Social & Mobile Worldwide in 2015, accessed on Jan. 2015. [Online].Available:http://wearesocial.net/blog/2015/01/digitalsocial-mobile-worldwide-2015

[4] J. Holcomb, J. Gottfried, and A. Mitchell. News Use Across Social Media Platforms, accessed on Nov. 2013. [Online]. Available: http://www.journalism.org/2013/11/14/news-use-across-social-mediaplatforms/

[5] D. Chu, A. Kansal, J. Liu, and F. Zhao, "Mobile apps: It's time to move up to CondOS," in Proc. 13th USENIX Conf. Hot Topics Oper. Syst., 2011, p. 16.

[6] B. D. Higgins et al., "Informed mobile prefetching," in Proc. 10th Int. Conf. Mobile Syst., Appl., Services, 2012, pp. 155–168.

[7] Y. Wang, X. Liu, D. Chu, and Y. Liu, "EarlyBird: Mobile prefetching of social network feeds via content preference mining and usage pattern analysis," in Proc. 16th ACM Int. Symp. Mobile Ad Hoc Netw. Comput., 2015, pp. 67–76.

[8] D. Rout, K. Bontcheva, D. Preo¸tiuc-Pietro, and T. Cohn, "Where's@ wally?: A classification approach to geolocating users based on their social ties," in Proc. 24th ACM Conf. Hypertext Social Media, 2013, pp. 11–20.

[9] K. Makice, Twitter API: Up and Running: Learn How to Build Applications with the Twitter API, Sebastopol, CA, USA: O'Reilly Media, 2009.

[10] R. McCreadie, I. Soboroff, J. Lin, C. Macdonald, I. Ounis, and D. McCullough, "On building a reusable Twitter corpus," in Proc. 35th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr., 2012, pp. 1113–1114.

[11] M. E. J. Newman, "Modularity and community structure in networks," Proc. Nat. Acad. Sci. USA, vol. 103, no. 23, pp. 8577–8582, 2006.

[12] P. H. Wright, "Interpreting research on gender differences in friendship: A case for moderation and a plea for caution," J. Social Pers. Relationships, vol. 5, no. 3, pp. 367–373, 1988.

[13] A. Ahmad and L.Dey, "A k-mean clustering algorithm for mixed numeric and categorical data," Data Knowl. Eng., vol. 63, no. 2, pp. 503–527, 2007.

[14] A. J. Nicholson and B. D. Noble, "Bread Crumbs: Forecasting mobile connectivity," in Proc. 14th ACM Int. Conf. Mobile Comput. Netw., 2008, pp. 46–57.

[15] L. Hong, R. Bekkerman, J. Adler, and B. D. Davison, "Learning to rank social update streams," in Proc. 35th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr., 2012, pp. 651–660.

[16] L. Bottou, "Stochastic gradient descent tricks," in Neural Networks: Tricks of the Trade. Berlin, Germany: Springer, 2012, pp. 421–436.

[17] P. Ferreira, M. McGregor, and A. Lampinen, "Caring for batteries: Maintaining infrastructures and mobile social contexts," in Proc. 17th Int. Conf. Human-Comput. Interact. Mobile Devices Services, 2015, pp. 383–392.

[18]D. Shamma, L. Kennedy, and E. Churchill, "Tweetgeist: Can the Twitter timeline reveal the structure of broadcast events," CSCW Horizons, pp.