Temporal-Spatial Mobile Application Usage Understanding and Popularity Prediction for Edge Caching

Ming Zeng, Tzu-Heng Lin, Min Chen, Huan Yan, Jiaxin Huang, Jing Wu, and Yong Li

Abstract

The explosive growth of smart devices and expansion of network services drives the flourish of mobile applications. Caching popular application services at the edge, BSs, or APs, which are closer to end users, could significantly improve the user experience and network capacity. To exploit this potential, it is critical to understand the traffic consumption and app usage pattern under BSs in the metropolitan area. Mobile big data collected from network interfaces facilitates the data-driven approach in characterizing these features. This article aims to design an edge caching strategy for app services based on the observed characteristics of BSs in terms of points of interest (POIs), logs, and traffic generated by various categories of apps. We first analyze the temporal characteristics of different categories of apps, and then further investigate the logs and traffic generated by app types under different BSs clustered by POIs. The top *N* popular app types in a given period of time under different BS clusters are predicted, which is helpful for network operators to know the traffic distribution of different app services over all BSs and design the edge caching scheme.

Introduction

Explosive growth of smart devices and cellular network users has been witnessed in the last decade. Cisco's white paper states that global mobile devices and connections reached up to 8 billion in 2016, and is forecast to be 11.6 billion by 2021 [1]. Meanwhile, much traffic is generated when users are interacting with mobile applications. Users have to go through the backhaul or core network and fetch the content stored in the remote app servers. This not only aggravates the burden of the core network, but also incurs long delay in accessing the content exhibited on those applications. It has been shown that the traffic loads of mobile applications distribute non-uniformly and vary over time over widespread base stations (BSs) [2]. This status quo increases the difficulty for mobile network operators to execute delicate network management strategies [3], such as access control, bandwidth allocation, and load balancing. However, the downlink peak transmission rate is required to reach 20 Gb/s, while latency should be reduced to 1 ms in fifth generation (5G) networks. Thus, caching app services at the edge of wireless networks, which has received considerable attention recently, is considered as a promising solution. Prefetching app content from remote servers through backhaul networks does reduce the transmission latency and ensures a smoother user experience when using apps.

In recent years, edge caching problem in cellular networks has aroused the inspirations of many researchers. Li *et al*. [4] formulated an analytical framework to minimize the average content provisioning cost by joint designing the storage allocation and content placement strategy. Zhang *et al*. [5] proposed to leverage cooperative caching and exploit caching diversity to explore delay-optimal caching scheme. Psomas *et al*. [6] introduced the use of content caching at relays in order to improve the performance of relay selection. Carlsson *et al.* [7] found that a caching scheme based on ephemeral popularity was inefficient as content that were added to the cache would not be requested again. Reference [7] developed an alternative edge caching policy based on a novel workload modeling approach to greatly reduce the inefficiency. However, existing works did not incorporate the traffic consumption pattern in realistic networks into the edge caching scheme design. In addition, as [8] showed, the traffic consumption pattern under a BS is strongly correlated to the location of the BS and points of interest (POIs) around. Inspired by this, we utilize POIs around BSs to predict app usage, and then design edge caching schemes for different BSs according to the different app usage.

Hence, the first and key step for caching scheme design is to understand the app usage under different BSs. Here, we aim to analyze the spatial-temporal features of app traffic along with characteristics of BSs including the number of users connected, traffic density, POIs in the coverage, and so on. Attributed to the convenience brought by using apps on smart devices, different kinds of apps are developed on both the iOS and Android platforms. Traffic generated by various apps accounts for a significant component of cellular traffic. Thus, understanding the service request in each BS from the app usage perspective is noteworthy. Nevertheless, it is also challenging to depict the characteristics of each cellular tower from the app usage perspective for the following reasons.

Digital Object Identifier: 10.1109/MWC.2018.1700330

Ming Zeng, Tzu-Heng Lin, Huan Yan, Jiaxin Huang, Jing Wu, and Yong Li are with Tsinghua University; Min Chen is with Huazhong University of Science and Technology.

First, it is complicated to filter data records generated by apps from the mobile big data collected from network interfaces since there are abundant logs generated by other data connection sessions rather than apps. Second, even as users access services through apps, some apps will invoke services of other apps, so the apps that offer the service and the apps by which the logs and traffic are generated may not be the same. Third, the number of BSs in a metropolitan area is tens of thousands, covering millions of people. To extract features that distinguish BSs and cluster a large number of BSs into meaningful groups is difficult. Fourth, the traffic consumption patterns of different kinds of apps are quite heterogeneous and vary with time and space. We do not have prior knowledge about the traffic consumption patterns of mobile apps. To overcome these challenges, we investigate how to understand cellular towers from the app usage perspective based on a credible cellular dataset collected by one of the three main mobile network operators in China.

Based on a large-scale anonymous flow record dataset with 9647 BSs and over 1,188,000 users in 6 days, the contributions of this article are threefold. First, we reveal the temporal characteristics of app usage under BSs that reflect most mobile users' regular life to some extent. Second, BSs are clustered into groups that characterize different functional regions in a city. Some interesting relationships are found between the BS clusters and category differentiated app usage. Last but not least, we propose a regression-based method to predict the top *N* popular app types in a given period of time under different BS clusters to facilitate operators to understand the traffic consumption and app usage over all BSs. The average hit rate reaches about 60 percent, which indicates high feasibility of the cache schemes in future wireless networks.

The rest of this article is organized as follows. In the next section, we describe the dataset that we utilize in this article and some necessary preprocessing steps. Following that, we analyze the temporal characteristics of app usage under all BSs. Characteristics of BSs are then further depicted, and we obtain some interesting observations. Following that, we aim to predict the top *N* popular app types in a given period of time under different BS clusters based on a regression model. Furthermore, we investigate the variation of average hit rate with the increase of *N*, which implies a tradeoff between cache size and hit rate when deploying edge cache in realistic cellular networks. Finally, we conclude the article in the final section.

Dataset and Preprocessing

Dataset Description

The dataset utilized in our work is a six-day anonymized flow record data collected by one of the main network operators in China. The time span is from April 21 to April 26, 2016. Recorded data of 9647 BSs deployed in one of the largest cities in China is collected using deep packet inspection (DPI) technology. Each entry consists of information about URL, device ID, start/end time of the flow, flow duration, BS ID, bytes received/sent, number of packets received/sent, domain, and user agent. The device ID is the encrypted international mobile subscriber identify (IMSI) that is the unique identification of a mobile device. Start/end time of the flow is formatted in GMT+8. BS ID denotes the BS the mobile device is accessing. Note that all of the apps we analyze use the HTTP protocol; thus, HTTPS does not affect our study.

Preprocessing

The dataset needs to be preprocessed because of the existence of abundant logs generated when users access network services through other portals instead of apps, such as web browsers. Meanwhile, there are also lots of logs generated by unknown apps. To reliably identify the apps that have generated the network service requests, we utilized a framework called SAMPLES [9]. This framework operates a supervised methodology on HTTP headers to automatically generate conjunctive rules for identification. It is shown that this system is able to identity 90 percent of apps with 99 percent accuracy on average. After filtering, the dataset utilized in this article contains 718,298,220 entries related to 305 apps with 1,188,266 users.

Temporal Characteristics of App Usage

It is impossible to distinguish BSs by analyzing single app usage. Nevertheless, by classifying apps into a couple of categories, BSs are supposed to have similar characteristics under which the category differentiated app usage shows similar distribution. Thus, the most popular top 305 apps are further classified into 19 categories manually, including games, video, news, and so on, as shown in Table 1.

Understanding the temporal characteristics of traffic consumption pattern has guiding significance for network bandwidth allocation and green communications. With the dataset, we are able to analyze various categories of app usage patterns under BSs in the time domain. We present the temporal characteristics of traffic consumed by different categories of app types under all BSs in Fig. 1. From the results, we obtain four basic observations as follows:

- The traffic consumed by various categories of apps changes in a one-day cycle except on weekends. This can be explained by the fact that most people have regular life on weekdays, while most of them change the way they access the Internet on the weekend. For example, on the weekend, some people stay in places where there are WiFi connections, so less traffic is consumed in the cellular networks.
- Traffic generated by fnancial apps is very little on weekends mainly because the stock market is closed, so people use financial apps less frequently.
- Traffic generated by most categories of apps is less on weekends. However, some types of apps such as life service, baby care, social, maps, video, estate, and fashion consume more traffic on weekends since people are prone to relax on weekends.
- The traffic consumed by financial apps and musical apps shows the bimodality. This observation shows that people like to listen to music on the commute home and to work. In addition, the traffic consumed by financial apps var-

It is impossible to distinguish BSs by analyzing single app usage. Nevertheless, by classifying apps into a couple of categories, BSs are supposed to have similar characteristics under which the category differentiated app usage shows similar distribution. Thus, the most popular top 305 apps are further classified into 19 categories manually.

Number	Category	Number	Examples	Number	Category	Number	Examples
1	Games	52	JellyBlast, HuanLeDouDiZhu	11	Medical	5	HaoDaiFu, Keep
2	Video	16	QQLive, Bilibili	12	Commute	19	DiDiTaxi, QiCheZhiJia
3	News	18	BaiduNews, JinRiTouTiao	13	Vehicle	4	QiCheHui, JiaKaoBaoDian
4	Social	27	Wechat, Weibo	14	Music	14	QingTingFM, BaiduMusic.
5	Online shopping	22	JingDong, Taobao	15	Maps	10	GaoDeMap, BaiduMap
6	Finance	20	ZiXuanGu, DaZhiHui	16	Reading	14	ZhangYue, iBook
7	Estate	4	AnJuKe, TuBoShu	17	Fashion	11	MeiTuXiuXiu, MoGuJie
8	Travel	11	XieChengLvXing, FeiChangZhun	18	Working	10	QQMail, 163mail
9	Life service	32	BaiduTieBa, DaZhongDianPing	19	Baby care	6	BeiBeiWang, YunYuGuanJia
10	Education	11	ZhiHu, YouDaoDict				

TABLE 1. App classification.

ies with the opening and closing of the stock market as the stock market in Mainland China opens from 9:30 a.m. to 11:30 a.m. and from 1:00 p.m. to 3:00 p.m.

Understanding Characteristics of BSs

The above analysis of traffic under all BSs helps us understand the app usage patterns of different categories in the time domain. We further investigate the app usage pattern under different BSs so that the network administrator can apply diverse resource allocation strategies on different BSs and recommend different categories of apps when mobile devices are in the coverage of different BSs.

Clustering BSs

Due to the large number of BSs, we need to divide BSs into groups to discover different categories' differentiated app usage patterns. We cluster BSs based on POIs located in the coverage of BSs. The coordinates (i.e., longitudes and latitudes) of 16 types of POIs are obtained through the application programming interface (API) offered by BaiduMap. The types of POIs include food, hotels, shopping, entertainment, gyms, schools, scenic spots, tourism development, banking, offices, companies, business districts, residential, life services, towns, and villages. We calculate the term frequency-inverse document frequency (TF-IDF) values [10] of POIs in the coverage of each BS as features used in the clustering algorithm, which is a statistical method usually used to evaluate the importance of a word in a document set. TF (i.e., term frequency) denotes the frequency at which a word occurs in a specified document. IDF (i.e., inverse document frequency) is inversely proportional to the number of documents that contain the word. Therefore, the TF-IDF value of a word expresses whether the word is distinctive for a document. Here, each POI type can be referred as a word, and each BS is regarded as a document. Then BSs are clustered into seven clusters including Suburb, Education, Attractions, Commercial, Urban Resident, Entertain, and Unknown using a *K*-means clustering algorithm [11]. Table 2 shows the number of BSs divided into different clusters, their corresponding percentage, and the average TF-IDF value of POIs in each cluster of BSs. Most BSs belong to functional regions, while 2.36 percent are located in regions with few living facilities. The TF-IDF values in pink show the most distinctive POI type in each BS cluster. For BS clusters "Entertain," "Education," "Scenic spot," "Commercial," "Urban resident," and "Suburb," the distinguishing POIs are "Shopping," "School," "Scenic spot," "Company," "Resident," and "Village," respectively. Although the average TF-IDF value of "Company" POIs is highest in BS Cluster "Unknown," we cannot confirm "Company" POIs are distinctive for cluster "Unkown," since the difference between the average TF-IDF value of "Company" POIs and the average TF-IDF values of other types of POIs is not very notable. Only some POI types are listed in Table 2 due to space limitations.

Temporal-Spatial App Type Distributions

The app usage under different clusters of BSs is investigated in this subsection. Figures 2a and 2b present the number of logs generated by different categories of apps and traffic consumed by different categories of apps under each cluster of BSs, respectively. Due to space limitations, we only label the names of the top three app types and their corresponding proportions. News apps generate the most logs and traffic in every cluster of BSs; thus, news apps can be regarded as the most popular app type in this city. Figure 2a shows that the log distribution of various app types in BS clusters "Suburb" and "Commercial" is similar, and BS clusters "Urban resident" and "Entertain" is similar, while BS cluster "Scenic spot" is different. Note that reading apps generate the third most logs in BS cluster "Education." Furthermore, from Fig. 2b, we can observe that the traffic distribution of various app types in BS clusters "Education," "Scenic spot," and "Entertain" is similar, and is also similar in BS clusters "Suburb" and "Commercial." Notably, estate apps consume the second most traffic in BS cluster "Urban resident".

The proportions of different categories of apps used are also different under different clusters of BSs within a given time duration. Figure 3 shows the distribution of the number of logs generated by different categories of app types under 6 BS clusters from 17:00 to 17:59. It is observed that commute apps generate the third most logs under BS clusters "Entertain," "Commercial," and "Scenic spot." The most distinctive POIs under BS clusters "Entertain," "Commercial," and "Scenic spot" are

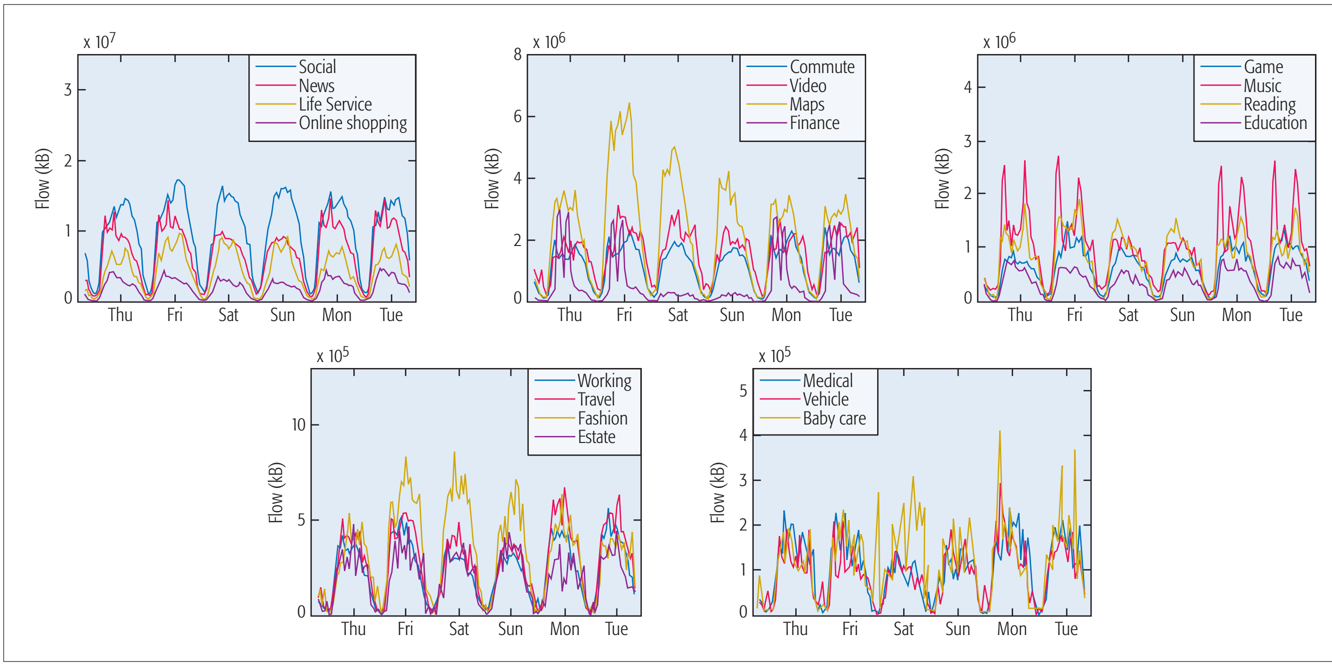

FIGURE 1. Temporal characteristics of traffic consumed by different categories of app types under all BSs.

shopping, company, and scenic spot, respectively. This observation agrees with our common knowledge that people are likely to leave shopping places, companies, and scenic spots at 17:00.

Temporal-Spatial Popular App Type Prediction under Different BS Clusters

Based on the temporal-spatial app type distributions under different kinds of BSs observed in the last section, operators can customize the application caching scheme for individual BSs. For example, news, estate, and reading apps can be cached for BSs that belong to cluster "Education," as these three app types generate the most logs and traffic in that BS cluster. However, predicting the app types that generate the most logs or consume the most traffic under a specific BS during a specific period of time (1 hour) can provide more significant guidelines to network operators. For example, operators should reserve resources for service requested by video apps and news apps containing streaming content. For service requested by game apps, the operators should guarantee low packet loss probability.

As observed earlier, the distribution of the number of logs generated by different categories of app types under each cluster of BSs is heterogeneous in a specific period of time. Meanwhile, the temporal characteristics of traffic consumed by different categories of app types differ on weekdays and weekends. We aim to predict the top *N* app types under BSs in each cluster during a specific period of time on April 26, 2016 by utilizing the flow record data on Aril 21, 22, and 25. We select the top five app types to predict because we find that the top five app types make up almost 50 percent of traffic under each BS cluster when preprocessing the dataset utilized in this article.

To predict the top five app types under BSs in each cluster during a specific period of time, we decompose the problem into two steps. First, we predict the traffic consumption of each app type under BSs in each cluster during a specific period of time. Then we rank the predicted traffic consumption and obtain the top five app types under BSs in each cluster (Fig. 4). Eighty percent of BSs' flow record data in each BS cluster is used as the training set, while 20 percent of BSs' flow record

BS cluster	No. of BSs	Percentage	Some POI types					
			Shopping	School	Scenic spot	Company	Resident	Village
Entertain	2037	21.12 %	0.260	0.00860	0.00856	0.119	0.0525	0.0125
Education	258	2.67 %	0.0733	0.274	0.0125	0.118	0.0587	0.0328
Scenic spot	217	2.25 %	0.0784	0.00741	0.312	0.103	0.0429	0.0298
Commercial	5372	55.69 %	0.101	0.00932	0.00563	0.304	0.127	0.000578
Urban resident	754	7.82 %	0.0945	0.00919	0.00404	0.0677	0.559	0.00557
Suburb	781	8.09 %	0.0481	0.00743	0.00731	0.223	0.0253	0.454
Unknown	228	2.36 %	0.00647	0.00791	0.00103	0.0152	0.00196	0.00926

TABLE 2. Average TF-IDF values of POIs in each BS cluster.

We apply a linear regression model to establish the relationship between the traffic consumed by different categories of app types under a specific BS in each time period and the traffic consumed by different categories of app types under the same BS in the same time period in historical weekdays.

data is used as the testing set to validate the performance of our prediction approach.

Since the traffic consumed by different categories of apps changes over the period of one day, the amount of traffic consumed by various categories of apps under BSs in the same cluster at the same time in historical weekdays or weekend is selected as the feature. Then we apply a linear regression model [12] to establish the relationship between the traffic consumed by different categories of app types under a specific BS in each time period and the traffic consumed by different categories of app types under the same BS in the same time period on historical weekdays. The weight parameters are learned for each BS cluster in each time period utilizing the data of BSs in the training set. The linear regression model trained is used to predict the traffic consumed by different categories of app types under BSs in the testing set in each time period. The predicted traffic consumption

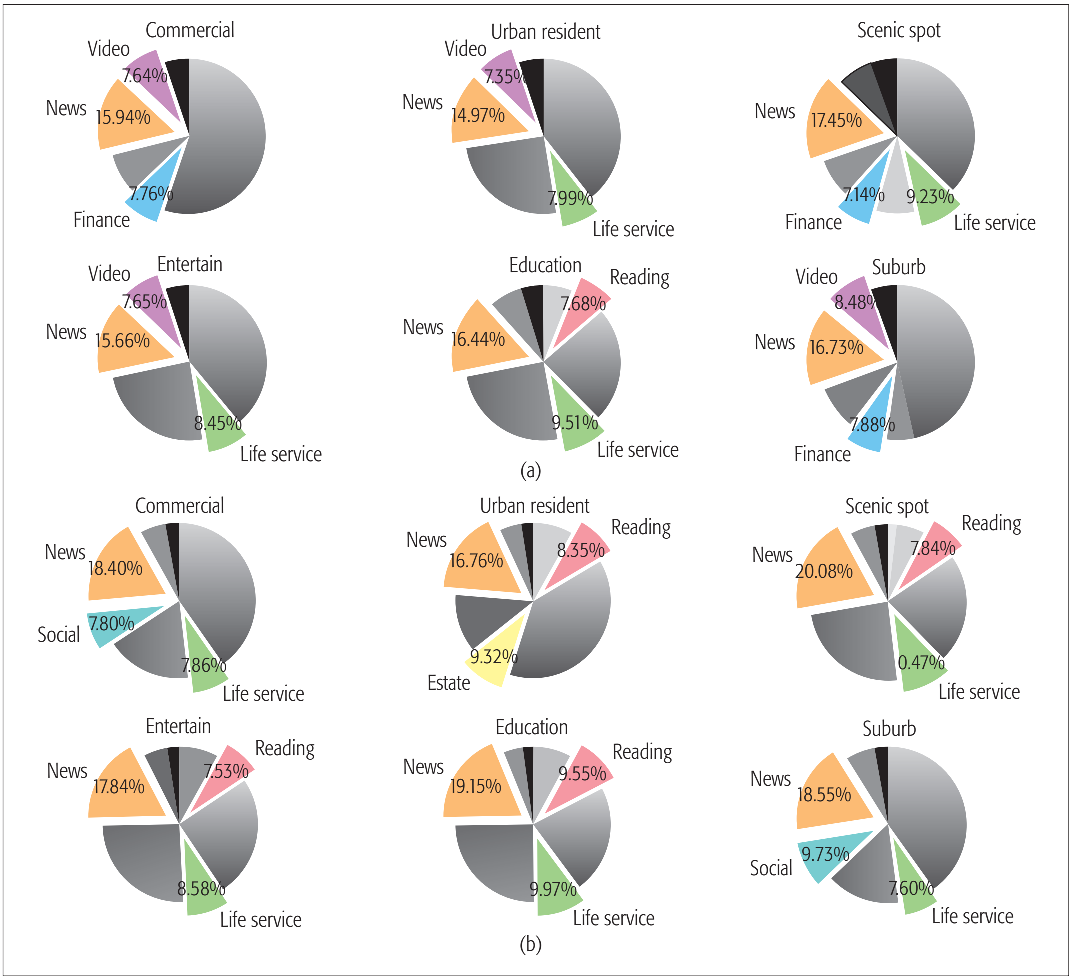

FIGURE 2. Distribution of the number of logs and traffic consumed by different categories of app types in each cluster: a) number of logs; b) traffic.

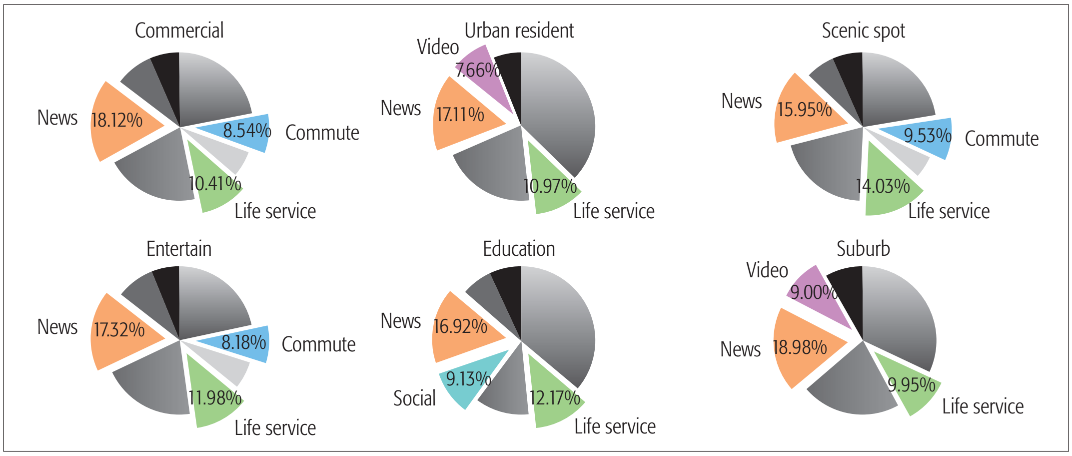

FIGURE 3. Distribution of the number of logs generated by different categories of app types under seven BS clusters from 17:00 to 17:59.

of different categories of app types is ranked to obtain the top five popular app types under each BS in each time period. The predicted top five app types are compared to the ground truth of BSs in the testing set. Hit rate [13] is used to evaluate the accuracy of our prediction method, which is defined as the ratio between the number of app types both contained in the prediction results and the ground truth and the number of app types predicted. Figure 5 shows the average hit rate of the top five app types predicted under each BS cluster in a specific time period. The median of hit rate in BS cluster "Entertain" is 61.4 percent, and the top five app types under some BSs can be predicted with 75 percent accuracy. The median hit rate is 59.4 percent in BS cluster "Education" and 70 percent for some BSs in this cluster. The median hit rate is 56.1 percent in BS cluster "Scenic spot and reaches up to 66.7 percent for some BSs. The median hit rate is 61.8 percent in BS cluster "Commercial," and reaches up to 76.7 percent. The median hit rates are 60 percent in both BS clusters "Urban resident" and "Suburb," and reach up to 73.3 an d 76.9 percent, respectively. The median hit rate is higher in a BS cluster with more training BSs.

Then we investigate the relationship between N and the average hit rate in each BS cluster. Naturally, when N increases, the average hit rate in each BS cluster increases. However, the increase of N will incur extra cost of cache. Thus, when deploying edge cache in realistic cellular networks, the trade-off between cache size and hit rate should be considered carefully. Meanwhile, for different BS clusters, to achieve the same hit rate, the cache size could be different. For example, for cluster "Scenic spot," to reach a hit rate around 60 percent, N should be at least 6, while for cluster "Commercial," it is enough to cache the top 5 app types. This implies that the cost of deploying edge cache in different BS clusters varies.

Conclusions and Future Work

In this article, we have understood the spatial-temporal application usage behaviors and predicted the top N popular applications over BSs based on a data-driven approach. The purpose of popular application prediction is to apply cache strategies at the edge of cellular networks to reduce transmission latency and improve user experience. Based on the finding that under different BS clusters the traffic distribution of different categories of app types is heterogeneous and varies over time, the application prediction algorithm is applied in each cluster of BSs. The average prediction hit rate reaches up to about 60 percent. Our study provides basic insights with a new method for network operators to understand the traffic consumption of applications and design cache schemes.

In our future work, it will be more practical to analyze the cache hit ratio at a finer granularity so that the specific popular content of apps will be cached. Since video accounts for a large proportion of traffic in app usage, in our ongoing work, we are extracting records related to video apps to predict the click rate of each video on the apps. A neural network is utilized to predict the popularity of videos under different BS clusters, and video caching schemes at BSs are designed to improve the average access rate of videos.

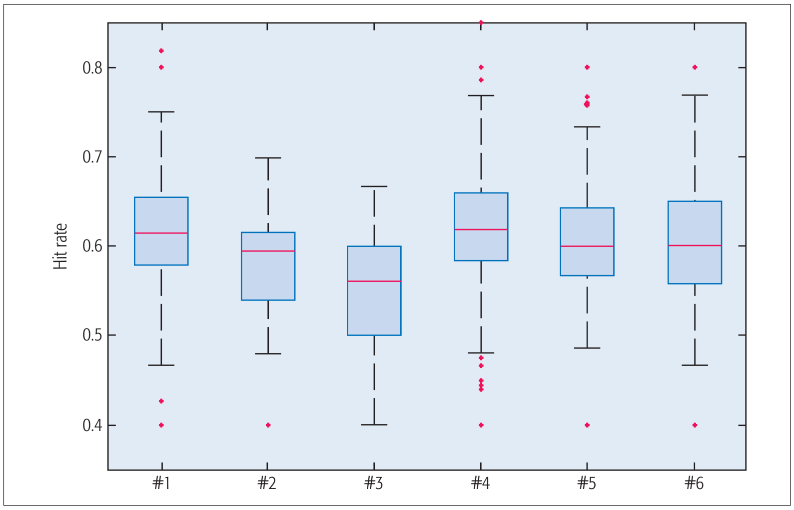

FIGURE 4. The average hit rate of the top five app types predicted under each BS cluster in each time period. #1 represents BS cluster "Entertain," #2 represents BS cluster "Education," #3 represents BS cluster "Scenic spot," #4 represents BS cluster "Commercial," #5 represents BS cluster "Urban resident," #6 represents BS cluster "Suburb."

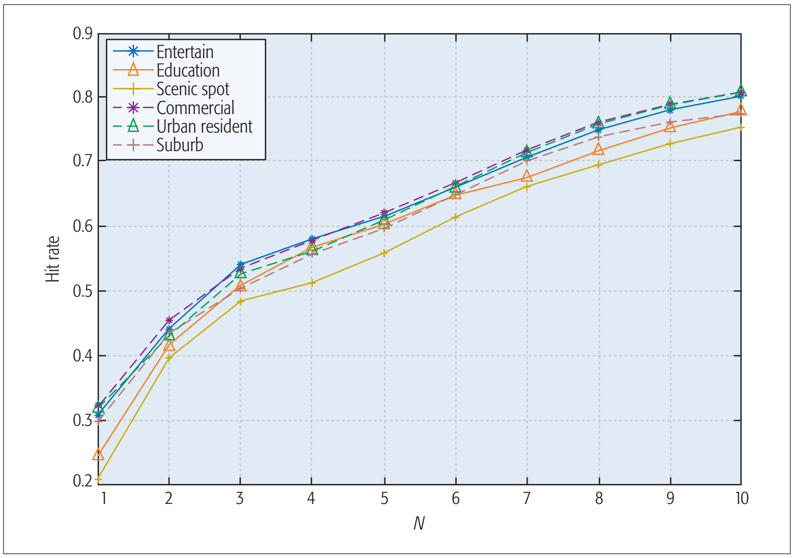

FIGURE 5. The average hit rate under each BS cluster varies with cache size.

Acknowledgments

This work was supported in part by the National Nature Science Foundation of China under 6171101425, 61621091, and 61673237, and the Research Fund of Tsinghua University-Tencent Joint Laboratory for Internet Innovation Technology.

References

[1] Cisco, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021," 2016, available: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobilewhite- paper-c11-520862.pdf.

[2] S. Zhou *et al.*, "On the Spatial Distribution of Base Stations and Its Relation to the Traffic Density in Cellular Networks," *IEEE Access*, vol. 3, 2015, pp. 998–1010.

[3] A. Imran and A. Zoha, "Challenges in 5G: How to Empower SON with Big Data for Enabling 5G," *IEEE Network*, vol. 28, no. 6, Nov./Dec. 2014, pp. 27–33.

[4] Q. Li et al., "Cooperative Edge Caching in Software-Defined Hyper-Cellular Networks," *IEEE JSAC*, vol. 35, no. 11, 2017, pp. 2596–2605.

[5] S. Zhang *et al.*, "Cooperative Edge Caching in User-Centric Clustered Mobile Networks," *IEEE Trans. Mobile Computing*, 2017.

In our future work, it be more practical to analyze the cache hit ratio in a finer granularity so that the specific popular content of apps will be cached. Since video accounts for a large proportion of traffic in app usage, in our on-going work, we are extracting records related to video apps to predict the click rate of each video on the apps.

[6] C. Psomas, G. Zheng, and I. Krikidis, "Cooperative Wireless Edge Caching with Relay Selection," *Proc. ICC 2017*, 2017, pp. 1–5.

[7] N. Carlsson and D. Eager, "Ephemeral Content Popularity at the Edge and Implications for On-Demand Caching," *IEEE Trans. Parrallel and Distributed Systems*, vol. 28, no. 6, 2017, pp. 1621–34.

[8] F. Xu *et al*., "Understanding Mobile Traffic Patterns of Large Scale Cellular Towers in Urban Environment," *IEEE Trans. Networking*, vol. 25, no. 2, 2015, pp. 225–38.

[9] H. Yao *et al*., "Samples: Self Adaptive Mining of Persistent Lexical Snippets for Classifying Mobile Application Traffic," *Proc. ACM MOBICOM*, 2015, pp. 439–51.

[10] G. Salton and C. Buckley, "Term-Weighting Approaches in Automatic Text Retrieval," *Info. Processing and Management*, vol. 24, no. 5, 1988, pp. 513–23.

[11] J. A. Hartigan and M. A. Wong, "A K-Means Clustering Algorithm," *Applied Statistics*, vol. 28, no. 1, 1979, pp. 100–08.

[12] A. Karlsson, "Introduction to Linear Regression Analysis," *J. Royal Statistical Society: Series A (Statistics in Society)*, vol. 170, no. 3, 2007, pp. 856–57.

[13] M. Deshpande and G. Karypis, *Item-Based Top- N Recommendation Algorithms*, ACM, 2004.

Biographies

Ming Zeng received her B.E. degree and Ph.D degree in communication and information systems from the University of Electronic Science and Technology of China, Chengdu, in 2009 and 2016, respectively. She is currently a postdoctoral researcher at Tsinghua University, Beijing, China. Her research interests include data-driven network optimization, intelligent transportation systems, fog computing, and spatial-temporal data mining.

Tzu-Heng Lin is currently an undergraduate student in the Department of Electronic Engineering, Tsinghua University. His current research interests include big data mining, user behavior analysis, and recommender systems.

Min Chen [M'2008, SM'2009] received his Ph.D. degree in communication and information systems from South China University of Technology, Guangzhou. From September 2009 to February 2012, he was an assistant professor with the School of Computer Science and Engineering, Seoul National University (SNU), Korea. He has performed postdoctoral research with the University of British Columbia, Vancouver, Canada, and served on the faculty of SNU in 2004. He is currently a professor with the School of Computer Science and Technology, Huazhong University of Science and Technology (HUST), Wuhan, China. He has published more than 180 papers, including 90 Science Citation Index (SCI) papers in the areas of the Internet of Things, mobile cloud, body area networks, healthcare big data, emotion-aware computing, robotics, and cyberphysical systems. He is the Chair of the IEEE Computer Society Special Technical Community on Big Data. He has received Best Paper Awards from the IEEE International Conference on Heterogeneous Networking for Quality, Reliability, Security, and Robustness (Qshine 2008) and IEEE ICC 2012.

Huan Yan received his B.S. degree in electronic information engineering from Nanchang University, China, in 2009, and his M.S. degree in communication engineering from Hangzhou Dianzi University, China, in 2012. He is currently working toward a Ph.D. degree in the Department of Electronic Engineering of Tsinghua University. His current research interests include network big data and software-defined networking.

Jiaxin Huang is currently an undergraduate student in the Department of Electronic Engineering, Tsinghua University. Her current research interests lie in data mining and user behavior modeling.

Jing Wu is an M.S. student in the Department of Electrical and Engineering of Tsinghua University. She received her B.S. degree in communication engineering from Central South University, Changsha, China, in 2016. Her current research interests include mobile big data mining and the Internet of Things.

Yong Li [M'2009] received his B.S. degree in electronic and information engineering from HUST in 2007, and his Ph.D. degree in electronic engineering from Tsinghua University in 2012. During July 2012 to August 2013, he worked as a visiting research associate at Telekom Innovation Laboratories (T-labs) and HK University of Science and Technology, respectively. During December 2013 to March 2014, he visited the University of Miami, Florida as a visiting scholar. He is currently an associate professor in the Department of Electronic Engineering of Tsinghua University. His research interests include mobile big data, device-to-device communication, software-defined networks, and fog computing. He has served as Technical Program Committee (TPC) Chair for several international workshops and conferences. He has also been a Guest-Editor for many journals, such as the *EURASIP Journal on Wireless Communications and Networking* and *IEEE Access*.