Mobile Surfing Pattern Analysis over Time and Location on a Big Access Record

Gengliang Zhu[1], Hao Sheng[1,*], Wenge Rong[1], Chao Li[1,2], Zhang Xiong[1,2]
[1]State Key Laboratory of Software Development Environment, School of Computer Science, Beihang University, Beijing 100191, P.R.China
[2]Shenzhen Key Laboratory of Data Vitalization, Research Institute in Shenzhen, Beihang University, Shenzhen 518057, P.R.China
zhugengliang@gmail.com, shenghao@buaa.edu.cn,w.rong@buaa.edu.cn, licc@buaa.edu.cn, xiongz@buaa.edu.cn
[*] Corresponding author

***Abstract*—The fast increasing WAP usage makes mobile-Internet a more and more important part of citizens' Internet life. As it is critical to mobile service market and city management, identifying surfing patterns of subgroups of mobile Internet users over time and based on locations will creates great value in Mobile Internet business and smart city management. In this paper, the author firstly examine the variation of surfing access, download and upload flow based on time and locations in which significantly and stable accessing peaks is found and verified. To detect the surfing behavior by individuals, the author traced the variation of surfing accessing over time and the variation of location of accessing and present them visually. The author finally obtain several surfing patterns that can precisely fit smaller subgroup of Mobile Internet users by hypothesis testing experiment.**

Keywords—mobile surfing; user pattern; time based; location based

I. Introduction

Understanding the temporal patterns of the mobile surfing behavior of citizens over time and based on location is critical for developing effective, efficient mobile Internet service. A well-known example is that the China Mobile Communication Corporation(CMCC) launched a special pricing package named *Day plus Night* which attracted a large quantity of mobile services users and proved to be a smart sale strategy. The profitable pricing package *Day plus Night* offers the user two kinds of data-service sub-package, the *Busy Time* sub-package and the *Spare Time* sub-package. In that pricing package, *Spare Time* stand for 23:00 pm to 7:00 am the next day and *Busy Time* stand for other time zones except *Spare Time*. The *Busy Time* sub-package contains a 600 Mega bytes surfing flow limitation. While it added up to 3 Giga bytes for the *Spare Time* sub-package. It was the surfing behaviors distribution statistics that help CMCC to design the pricing package that successfully matched the need of mobile services users.

The mobile Internet has quickly penetrated peoples life and has changed the world in a significant manner [1]. Therefore, mobile applications were getting more popular and more fruitful, which help people to obtain information, communicate with others and take amusement anytime and anywhere. Since most of the mobile applications are based on mobile Internet, understanding the mobile surfing pattern will deeply help to produce effective and appealing mobile applications. As for city managers and telecom operators, the mobile surfing patterns could be taken into account of making the strategy of load balancing, resource allocation and security monitoring. There are three main factors that critical to the mobile surfing pattern, time, location and content. The patterns only about time and location are called non-content based surfing pattern. Content based surfing pattern studies are mainly focusing on the content understanding, especially text understanding. Content based studies aim at discovery mobile Internet users' intent, emotion and the connection between the users. However, the study methods of content based studies are more general for surfing behaviors both on mobile phone and personal computer. Typical features of mobile surfing are time variation and location variation. For that non-content based properties show people's moving state and mobile surfing habit during the day.

In this study, the author analyzed the regularities over time and location on a very large dataset that recorded the features of surfing behaviors of the users in city Beijing. In the first part in this paper, the author analyzed the distribution of surfing access and surfing flow hourly. The result shows that there are stable surfing peaks and troughs of each day. This initial analysis suggested that there might be some stable patterns in mobile surfing behavior or surfing habits which can help us to get an insight of mobile Internet usage and to predict user behaviors. To get more details, the author analyses the download and upload flow usage over time and trace the mobile surfing behaviors of each user in the dataset. For simply description, the author divide the day into 7 time slots and define surfing energy as well as relative surfing energy to describe how active one user is in a certain time slot. By counting the surfing energy of each user, the author selected several significant patterns that describe the features of corresponding subgroups of the users.

This study further analysed the location features influence for the surfing behavior of citizens. Integrally, the author visualized the surfing access distribution on the map base on location and find out some stable accessing hot spots during the city. By examining the variance of the access records of each region in the gridding map, the author find that the surfing behavior tend to be more dispersive at noon and afternoon while more concentrated at morning and night.

Finally, by tracing the location variation of each user in the dataset, it shows that 85% of the users in the dataset are not recorded location change in one day. As the results of time based analysis and location based analysis were combined, the subgroups of users that have stable patterns over time and based on locations became smaller but more precise. Our findings may help to enable a more detailed analysis of the mobile Internet usage record data of massive users as time features and location features are considered.

II. Related works

Mobile network recently draws the interests from many researchers as it gradually plays an important role in daily life [2]. Halvey [3] identified the features of mobile web surfing. The analysis showed the surfing pattern of mobile Internet usage put up strong regularities that found in traditional web surfing, so it could be possible to identify the mobile Internet usage pattern by statistical analysis over surfing behavior. He also analyzed a dataset of mobile-Internet surfing information [4] covering 3.7M individuals for 4 weeks and predicted user navigation models using Markov Models [5] [6]. Yamakami used time slot counting in mobile click stream data to measure the regularities of surfing behavior. He obtained almost 80% accuracy of prediction [7]. Yamakamis study focused on the re-accessing behaviors of the mobile web users. However prediction accuracy with different time slot size showed similar value. The time slots were too big that the regularities found were not detailed enough. He extended this study in 2008,using clustering method to identify three subgroups in mobile Internet users which assume that there were three kind of user: (a)always on, (b)night-prime time and (c)others [1]. The author suggested the division of three subgroups would detailed information and more precise patterns could be found [8].

The above studies all used time based analysis methods to identify user patterns, which is a method that widely used in data mining research jobs. Beitzel's team examine query traffic on an hourly basis by matching it against list of queries that had been topically categorized by human editors [9]. The purpose of the analysis was to improve retrieval effectiveness and efficiency. Onnela also used similar time based analysis methods to measure the weighted network of mobile one-to-one communication [10] [11]. His study concerned more on the structure of the network and how it evolved. Jiang Focused on the calling patterns of individuals with a Weibull duration distribution and filled the users into three anomalous clusters and an ordinary one [12].

However, the location information recently draws the intention of researchers, especially in search behavior studies and recommendation in social network or mobile network. Teevan emphasize the importance of location and time in mobile local search behavior. The two features have high relevance to peoples calling and searching action [13] [14]. The author believe it would show similar relevance in mobile surfing patterns identifying. Ratti reviewed and introduced the potential of mobile device location determination technology and used the mapped cell phone usage at different time to represent the intensity of urban activities and their evolution [15]. As the structure and content of the dataset of each study are different, this paper is going to study the detailed patterns of mobile surfing considering time influence as well as location variation [16].

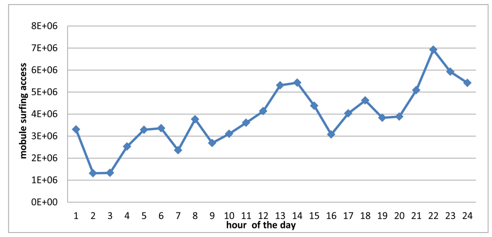


Fig. 1: Mobile surfing access flow based on different hour of the day

III. Analysis of Mobile Surfing Pattern

A. Dataset

In order to determine the mobile users surfing pattern in downtown city and divide users into several group that represent different user pattern in mobile surfing, we analyzed a enormous data set of mobile Internet records which provided by a major telecom operator in China. The data was collected from May 14th 2012 to May 20th 2012 and covered all users using the telecom operators service in Beijing. There are totally 97.6 million records each day and almost 3 million users are involved. In the data set, each record contains the user id(encrypted phone number), the geo-location of the mobile device, the surfing stating time and finishing time, the upload flow and download flow from mobile Internet in an access, the navigation web URL and content type. However, the web address or ip address of each record is sparse, which makes it more difficult to analyze the content of each surfing action. Besides,users's phone number were partly hidden for legality and security.

B. Time-Based Regularities in Surfing Behaviors

1) regularities on surfing access and flow: First, timing feature shows the most obvious variation in mobile surfing behaviors. If one looks at the surfing distribution count at different hour in a full day, he can find variation during 24 hours of the day and find clear peaks and troughs. According to the data set described above, it reaches the access peak at noon about 11:00 am to 1:00 pm and 9:00 pm to 11:00 pm, especially at night when the author counted over 13 million surfing access in two hours in Beijing, which is the peak surfing frequency of the day. Interestingly but unsurprisingly the statistical graphic of different day that counted from the data set almost have the same shape and the same variation tendency. So the author supposed that there is a stable pattern in a period of time such as half a year or 10 months at the macro level. To attain more detail of the time-based mobile surfing pattern, the author selected several features about mobile surfing that some of them differ from the previous work [5]. The selected features contains upload internet traffic, download internet traffic, total internet traffic, surfing time duration. To find out more regularities about surfing pattern based on time, the author split the day time into smaller granularity period. When the time slot was set to 1 hour, more details are showed from the statistics.

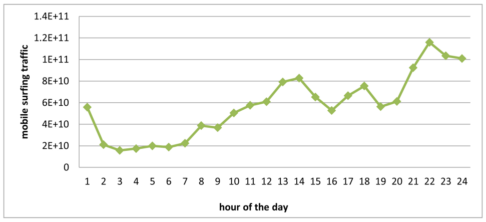


Fig. 2: Mobile surfing access count based on different hour of the day

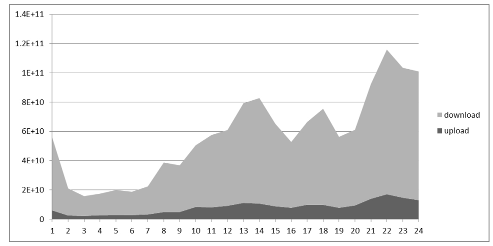


Fig. 3: The distribution of download and upload flow

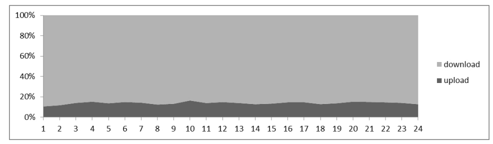


Fig. 4: The proportion of download and upload flow

As is shown in Figure 1 and Figure 2, when the access traffic is considered, the distribution curve is more smooth than that only access frequency was counted and clearly shows three surfing peaks of the day, which also can be figured out in the access frequency figure. The first peak appear at 13:00 to 14:00, thats the noon break time which easily explain why the mobile surfing behavior increases. The second peak appears at 17:40 to 18:50 and a decreasing trend until 20:00 and then the access increases again, reaching the third peak at 22:00. So, the author designed more experiments to find out why the second peak appears at 18:00 and the third peak appears at 22:00 and is there some connections between surfing pattern and the life pattern of the WAP users.

2) download and upload regularities: Before that, the author calculated the distribution of the total download flow and upload flow each hour of the day on average. As seen in Figure 3 and Figure 4, Figure 3 shows the average flow each hour of the day as the download flow and upload flow accumulated. While Figure 4 shows the percentage rate between download flow and upload flow. Its clearly shows that the upload flow almost keeps the same percentage rate (16%) of the total flow each hour of the day. This interesting result tells that mobile Internet users currently may use similar web surfing content, such as Web pages, map services or social networking applications. To verify this, the author selected 2000 records of the data that with surfing content and count the download and upload flow percentage rate. As seen in Table I, when the user was taking a media accessing action, the upload flow only take 7% to the totality and it would take an important influence to the total percentage rate for video or music consumes most flow than other mobile Internet service. As for mobile Internet games, when the user was playing online-games using mobile Internet, the user consumed more upload flow, which takes 34% of the total flow. Also, the map service need 11% upload flow. The Web page access, social networking applications and other items(the author can not identify which kind of service the URLs belong) in all take a percentage of nearly 16% upload flow. The result shows that when the users was using mobile Internet, they were less tend to take media access action or play Mobile online games at any time zone of the day, which caused the download-upload flow percentage rate keeps stable at 16%. Since that, we can infer that the distribution of download flow and upload flow is not a significant feature to the three peaks of mobile surfing of the day.

3) Surfing behavior of single user: Since it can not find reasonable regularities from holistic statistics analysis, the author further analyzed the time-based surfing behavior of each single mobile mobile Internet users. As Yamakami [1] divided the mobile Internet users into three subgroups, the author also wanted to do a classification to several groups that the users in each group have similar surfing pattern. However, the classification is described in next section when the location features are added. According to the regularities that found in time-based mobile surfing flow distribution, the author divided the day into 7 time slots, as seen in II. The three surfing peaks respectively appear at time slot 3,5 and 7. Then the author calculated the surfing behavior of each user in each time-slot.

To determine the activity level and information receiving size of a single user at a certain time slot, the author define surfing energy Se. As can be seen from Equation 1, U_i is the certain user, ST is the surfing duration of once access and F is the flow that the access consumed. The one plus at each access in Equation 1 is the Laplace Smoothing Factor, helping to improve the performance of Se when St of F is zero, which appeared a lot of time in the raw data. To normalize Se, Equation 2 is given to concern on the surfing behavior based on time slot.

TABLE I: The download and upload proportion of mobile Internet service

WAP service	Avg Download(%)	Avg Upload
Media Access(video or music)	93.13%	6.87%
Mobile Online Game	65.53%	34.47%
Web Page	83.64%	16.36%
Sns Application	84.47%	15.53%
Other	83.79%	16.21%

TABLE II: The division of time-slot

number	Time slot	User Type
1	1:00-9:00	midnight WAP user
2	9:00-12:00	forenoon WAP user
3	12:00-14:00	noon WAP user
4	14:00-17:00	afternoon WAP user
5	17:00-19:00	dusk WAP user
6	19:00-21:00	early at night WAP user
7	21:00-1:00	late at night WAP user

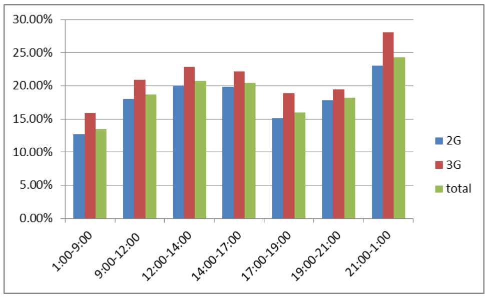


Fig. 5: The proportion of users that have used mobile Internet at each duration

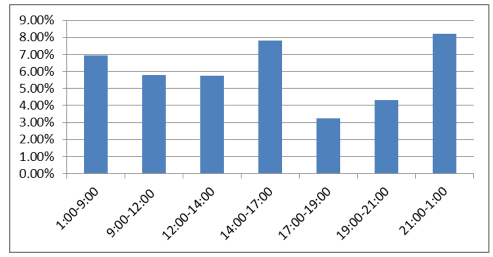


Fig. 6: The proportion of users that only access in one duration

$$Se_{u_i}(t) = \sum (ST_{u_i,t} \cdot F_{u_i,t} + 1) \tag{1}$$

$$RSe_{u_i}(t) = \frac{\sum (ST_{u_i,t} \cdot F_{u_i,t} + 1)}{\frac{1}{\sum_n \sum_t I(ST_{u_i,t} \cdot F_{u_i,t} > 0)} \sum_n \sum_t (ST_{u_i,t} \cdot F_{u_i,t})} \tag{2}$$

After the Se and RSe of each user were calculated, we can decide if a user is especially surfing active at a certain time slot and measure the time-based surfing pattern of each user by examining if the user access RSe is significantly larger at certain time slot. When doing the calculation, the author also took the RAT type into account. As seen in Figure 5, the result was convert to percentage rate. The sum of the rates is higher than 100% because there is overlap in the significant time slot of each user. As Figure 4 shows over 23% of the users in the record surf the Internet by mobile Internet late at night(21:00-1:00) which is the third access peak duration, and thats the largest rate. The second largest group is the noon mobile Internet user group as well as the duration of the first peak of surfing access. As for the dusk mobile Internet user group(18%), it shows that it was a smaller group but caused the second access peak. Whats more, it can tell that the group of users that tend to surf the mobile Internet at the day(morning, noon and afternoon) are bigger than the user group that surf the WAP at dusk or early at night. The author believe it may have some connection to peoples energy or the desire of obtaining information, they are more active at the first half of the day, then get tired at afternoon, dusk and early at night and the energy recovered late at night that build the largest access group. The author also respectively counted the user group of 2G and 3G given by Figure 4. It can be seen that the 3G user subgroups percentage rates are all higher than those of 2G. That shows the 3G groups has bigger overlaps than the groups of 2G, which means the users with 3G mobile service is more active around the day and they are more probable to surf the WAP more times a day. In a continuity of time, the 2G groups and 3G groups keeps the similar variation trend as the total group. However, 2G groups have bigger variance than 3G groups. So it can be considered that the 2G access information contains more details of the surfing regularities.

In the last section, the author calculated the subgroups of users at each time slot that divided the day into 7 durations. As well as the download, upload flow statistics, the subgroups of each time slot can not clearly tell how the three peaks appeared. So the author examine the combination of each time slot under the circumstance that the combination of time slots match the common sense of citizen life pattern. Easy to think of the three peaks of surfing access and assume the first pattern that a part of the mobile user would using the mobile Internet while their spare time, including the noon break(12:00-1400), before or just after getting off work(17:00-19:00) and before their sleep(21:00-1:00). To verify the patterns, the author counted the percentage rate of the user group that was significantly active in all those three time slot. Other patterns such as use mobile Internet all around the day , just use mobile Internet at night and only use mobile Internet at work time are also counted.

Firstly, the author think about examining the users that only access in one duration, and it turned out that 42% of the users in the data only use mobile Internet in one of the above 7 durations the author defined. To get more details, the ratio of user that only use mobile Internet at one duration was given by Figure 6. Over 8.2% of the users only using mobile Internet late at night(21:00-1:00), 7.8% of the users only using mobile Internet at afternoon(14:00-17:00). Since there are almost 1 million users in the dataset, the group that only surf late at night contains over 80000 persons. Interestingly, the author found 6.9% of the uses only have their surfing records at duration 1, which means their mobile equipment only connected to mobile Internet at midnight or the vary morning.

Secondly, the author concern about the left 58% users in the dataset, who would use mobile Internet in at least two duration. The results show at Table III. As can be seen in Table 5. 37.4% of the users only surf at the day time and 19.0% of the users only surf at night, which means 43.7% of the users surf the mobile Internet both at day and night and 14.3% of the users just surf the mobile Internet at day time or at night but at least through two durations. After that, the author found nearly 5% of the users have surfing record (Flow rate is higher than a threshold) at every time slot and the author take this an appreciable mobile surfing pattern.

Thirdly, the author concern about the three time slot the three access peaks appears. According to the result in Table 3, just 1% of the users surf only at the three peak duration (noon, dusk and late at night). If surfing actions at other duration are allowed, under the condition that the user consumes most flow at the 3 duration that peaks appear, the result become 6.9%. For comparison purposes, the author calculated the combination of any other 3 time slots, keeping the condition unchanged, and found out the highest ratio of the users that consume their access flow at these three time slots was 4.8%. In the totally 35 three-duration combinations, the author used Chi-square statistic to reveal that the group that tend to surf the mobile Internet at noon, dusk and late at night is reliably higher (p¡.05). The analysis suggest that the three durations surfing access peaks appears are significant.The pattern based on these three durations is remarkable. In the next section, the author is going to integrate location-based analysis with time-based analysis result over an undifferentiated set.

C. Analysis with Location Factor Added

1) Mobile Surfing Location Distribution in Different Time of a Day: To deeply gain an insight into the mobile users surfing pattern in urban city, the author conducted the analysis of the locations of each surfing record and each mobile Internet user in the dataset to find out the location features contribution to the surfing behavior.

Firstly, the author examined the city of Beijing as a matrix on the map. When the granularity was shrunk into 128 times 128, any two main roads with the same direction would not appear in one region of the matrix. Taking the urban area into consideration, each geographic region was defined as 0.01 longitude unit time 0.01 latitude unit. Then the author calculated the mobile Internet access flow of all the regions at each hour of the day. To visualize the result, the author put the statistics on a map software. As Figure 7 shows, the colors from green to red indicate the level of surfing access flow. Green region is recorded less surfing access flow and the red region recorded more. As can be seen in Figure 7, not all the regions in the matrix was colored because there may be no base station in those area and the location of each surfing record is logged by the base station owned by the telecom operator. Figure 7 shows the access distribution of four time interval, 9:00, 12:00, 18:00 and 22:00. The latter three (7b, 7c, 7d) respectively correspond the three surfing access peaks that found in the last section. (From the statistics, the author found 47 hot spots in the matrix which are 31% more than other regions that colored on average.) As Figure 7 shows, at midnight and at morning the access flow is generally lower for all the regions. At noon and afternoon including the dusk, there are more hot spots that recorded pretty high access flow and other regions access flow also increase. The access flow increases to maximum value at night especially 22:00 to 23:00. So the author calculated the variance of the matrix of each hour, the normalized result is given by Figure 8. As Figure 8 shows, the variances at the middle of the day (12:00 to 15:00) are significantly higher. From Figure 7 and the variance statistics, it shows that users more concentrated at morning and the night, while more dispersive at the working time. That

TABLE III: The proportion of the users of each defined pattern

pattern description	count	proportion
Surfing only in one duration	416051	42.12%
Only surfing at the day time	369230	37.38%
Only surfing at night	187381	18.97%
Surf but not only at the three peak duration	78133	7.91%
Surfing only at the three peak duration	10174	1.03%
Surfing all the 7 duration	48302	4.89%

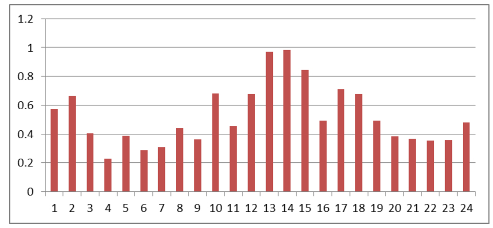


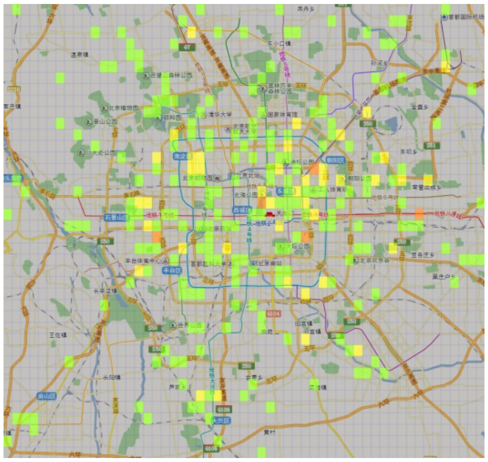
Fig. 8: Variance of surfing access flow of city matrix

means people tend to gather together at the day and disperse at night. So the first and the second surfing peaks appears when the users are at a concentrated state and the third peak appears when the users are at a dispersive state. Its easy for us to suggest that there is some working districts that people flock together at the day, and there is some living districts that people stay at night.

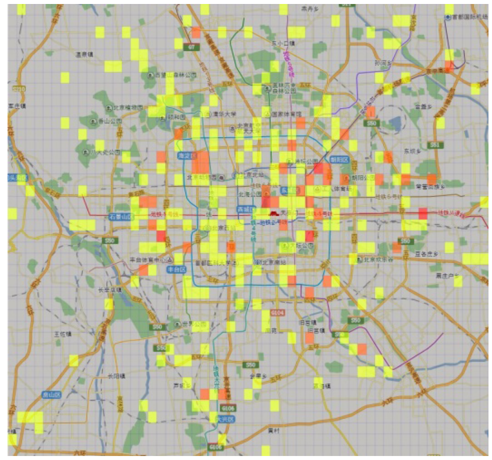
Secondly, the author traced the surfing location variation of each user in the dataset and tried to find out some regularities that parameterized by time features and location features. As 9(a) shows, 84.7% of the users in the data set were just recorded one location, 8.7% were recorded having two locations and 6.7% were recorded more than 2. That means most people do not tend to surf the mobile Internet at different at one day. In the calculation, the author took the two records as they are at different location only when their Manhattan Distance in the matrix is bigger than 3 for the purpose that avoid connecting two base stations at the same location. Since last section shows that 42% of the users in the dataset just surf mobile Internet at one duration and totally 56.4% of the users only surf mobile Internet at day time or at night, the author mainly calculated the subgroup that the surf but not only at the three peak duration which takes 7.91% of the entire group. The result is given by 9(b), which shows nearly half of the users in that subgroup have the same location data in their records,32% of the user have two locations at day time and night as well as 15% users own two more locations in their record. Converting to real percentage, 2.5% (over 25000) of the total users, have a surfing behavior pattern that tend to surf the mobile Internet at different place respectively at day time and at night. 1.5% users of the subgroup recorded more than 2 locations.

IV. Conclusion

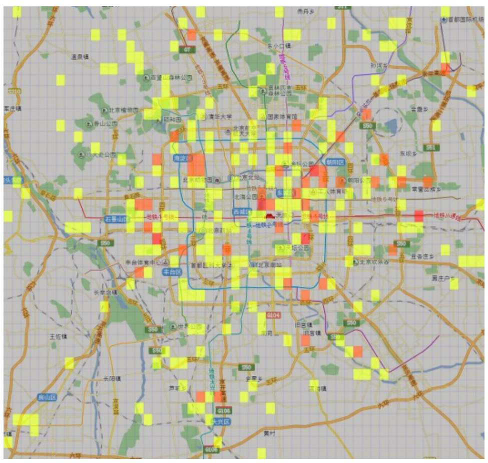
This study focuses on investigating the surfing patterns when people use mobile mobile Internet service to access to the Internet on a large dataset of mobile Internet surfing records on containing the information of surfing time and location. The purpose of the study is to understand the details of how users surfing behavior change according to time variation and location variation and providing helpful information for smart city management. In this study, the author firstly examine the variation of surfing traffic based on time and find three significantly accessing peaks respectively at noon, at dusk and late at night. To inquire the features that influence the distribution of the mobile Internet usage, the author counted the download and upload flow of surfing access base on time and found that the proportion of upload and download flow is stable over time despite continuing fluctuation. In addition, the author also found that people were less to access



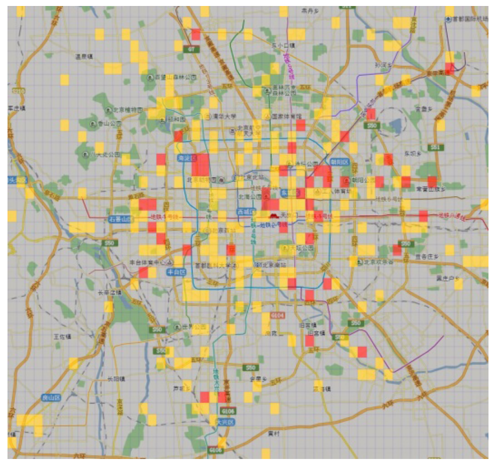
(a) 9:00 am-10:00 am



(b) 12:00 am-13:00 pm

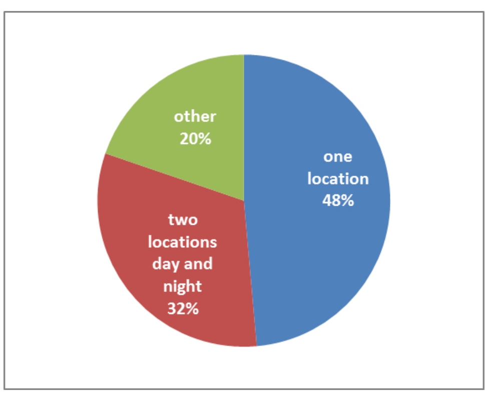


(c) 18:00 pm-19:00 pm

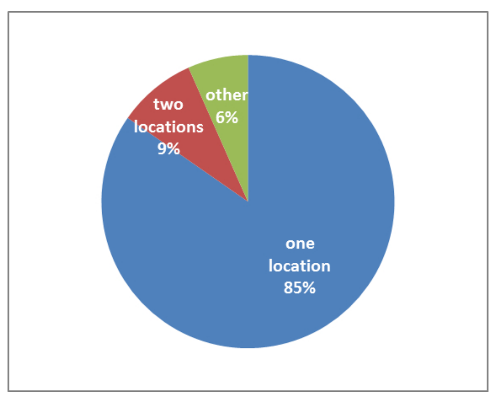


(d) 22:00 pm-23:00 pm

Fig. 7: Surfing frequency statistics on the gridding map. Deep color means more access record counted.



(a) location change of all users



(b) location change of users that caused the three surfing peaks

Fig. 9: Location change pattern distribution

media or game content when they using mobile Internet. However the conclusion seems weak because the subset for this experiment is much smaller. To watch the surfing behavior in a microscopic way, the author traced each user in the dataset of their surfing action based on time and find that 3G users tend to be more active than 2G users and quite a proportion of the users only have surfing records in one duration of the day among the seven durations the author defined. Besides, the author find several subgroups of the users that each of them has stable surfing pattern over time.

This study further investigates the changes in location when users are using mobile Internet by gridding the city map and count the access records of each region. The author visualize the variation of surfing access based on time and location and find that there is some hot spots in the city that basically have much higher surfing access rate than other regions. Additionally, the author find that people use mobile Internet more dispersedly at noon, afternoon and the dusk according to the variance result. Finally, when tracing the trajectory variation of each single user, the author find that most users were not recorded location changes in one day and only a little part of the users have a stable pattern that surfing at day time and at night respectively at different locations.

There are extensions that can be made to this work. Firstly, as the URL data in the dataset was sparse and its hard to identify the content type of each URL, the content based analysis is weak in this study. If the content can be precisely identified, it would help us to take a deep insight to the mobile surfing behavior. Secondly, based on the analysis in this study, some learning algorithm and predictive model can be used to predict the trend of the mobile surfing over time and based on location.

ACKNOWLEDGMENT

This work was partially supported by the Natural Science Foundation of China (NSFC) under grant ♯61332018, the National High Technology Research and Development Program of China (No.2013AA01A603), the National Science & Technology Pillar Program (2012BAH07B01), and the Open Fund of the State Key Laboratory of Software Development Environment under grant ♯SKLSDE-2015ZX-21.

REFERENCES

[1] T. Yamakami, "A user-perceived freshness clustering method to identify three subgroups in mobile internet users," in *Multimedia and Ubiquitous Engineering, 2008. MUE 2008. International Conference on*. IEEE, 2008, pp. 570–575.

[2] R. Baeza-Yates, B. Ribeiro-Neto *et al.*, *Modern information retrieval*. ACM press New York, 1999, vol. 463.

[3] M. Halvey, M. T. Keane, and B. Smyth, "Mobile web surfing is the same as web surfing," *Communications of the ACM*, vol. 49, no. 3, pp. 76–81, 2006.

[4] K. Yu, D. He, Y. Dou, and Z. Lei, "Cdma mobile internet user behavior analysis based on rp interface," in *Web Information Systems and Mining*. Springer, 2011, pp. 114–123.

[5] M. Halvey, M. T. Keane, and B. Smyth, "Time based patterns in mobile-internet surfing," in *Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM, 2006, pp. 31–34.

[6] M. J. Halvey and M. T. Keane, "Analysis of online video search and sharing," in *Proceedings of the eighteenth conference on Hypertext and hypermedia*. ACM, 2007, pp. 217–226.

[7] T. Yamakami, "Regularity analysis using time slot counting in the mobile clickstream," in *Database and Expert Systems Applications, 2006. DEXA'06. 17th International Workshop on*. IEEE, 2006, pp. 55–59.

[8] C.-H. Yun and M.-S. Chen, "Mining mobile sequential patterns in a mobile commerce environment," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 37, no. 2, pp. 278–295, 2007.

[9] S. M. Beitzel, E. C. Jensen, A. Chowdhury, D. Grossman, and O. Frieder, "Hourly analysis of a very large topically categorized web query log," in *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2004, pp. 321–328.

[10] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási, "Structure and tie strengths in mobile communication networks," *Proceedings of the National Academy of Sciences*, vol. 104, no. 18, pp. 7332–7336, 2007.

[11] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, M. A. De Menezes, K. Kaski, A.-L. Barabási, and J. Kertész, "Analysis of a large-scale weighted network of one-to-one human communication," *New Journal of Physics*, vol. 9, no. 6, p. 179, 2007.

[12] Z.-Q. Jiang, W.-J. Xie, M.-X. Li, B. Podobnik, W.-X. Zhou, and H. E. Stanley, "Calling patterns in human communication dynamics," *Proceedings of the National Academy of Sciences*, vol. 110, no. 5, pp. 1600–1605, 2013.

[13] J. Teevan, A. Karlson, S. Amini, A. Brush, and J. Krumm, "Understanding the importance of location, time, and people in mobile local search behavior," in *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*. ACM, 2011, pp. 77–80.

[14] V. S. Tseng and K. W. Lin, "Efficient mining and prediction of user behavior patterns in mobile web systems," *Information and software technology*, vol. 48, no. 6, pp. 357–369, 2006.

[15] C. Ratti, S. Williams, D. Frenchman, and R. Pulselli, "Mobile landscapes: using location data from cell phones for urban analysis," *ENVIRONMENT AND PLANNING B PLANNING AND DESIGN*, vol. 33, no. 5, p. 727, 2006.

[16] Z. Dou, R. Song, and J.-R. Wen, "A large-scale evaluation and analysis of personalized search strategies," in *Proceedings of the 16th international conference on World Wide Web*. ACM, 2007, pp. 581–590.