Predicting Mobile Users Traffic Behavior Using Machine Learning

Abdulrahman Alamoudi
Electrical and Computer Engineering
Georgia Institute Of Technology
Atlanta, USA
aalamoudi31@gatech.edu

Ali Payani
Electrical and Computer Engineering
Georgia Institute Of Technology
Atlanta, USA
apayani@gatech.edu

Faramarz Fekri
Electrical and Computer Engineering
Georgia Institute Of Technology
Atlanta, USA
faramarz.fekri@ece.gatech.edu

***Abstract*—With the size of wireless data growing, mobile users and network operators are continually struggling to reduce their data costs and network operation costs, respectively. A promising approach to alleviate the cost on network providers and mobile users is to predict mobile users traffic behavior. In this paper, we develop a machine learning platform to forecast mobile users' web traffic behaviors. We first analyze users detail traffic records and observe that the traffic patterns has strong dependency on the time of access, location, time intervals (or frequency) between the user's web access, and popularity of webpages. Then we develop a Recurrent Neural Network (RNN) with Long Short Term Memory (LSTM) to leverage the above dependencies as features to predict users' web browsing activities. Using real world data from mobile users, we show that our LSTM model can predict the type of the web domain a user will access with more than twice the accuracy of the Multiple Additive Regression Tree, when averaged over all users. Also, we compared our model's performance with a conventional neural network, Multilayer Perceptron, under the same complexity and constraints and we exhibit that our model accurately predict higher than MLP by 15 %. Then we apply our model to predict the type of the domain will be accessed within a time window and we show that our model can predict accurately by more than 70 %. Moreover, we extended our framework to predict the next web domain to be accessed in a mobile cell tower, where we aggregate traffic behavior from diverse group of mobile users for learning. As such, we resort to clustering methods to group the users of a tower to similar super user groups. Then we develop an LSTM model for each super user separately to predict the web domain traffic activities for the cell tower. Our experiments show that the clustering method can improve the prediction performance by more than 90% relative to no clustering method. Our proposed traffic forecasting framework can potentially be used to prefetch the contents when the network is cheap or when the user has access to WiFi.**

***Index Terms*—Algorithms, Wireless network, Prediction, Recurrent neural networks, WiFi**

I. Introduction

An exponential increase in wireless data has become a huge burden on cellular networks. This growth is expected to continue in coming years to reach 49.0 exabytes per month [1]. This increase will not only cause higher pressure on the transmission of the data but also impact the efficiency of the network.Wireless network operators are continually exploring different strategies for optimizing how their spectrum is used. Although many website types are dynamic, predicting and prefetching a mobile user contents on cheaper and better period of connectivity on time shifted fashion is a promising solution. There are significant opportunities for more effective fetching of content, if the content can be fetched ahead of time, sometimes by several hours. There are multiple scenarios in which such opportunities arise. In one scenario, consider a smartphone user who is connected to faster/cheaper WiFi networks (\$0.06/GB) for certain times of a day, and connected to slower/expensive Cellular networks (\$10/GB) for the other times. In another scenario, consider the same cellular network experiencing significant peak to trough ratios across time. For instance, In [2], through an analysis of cellular usage of 150000 users in a major International metropolitan city, the authors show that the peak to trough ratio is 9 : 1 (the peak and trough times are approximately 12 noon and 5 a.m. respectively). In both scenarios, it would be desirable to fetch content during the cheaper/faster/better connectivity for a future time when the user is likely to be connected to expensive/slower/worse networks. Likewise, predicting the next action in advance at the cell tower will help to mitigate the burden of transmission.

Prior contributions focused on caching part of the cloud services into mobile phone and dealt with correlating information of sets of websites and prefetch them once any website in those sets has been accessed [3]. However, this contribution deals with popular search queries and loaded to the phone while charging at a specific time frame. Moreover, [4] focused on web content based on analyzing user behavior but this kind of placement has an apparent problem that sequential aspects of the user behavior is not fully utilized. Although some studies focused on modeling sequential behavior of user activities using different types of Markov chain models. The rational is that highly sequential aspects of a user's activity is not fully utilized by Markov models, [5]–[7]. In [8]–[12] contributions are focused on predicting the most requested content that has a higher probability on different cell towers. These methods did not incorporate the location and access pattern diversity. In [9], [13], [14], the location of cell towers are implemented in order to provide collaboration between cell towers. To address these shortcomings, we propose a machine learning approach based on Deep Neural Networks (DNN) which can effectively combine the statistical dependencies and sequential behavior of users traffic activities in mobile device for an accurate prediction.

The contributions of the paper are summarized as follows:

- We perform an analysis on real world mobile data that could influence the prediction task. For example we show that the user behavior in cellular networks depends on time and location of its access.
- We formulate web content prediction problem for a mobile user via LSTM deep networks.
- We extend our LSTM framework for forcasting the web-type traffic activities in cell towers by using clustering method.

II. User Web Browsing Behavior Analysis

The essence of our work is to understand the underlying statistics of users behavior in a wireless network in order to model them effectively in our proposed prediction machinery. We start our study by describing the dataset that we used and continue with the analysis of the mobile Internet user's browsing behavior within our dataset.

A. The real world dataset

In this study, we use a dataset that contains web access log of the mobile Internet users for the period of one month. This data is collected by a wireless provider from over 1000 cell tower locations. Each access log in the dataset contains the anonymized subscriber ID, session time, cell tower ID, website and finally, traffic volume in kilobytes.

B. Detailed Analysis

People are creatures of habit and usually follow a predictable routine. For example, most people commute to work, have their lunch break, go back home at certain hours during a weekday and they may follow different pattern in weekends. As expected, our data show that for most mobile users the content depends on features such as time of access and the location. Moreover, some of the web domains are more likely than others. For example, news and sport related websites are more frequent than a typical blog page on the Internet. In order to model the user behavior accurately we would need to incorporate all these relevant features into our model. To examine how these features change the distribution of likely website visits, we have depicted some of these feature for random mobile Internet users. Fig. 1a shows the website IDs that were accessed by a mobile user versus time of the day. The access has a peek around 6 AM and the activity is low until 4 PM. More importantly, the distribution of the websites changes considerably based on the time of the day. Similarly, Fig. 1b shows how different locations can associate with different contents and as such considering the current location of the user would be beneficial in our prediction model. Further, this data strongly suggests that there are highly repetitive patterns in the web access of the mobile users, i.e., a relatively small number of unique websites account for most of the usage as shown in Fig. 1c. While there were thousands of distinct websites within the dataset, as Fig. 1c indicates, the 300 most popular websites account for approximately 80% of all the website accesses among all users.

TABLE I: Features Used In The Prediction Model

ID	Feature
1	previous accessed website $[1, N_{website}]$
2	location of the Access $[1, N_{location}]$
3	duration of the access (min)
4	time of the day (min)
5	is weekend $[0, 1]$
6	time since last Access for each website (min)
7	average time between consecutive accesses (min)
8	website popularity $[0, 1]$

C. Input Feature

In addition to the raw information obtained from the dataset, we extracted intuitive features and used them in our model as shown in Table I. For instance, in item 8 in the Table, we used a Bayesian estimator with a uniform prior (Laplace estimator) for estimating the popularity of the target websites. This estimate for each record was calculated using all the previous records from the same user. For multi-valued discrete feature (e.g, website ID), we used one-hot representation. In other words, for a feature which has N possible states, we use a vector of size N such that only one of the elements corresponding to one of N values is 1 and the other elements are set to zero.

III. User Behavior Prediction

We use deep neural networks as a learning model. This model is designed to predict a user's next webpage access given its activity history $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_t\}$ where $\mathbf{x}_i$ is the feature vector at time i. Those features are used in our model to estimate the conditional probability distribution of the websites $P(w_{t+1}|..., \mathbf{x}_{t-1}, \mathbf{x}_t)$ in the next time stamp $t+1$. We assume that the set $\{w_1, w_2, ..., w_N\}$ are the list of N possible website choices for a user. Then the website $w_j \in \{w_1, w_2, ..., w_N\}$ with the highest probability is predicted as the top choice for the next access.

A. User Behavior Prediction Structure

In this work we use Long Short Term memory model (LSTM) [15]. The LSTM method is widely used as a learning tool over time series. In our proposed model, instead of using the standard LSTM recurrent cells, we intend to use a simpler gated design to increase the efficiency and to reduce the computational cost for design on mobile devices. LSTM cells are a good choice for designing recurrent neural networks especially because of their ability to learn long-term dependencies. This ability is due to the gated design of the cells which allows for controlling the amount of information that is read (written)

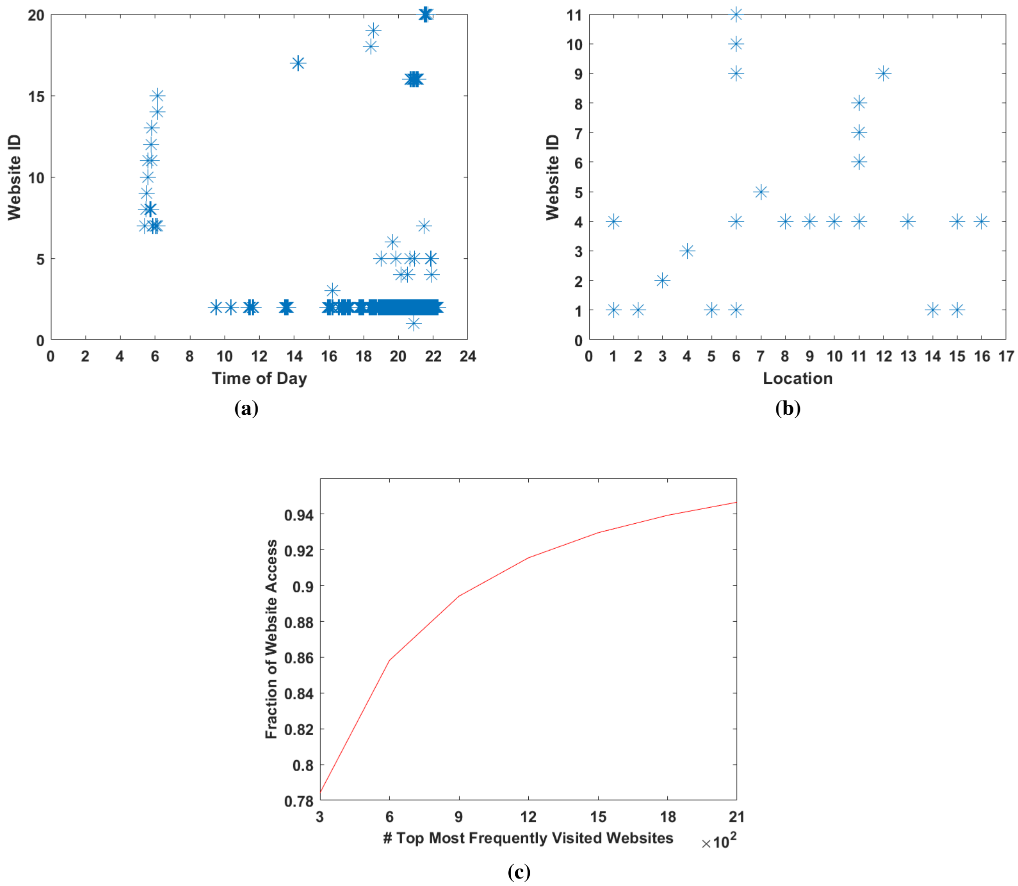

Fig. 1: Dataset Analysis. (a) Absolute Time Of Access; (b) Location Of Access; (c) Repeatability Of Websites Visits

from (to) cell memories. The cell output is also controlled by another gate which requires us to keep the state of the last cell output in addition to the cell state memory. In our proposed multi-layered recurrent design, we aggregate all the hidden states using a fully connected output layer, thereby eliminating the need for an extra state and the output gate, which results in a simplified gated recurrent network depicted in Fig. 2. The model is built based on the training set and validated using the testing set. The data are divided into training and testing data where about 20% is used for testing. Each input of the LSTM model at time t involves a feature vector $\mathbf{x}_t$. The recurrent backpropagation based on the gradient is used for training. In this framework, $\mathbf{z}_t$ encompasses the context $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_t\}$ and used to quantify the conditional distribution of the next website. In the following, we use lowercase bold letters to indicate vectors and uppercase bold letters for matrices. More formally, the LSTM formulation is shown below in (1) to (4):

$$\mathbf{h}_t^1 = \phi^h(\mathbf{W}^{h_1 h_1}\mathbf{h}_{t-1}^1 + \mathbf{W}^{x h_1}\mathbf{x}_t + \mathbf{b}^{h_1}) \quad (1)$$

$$\mathbf{h}_t^i = \phi^h(\mathbf{W}^{h_{i-1} h_i}\mathbf{h}_t^{i-1} + \mathbf{W}^{h_i h_i}\mathbf{h}_{t-1}^i + \mathbf{W}^{x h_i}\mathbf{x}_t + \mathbf{b}^{h_i}) \quad (2)$$

where $i \in \{2, ..., L\}$

$$P(w_{t+1}|..., \mathbf{x}_{t-1}, \mathbf{x}_t) = P(w_{t+1}|\mathbf{z}_t) \quad (3)$$

where

$$\mathbf{z}_t = \sum_{l=1}^{L} \mathbf{W}^{h_i z}\mathbf{h}_t^l + \mathbf{b}^z \quad (4)$$

Here $\mathbf{W_S}$ and $\mathbf{b_S}$ are weight matrices and bias vectors, respectively. Further, ϕ^h is a non-linear activation function that is applied element-wise. We fix the weights of layers $\{\mathbf{h}^1, \mathbf{h}^2, ..., \mathbf{h}^L\}$ across the temporal dimension but we use different weights for each layer. Based on this model, the probability distribution for each website at each time stamp is calculated by applying SoftMax activation in the output layer as

$$P(w_{t+1} = w_j|\mathbf{z}_t) = \frac{exp(z_t^{w_j})}{\sum_{i=1}^{N} exp(z_t^{w_i})} \quad (5)$$

where $j \in \{1, ..., N\}$

We choose the negative log of likelihood as the cost function in (6). Then, we use Adaptive Moment Estimation (ADAM) [16] with learning rate of 10^{-3} for minimization of the cost.

$$-log(P(w^N) = -\sum_{t=0}^{N-1} log(P(w_{t+1}|\mathbf{z}_t) \tag{6}$$

B. Time Interval User Behavior Prediction

We further extend our model to predict a user's next webpage access in the next time stamp within specified time interval given its activity history as feature vectors. This model estimates the conditional probability distribution of the websites including null activity in the next time stamp. We assume that the set $\{w_0, w_1, ..., w_N\}$ are the list of N possible website choices for a user where w_0 indicates no activity in the next time stamp. Then the most probable website is predicted. In this work we design our model to predict the next domain within an hour and half hour time window and in order to build this model we divided our data set into time frames and inserted null domain when no activity is depicted. Then, we use the model we trained for the individual users to learn the user traffic behavior in the next time interval.

IV. Prediction Model in Cell Tower

Next, our goal is to learn the webpage traffic behavior in the cell tower using which we can predict the next webpage access by users from that cell tower. Note that to develop a reliable prediction engine for the cell tower, it is not helpful to just use the model we trained for the individual users, as the solution would not scale. Moreover, building a single model for a cell tower using all mobile users can be problematic. This is explained in the following: First, note that the web access activities in the tower are the interleaved traffic activities of mobile users; potentially causing confusion in the machine learning model. For example, suppose there are two users connected to a cell tower. Let us assume user 1 would always access w_j after accessing w_i, but user 2 would always access w_i after w_j. Then from the cell tower point of view, sometimes w_j occurs after w_i but some other times the order is reversed. Therefore, the time series from the aggregate traffic pattern from all users of a cell tower could introduce significant challenges to the machine learning model. As such, we propose to group the users in a cell tower based on their similarity in webpage traffic activities into several clusters (or super users). In order to build these clusters we applied Hellinger distance as similarity measure between two users. Hellinger distance measures the closeness of two probability distributions. First, for each given user we estimated the probability distribution over all the websites in a specific cell tower. Let $\mathbf{U} = [\mathbf{u_1}, \cdots, \mathbf{u_M}]$ be the set of distributions for M mobile users in a cell tower. Let $\mathbf{u_i} = \left[p_i^1, \cdots, p_i^N\right]$ be i^{th} user distribution where p_i^l is the estimated probability for website l in a set of N websites. Then, We define a similarity index between user $\mathbf{u_i}$ and user $\mathbf{u_k}$ as

$$s_{ik} = H(\mathbf{u_i}, \mathbf{u_k}) = \frac{1}{\sqrt{2}}\sqrt{\sum_{j=1}^{N}(\sqrt{p_i^j} - \sqrt{q_k^j})^2} \tag{7}$$

After that, we used Hellinger distance as a cost function in K-mean clustering where we cluster users with similar probability distribution in a same cluster and we resort on squared euclidean distance to compute the center of those clusters. After that, we use 1000 repeated clustering with new initial center position for each cluster with objective of returning a solution with the lowest sum of point to center distances. Fig. 3 shows cluster assignments for a random cell tower where each point corresponds to a user in the same cell tower. In this model, we form different number of clusters and design a predictive LSTM model for each cluster separately. Again, the data are divided into training and testing data where testing is 20% of the data. We feed our model to various features in order to maximize the correct predicted websites. We use the same set of features as the predictive model for users as in Table I except the location feature due to its irrelevancy. Further, in order to enhance the prediction accuracy, we train our models based on each user data separately as shown in Fig. 4. For clarity purpose, we feed our model with all features of users in sequence without interleaving. In the test phase, we time order the activities from different users. We trained our model by minimizing the negative log-likelihood as in (6) using ADAM and learning rate of 10^{-3}.

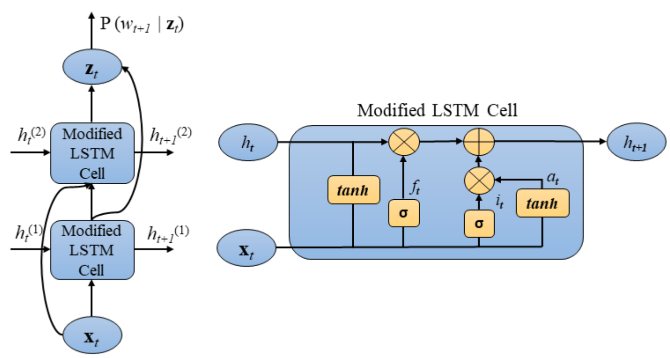

Fig. 2: LSTM Structure Used For User Model

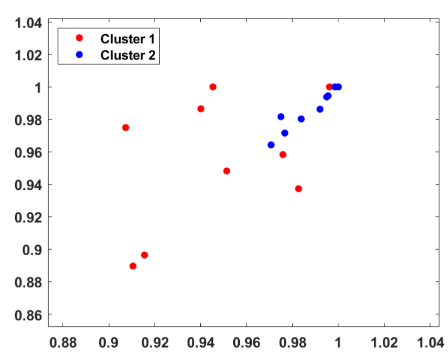

Fig. 3: Cluster Assignments for a Random Cell Tower

TABLE II: Features Selection In User And Cell Tower Models

Feature Missing[a]	Accuracy (%)	
	User	Cell Tower
none	64	21
previous accessed website	54	3
location of the access	62	N/A
duration of the access	61	16
time of the day	61	10
is weekend	63	15
time since last access for each website	64	35
average time between consecutive accesses	60	17
website popularity	62	16

[a]Feature that is dropped in the evaluation process.

TABLE III: Hyperparameters In User And Cell Tower Models

Parameter	User model	Cell Tower model
number of hidden layers	2	2
number of hidden units	100	1000
batch size	100	200
epochs	200	400
initial learning rate (α)	10^{-3}	10^{-3}
exponential decay rates $[\beta_1, \beta_2]$	$[0.9, 0.99]$	$[0.9, 0.99]$
denominator offset (ϵ)	10^{-8}	10^{-8}

V. Evaluation

A. User Behavior Prediction

We evaluated our purposed model on access logs of more than 1000 users. Our model was evaluated on each individual user. Test set was used to measure the accuracy of the prediction system. In order to achieve the best predictive result in modeling the data we applied feature selection to study the effect of different features on our model. Table II shows that when feeding our model with all features we accomplished the best average prediction accuracy among all users. It is noted that when we eliminated the previously accessed website, the accuracy of our predictive model was decreased significantly. This implies that the previous accessed website has high influence on our model. Next we studied the effect of tuning different parameters on our predictive model. We noted that increasing the size of our model leads to drop on the accuracy. Further,Table III shows structural design and hyper parameters selection of our model. We evaluated the performance of our model with unmodified LSTM model for top 1,top 3 and top 5 as shown in Fig. 6a. We observed that our modified LSTM accuracy was approximately as accurate as LSTM model. We also evaluated the results using Multiple Additive Regression Tree (MART) as a benchmark for both top 1 and top n most probable candidates. Fig. 6b depicts the average prediction accuracy of each model for Top 1, Top 3 and Top 5 websites. The results show superiority of our model where the accuracy was 64 % in comparison to 30% in the Multiple Additive Regression Tree. In another experiment, We compared our proposed model with a Multilayer Perceptron (MLP) to measure the sequential nature of our data. We noticed that our model outperform MLP as shown in Fig. 6c. Then we measured the prediction accuracy of our model within time interval. we evaluated prediction accuracy of our model within the next hour and half hour. Fig. 6d demonstrates impressive performance of the proposed model within the next hour and half hour.

B. Cell Tower Behavior Prediction

We evaluated our purposed model on test data at a cell tower. First, we examined the feature selection on our model.

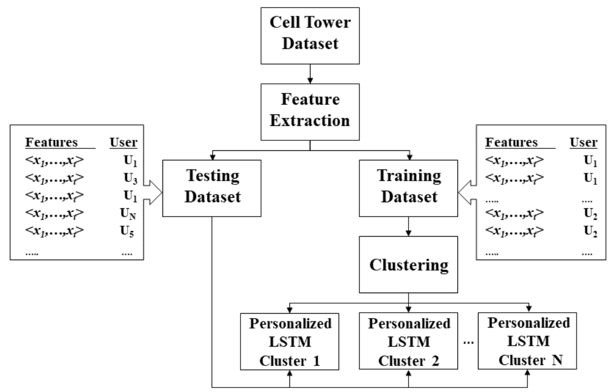

Fig. 4: Cell Tower Model Diagram

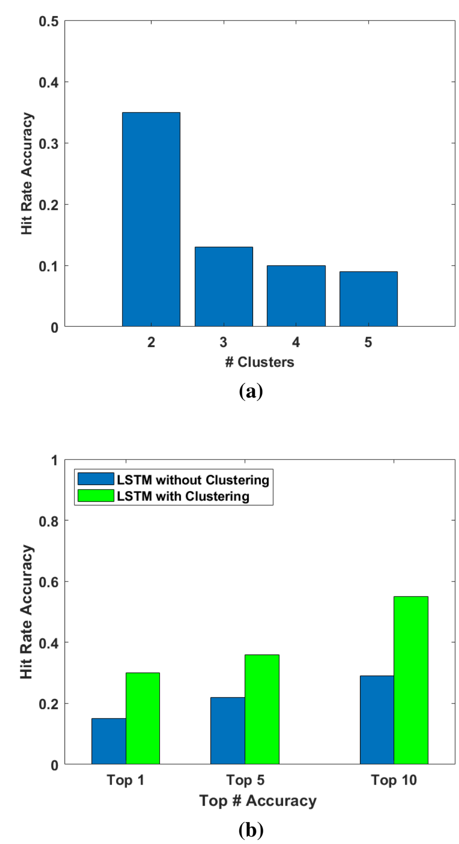

Fig. 5: Cell Tower Behavior Model Result. (a) LSTM With Clustering Performance For Different Number of Clusters; (b) LSTM With Clustering Performance In Comparison With LSTM Without Clustering

We evaluated the sensitivity to different sets of features as shown in Table II. It is noted that time of day and previously accessed website have significant influence on our model. We observed that ”time since last access for each website ” negatively impacts our model’s performance. Thus, we feed our models with all the features except ”time since last access for each website”. Then we evaluated our model on different parameters and we found that as we expand the size of the parameters, our model’s accuracy is decreased. Moreover, Table III shows the parameters we used in our model for the cell tower. We implemented our model on different number of clusters in order to obtain the best predictive result as shown in Fig. 5a. We noticed that the accuracy of our model when implemented on more than two clusters is diminished dramatically and that is because insufficiency and limitation of our dataset. Our model is evaluated on each cluster separately and then is reported over the average prediction accuracy. We evaluated our result on cell tower without clustering as a benchmark. We also reported the average accuracy of the prediction for Top 1 , Top 5 and Top 10 most probable websites for cell tower when two clusters are used. Fig. 5b shows that performance of clustering method outweighed the performance of the cell tower model without clustering.

VI. Conclusion

Understanding and utilizing user behavior in web browsing is a key element to mitigate the cost on users and network providers as well as alleviate the bottleneck of increasing traffic in wireless networks. In this paper we proposed a dynamic behavior based prediction of users traffic activities. The major contribution when compared to other works is to enhance prediction of the traffic in wireless networks by fully exploiting the sequential dependencies. Specifically, we introduced two prediction algorithms, one for each mobile user and another for each cell tower. We analyzed detail records of users and we explored their browsing statistics in our model. Our evaluation result suggests that, when averaged over all users, we can accurately predict the next webpage 64% of times. This accuracy increases to 83% if we output top 5 webpages. Further, our models showed that we can predict 68 % accurately within the next hour time window and 73 % within the next half hour time window. Those results leap to 87 % and 90 % respectively. In the cell tower scheme, the predictive model based on clustering showed an impressive performance improvement in comparison to the predictive model without clustering. Our result showed that using clustering we can predict the next webpage with accuracy as high as 35% of times.

References

[1] Cisco., “(2016) cisco visual networking index: Global mobile data traffic forecast update,” Tech. Rep., 2016-2021 white paper.

[2] “In 2017, how much low-band, mid band and high band spectrum,” https://www.fiercewireless.com/wireless/2017-how-much-low-mid-and-high-band-spectrum-do-verizon-at-t-t-mobile-sprint-and-dish-own, accessed: 2019-02-10.

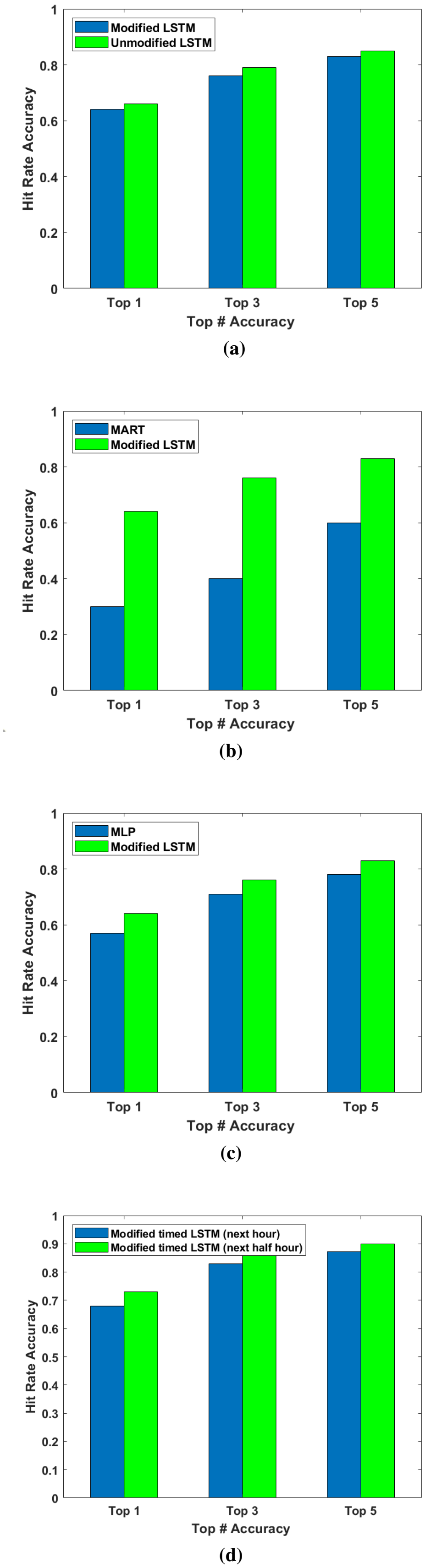

Fig. 6: User Behavior Model Result. (a) Modified LSTM Accuracy with Unmodified LSTM Accuracy; (b) Modified LSTM Performance In Comparison With MART (c) Modified LSTM Performance In Comparison With MLP (d) Modified LSTM Performance within The Next Hour and Next Half Hour

[3] E. Koukoumidis, D. Lymberopoulos, K. Strauss, J. Liu, and D. Burger, "Pocket cloudlets," *SIGARCH Comput. Archit. News*, vol. 39, no. 1, pp. 171–184, Mar. 2011. [Online]. Available: http://doi.acm.org/10.1145/1961295.1950387

[4] D. Lymberopoulos, O. Riva, K. Strauss, A. Mittal, and A. Ntoulas, "Pocketweb: Instant web browsing for mobile devices," *SIGARCH Comput. Archit. News*, vol. 40, no. 1, pp. 1–12, Mar. 2012. [Online]. Available: http://doi.acm.org/10.1145/2189750.2150978

[5] and, "A new markov model for web access prediction," *Computing in Science Engineering*, vol. 4, no. 6, pp. 34–39, Nov 2002.

[6] V. N. Padmanabhan and J. C. Mogul, "Using predictive prefetching to improve world wide web latency," *SIGCOMM Comput. Commun. Rev.*, vol. 26, no. 3, pp. 22–36, Jul. 1996. [Online]. Available: http://doi.acm.org/10.1145/235160.235164

[7] M. Deshpande and G. Karypis, "Selective markov models for predicting web page accesses," *ACM Trans. Internet Technol.*, vol. 4, no. 2, pp. 163–184, May 2004. [Online]. Available: http://doi.acm.org/10.1145/990301.990304

[8] Z. Chen, J. Lee, T. Q. S. Quek, and M. Kountouris, "Cooperative caching and transmission design in cluster-centric small cell networks," *IEEE Transactions on Wireless Communications*, vol. 16, no. 5, pp. 3401–3415, May 2017.

[9] E. Baştuğ, M. Bennis, M. Kountouris, and M. Debbah, "Cache-enabled small cell networks: modeling and tradeoffs," *EURASIP Journal on Wireless Communications and Networking*, vol. 2015, no. 1, p. 41, Feb 2015. [Online]. Available: https://doi.org/10.1186/s13638-015-0250-4

[10] B. Zhou, Y. Cui, and M. Tao, "Stochastic content-centric multicast scheduling for cache-enabled heterogeneous cellular networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 9, pp. 6284–6297, Sep. 2016.

[11] C. Yang, Y. Yao, Z. Chen, and B. Xia, "Analysis on cache-enabled wireless heterogeneous networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 1, pp. 131–145, Jan 2016.

[12] D. Liu and C. Yang, "Cache-enabled heterogeneous cellular networks: Comparison and tradeoffs," in *2016 IEEE International Conference on Communications (ICC)*, May 2016, pp. 1–6.

[13] S. Tamoor-ul-Hassan, M. Bennis, P. H. J. Nardelli, and M. Latva-Aho, "Modeling and analysis of content caching in wireless small cell networks," in *2015 International Symposium on Wireless Communication Systems (ISWCS)*, Aug 2015, pp. 765–769.

[14] B. B. Nagaraja and K. G. Nagananda, "Caching with unknown popularity profiles in small cell networks," in *2015 IEEE Global Communications Conference (GLOBECOM)*, Dec 2015, pp. 1–6.

[15] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997. [Online]. Available: https://doi.org/10.1162/neco.1997.9.8.1735

[16] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *CoRR*, vol. abs/1412.6980, 2014. [Online]. Available: http://arxiv.org/abs/1412.6980