Traffic Prediction Based Power Saving in Cellular Networks: A Machine Learning Method

[1]Sheng Zhang, [2,3,4]Shenglin Zhao, [4]Mingxuan Yuan, [4]Jia Zeng,
[1]Jianguo Yao, [2,3]Michael R. Lyu, [2,3]Irwin King *
[1]Shanghai Key Laboratory of Scalable Computing and Systems, Shanghai Jiaotong University, Shanghai, China
[2]Shenzhen Key Laboratory of Rich Media Big Data Analytics and Application, Shenzhen Research Institute, The Chinese University of Hong Kong, Shenzhen, China
[3]Department of Computer Science & Engineering, The Chinese University of Hong Kong, Hong Kong
[4]Huawei Noah's Ark Lab, Hong Kong
zhangsheng1730@sjtu.edu.cn,slzhao@cse.cuhk.edu.hk,{yuan.mingxuan,zeng.jia}@huawei.com
jianguo.yao@sjtu.edu.cn,{lyu,king}@cse.cuhk.edu.hk

ABSTRACT

In smart cities, green cellular networks play a crucial role to support wireless access for numerous devices anywhere and anytime with efficiency and sustainability. Because base stations (BSes) consume more than 70% of overall cellular network infrastructure energy, saving the power consumption of BSes is the key task to build a green cellular network. Except for low power design of the BS hardware and software, the traffic-driven BS sleeping operation is an economical way to improve existing cellular networks, which can reduce the BS power consumption at low traffic load. However, prior BS sleeping strategies establish on the static temporal characteristics of traffic load, which ignore the fact that network traffic is influenced by many factors such as time, human mobility, holiday, weather, etc. Hence, prior traffic estimation is coarse, and the BS sleeping strategies cannot apply to the changing network traffic. In this paper, we exploit a machine learning method to estimate the BS traffic and propose a BS sleeping strategy based on predicted traffic for power saving in the cellular network. We analyze network traffic in multi-views: temporal influence, spatial influence, and event influence. Then, we propose a multi-view ensemble learning model to predict network traffic load, which learns the traffic in multi-views and combine the results with ensemble. Furthermore, we formulate a BS sleeping strategy based on the predicted traffic load. Finally, we evaluate our traffic prediction algorithm on real cellular network data. The evaluation shows that our traffic prediction algorithm improves about 40% than state-of-the-art machine learning methods. Also, we evaluate the proposed BS sleeping strategy, which yields about 10% more energy savings and less device damage than the competitors in the simulated environment.

*This work is done when the second author interns in Huawei Noah's Ark Lab.

SIGSPATIAL'17, November 7–10, 2017, Los Angeles Area, CA, USA

ACM ISBN 978-1-4503-5490-5/17/11…$15.00
https://doi.org/10.1145/3139958.3140053

CCS CONCEPTS

• **Networks → Traffic engineering algorithms**; **Network management**; • **Information systems → Spatial-temporal systems**; **Decision support systems**;

KEYWORDS

Smart City, Green Cellular Network, Network Traffic Prediction, Multi-view Learning, Spatio-Temporal Data Analysis

1 INTRODUCTION

The cellular networks play a role of cornerstones in a smart city. The huge energy consumption in cellular networks encourages to build a green cellular network due to the economic and ecological concern [41]. The world's power consumption of cellular networks in 2016 is estimated up to $4\text{-}5*10^{11}$ Kwh [9, 33]. Hence, even 0.1% power saving in cellular networks is $4\text{-}5*10^{8}$ Kwh, which costs about 50 million US dollars [39]. In addition, the huge energy consumption will contribute to a surge in the emission of green house gases and exacerbate global warming. To address this matter of wallet and planet, the green cellular network is proposed to reduce energy cost. To build a green cellular network, the first issue is to enhance the critical subsystem that dominates the energy consumption, namely the base station (BS) subsystem. BSes consume more than 70% of overall infrastructure energy [5, 9, 29]. Hence, power saving for BSes is an important task for green cellular networks and the smart city [31, 41].

In addition to low power design of the BS hardware and software, the traffic-driven BS sleeping strategy is a recent research hot spot to improve existing cellular networks, which switches the BSes to low energy consumption mode, namely the "sleep" mode, according to the traffic patterns [29, 38]. The power consumption of a BS contains two parts: the functional component that consists of remote radio units (RRUs), base band units (BBUs), feeder and antenna array, and the auxiliary component that consists of the cooling system, power supply, and monitoring system [29]. When the traffic load of a region is low, we can switch some BSes in the region to the sleep mode. The original communication requests to the slept BS will be satisfied by the nearby BS, exerting few delays for users. Such a BS sleeping strategy is proposed and implemented in [29, 38]. The slept BS will save the power consumption and reduce the whole consumption up to 50% [29, 38].

However, previous work [5, 29, 38] designs the BS sleeping strategy based on the static temporal traffic characteristics. For instance, Cao et al. [5] assume the traffic patterns of a BS in each day are similar; Peng et al. [29] consider that the one-week traffic pattern in BSes keeps the same for each week. In fact, the temporal traffic pattern of a BS is not static. Other factors such as holiday and special conditions (e.g., rainy days) will affect the traffic pattern. Hence, we aim to seek a better traffic estimation method, which will motivate a more efficient power saving strategy.

In this paper, we propose a traffic prediction based power saving method in cellular networks, which adapts to the dynamic traffic changes. First, we conduct traffic data analysis in multi-views, showing how the temporal influence, spatial influence, and events affect the traffic. From the temporal perspective, we demonstrate the BS data characteristics in three aspects: autocorrelation, trends, and seasonality. From the spatial perspective, we analyze the nearby BSes' data correlations and show how the human mobility affects the BS traffic. From the event perspective, we illustrate how the events such as holiday affect the BS traffic. Then, we propose a multi-view ensemble learning model to predict the BS traffic. In particular, we (1) propose long and short term SARIMA [4] models to learn the temporal influence; (2) propose a spreading model to capture the spatial influence; (3) leverage the decision tree [3] model to learn the event influence; (4) propose a top-K regression tree model to learn the historical pattern for capturing the residual. Furthermore, we employ the random forest to ensemble the results. Finally, we formulate a BS sleeping strategy based on the predicted traffic load. Compared with prior work [5, 29], using the dynamic strategy based on traffic prediction can help us make better decisions on cellular infrastructure to reduce cost. We evaluate our traffic prediction algorithm on real cellular network data and apply the BS sleeping strategies in a simulated environment. The evaluation shows that our traffic prediction algorithm improves the performance by 40% compared with the state-of-the-art prediction methods and our BS sleeping strategy yields up to 10% more savings compared with the competitive baselines.

The contributions of this paper are summarized as follows.

(1) We propose a traffic prediction based BS sleeping system. A recent study [5] proposes a new practical way to sleep the BSes, which partially shuts down the carriers in BS and minimizes effects on the quality of service. However, the strategy in [5] is based on a static traffic temporal characteristic, which oversimplifies the real scenario. To this end, we propose a machine learning method to predict the BS traffic and establish strategies to shut down BS carriers according to the predicted traffic.
(2) To the best of our knowledge, this is the first work that systematically analyzes the influential factors for the BS traffic. Moreover, we propose a multi-view ensemble learning model for the traffic prediction task. Specifically, we establish appropriate models for each view, including temporal influence, spatial influence, event influence, and residual. Then, we leverage the random forest to ensemble the results. The evaluation on a real data set shows that our proposed method outperforms state-of-the-art traffic prediction models by 40% regarding accuracy.
(3) We build a BS sleeping strategy based on the predicted traffic for power saving in cellular networks. In particular, we take advantage of our accurate prediction result to choose the best point to sleep the energy-wasting BSes. Unlike the model proposed in [5], which requires the number of operations to be under a certain amount, we explicitly make the trade-off by choosing a suitable measurement of operation cost, so as to dynamically optimize our total cost under different BS structures. We evaluate our BS sleeping strategy in a simulated environment. The evaluation shows our BS sleeping strategy yields up to 10% more energy savings and less device damage.

2 RELATED WORK

In this section, we first review the literature of green cellular networks. Then, we demonstrate the progress of BS sleeping strategies. Finally, we show related studies for network traffic prediction.

Green cellular networks. To reduce the green house gas emissions and the operating cost, green cellular networks have become an important subject in the area of Information and Communications Technology (ICT) for smart city and green city. Insofar various approaches are proposed to reduce the energy consumption for green cellular networks, which can be broadly classified into four categories [38]. The first category aims to improve the hardware (e.g., power amplifier) energy efficiency [11, 14]. The second category works on the radio transmission process, physical or MAC layer, cognitive radio transmission, channel coding, and resource allocating for signaling [6, 17, 40]. The third category works in the direction of reducing the energy consumption by deploying small cells such as micro, pico, and femto cells to share the heavy traffic [1, 7, 16]. The fourth category covers the approach that selectively switches off some resources at low traffic load hours [2, 5, 23, 29]. The proposed approaches attempt to save energy by monitoring the traffic load on the network and then decide whether to switch off (i.e., sleep mode) or switch on (i.e., active mode) certain elements in the cellular network, which can reduce unnecessary energy consumption at low traffic load. In this paper, we focus on the fourth category—we propose a traffic prediction based power saving method in cellular networks.

BS sleeping strategies. Sleeping BSes is an economical operation to improve the existing cellular networks for power saving. There are two streams for BS sleeping strategies: Markov Decision Process (MDP)-based strategy [13, 18, 30] and traffic data-driven strategy [5, 19, 29, 34]. The MDP-based strategy assumes a distribution on the arrival, departure or handoff of a user and then yields a traffic load distribution on each BS. For instance, [18] assumes the Gamma distribution on the traffic load in BSes. Furthermore, a state vector is constructed to depict the BS traffic load situations. The actions space is a binary vector depicting the BS sleep mode. In addition, a transition probability expresses the state transitive possibility, and the cost or reward value in the MDP process is used to measure the energy consumption and quality of service. To solve the MDP problem, researchers propose a series of reinforcement learning method to control the BSes, such as Actor-Critic strategy [18] and Q-learning [13, 30]. Three reasons make the existing MDP-based strategies hard to use in practice: 1) traffic load in BSes is more complicated in real scenarios, beyond a presumed distribution; 2) the huge state space results in high complexity; 3) real cellular networks cannot afford frequent switch-off operations. The drawbacks in MDP-based strategies inspire the traffic data-driven strategy. This strategy attempts to analyze the traffic load history on BSes and predict the traffic load on BSes. Then, the BS controller can switch off some BSes according to the traffic load using some heuristic algorithms. The advantages lie in that: 1) the traffic is estimated based on real data; and 2) the complexity of the control strategy is low so that it is easy to implement in practice. However, the traffic prediction in prior work is coarse, and hence the sleeping strategy cannot adapt to the traffic variance.

Traffic prediction methods. Mobile traffic analysis attracts much attention with the increase of smart phones[25]. The work in [36] shows that the traffic demand exhibits spatial and temporal patterns, which help to predict the traffic load. Previous studies [28, 36] observe that the traffic at BSes presents a repetitive daily pattern over different weekdays, namely low demand at night and high demand during the day. This pattern is further verified in later work in hourly scale [15, 27]. Although the night and day hours form two distinct categories with the most different mobile usage behaviors, the variance among different daytime hours is also noticeable: traffic peaks always appear at day time but the peak time and the number of peaks vary at different areas [35, 36, 44]. In addition, researchers in [35, 36] also observe that weekdays and weekends yield distinct traffic demands: the traffic load on weekends are lower than that on the weekdays. In addition to the temporal pattern of traffic demand, the spatial patterns also exist. Researchers in [10, 26] observe that nearby regions take the similar average mobile demands during working days but a huge variability during weekends. The temporal and spatial regularity of aggregate mobile traffic can be used to predict the future network load [32]. However, the observed temporal and spatial characteristics are static, oversimplifying the traffic prediction task. For instance, although we observe that the traffic is low at night and high at daytime, there is a noticeable variance among different days. To capture the temporal and spatial dynamics, we turn to a machine learning method to predict the traffic in hourly scale. Machine learning methods have been used to some similar tasks such as city traffic prediction, business check-in prediction, and bike demand prediction [12, 20, 42, 43, 45] and turned out to be successful.

3 PRELIMINARIES

In this section, we first demonstrate the structure of a BS, which determines how to design a strategy for power saving. Then, we show the empirical study on BS traffic data, which motivates our prediction algorithm and the BS sleeping strategy.

3.1 BS Structure

A typical BS in the cellular network, also known as Node B, usually consists of a communication subsystem and a supporting subsystem. The communication subsystem includes antenna array, RRU, BBU along with the fiber-optic cable connecting BBU and RRU as illustrated in Figure 1. Each BS may contain several RRUs near the antennas to provide larger coverage and capacity. BBU is the main unit, in charge of all the communication functionalities. The supporting subsystem includes the cooling system, the environment monitor, and other auxiliary devices.

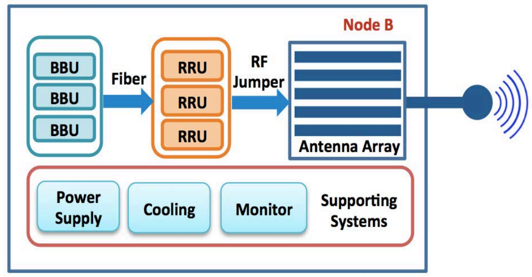


Figure 1: BS structure

The energy consumption model of each BS is composed of two categories [29]. One is from the traffic loads transmission, whose power consumption P_t is linear with the traffic load L, formulated as $P_t = P_\alpha \cdot L + P_\beta$, where P_α is a coefficient and P_β is a fixed consumption for a BS. The variance of P_t comes from the RRU and BBU components. For example, RRU has to support more active links if the traffic loads are heavy. The coefficient P_α largely depends on the transmission distance because one BS will cost more energy to serve the traffic loads from a longer distance. The other energy consumption P_s comes from the fixed energy consumption, owing to the supporting subsystem and some communication modules, especially the cooling system. The consumption P_s mainly depends on the working environment and is almost invariant to the traffic loads. The constant power consumption contains P_β and P_s, which contributes to the overall power consumption as high as 50 percent [29]. Therefore, our power saving strategy aims to save the energy at low traffic load.

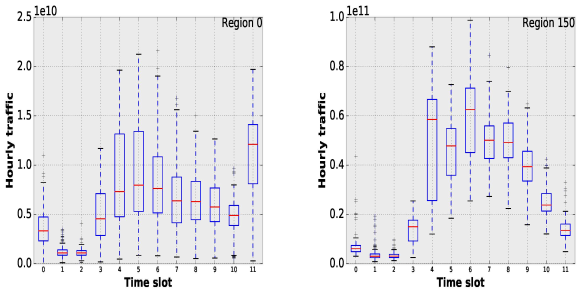
3.2 Empirical Data Analysis

Our empirical analysis is based on 3G BS traffic data collected from one of the main operators in a big city in China from Jan. 1, 2016 to Jun. 30, 2016. Following [29], we cluster the BSes into different grids, also referred as regions [1], according to their geographical relations to ensure that communication requests can interchange among different BSes in a region. Then, we analyze the BS traffic data for each BS region in three aspects: temporal influence, spatial influence, and event influence. The volume of the traffic is calculated in the byte.

3.2.1 Temporal Influence. The traffic data is a typical time series. We focus on three important features: autocorrelation, trend, and seasonality to analyze the traffic time series data.

First, we analyze the data autocorrelation. We cut one day into 12 intervals to observe the traffic data at each interval of two hours. Figure 2 shows two typical BS regions' distribution on each interval. We observe that two consecutive intervals' traffic data contain similar statistical features: median, mean, and deviation. For instance, in region 0 and 150, at 8th and 9th intervals, their statistical features

[1] We alternately use "grid" and "region" in this paper.



(a) Region 0 (b) Region 150

Figure 2: Autocorrelation demonstration of traffic data

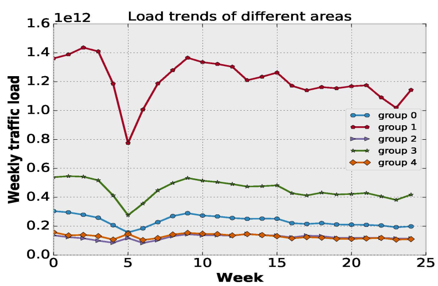


Figure 3: Weekly change trend of traffic data

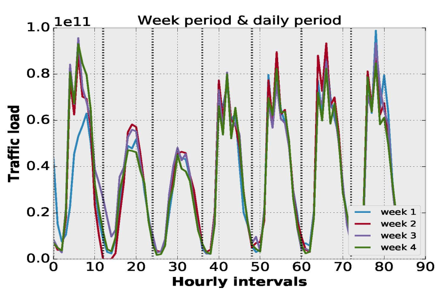
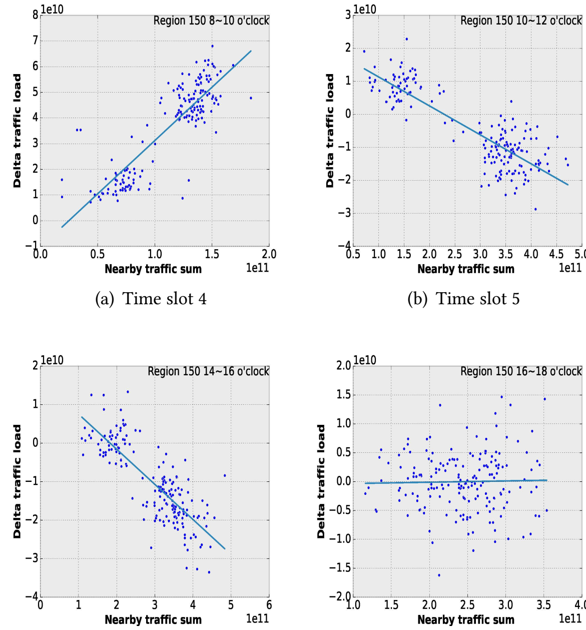


Figure 4: Weekly and daily pattern of traffic data

are quite similar. It enlightens us on the idea that we can infer the current interval's traffic from the last.

Second, we bag the BS regions into 5 groups according to their locations—group 1, 2, and 3 are in the urban area and group 4 and 5 are in the rural area. Then, we aggregate each week's traffic data to see how the traffic data change in a long term, as shown in Figure 3. We observe that there is a downward trend. This phenomenon may be caused by the fact that the 4G technology is gradually replacing the 3G network. We also observe that there is a sharp change during the third to seventh week, i.e., decrease in group 1, 2, 3 and increase in group 4, 5, which is attributed to the Chinese Spring Festival.

Last, we analyze the data seasonality from two time scales: day and week. As shown in Figure 4, the traffic data show periodicity on day and week. Most of regions take high traffic load in the day time but low traffic load at night. In addition, the regions in the urban area take high traffic load on weekdays but low traffic load on weekends.



(a) Time slot 4 (b) Time slot 5

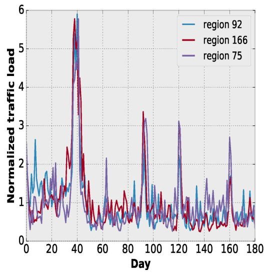
(c) Time slot 7 (d) Time slot 8

Figure 5: Spatial traffic pattern

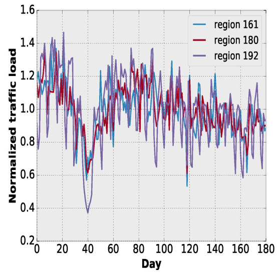
3.2.2 Spatial Influence. Spatial influence plays an important role in the traffic data analysis. Two facts inspire the spatial influence. First, different regions attract a different number of users. For instance, the urban central districts attract much more people than the rural areas. The spatial diversity inspires us to establish a model for each BS region correspondingly. Second, the human mobility results in the traffic transferring to nearby regions. We portrait data at four different time intervals, of region 150 from 8 am to 8 pm, shown in Figure 5. The horizontal axis of Figure 5 is the total load in the neighborhood at the previous time. It shows that at the working time, loads of region 150 are correlated with the last time nearby traffic load. When it comes to rest time, this relation is not obvious. Most other regions in the urban area also follow this pattern. This kind of traffic flow resulted from the human mobility inspires us to capture the spatial influence for BS traffic prediction.

3.2.3 Event Influence. The event influence causes the unusual traffic change. For instance, in Figure 3, the Chinese Spring Festival makes the traffic of BS regions in the urban area unusually decrease. Moreover, we observe events such as holiday, weather, concert, news will affect the traffic data. In the following, we take the holiday and weather as examples to analyze the event influence for traffic load.

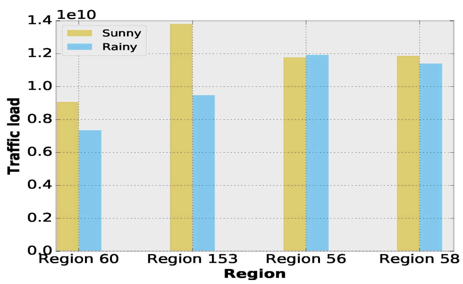
Figure 6(a) and 6(b) show how holidays affect the BS traffic data. Those peaks in Figure 6(a) and their corresponding troughs in Figure 6(b), respectively, appear in the Spring Festival, Qingming Festival, Labor Day, etc. We observe that, for regions located in the rural areas of the city, holiday events are main reasons behind the peak load. However, these events bring other regions traffic load to sink. Moreover, these holidays contribute ten percents overall volatility to the series data.



(a) Holidays in rural areas



(b) Holidays in urban areas



(c) Weather Impact

Figure 6: Event impact

Figure 6(c) gives details about how the weather affects the BS traffic data. For region 20 and region 153, rainy conditions cause nearly 20% drop of traffic load. However, this phenomenon is not significant in region 56 and 58. Given that different regions response to different kinds of events vary a lot and the huge impacts of events, we decide to extract event information and establish event models for each region.

4 TRAFFIC PREDICTION

In this section, we propose a multi-view ensemble framework to estimate the traffic data of each BS grid. First, we analyze each influential factor and establish a model to predict the traffic. Then, we leverage the random forest model to combine the influences and yield the final result.

4.1 Learning Temporal Influence

We leverage the SARIMA model [4] to learn the temporal influence: auto-correlation, trend, and seasonality. The SARIMA model is a classic time series forecasting model. We assume most regions' time series hold a similar time structure. Hence, we entitle every region with the same SARIMA formulation but set the formulation with different parameter values. Because the traffic data exhibit seasonality on different scales: day and week, we propose two SARIMA models to capture the temporal influence: long-term SARIMA and short-term SARIMA.

The long-term SARIMA deals with daily-scaled relationships, treating seasons on the level of week. We sample traffic data at specific time slot of each day and estimate the traffic at the same time slot in the future. For example, we sample the traffic data at the time slot from 8:00 am to 10:00 am of each day, and predict the traffic data at the time slot from 8:00 am to 10:00 am of one future day. The Eq. (1) is our long-term SARIMA formula. B is a time backshift operator. The seasonal period is set to seven, with AR and MA parameters both being one for simplicity. The integrated parameter is set to one to eliminate the trend for the stability concern.

$$\underbrace{(1-\alpha_1 B)}_{\text{AR}}\underbrace{(1-\alpha_2 B)}_{\text{Seasonal-AR}}\underbrace{(1-B)}_{\text{Diff}}\underbrace{(1-B^7)}_{\text{Seasonal-Diff}}\ y_t = \underbrace{(1+\theta_1 B)}_{\text{MA}}\underbrace{(1+\theta_2 B^7)}_{\text{Seasonal-MA}} e_t \quad (1)$$

The short-term SARIMA tackles hourly-scaled relationships, treating seasons on the level of day. Comparing with long-term SARIMA that samples data in the same time slot of each day, the short-term SARIMA uses all time slots' data of one day. For the short-term SARIMA, shown in Eq. (2), we set the parameter to $(3, 1, 3) * (1, 1, 1)_{12}$. The AR (3) and MA (3) contains one-quarter of daily data, which is enough for catching the hour structure of daily bandwidth load and not too heavy to compute.

$$\begin{aligned}&\underbrace{(1-\alpha_1 B-\alpha_2 B^2-\alpha_3 B^3)}_{\text{AR}}\underbrace{(1-\alpha_4 B^{12})}_{\text{Seasonal-AR}}\underbrace{(1-B)}_{\text{Diff}}\underbrace{(1-B^{12})}_{\text{Seasonal-Diff}}\ y_t \\ &\quad = \underbrace{(1+\theta_1 B+\theta_2 B^2+\theta_3 B^3)}_{\text{MA}}\underbrace{(1+\theta_4 B^{12})}_{\text{Seasonal-MA}} e_t\end{aligned} \quad (2)$$

The reason why we do not combine these two models lies in different characteristics between intervals. Night slots in sleeping hours have a little deviation, which largely depend on daily-scaled data covered by long-term SARIMA. Busy hours such as 8th and 9th slots are related to previous time slots data covered by short-term SARIMA. Finally, we ensemble these two models to capture the temporal influence in the multi-view learning model.

4.2 Learning Spatial Influence

In this section, we propose a spreading model to capture the spatial influence resulted from the human mobility. We assume that the human mobility carries the mobile load to flow between BS grids. Moreover, we assume the traffic data is affected by the traffic of neighboring BS grids at the last time interval. Hence, we take each neighboring grid's load as our features and employ Eq. (3) as our model formulation. P denotes the decay rate to transform nearby regions' load into the target region. Therefore, the traffic y_t^r of the target BS region r at time t can be estimated by the linear combination of decayed nearby BS regions' traffic and itself at last time slot,

$$y_t^r = \sum_{k \in \{r\} \cup \{nearby\}} P_t^k * y_{t-1}^k. \quad (3)$$

4.3 Learning Event Influence

The key to learning event influence is to discover the events that may affect the network traffic. In Section 3.2, we have shown that holiday and weather will affect the network traffic. In addition, we observe that the hot news and the entertainment event such as concert affect the network traffic as well. Because the traffic data collected from the BSes do not contain the event information, we cannot directly capture the event influence. To resolve this challenge, we exploit the external information from the Internet to

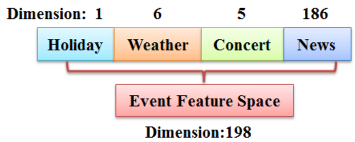


Figure 7: Demonstration of event feature

construct the event feature and learn the event influence through a decision tree regression model since it has turned out being successful to exploit external information to improve a machine learning task performance [37].

In particular, we establish the event feature from the external information, shown in Figure 7. For *holiday* event, we observe that the span of the holiday is an important factor to affect the traffic load. We construct a one-dimensional feature to represent the holiday event, which is an integer reflecting the span of the holiday. For *weather* event, we employ the crawled information to construct a six-dimensional feature, including the highest temperature, the lowest temperature, weather condition (e.g., rainy, sunny, etc.), air pressure, wind direction, and wind strength. For *concert* event, we employ the crawled information to establish a five-dimension feature, including the number of concert and detailed information (i.e., price and the distance to the BS grid) for two kinds of concerts: the most expensive and the nearest. For *news* event, we choose the SINA news as our source to collect the news information reported at each time slot. We use the tf-idf [22] to analyze words in the news documents and find six words with the highest weights. Then, we project each word to a 30-dim vector through a Word2Vec library [24] pre-trained on the whole SINA news data. Plus the news categories (six options), we obtain a 186-dimension feature the represent the new event. Based on the constructed event feature, we employ a decision tree [3] to estimate the traffic load.

4.4 Learning Historical Patterns

We propose a top-K regression tree method to learn the historical pattern to capture the residual not covered by the proposed models for temporal influence, spatial influence, and event influence. We predict each BS grid's traffic load at time interval t globally using its local context feature. First, we construct a BS region's local context feature X_c^t using its previous three time intervals' load data and nearby five BS regions' load data. Next, we calculate the Pearson correlation to measure the similarity of target BS grid at t with historical records of all BS regions before time t. Formally, for the target BS grid g_i^t at time t with local context feature X_{ci}^t, its similarity with BS grid g_j^p at time p ($p < t$) with local context feature X_{cj}^p can be measured as follows,

$$Sim(g_i^t, g_j^p) = \frac{Cov(X_{ci}^t, X_{cj}^p)}{\sigma(X_{ci}^t)\sigma(X_{cj}^p)}, \tag{4}$$

where $Cov(\cdot)$ and $\sigma(\cdot)$ denote the covariance and standard deviation. Then, we select the top K historical records of g_j^p and then concatenate their context features X_{cj}^p and similarity value $Sim(g_i^t, g_j^p)$ with original feature X_{ci}^t to construct a new feature. Finally, we employ a regression tree algorithm [3] to train and predict the BS grid's traffic load at time interval t. It is important to notice that we calculate the similarity among all regions so we discover the historical patterns globally. In addition, j is allowed to equal to i, which means we also consider self historical similarity.

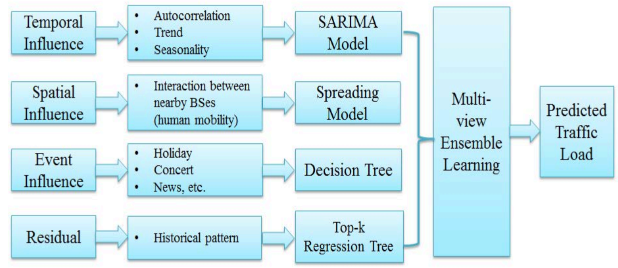


Figure 8: Multi-view ensemble learning framework

4.5 Multi-view Ensemble Learning Model

In this section, we leverage the random forest model to ensemble results from established multi-view models for the BS traffic. The random forest model is stable and fast to compute, thus it is wildly used in the industry. Since our solution space may be complicated, with every model contributing to traffic load dependent on other models, we could take advantage of the random forest [21] to decide feature compositions. The process is shown in Figure 8. First, we feed every model with its corresponding features to let it learn its parameters. Next, we view each model as a domain expert and fetch its prediction value, together with its model features, to let the forest study every model's role in the load contribution. Finally, we ask the random forest for the ultimate prediction result.

5 SLEEPING STRATEGY

BS sleeping strategy aims to shut down some carriers in RRU according to the traffic load for power saving. In this paper, the sleep mode is a general concept for BSes with some power consuming components off, rather than the specific state with only power on. The major concern of this sleeping strategy is the trade-off between the following three factors: the energy saving, operation cost, and the quality of service. For example, frequently switching on and off equipment according to the demand, would greatly reduce the energy cost while providing a more reliable service but sharply increase the operation cost. Adopting a more conservative strategy to stabilize the number of on-load devices would decrease destructive operations but become incapable of meeting the service quality requirement.

For generality, we estimate the traffic and take actions on each time interval Δt, and cut the strategy execution time $[0, T]$ into N copies, namely $N = \lceil \frac{T}{\Delta t} \rceil$. Then, we give the formal definitions of three factors explicitly.

Definition 5.1 (Quality of service). Quality of service is used to measure the communication quality, which can be reflected from the data missing rate. Formally, the quality of service during time $[0, T]$ is measured by the total missing data, $Q(T) = \sum_{k=1}^{N} \frac{max(L_k - C_k, 0)}{L_k}$, where L_k and C_k denote the region load and load capacity at the k-th interval, respectively.

Definition 5.2 (Operation cost). Operation cost is used to measure the device damage resulted from the operations (sleeping or activating), which can be reflected from the frequency of operations. Formally, the operation cost during time $[0, T]$ is measured by the total operations, $W(T) = F(\sum_{k=1}^{N} \lceil|\frac{C_k - C_{k-1}}{\Delta C}|\rceil)$, where $F(n)$ symbols for the accumulation function of device damage after n operations, C_k denotes load capacity at the k-th interval, and ΔC denotes the incremental capacity for one operation.

Definition 5.3 (Energy consumption). Energy consumption is the total consumed power during time $[0, T]$, formally defined as $E(T) = \sum_{k=1}^{N} P(min(C_k, L_k))$, where $P(\cdot)$ is the function of energy with the traffic load, determined by the BS structure shown in Section 3.1.

Based on the definitions, we can establish an optimization model for the sleeping strategy based on the predicted BS grid traffic. First, we assume BS energy function P has the form of $P(L) = \alpha L + \beta\lceil\frac{L}{\Delta C}\rceil$, where l is the traffic load, ΔC is the incremental traffic capacity from opening a carrier, and α and β are parameters modeling how the traffic load and each carrier operation affect the energy consumption. Further to notice, during the planning time, we cannot know L_t before hand, so we adopt the predicted value M_t instead. Next, we consider the quality of service. Because the operator would not like to receive any quality loss. We assume there is no missing data, $Q(T) = 0$. To fulfill this, we set a threshold e.g. 99% of the current capacity that once the actual load is beyond it, we switch on another BS to serve. This can help us to correct the following prediction, but increase the operation cost due to false predictions. For the constraint of operation cost $W(T)$, we observe $F(n)$ can be viewed as the depreciation of fixed asset after n operations. Hence, by estimating the total operations in the life time of the device, we get the discounted value to every operation. Consequently, this brings us the convenience to combine $W(T)$ and $E(T)$, because they are under the same measurement of money. Finally, this gives us Equation 5 and the solution to it, namely Algorithm 1.

$$\begin{aligned}
\min \quad & \sum_{k=0}^{N} P(min(C_k, M_k)) + F(\sum_{k=1}^{N} \lceil|\frac{C_k - C_{k-1}}{\Delta C}|\rceil) \\
\text{s.t.} \quad & \sum_{k=0}^{N} max(M_k - C_k, 0) = 0 \\
& C_k = n\Delta C, n \in \mathbb{N}
\end{aligned} \tag{5}$$

Algorithm 1: Strategy of Sleeping Device

Data: Predicted Load M, Current Time Index k, Last Time Activating Device Number x_{k-1}
Result: Next Time Activating Device Number x_k
μ is current cost for each operation;
$ShouldOpen_k = \lceil\frac{M(k\Delta t)}{C}\rceil$;
$x_k = max(x_{k-1}, ShouldOpen_k)$;
while $x_k > max(0, ShouldOpen_k)$ **do**
  find next time τ when $x_k \leq ShouldOpen_\tau$;
  if $2 * \mu > \beta(\tau - k)$ **then**
    $x_k += 1$;
    break ;
  end
  $x_k -= 1$;
end
return x_k

6 EXPERIMENTS

In this section, we conduct experiments to evaluate the performance of our proposed traffic prediction model and the dynamic BS sleeping strategy. So our experiments contain two parts: traffic prediction performance and sleeping strategy comparison. We aim to answer the following questions. 1) Does the proposed traffic prediction model perform better than state-of-the-art models? 2) Does the multi-view learning mechanism improve the prediction? 3) Does the proposed BS sleeping strategy work well?

6.1 Prediction Performance

6.1.1 Experimental Setting. We conduct the traffic prediction experiment on a real data set from one of the main operators in a big city of China. The data contain the traffic information of all BSes from Jan. 1, 2016 to Jun. 31 2016. We build our test set on the last 28 days of 181 available data entries. Moreover, the traffic load is calculated in Gigabyte unit. The time interval for traffic estimation and BS operation is two hours, and we evaluate the traffic prediction performance with the averaged root mean square error (ARMSE). ARMSE is the average RMSE for each region g, defined as follows,

$$ARMSE = \frac{\sum_{g=0}^{R-1}\sqrt{\frac{\sum_{i=1}^{N}|True - Predict|^2}{N}}}{R}, \tag{6}$$

where N is the number of total intervals and R is the number of regions.

6.1.2 Model Comparison. In this paper, we propose the multi-view ensemble learning model (MELM) to predict the BS traffic. In order to show how each view model affects the traffic, we propose three models for comparison: MELM (H), MELM (HT), and MELM (HTS), which capture the "historical patterns", "historical patterns and temporal influence", and "historical patterns, temporal influence, and spatial influence", respectively. Moreover, we compare our proposed methods with state-of-the-art statistical time series forecasting models: VAR and SARIMA [4]. We set their parameters as VAR(6) and SARIMA$(12, 1, 12) * (1, 1, 1)_{84}$ for the best performance. Also, we compare with machine learning based forecasting models: HMM [8] and ST-MVL [42].

Table 1 shows the experimental results. We observe that our proposed models outperform the baselines. MELM (H) performs better than the baseline methods because the temporal and spatial information is incorporated into the historical patterns. MELM (H) could find the right period by choosing the most correlated time

[2] The improvement is measured over the best baseline—SARIMA.

Table 1: Prediction Model Performance

Prediction Model	ARMSE	Improvement [2]
ST-MVL	23.79	–
SARIMA	21.85	0
VAR	22.42	–
HMM	24.70	–
MELM (H)	17.18	21%
MELM (HT)	15.91	27%
MELM (HTS)	13.48	38%
MELM	**13.20**	**40%**

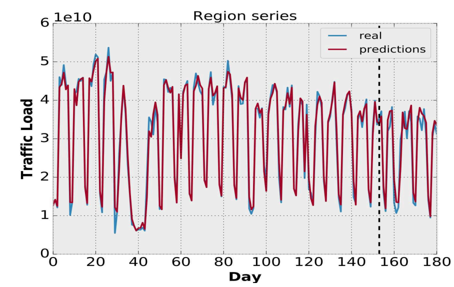


Figure 9: Catching Periods

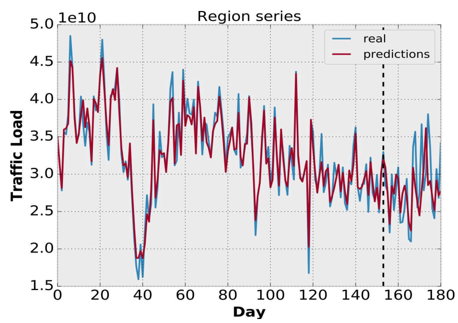


Figure 10: Catching Trends

point and notice the sudden change by watching the most related region traffic, also it can capture the temporal and spatial information hidden in the holiday pattern.

We also observe MELM (HT) is boosted up by around 8% than MELM (H) with the help of the temporal model. Further, MELM (HTS) improves MELM (HT) around 15% by capturing the spatial influence resulted from human mobility. In our experiment, the event influence only facilitates the performance 2%, because there is no significant events in our test data (data in June 2016).

In addition, we conduct experiments to show how our models capture the influence of each view in a traffic time series. Figure 9 shows how the proposed model catches the temporal period of the traffic data in a selected BS grid. The black line in the figure separates the training area and testing area. We can see that our prediction framework can recognize the period information. This functionality is attributed to the SARIMA model with the appropriate seasonality setting. The appropriate seasonality setting helps our prediction get higher accuracy.

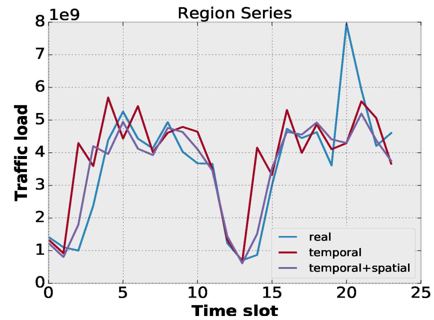


Figure 11: Exploiting Spatial Relations

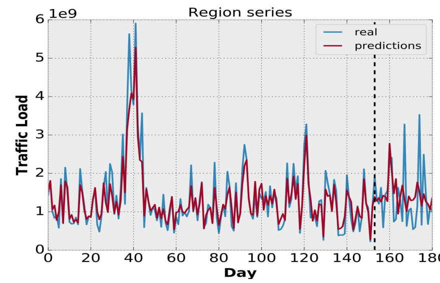


Figure 12: Catching Events

Figure 10 shows how the proposed model catches the trend effect in a selected BS grid. Similar to Figure 9, the black line in the figure separates the training area and testing area. We can observe our fitting line follows the downward trend and offers us a decent prediction, which benefits from the flagged integrated parameter in the SARIMA model.

Figure 11 shows a two-day-length time series of a selected region. We can observe that spatial information can improve our prediction model by correcting the false pulse of the temporal prediction at 2nd, 4th, 6th slots in the prediction days. This correction may follow the constraints of nearby load.

Figure 12 shows how the proposed model to catch the events, especially the holiday events in a time series of a selected region. Similar to Figure 9, the black line in the figure separates the training area and testing area. We notice that on the 160th day in our test days, the region met the Dragon Boat Festival, thus incurring high traffic load. The event view helps the prediction framework notice this event, and teach it to give a high expectation to coming hours.

6.2 Sleeping Strategy Evaluation

6.2.1 Experimental Setting. We evaluate our sleeping strategy in a simulated environment because taking actions on real BSes is not practical. Following [5, 29], we ignore the diversity of BSes and assume all BSes share the same configuration. Moreover, we assume the BSes in the same grid can replace each other when serving user clients.

We set $2100W$ as BS maximum output and $94.8W$ as the minimum energy cost only keeping power-on, while the former value is doubled in its load dependency part compared with [38] to achieve the maximum coverage in the activation mode according to [29], and the latter is consistent with the [38]. Therefore, the energy

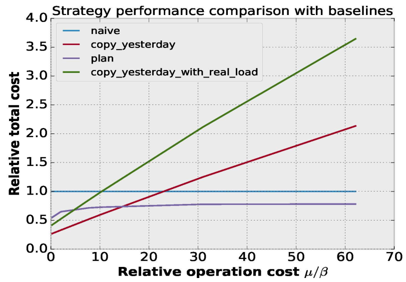


Figure 13: Performance comparison with baseline strategies

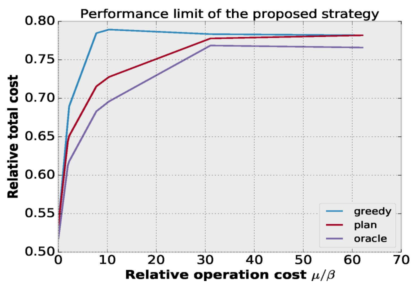


Figure 14: Demonstration of performance limit

consumption function with the traffic load L can be denoted as, $p(L) = 1200L + 805.2 + 94.8$, while the 805.2 can be interpreted as the fixed power used in the activation mode. Moreover, the energy cost is calculated in 1 RMB per Kwh. The capacity of each BS is decided by its equipment setting, each of them with three sectors, each sector fitted with two carriers. For each BS, we roughly evaluate it as 400 thousand RMB. In the strict condition, if every station can only bear one activating operation and corresponding one sleeping operation during each day through its ten years life time, we can distribute its value evenly into every operation cost μ, namely 55 RMB. Notice that, our model cares about the relative value of the operation cost to the energy saving $\frac{\mu}{\beta}$, rather than the absolute one μ, where β is the marginal revenue of turning a base station the into sleeping mode. The reasonable value of $\frac{\mu}{\beta}$ might be located between the $[5.18, 62.11]$, where left endpoint symbols for the chance to sleep BS in each time slot during its ten years life, and the right symbols for the constraint to sleep BS only once in each day.

6.2.2 Strategy Comparison. We compare our traffic prediction based strategy with three baseline strategies. The first one is a *naive strategy* that keeps all base stations in the activation mode. The second one named *copy yesterday* (CY) is using yesterday load as the prediction value to minimize the energy cost. The third one, *copy yesterday with real load* (CYR), is similar with the previous one but neglects the cap effect by the false prediction, which always uses the yesterday real load to predict. When the interval is half a day, the CY strategy is similar to the static strategies in [5, 29]. Without loss of generality, we compare ou strategy with the proposed three baselines.

Figure 13 demonstrates the experimental results of different BS controlling strategies. The total cost contains the energy consumption and the operation cost. Moreover, all cost values in Figure 13 are normalized over the cost of the naive strategy. From the experimental results, we obtain the following observations. **1)** Our strategy named *plan* mostly performs better than the baseline strategies. When the relative cost ($\frac{\mu}{\beta}$) approaches to zero, namely the operation cost is negligible, the *plan* strategy achieves the optimization upper bound, 47% power saving over the naive way. The reason why the *plan* strategy is worse than CY and CYR when the operation cost is near zero, is because CY and CYR miss more load requests compared with our strategy. The number of slots failing to offer enough capacity of CYR is 24% greater than ours, while the number of CY is 38% greater than ours. Also, the number of missing operations to activate the BS to satisfy the demand of CYR is 42% greater than ours. Therefore, in this case, we consider our strategy keeps far better service quality at less cost of energy savings. When we consider the operation cost, the *plan* strategy mostly performs much better than CY and CYR, especially when the relative cost ($\frac{\mu}{\beta}$) is larger than 20. **2)** We observe CY and CYR nearly form linear functions, because they both ignore the operation cost, thus keep a fixed operation number under any conditions. This explains the result that, when the operation cost is large enough, their performance will be worse than the naive strategy. Also, the result that the performance of the CY is better than CYR needs to be further clarified. Due to the cap effect of CY, once failing to fulfill the requesting load, it will miss all the load beyond that level so forth. Hence, its better performance is at the cost of missing more demanding load.

Moreover, we compare our proposed strategy with another two extreme strategies shown in Figure 14. The first is the greedy algorithm, which sleeps BSes as many as possible when marginal energy saving is larger than marginal operation cost. The second is the best strategy named *oracle*, assuming we can observe the real future load. Since they use the same prediction algorithm, the three strategies obtain the same total cost when the operation cost approaches zero. When the relative cost ($\frac{\mu}{\beta}$) locates near 10, which our strategy outperforms the greedy one by almost 10%. The *greedy* strategy fails to see the coming future load to take a conservative policy to keep stations activated, while *plan* exploits future information to sleep those unnecessary stations to save energy. These two strategies also converge since both of them will degenerate into the naive strategy, when the operation cost is very large. Generally, our strategy performs near to the *oracle* strategy with a gap of 5% in terms of the cost saving, which verifies the effectiveness of our traffic prediction algorithm.

7 CONCLUSION

In this paper, we propose a traffic prediction based BS power saving system in cellular networks. First, we partition the BSes of a city into different grids according to their geographical locations and service coverage. Next, we use a multi-view learning method to study the spatial and temporal pattern of each BS time series, and consider the event and historical influence as well. Then, we employ the ARMSE as our performance metric to evaluate the traffic prediction algorithm. The experiment shows that our prediction method outperforms the best competitor by around 40% in terms of accuracy. Moreover, on basis of a dynamic traffic prediction, we design a sleeping strategy to save the energy consumption by sleeping BSes at low traffic load. Our strategy explicitly takes the operation cost into consideration. Finally, we evaluate our strategy in a simulated environment and the experimental results demonstrate that our strategy yields more energy savings.

8 ACKNOWLEDGEMENTS

This work was partially supported by the Research Grants Council of the Hong Kong Special Administrative Region, China (No. CUHK 14203314 and No. CUHK 14234416 of the General Research Fund). In addition, Jianguo's work was supported in part by the Program for NSFC (No.61572322) and the STCSM project (No.16QA1402200).

REFERENCES

[1] Oliver Arnold, Fred Richter, Gerhard Fettweis, and Oliver Blume. 2010. Power consumption modeling of different base station types in heterogeneous cellular networks. In *Future Network and Mobile Summit, 2010.*
[2] Imran Ashraf, Federico Boccardi, and Lester Ho. 2011. Sleep mode techniques for small cell deployments. *IEEE Communications Magazine* (2011).
[3] Leo Breiman, Jerome Friedman, Charles J Stone, and Richard A Olshen. 1984. *Classification and regression trees.* CRC press.
[4] Peter J Brockwell and Richard A Davis. 2016. *Introduction to time series and forecasting.* springer.
[5] Bin Cao, Jing Fan, Mingxuan Yuan, and Yanhua Li. 2016. Toward accurate energy-efficient cellular network: switching off excessive carriers based on traffic profiling. In *Proceedings of the 31st Annual ACM Symposium on Applied Computing.*
[6] Shuguang Cui, Andrea J Goldsmith, and Ahmad Bahai. 2004. Energy-efficiency of MIMO and cooperative MIMO techniques in sensor networks. *IEEE Journal on selected areas in communications* (2004).
[7] Aleksandar Damnjanovic, Juan Montojo, Yongbin Wei, Tingfang Ji, Tao Luo, Madhavan Vajapeyam, Taesang Yoo, Osok Song, and Durga Malladi. 2011. A survey on 3GPP heterogeneous networks. *IEEE Wireless Communications* (2011).
[8] Weiwei Duan and Yao-Yi Chiang. 2016. Building knowledge graph from public data for predictive analysis: a case study on predicting technology future in space and time. In *SIGSPATIAL International Workshop on Analytics for Big Geospatial Data.*
[9] Gerhard Fettweis and Ernesto Zimmermann. 2008. ICT energy consumption-trends and challenges. In *Proceedings of the 11th international symposium on wireless personal multimedia communications.*
[10] Fabien Girardin, Andrea Vaccari, Alexandre Gerber, Assaf Biderman, and Carlo Ratti. 2009. Towards estimating the presence of visitors from the aggregate mobile phone network activity they generate. In *Intl. Conference on Computers in Urban Planning and Urban Management.*
[11] Congzheng Han, Tim Harrold, Simon Armour, Ioannis Krikidis, Stefan Videv, Peter M Grant, Harald Haas, John S Thompson, Ivan Ku, Cheng-Xiang Wang, and others. 2011. Green radio: radio techniques to enable energy-efficient wireless networks. *IEEE communications magazine* (2011).
[12] Minh X Hoang, Yu Zheng, and Ambuj K Singh. 2016. FCCF: forecasting citywide crowd flows based on big data. In *SIGSPATIAL.*
[13] Nahina Islam, Sithamparanathan Kandeepan, and James Scott. 2015. Energy efficiency of cellular base stations with ternary-state transceivers. In *ICSPCS.*
[14] Jinseong Jeong, Donald F Kimball, Myoungbo Kwak, Paul Draxler, Chin Hsia, Craig Steinbeiser, Thomas Landon, Oleh Krutko, Lawrence E Larson, and Peter M Asbeck. 2009. High-efficiency WCDMA envelope tracking base-station amplifier implemented with GaAs HVHBTs. *IEEE journal of solid-state circuits* (2009).
[15] Ram Keralapura, Antonio Nucci, Zhi-Li Zhang, and Lixin Gao. 2010. Profiling users in a 3g network using hourglass co-clustering. In *Proceedings of the sixteenth annual international conference on Mobile computing and networking.*
[16] Chang Li, Jun Zhang, and Khaled Ben Letaief. 2013. Energy efficiency analysis of small cell networks. In *ICC.*
[17] Ming Li, Pan Li, Xiaoxia Huang, Yuguang Fang, and Savo Glisic. 2015. Energy consumption optimization for multihop cognitive cellular networks. *IEEE Transactions on Mobile Computing* (2015).
[18] Rongpeng Li, Zhifeng Zhao, Xianfu Chen, Jacques Palicot, and Honggang Zhang. 2014. TACT: A transfer actor-critic learning framework for energy saving in cellular radio access networks. *IEEE transactions on wireless communications* (2014).
[19] Rongpeng Li, Zhifeng Zhao, Xuan Zhou, and Honggang Zhang. 2014. Energy savings scheme in radio access networks via compressive sensing-based traffic load prediction. *Transactions on Emerging Telecommunications Technologies* (2014).
[20] Yexin Li, Yu Zheng, Huichu Zhang, and Lei Chen. 2015. Traffic prediction in a bike-sharing system. In *SIGSPATIAL.*
[21] Andy Liaw and Matthew Wiener. 2002. Classification and regression by randomForest. *R news* 2, 3 (2002), 18–22.
[22] Christopher D Manning, Prabhakar Raghavan, Hinrich Schütze, and others. 2008. *Introduction to information retrieval.*
[23] Marco Ajmone Marsan, Luca Chiaraviglio, Delia Ciullo, and Michela Meo. 2013. On the effectiveness of single and multiple base station sleep modes in cellular networks. *Computer Networks* (2013).
[24] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *NIPS.*
[25] Diala Naboulsi, Marco Fiore, Stephane Ribot, and Razvan Stanica. 2016. Large-scale mobile traffic analysis: a survey. *IEEE Communications Surveys & Tutorials* (2016).
[26] Diala Naboulsi, Razvan Stanica, and Marco Fiore. 2014. Classifying call profiles in large-scale mobile traffic datasets. In *INFOCOM.*
[27] Eduardo Mucelli Rezende Oliveira, Aline Carneiro Viana, Kolar Purushothama Naveen, and Carlos Sarraute. 2015. Measurement-driven mobile data traffic modeling in a large metropolitan area. In *PerCom.*
[28] Utpal Paul, Anand Prabhu Subramanian, Milind Madhav Buddhikot, and Samir R Das. 2011. Understanding traffic dynamics in cellular data networks. In *INFOCOM, 2011 Proceedings IEEE.*
[29] Chunyi Peng, Suk-Bok Lee, Songwu Lu, Haiyun Luo, and Hewu Li. 2011. Traffic-driven power saving in operational 3G cellular networks. In *Proceedings of the 17th annual international conference on Mobile computing and networking.*
[30] Louai Saker, Salah-Eddine Elayoubi, Richard Combes, and Tijani Chahed. 2012. Optimal control of wake up mechanisms of femtocells in heterogeneous networks. *IEEE Journal on Selected Areas in Communications* (2012).
[31] Mohammad Salahuddin, Khorshed Alam, and Ilhan Ozturk. 2016. The effects of Internet usage and economic growth on CO2 emissions in OECD countries: A panel investigation. *Renewable and Sustainable Energy Reviews* (2016).
[32] M Zubair Shafiq, Lusheng Ji, Alex X Liu, and Jia Wang. 2011. Characterizing and modeling internet traffic dynamics of cellular devices. In *SIGMETRICS.*
[33] Ward Van Heddeghem, Sofie Lambert, Bart Lannoo, Didier Colle, Mario Pickavet, and Piet Demeester. 2014. Trends in worldwide ICT electricity consumption from 2007 to 2012. *Computer Communications* (2014).
[34] Guoxiang Wang, Caili Guo, Shengsen Wang, and Chunyan Feng. 2013. A traffic prediction based sleeping mechanism with low complexity in femtocell networks. In *ICC.*
[35] Yi Wang, Michalis Faloutsos, and Hui Zang. 2013. On the usage patterns of multimodal communication: Countries and evolution. In *Computer Communications Workshops (INFOCOM WKSHPS), 2013 IEEE Conference on.*
[36] Carey Williamson, Emir Halepovic, Hongxia Sun, and Yujing Wu. 2005. Characterization of CDMA2000 cellular data network traffic. In *Local Computer Networks, 2005. 30th Anniversary. The IEEE Conference on.*
[37] Fei Wu, Hongjian Wang, and Zhenhui Li. 2016. Interpreting traffic dynamics using ubiquitous urban data. In *SIGSPATIAL.*
[38] Jingjin Wu, Yujing Zhang, Moshe Zukerman, and Edward Kai-Ning Yung. 2015. Energy-efficient base-stations sleep-mode techniques in green cellular networks: A survey. *IEEE communications surveys & tutorials* (2015).
[39] Liu Ximei. 2016. http://www.escn.com.cn/news/show-334775.html. (2016).
[40] Cong Xiong, Geoffrey Ye Li, Shunqing Zhang, Yan Chen, and Shugong Xu. 2012. Energy-efficient resource allocation in OFDMA networks. *IEEE Transactions on Communications* (2012).
[41] Zheng Yang, Jonathan Roth, and Rishee K Jain. 2016. Data-driven benchmarking of building energy performance at the city scale. In *SIGSPATIAL Workshop on Smart Cities and Urban Analytics.*
[42] Xiuwen Yi, Yu Zheng, Junbo Zhang, and Tianrui Li. 2016. ST-MVL: Filling Missing Values in Geo-Sensory Time Series Data. In *IJCAI.*
[43] Ming Zeng, Tong Yu, Xiao Wang, Vincent Su, Le T Nguyen, and Ole J Mengshoel. 2016. Improving Demand Prediction in Bike Sharing System by Learning Global Features. *Machine Learning for Large Scale Transportation Systems (LSTS)@ KDD-16* (2016).
[44] Ying Zhang and Ake Årvidsson. 2012. Understanding the characteristics of cellular data traffic. *ACM SIGCOMM Computer Communication Review* (2012).
[45] Shenglin Zhao, Irwin King, Michael R Lyu, Jia Zeng, and Mingxuan Yuan. 2017. Mining Business Opportunities from Location-based Social Networks. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval.* ACM, 1037–1040.