coMobile: Real-time Human Mobility Modeling at Urban Scale Using Multi-View Learning

Desheng Zhang
zhang@cs.umn.edu
University of Minnesota, USA

Juanjuan Zhao, Fan Zhang
{jj.zhao, zhangfan}@siat.ac.cn
SIAT, Chinese Academy of Sciences

Tian He
tianhe@cs.umn.edu
University of Minnesota, USA

Abstract

Real-time human mobility modeling is essential to various urban applications. To model such human mobility, numerous data-driven techniques have been proposed. However, existing techniques are mostly driven by data from a single view, e.g., a transportation view or a cellphone view, which leads to over-fitting of these single-view models. To address this issue, we propose a human mobility modeling technique based on a generic multi-view learning framework called coMobile. In coMobile, we first improve the performance of single-view models based on tensor decomposition with correlated contexts, and then we integrate these improved single-view models together for multi-view learning to iteratively obtain mutually-reinforced knowledge for real-time human mobility at urban scale. We implement coMobile based on an extremely large dataset in the Chinese city Shenzhen, including data about taxi, bus and subway passengers along with cellphone users, capturing more than 27 thousand vehicles and 10 million urban residents. The evaluation results show that our approach outperforms a single-view model by 51% on average.

Categories and Subject Descriptors

H.4 [**Information System Application**]: Miscellaneous

Keywords

Human Mobility, Model Integration

1 Introduction

Nowadays, we are in a rapid process of urbanization where more than half of people in the world has moved to urban areas [20]. To ensure urban sustainability, how to capture human mobility at urban scale is one of the fundamental challenges we need to address. Such human mobility has many real-world applications, e.g., urban planning, transportation, social networking, and location based services [19]. To capture generic human mobility patterns, several theoretical models have been proposed, e.g., the gravity model and the radiation model [15]. However, a key drawback of these theoretical models is that they cannot capture human mobility at fine spatiotemporal granularity, e.g., mobility at small region levels in *real time*.

SIGSPATIAL'15, November 03 - 06, 2015, Bellevue, WA, USA

ACM 978-1-4503-3967-4/15/11...$15.00
http://dx.doi.org/10.1145/2820783.2820821

Recently, thanks to upgrades of urban infrastructures, many real-time location-tracking devices become available, e.g., cellphones, onboard GPS devices and smartcards. These devices generate massive real-time location data, which hold the key potential to revolutionize real-time human mobility modeling. Based on these real-time data, several data-driven models have been proposed, e.g., driven by data from cellphones [11], smartcards [16], taxis [6], buses [3], or subways [7]. However, a common feature of these models is that they capture mobility only from one view, e.g., a cellphone view or a transportation view. These single-view models are sufficient if single-view data are complete, but in reality this is not the case. From the cellphone view, the models driven by cellphone data cannot capture residents without cellphone data, e.g., residents who do not have cellphones and residents who have cellphones but do not use their cellphones during our modeling time; similarly, from the transportation view, the models driven by one kind of transportation data, e.g., taxi, cannot capture the passengers who use other transportation modes, e.g., bus and subway, and further there is no urban infrastructure that can capture private vehicles at urban scale. To our knowledge, no data-driven urban human mobility models are based on a complete view so far. As a result, these single-view human mobility models essentially use residents captured by these single views as a sample to study all residents, which inevitably leads to a bias and thus over-fitting of their models, as shown in Section 2.

To address this issue, we aim to combine different views for multi-view modeling. Each view is incomplete to capture mobility by itself, but one view is often complementary to others, e.g., the cellphone view can capture some private-vehicle passengers, whereas the transportation view can capture some inactive cellphone users. But a view's ability to capture human mobility is unknown *a priori* and is highly dynamic based on spatiotemporal contexts. As a result, such dynamic view completeness makes multi-view human mobility modeling extremely challenging.

In this work, we propose coMobile, a generic framework to capture human mobility with a multiple-view learning technique. In coMobile, we first design a single-view learning technique based on context-based tensor decomposition to improve completeness of single-view models. Then, we integrate those improved single-view models together by formulating a convex optimization to obtain the ground truth of urban mobility. Mostly importantly, we implement coMobile based on extremely large datasets in the Chinese city Shenzhen with cellphone data and transportation data including taxis, buses, and subways. In particular, the key contributions of the paper are as follows:

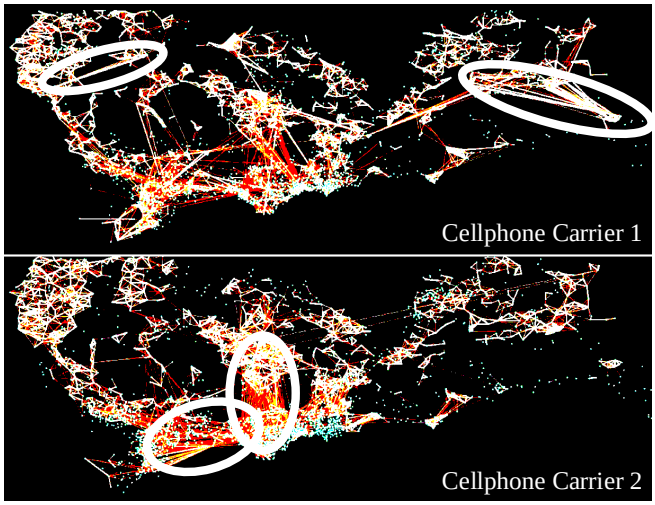

Fig 1. Models Driven by Two Carrier's CDR Data

- We propose the first multi-view learning framework for human mobility to integrate incomplete yet complementary knowledge from individual views. To our knowledge, the proposed model is the only human mobility model driven by more than one view, which aims to address over-fitting of single view models. It is challenging to apply multi-view learning in human mobility modeling, because data-driven views are mostly incomplete to urban-scale mobility.
- We design a single-view learning technique based on context-aware tensor decomposition with both real-time and historical data to improve completeness of single-view models. This technique addresses data sparsity challenges of particular views to improve their completeness. In particular, we use a cellphone-view model as an example to show how we extract three contexts, i.e., cellphone user density, calling location patterns, and calling time patterns, based on historical data for joint tensor decomposition.
- Based on improved single-view models, we formulate a multi-view modeling problem by designing a joint optimization, which minimizes overall weighted deviation from observed mobility to the ground truth. To solve this optimization, we propose an iterative learning process to alternatively update ground truth and view completeness until no further improvement can be made for the objective function. We formally prove the convexity of the joint optimization and the convergence of our iterative learning.
- We implement our multi-view human-mobility model based on two datasets in the Chinese city Shenzhen, with 10 million cellphone users and 16 million smartcard users involved. To our knowledge, this is one of the largest human mobility models driven by real-world datasets. We evaluate our model by comparing it to a single-view model, and results show that we reduce error rates by 51% on average.

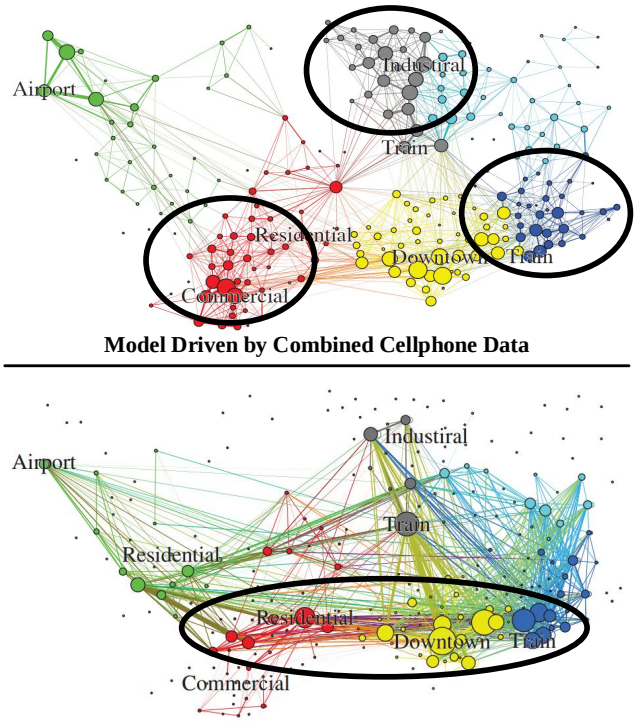

Fig 2. Models Driven by Cellphone and Transit Data

2 Motivation

We first show the drawback of single-view models and the opportunity of multi-view models.

2.1 Drawback of Single-View Models

We give two comparisons: (i) models driven by two different cellphone views; (ii) models driven by a cellphone view and a transportation view.

As in Fig. 1, we first compare models driven by two one-day CDR (call detail records) datasets from two carriers in Shenzhen. This kind of models driven by single-carrier data is mostly used for human mobility modeling [9]. A point indicates a spatial unit covered by a cell tower, and an edge linking two points together indicates the mobility between them. We only show the major mobility for the clarity of the figure. As shown by the circles, we found that each model can capture some unique mobility that cannot be captured by the other, which leads to over-fitting of these models driven by CDR data from single cellphone carriers.

We combine the CDR data from different carriers, and obtain a model driven by combined CDR data. Similarly, we combine data from different urban transportation, i.e., taxi, bus and subway, together, and then obtain a model driven by the transportation data. Due to different spatial granularity (details in Section 3.1), we use a urban-region-based model to show captured mobility in the morning rush hour. As in Fig. 2, every point indicates a region in the Shenzhen urban area; every edge linking two regions together indicates the mobility volume between them. The size of a vertex indicates associated mobility, and the different color indicates urban districts. As shown by the circles, we also found that each model can capture some mobility that cannot be captured by the other.

Cellphone Dataset		Taxicab Dataset		Bus Dataset		Smartcard Dataset	
Beginning	2013/10/1	Beginning	2012/1/1	Beginning	2013/1/1	Beginning	2011/7/1
# of Users	10,432,246	# of Taxis	14,453	# of Buses	13,032	# of Cards	16,000,000
Size	1 TB	Size	1.7 TB	Size	720 GB	Size	600 GB
# of Records	19 billion	# of Records	22 billion	# of Records	9 billion	# of Records	6 billion
Format		Format		Format		Format	
SIM ID	Date&Time	Plate ID	Date&Time	Plate ID	Date&Time	Card ID	Date&Time
Cell Tower ID	Activities	Status	GPS&Speed	Stop ID	GPS&Speed	Device ID	Station ID

Fig 3. Cellphone and Transportation Data

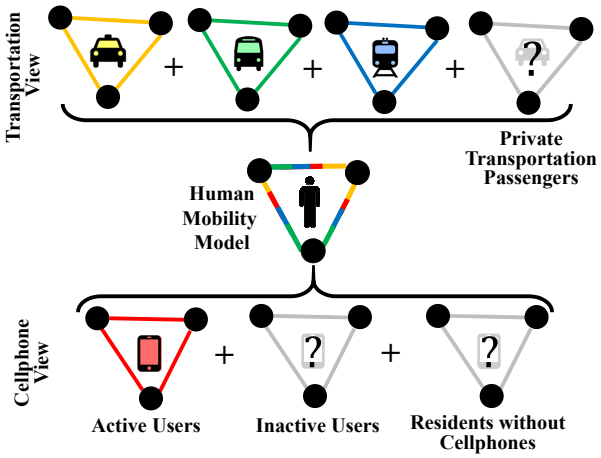

Fig 4. Multi-View Modeling

2.2 Opportunity of Multi-View Models

Due to the limitation of the single-view models, we are motivated to combine two separate views together in order to design a multi-view model for human mobility.

As shown by Fig. 4, from the transportation view, we aim to combine four independent models (i.e., four triangles) driven by data from taxis, buses, subways, and private vehicles for human mobility modeling. But currently there is no urban infrastructure that can capture private transportation in real time at urban scale. Some efforts have been made by the research community to install GPS devices in the private vehicles to study human mobility [21], but only limited private vehicles are involved.

Alternatively, we can design a model driven by cellphone CDR data as in Fig. 4. But there are two challenges. (i) Some cellphone users would not use their cellphones (i.e., being inactive) during the time we perform modeling. To address this issue, we design a technique based on tensor decomposition with correlated contexts to infer locations of inactive cellphone users in Section 4. (ii) Some urban residents who opt out of allowing their CDR data used for other purposes or do not have cellphones at all. Therefore, for these residents, we cannot capture their mobility.

As a result, neither the transportation view nor the cellphone view is complete by itself, but one view is often complementary to another. For example, the model driven by cellphone data can provide some mobility about residents using private transportation; whereas the model driven by transportation data can provide some mobility about residents without cellphone CDR data. It motivates us to design an effective modeling technique to combine these two views for better mobility modeling.

3 Preliminary

In this section, we first introduce the data we collected for multi-view modeling, and then we present a concept called mobility graph to capture the real-time human mobility, and finally we give the architecture of coMobile.

3.1 Multi-View Data

We have been working with several service providers and the Shenzhen Transport Committee (hereafter STC) for data access of urban infrastructures. We consider two kinds of data, i.e., cellphone data and transportation data, as two individual views to model human mobility. A summary of these data is given by Fig. 3. The heat map of their spatial granularity is given by Fig. 5 with an area of 14×5 km^2.

Cellphone View: Cellphone CDR (call detail records) data are used to infer cellphone users' locations at cell tower levels. We utilize CDR data through two major operators in Shenzhen with more than 10 million users. The CDR data give 220 million locations per day.

Transportation View: Data from three kinds of transportation modes, i.e., taxi, subway and bus, are used to detect transportation passengers' locations. We study transportation data through STC to which taxicab, bus and smartcard companies upload their data in real time.

- **Taxi data** are used to infer taxi passengers' origins and destinations based on status (i.e., pickups and dropoffs) at GPS location levels. They account for 14 thousand taxis, each of which generates 2 records/min.
- **Smartcard data** are used to infer origins and destinations of residents with smartcards used to pay bus and subway fares, which capture more than 10 million rides and 6 million passengers per day. In particular, there are two kinds of smartcard readers: (i) a total of 14,270 onboard mobile readers in 13 thousand buses capturing 168 thousand bus passengers per hour, and (ii) a total of 2,570 fixed readers in 127 subway stations capturing 60 thousand subway passengers per hour. Smartcard data and subway map data are used together to detect subway passengers' origins and destinations at subway station levels.
- **Bus data** are used to infer bus passengers' origins and destinations along with smartcard data (showing that a passenger uses a smartcard at a bus station) at 4849 bus station levels. They account for all 13 thousand buses, each of which generates 2 records/min.

Our endeavor of consolidating the above data enables extremely large-scale real-time urban phenomenon rendering, e.g., human mobility, which is unprecedented in terms of both quantity and quality.

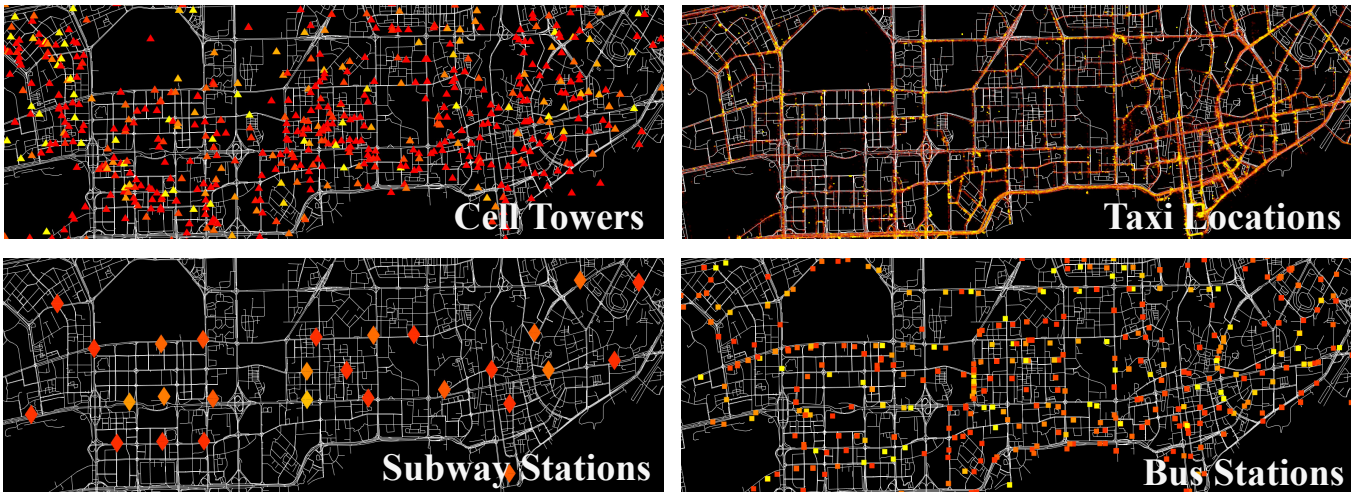

Fig 5. Data Spatial Granularity

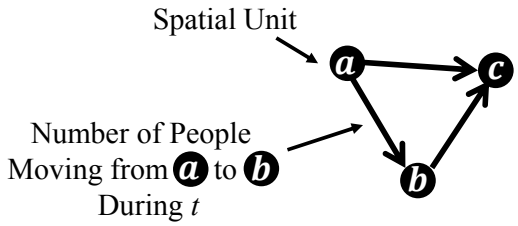

Fig 6. Mobility Graph

3.2 Mobility Graph

In this work, we use Mobility Graph to capture human mobility in real time at urban scale, which is a time-varying graph where a vertex indicates a spatial unit (e.g. a urban region or a street block) and a weight of an edge linking two vertices indicates the mobility volume between them. Due to its time-varying nature, a mobility graph G_t is associated with a time period t (e.g., 4-5PM), which shows the mobility during this particular time period.

Fig. 6 gives a simplified example of a mobility graph with only 3 vertices. The number of people moving between different spatial units, i.e., weights of edges, should include people associated with a particular view, e.g., the cellphone view or the transportation view. In this work, our main objective is to obtain mobility graphs based on single-view modeling, and then to combine them together by multi-view modeling for a comprehensive human mobility graph.

3.3 coMobile Framework

We introduce our coMobile Framework by Fig. 7. From the bottom, we have urban data generated by urban infrastructures, e.g., cellphone data and transportation data, which are introduced in Section 3.1. Based these two kinds of data, we design two single-view models capturing mobility patterns of cellphone users and urban transportation users by two mobility graphs, which are introduced in Section 4. Then, we present our multi-view learning to integrate single-view models for more complete human mobility modeling, which is introduced in Section 5. Finally, the obtained human mobility model can be used in many applications, e.g., ridesharing and transit energy inference.

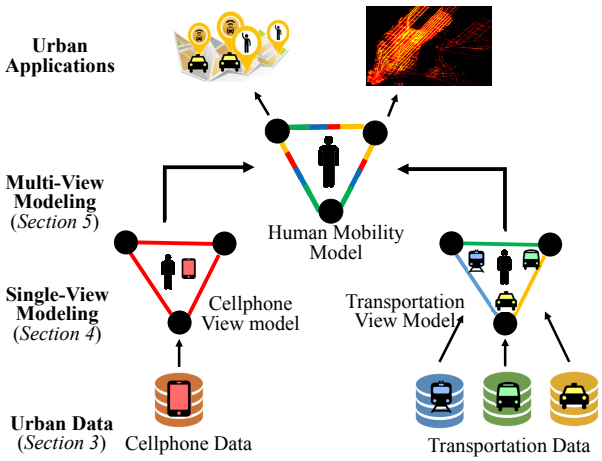

Fig 7. coMobile Framework

Note that we only consider two specific views in coMobile but it can be generalized to more views if more data are available. In coMobile, we first generate single-view models and then combine them together at model levels, instead of raw data levels (e.g., using multi-source raw data to directly design a multi-view model). This is because in many applications due to privacy issues, raw data are not available, and only high-level single models can be used as input. Our coMobile is still applicable to this situation.

4 Single-View Mobility Modeling

We introduce how to model urban mobility based on two single views, i.e., a cellphone view and a transportation view.

4.1 Cellphone-View Modeling

As introduced earlier, the key challenge to model human mobility based on cellphone data is that inactive cellphone users or residents without cellphones do not generate any CDR data. As a result, we cannot model their mobility to obtain mobility graph. For residents without cellphones, the solution is limited although the model based on transportation can capture some of them. In this subsection, we focus on inactive cellphone users to infer their mobility by an observation that inactive cellphone users who did not use their

cellphones today may use their cellphones before during the similar trips [8]. Accordingly, we formulate a tensor decomposition problem to infer mobility of both active and inactive users based on real-time and historical data.

4.1.1 Tensor Construction

We infer locations of cellphone users for specific time slots by a three dimensional tensor $\mathcal{A} \in \mathbb{R}^{N \times K \times M}$.

- A cellphone user dimension indicates individual cellphone users differentiated by SIM IDs: $[u_1, ..., u_N]$.
- A time slot dimension indicates specific time windows (e.g., one hour window from 5PM to 6PM): $[t_1, ..., t_K]$.
- A spatial unit dimension indicates specific spatial units (e.g., a urban region): $[s_1, ..., s_M]$.
- An entry $\mathcal{A}(n,k,m)$ indicates the number of CDR records a user n has in a spatial unit m during a slot k.

With our cellphone data, we fill this tensor $\mathcal{A}$, and then obtain all cellphone users' locations with a specific spatiotemporal partition. However, a key challenge is that the tensor $\mathcal{A}$ is sparse because for inactive cellphone users, their corresponding entries are empty due to lacking CDR data.

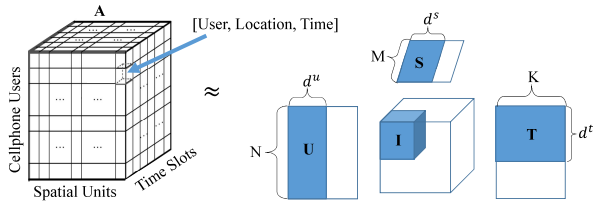

Fig 8. Tensor Decomposition

A common approach to address this issue is to use tensor decomposition. As in Fig. 8, we have a tensor with three dimensions indicating cellphone users, spatial units, and time slots. An entry denotes a tuple [user, location, time]. But this tensor is sparse due to inactive cellphone users. Based on the classic Tucker decomposition model [13], we decompose $\mathcal{A}$ into a core tensor I along with three matrices, $\mathcal{U} \in \mathbb{R}^{N \times d^u}$, $\mathcal{S} \in \mathbb{R}^{M \times d^s}$, and $\mathcal{T} \in \mathbb{R}^{K \times d^t}$. $\mathcal{U}$, $\mathcal{S}$, and $\mathcal{T}$ infer correlations between different cellphone users, different spatial units, and different time slots, respectively. d^u, d^s and d^t are the number of latent factors and very small.

The following objective function is used to optimize the decomposition.

$$||\mathcal{A} - I \times \mathcal{U} \times \mathcal{S} \times \mathcal{T}||^2 + \lambda(||I||^2 + ||\mathcal{U}||^2 + ||\mathcal{S}||^2 + ||\mathcal{T}||^2)$$

where the first term is to measure the error of decomposition and the second term is a regularization function to avoid over-fitting. $||\cdot||^2$ denotes the l_2 norm and λ is the parameter to control the contribution of the regularization function. By minimizing this objective function, we obtain the optimized I, $\mathcal{U}$, $\mathcal{S}$, and $\mathcal{T}$ by the sparse tensor $\mathcal{A}$, which is given by cellphone data. As a result, we use $I \times \mathcal{U} \times \mathcal{S} \times \mathcal{T} = \mathcal{A}$' to approximate $\mathcal{A}$ where $\times$ is the tensor-matrix multiplication.

However, a key challenge for this decomposition is that $\mathcal{A}$ is over sparse especially under fine spatiotemporal partition, which leads to poor performance of decomposition. To address this issue, in this work, we propose a technique to use historical cellphone data to establish correlated contexts that improve the performance of the decomposition.

4.1.2 Context Extraction

To provide additional information for the decomposition, we use the historical cellphone data to extract three contexts, i.e., cellphone user density, calling location pattern, and calling time pattern. We use three matrices to denote these three contexts as in Fig. 9.

- Cellphone User Densities are given by a matrix $\mathcal{B}$ where a row denotes a spatial unit; a column denotes a time slot; an entry denotes the average CDR record count in this spatial unit for this time slot over a period of historical time.
- Calling Location Patterns are given by a matrix $\mathcal{C}$ where a row denotes a spatial unit; a column denotes a cellphone user; an entry denotes a cellphone user's CDR record count in this spatial unit given a period of historical time.
- Calling Time Patterns are given by a matrix $\mathcal{D}$ where a row denotes a time slot; a column denotes a cellphone user; an entry denotes a cellphone user's CDR record count in this time slot given a period of historical time.

All the matrices $\mathcal{B}$, $\mathcal{C}$, and $\mathcal{D}$ can be obtained by a set of historical cellphone data.

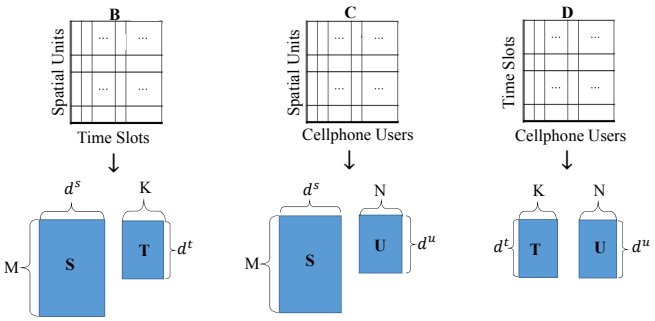

Fig 9. Context Matrix Factorization

4.1.3 Context-based Tensor Decomposition

We present a joint tensor decomposition based on the three extracted context matrices. In particular, we design the objective function as follows.

$$\min_{I,\mathcal{U},\mathcal{S},\mathcal{T}} \mathbf{L}(I,\mathcal{U},\mathcal{S},\mathcal{T}) = ||\mathcal{A} - I \times \mathcal{U} \times \mathcal{S} \times \mathcal{T}||^2 + \lambda_1||\mathcal{B} - \mathcal{S} \times \mathcal{T}||^2 + \lambda_2||\mathcal{C} - \mathcal{S} \times \mathcal{U}||^2 + \lambda_3||\mathcal{D} - \mathcal{T}^T \times \mathcal{U}||^2 + \lambda_4(||I||^2 + ||\mathcal{U}||^2 + ||\mathcal{S}||^2 + ||\mathcal{T}||^2). \quad (1)$$

where the first term is to measure the error of decomposing $\mathcal{A}$; the second, third, and forth terms are to measure the error of factorizing matrix $\mathcal{B}$, $\mathcal{C}$, and $\mathcal{D}$, respectively; the last term is to avoid over-fitting. In our setting, $d^u = d^s = d^t$. λ_1, λ_2,

λ_3, and λ_4 are preset parameters to indicate term weights. We normalized all values to [0, 1] for the decomposition.

In this objective function, $\mathcal{A}$ and $\mathcal{B}$ share $\mathcal{S}$ and $\mathcal{T}$; $\mathcal{A}$ and $\mathcal{C}$ share $\mathcal{S}$ and $\mathcal{U}$; $\mathcal{A}$ and $\mathcal{D}$ share $\mathcal{U}$ and $\mathcal{T}$. Since $\mathcal{B}$, $\mathcal{C}$, and $\mathcal{D}$ are not sparse, they lead to accurate $\mathcal{S}$, $\mathcal{T}$ and $\mathcal{U}$, which increases the performance of decomposing $\mathcal{A}$. As a result, the historical cellphone user calling patterns are transferred into the decomposition of $\mathcal{A}$, which leads to an accurate tensor decomposition.

Because this objective function does not have a closed-form solution to find the global optimal I, $\mathcal{U}$, $\mathcal{S}$, and $\mathcal{T}$, we use an element-wise optimization algorithm as a numeric method [12] to obtain a local optimal solution. Finally, after we obtain I, $\mathcal{U}$, $\mathcal{S}$, and $\mathcal{T}$, we use $I \times \mathcal{U} \times \mathcal{S} \times \mathcal{T} = \mathcal{A}'$ to obtain cellphone mobility graph G^C of all cellphone users.

4.2 Transportation-View Modeling

Based on our transportation data, we model human mobility by three transportation modes, i.e., taxi, bus and subway. Given attributes of our transportation data, we directly obtain origins and destinations of taxi, bus and subway passengers at GPS, bus station, and subway station levels. In this work, we use a space alignment technique where we assign taxi GPS locations, bus stations, and subway stations into corresponding spatial units based on a specific spatial partition of urban areas. Thus, for a pair of spatial units, e.g., from an airport to a train station, we aggregate all the above passengers who traveled between these two spatial units to obtain a mobility volume during a particular time period, because these three kinds of transportation modes are independent from each other. Thus, from the transportation view, obtaining transportation mobility graph G^T is straightforward.

Our context-aware tensor decomposition can also be used to improve completeness of the transportation-view model since we have missing data issues (e.g., GPS records) as well. The process is conceptually similar to the tensor decomposition for the cellphone-view model, which is omitted due to the space limitation.

Further, we did not consider private vehicles in our transportation view due to lack of private vehicle data. However, some urban residents using private transportation would be captured by multi-view learning, which is introduced as follows.

5 Multi-View Mobility Modeling

In this section, based on single-view modeling, we introduce multi-view modeling in coMobile. Even though our data can only form two views to obtain two mobility graphs, i.e., the cellphone mobility graph G^C and the transportation mobility graph G^T, we aim to tackle a more generic problem, i.e., multi-view modeling, and thus double-view modeling is a concrete example of multi-view modeling.

We first formulate a joint optimization problem for multi-view human mobility modeling, and then we develop an iterative learning processing to solve this problem, and finally we theoretically analyze the performance of modeling in terms of convexity and convergence.

5.1 Joint Optimization

The main objective of our multi-view modeling is to obtain a comprehensive human mobility graph G^H for a given time period based on several single-view mobility graphs, e.g., G^C and G^T. Because we have the same spatial partition for different mobility graphs, they have the same number of edges and vertices, and the key difference is edge weights. Since different edges are independent in a human mobility graph, we use one edge ab in a human mobility graph G^H as an example to show how we obtain the human mobility from one spatial unit a to another spatial unit b by our multi-view technique, and combine different edge weights together to obtain a complete human mobility graph G^H.

For a specific edge ab in G^H, the volume of passengers traveling from a spatial unit a (e.g., an airport) to b (e.g., a train station) during a time period t (e.g., 4-5PM) is $x^*_{ab\cdot t}$, which is the unknown ground truth we want to infer. Assuming we have V different views, which leads to V different mobility graphs that are incomplete by themselves yet complementary to each other. For a specific view $v \in [1, V]$, we use $x^v_{ab\cdot t}$ to indicate the weight of the edge ab during t in the mobility graph G^v; for a specific view $v \in [1, V]$, we use $w^v_{ab\cdot t}$ to indicate the completeness degree of this view during a time period t from this edge ab of G^v. The completeness degree of a view quantifies its capability to capture human mobility. The stronger the capability is, the higher the degree is. Under different spatiotemporal contexts, the completeness degree of the same view is different. We use a vector $\mathcal{W}_{ab\cdot t} = \{w^1_{ab\cdot t}, ..., w^v_{ab\cdot t}, ..., w^V_{ab\cdot t}\}$ to indicate completeness degrees for all V views.

In coMobile, based on the above definitions, V and $x^v_{ab\cdot t}$ are given in advance by the datasets; whereas $x^*_{ab\cdot t}$ and $\mathcal{W}_{ab\cdot t}$ are unknown. Therefore, we present a joint optimization to obtain optimal $x^*_{ab\cdot t}$ and $\mathcal{W}_{ab\cdot t}$ together. The basic idea behind our multi-view learning is that a view with a higher completeness degree provides more comprehensive information, so the ground truth should be close to mobility observed by a view with a higher completeness degree. As a result, we should minimize the deviation from mobility observed by a view v to the ground truth $x^*_{ab\cdot t}$ (unknown), proportionally to its completeness degree $w^v_{ab\cdot t}$ (also unknown). Therefore, we develop the following objective function for multi-view learning.

$$\min_{x^*_{ab\cdot t}, \mathcal{W}_{ab\cdot t}} \mathbf{F}(x^*_{ab\cdot t}, \mathcal{W}_{ab\cdot t}) = \sum_{v=1}^{V} [w^v_{ab\cdot t} \cdot \mathbf{D}(x^*_{ab\cdot t}, x^v_{ab\cdot t})], \quad \text{s.t.}, \mathbf{R}(\mathcal{W}_{ab\cdot t}) = 1. \tag{2}$$

$\mathbf{D}(x^*_{ab\cdot t}, x^v_{ab\cdot t})$ is a distance function that describes the distance between $x^*_{ab\cdot t}$ and $x^v_{ab\cdot t}$. Therefore, the term $\sum_{v=1}^{V} [w^v_{ab\cdot t} \cdot \mathbf{D}(x^*_{ab\cdot t}, x^v_{ab\cdot t})]$ indicates the overall weighted distance between the observed mobility and the ground truth. We aim to find the optimal $x^*_{ab\cdot t}$ and $\mathcal{W}_{ab\cdot t}$ that minimize this overall weighted distance under a constraint.

$\mathbf{R}(\mathcal{W}_{ab\cdot t})$ is a constraint function, which gives the distribution of view completeness. Without this constraint, the optimization problem is unbounded. For the sake of simplicity, we set $\mathbf{R}(\mathcal{W}_{ab\cdot t}) = 1$. Other constraint functions can

also be used since we can divide $\mathbf{R}(\mathcal{W}_{ab\cdot t})$ by a constant.

The rationale behind this function is that for a more-complete view, we have a high penalty if the mobility observed from this view has a longer distance to ground truth. In contrast, for a less-complete view, we have a low penalty if the mobility observed from this view has a longer distance to ground truth. Thus to minimize the objective function, ground truth relies on the more complete views.

5.2 Iterative Learning

We develop an iterative learning technique based on the block coordinate descent [2] to solve this optimization. Since in our objective function we have two sets of variables, i.e., both the ground truth $x^*_{ab\cdot t}$ and the view completeness degree $\mathcal{W}_{ab\cdot t}$, we aim to iteratively yet alternatively optimize these two sets of variables until the result converges. In particular, we optimize the value of one set to minimize the objective function while keeping the value of the other set fixed, and then we swap the fixed variable and the optimized variable to continue this process until the result converges. Fig. 10 gives the description of our iterative technique.

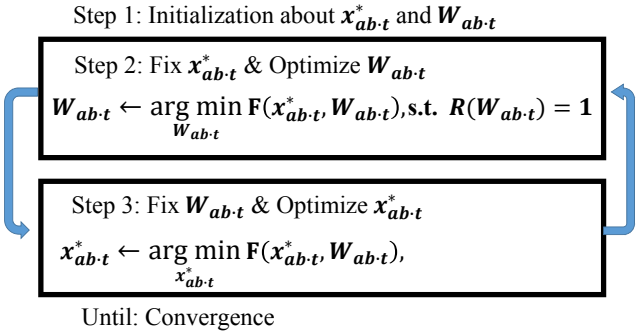

Fig 10. Iterative Multi-View Learning

In Step 1, we first initialize $x^*_{ab\cdot t}$ and $\mathcal{W}_{ab\cdot t}$ based on the average value of $x^*_{ab\cdot t}$, because the initialization does not affect the final results based on the property of the block coordinate descent [2]. In Step 2, we first fix the initialized $x^*_{ab\cdot t}$, and then find the optimal $\mathcal{W}_{ab\cdot t}$ that minimizes the objective function. In Step 3, with this optimized $\mathcal{W}_{ab\cdot t}$, we fix it and then find the optimal $x^*_{ab\cdot t}$ that minimizes the objective function again. Then, with this optimized $x^*_{ab\cdot t}$, we go back to Step 2 to fix $x^*_{ab\cdot t}$ again, and then to further optimize $\mathcal{W}_{ab\cdot t}$. This is an iterative process to alternatively optimize $x^*_{ab\cdot t}$ and $W_{ab\cdot t}$ until the result converges.

Based on the property of the block coordinate descent [2], the convergence of the above iterative process is based on the distance function and constraint function used. As follows, we theoretically analyze the performance of our technique in terms of convexity and convergence.

5.3 Theoretical Analyses

We use Negative Log Function as our constraint function:

$$\mathbf{R}(\mathcal{W}_{ab\cdot t} = \{w^1_{ab\cdot t}, ..., w^v_{ab\cdot t}, ..., w^V_{ab\cdot t}\}) = \sum_{v=1}^{V} \exp(-w^v_{ab\cdot t}).$$

This negative log function maps a number between 0 and 1 to a number from 0 to ∞, which enlarges the difference between different view completeness degrees for better modeling.

Further, we use Normalized Squared Loss function as our distance function given as

$$\mathbf{D}(x^*_{ab\cdot t}, x^v_{ab\cdot t}) = \frac{(x^*_{ab\cdot t} - x^v_{ab\cdot t})^2}{\mathrm{STD}(x^1_{ab\cdot t}, ..., x^v_{ab\cdot t}, ..., x^V_{ab\cdot t})}.$$

This normalized squared loss is an effective method to measure the distance between two variables and consider the distribution of $x^v_{ab\cdot t}$ at the same time.

As follows, we prove the convexity and convergence of our iterative learning with the above two functions.

THEOREM: If the negative log function and the normalized squared loss function are used, then convergence of our iterative process in Fig. 10 is guaranteed.

PROOF: Based on the convergence proposition on the block coordinate descent [2], the iterative process converges to a stationary point, if the optimizations in Steps 2 and 3 are convex. Thus, the rest of our proof has 2 steps: (i) in Step 2, if $x^*_{ab\cdot t}$ is fixed, the optimization for $\mathcal{W}_{ab\cdot t}$ is convex; (ii) in Step 3, if $\mathcal{W}_{ab\cdot t}$ is fixed, the optimization for $x^*_{ab\cdot t}$ is convex.

To prove the convexity of Step 2, we use another variable $y_v = \exp(-w^v)$. Therefore, the optimization problem becomes a new function with only one variable of y_v.

$$\min_{y_1,...,y_v,...,y_V} \mathbf{F}(y_1, ..., y_v, ..., y_V) = \sum_{v=1}^{V} [-\log(y_v) \cdot \mathbf{D}(x^*_{ab\cdot t}, x^v_{ab\cdot t})],$$

$$\text{s.t.}, \sum_{v=1}^{V} y_v = 1.$$

With this new variable y_v, we have a linear constraint function and a linear objective function (i.e., a linear combination of negative logarithm functions). Therefore, both the constraint function and objective function are convex, which leads to the fact that any local optimal solution is also the global optimal solution for Step 2.

To prove the convexity of Step 3, we treat the objective function as an unconstrained optimization with only one variable. In Step 3, since the normalized squared loss function is convex, the objective function is a linear combination of convex functions, which makes it convex. □

Note that other constraint and distance functions can also be used in our iterative process but may not lead to the convexity of the optimization problem, and thus the convergence of the iterative process cannot be guaranteed.

6 Implementation and Evaluation

In this section, we first introduce our implementation of coMobile based on data from the Chinese city Shenzhen. Then, we present our evaluation by comparing coMobile and a single-view model to the ground truth.

6.1 coMobile Implementation

We implement coMobile based on cellphone and transportation data in Shenzhen introduced in Section 3. Since our paper concentrates on modeling, we briefly introduce our data-related issues during our implementation. We establish a secure and reliable transmission mechanism, which feeds our server the data collected by STC and service providers with a wired connection. As shown in Section 3,

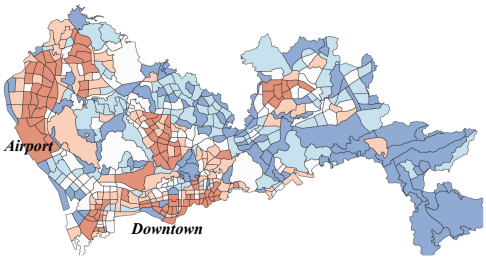

Fig 11. Shenzhen Urban Partition

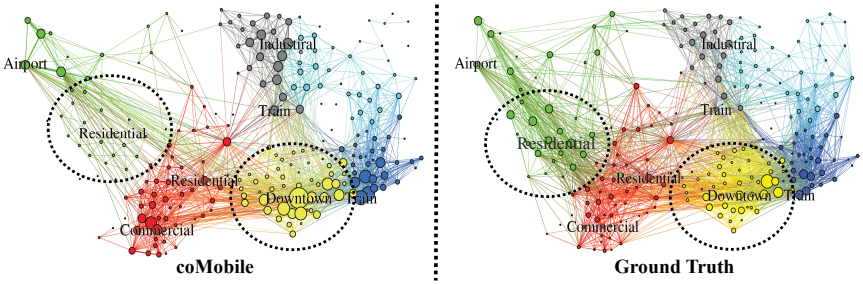

Fig 12. Mobility Graphs in Shenzhen

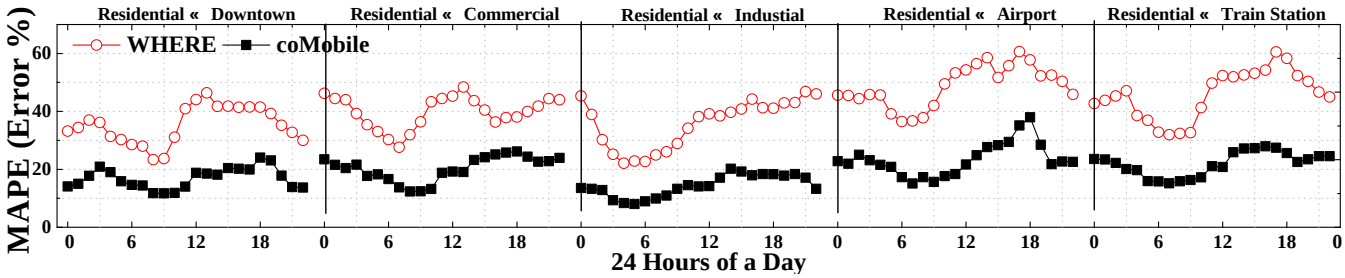

Fig 13. MAPE under One Hour Slot for 24 Hours of a day

we have been storing a large amount of data, requiring significant efforts for the daily management. We utilize a 34 TB Hadoop Distributed File System (HDFS) on a cluster consisting of 11 nodes, and each of them is equipped with 32 cores and 32 GB RAM. For daily management, we use the MapReduce based Pig and Hive. Because of the extremely large size of our data, we have been finding several kinds of errant data, *e.g.*, duplicated data, missing data, and data with logical errors. To address these issues, we conduct a detailed cleaning process to filter out errant data.

For real-world implementation, we have to decide the spatiotemporal partition for the mobility graph, which decides the spatiotemporal granularity of our model. For example, we have more than 110 thousand road segments, 496 urban regions, and 10 urban districts in Shenzhen, and we can capture the mobility with one of those three spatial partitions for every 15 mins, 30 mins, 60 mins, or even longer. Due to the spatial resolutions of our data (especially for bus, subway, and cellphones), we use a urban-region partition proposed by Shenzhen government as our spatial partition, which is given by Fig. 11. Different colors indicate different population density. Based on this partition, we implement our multi-view mobility modeling technique coMobile based on two views. A human mobility graph obtained by coMobile for major urban areas during the evening rush hour at region levels is given by the left of Fig. 12.

6.2 coMobile Evaluation

6.2.1 Evaluation Methodology

Based on our implementation, we compare coMobile with a single-view human mobility model called WHERE. WHERE [9] is a model driven by cellphone data, and it is based on spatial and temporal probability distributions of human mobility and produces synthetic cellphone records as the inferred mobility. We compare these two models with the inferred ground truth. In this project, to infer the ground truth, we introduce another new cellphone related dataset for the evaluation. Different from regular CDR data, this dataset logs locations of all cellphone users at cell tower levels for every 15 mins even without activities. We use the mobility graph obtained from this dataset as the ground truth, which is given in Fig. 12. By a visual comparison, we found that we underestimate the mobility at residential areas and overestimate the mobility at downtown areas.

We utilize three months of data to evaluate these two models. We use Mean Average Percent Error (MAPE) in a time slot as a metric to test those two models $\text{MAPE} = \frac{100}{n}\sum_{i=1}^{n}\frac{|\bar{\mathbf{T}}_i - \mathbf{T}_i|}{\bar{\mathbf{T}}_i}$, where $n = 496 \times 496 = 246016$ is the total number of region pairs, i.e., the total number of edges in a mobility graph; $\mathbf{T}_i$ is the inferred mobility between a region pair i; $\bar{\mathbf{T}}_i$ is the ground truth of the mobility between a region pair i. An accurate model yields a small MAPE, and *vise versa*. We use 90 days of data, leading to 90 experiments. The average results were reported.

We investigate the impact of different contexts by adjusting three model parameters, i.e., λ_1, λ_2, and λ_3, which control contributions of different contexts in our tensor decomposition with Eq.(1). The default setting is $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \frac{1}{4}$ where we consider all contexts and the regularization term equally. Further, we investigate the impact of historical cellphone data on the model performance in terms of extracting correlated contexts.

6.2.2 Evaluation Results

We compare two models' inferring accuracy in terms of MAPE values by (i) a low level comparison on five particular region pairs, (ii) a high level comparison on all 246016 region pairs, (iii) different lengths of slots, and (iv) different amount of historical data.

Fig. 13 plots the MAPE under one hour slots with the two-way mobility between a residential region and five other regions. We found that coMobile outperforms WHERE in general. This is because WHERE only uses the cellphone data to model the human mobility from the cellphone view

alone; whereas coMobile uses two views to model the human mobility, which leads to better performance. We also found that the performance gain between coMobile and WHERE is lower during the rush hour. One of the possible explanations is that the repeatable mobility patterns are higher during the rush hour, so all models have better performance. Comparing the five region pairs, we found that for the commuting region pairs (e.g., between the residential region and the industrial, commercial or downtown regions), all models have better performance than the region pairs on which the residents go for travel (i.e., between the residential region to the airport or train station regions). This is due to the fact that repeatable pattern for travel is limited.

Fig. 14 gives the MAPE on all region pairs under one hour slots. We found that all two models have higher MAPE than the MAPE we observed in Fig. 13. This is because the urban mobility is highly dynamic between various regions pairs, many region pairs have very limited mobility, which leads to high MAPE. But we also found that relative performance between these two models is the same as in Fig. 13. coMobile is better than WHERE, which shows the advantage of using multi-view learning to model the human mobility. coMobile outperforms WHERE by 51% in terms of MAPE, resulting from its multi-view learning from both cellphone data and transportation data.

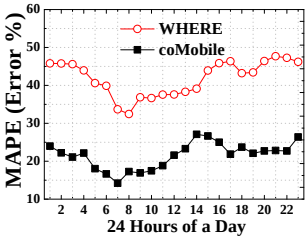

Fig 14. Hourly MAPE

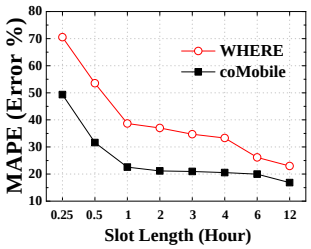

Fig 15. Effects of Lengths

Fig. 15 plots the MAPE of coMobile and WHERE with different slot lengths from 15 mins to 12 hours. Basically, the MAPE of both models reduces with the increase of the modeling lengths. This is because the mobility in a longer time slot is much more stable. coMobile significantly outperforms WHERE when the slot length is short. This is because the transportation data can capture lots of mobility during a short time period. We notice that the slot length becomes longer than 6 hours, both coMobile and WHERE have the similar performance, because in a long time slot, the cellphone data alone is capable of inferring mobility.

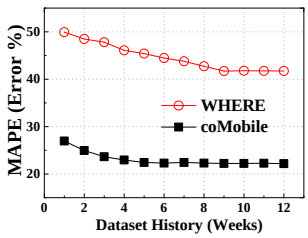

Fig 16. Historical Data

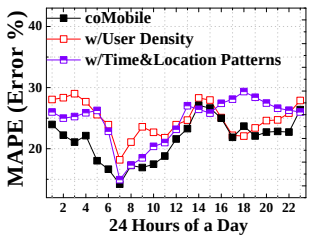

Fig 17. Contexts

Fig. 16 shows how much historical information is necessary for coMobile and WHERE. As expected, the longer the time, the lower the MAPE for both models, the better the performance. But a too long history does not help much, especially for coMobile whose MAPE became stable when the historical data are longer than 4 weeks. It shows that coMobile does not reply on long-term historical cellphone data, thanks to the transportation view. But WHERE needs a longer historical period of data, i.e., 9 weeks, before its MAPE becomes stable.

Fig. 17 shows the impact of two contexts, i.e., cellphone user densities and calling time&location patterns as introduced in Section 4.1.2. In particular, we set $\lambda_2 = \lambda_3 = 0$ and $\lambda_1 = \lambda_4 = \frac{1}{2}$ to obtain a model called coMobile w/ User Density, which only considers the cellphone user density as a context. Similarly, we set $\lambda_1 = 0$ and $\lambda_2 = \lambda_3 = \lambda_4 = \frac{1}{3}$ to obtain a model called coMobile w/ Time&Location Patterns, which only considers time&location patterns as contexts. We compare them with coMobile, which considers all contexts. In generally, coMobile outperforms the other two models. We found that for the early morning, considering time&location patterns is better than considering user density; while for the late night, considering user density is better than considering time&location patterns. Also, during some slots in the afternoon or evening, e.g., 14:00, 15:00 and 18:00, it leads to better performance if we do not consider certain contexts.

In short, we have the following observations. (i) As in Fig. 13, the accuracy of human mobility modeling is highly depended on both locations and time of day. (ii) As in Fig. 14, both models have better performance in the morning rush hour in general due to the predicability of morning commutes, and coMobile outperforms WHERE during all times. (iii) As in Fig. 15, the length of slots has significant impacts on performance of all models. (iv) As in Fig. 16, how much historical data to be used by coMobile does not significantly affect the performance of coMobile. (v) As in Fig. 17, the same contexts have different effects according to the time of day, but considering them together leads to better average performance.

7 Related Work

Modeling the human mobility in urban scales is crucial for mobile applications, urban planning and social networks [20]. However, almost all existing models are driven by single views. We made the first attempt to model the human mobility with multi-source data [18], but our previous work was to use transportation data to adjust the modeling process based on cellphone data, and did not treat these two kinds of data equally as two views. As follows, we summarize the related work by different views.

Cellphone View: Modeling from the cellphone view based on call detail records (CDR) is the most common method, e.g., modeling how residents move around the cities [9]; estimating cellphone users' travel range [10]; predicting where cellphone users will travel next [5]. However, the models from cellphone views are mostly biased against a certain group of residents, leading to inaccurate analyses. To our knowledge, we are the first to combine data from more than one carrier to model the human mobility.

Transportation View: Transportation data are another

important data source for human mobility, e.g., bus data [3], subway data [14], taxicab data [6], and private vehicle data [7]. However, the models driven by data from one kind of transportation are mostly biased against the passengers using other transportation. To our knowledge, there is no model driven by more than one transportation mode, and we are the first to combine data from three kinds of transportation for mobility modeling.

Other Views: Other data generated by urban residents have also been used to study human mobility, e.g., social networks or mobile *ad hoc* networks, i.e., with check-in data [4] and proximity data [1]. However, the number of residents captured by these views is often extremely limited compared to the cellphone data and transportation data, which leads to a bias that cannot be quantified.

In summary, almost all human mobility modeling is based on single views, which are often incomplete in terms of capturing the human mobility at urban scale in real time. Such a shortcoming motivates us to take a multi-view approach, which uses incomplete yet complementary views to model the human mobility.

8 Discussion

We provide some discussion about coMobile as follows.

Privacy Protections. While the data for the human mobility study have the potential for great social benefits, we have to protect the privacy of the residents involved for wider applications. We took two active steps for privacy protections. (i) Anonymization: All data analyzed are anonymized by the service providers who were not involved in this project, and all identifiable IDs, such as SIM card IDs, are replaced by a serial identifier during the analyses. (ii) Aggregation: the mobility patterns obtained by coMobile are given at aggregated results with a mobility graph in a large spatiotemporal partition. We do not focus on individual residents during the analyses.

Public Data Access. Accessing empirical datasets is vital to the geographic information system research, but such datasets are usually not available for the fellow researchers due to the privacy issues. As an initiative step, the partial aggregated data used in this work have been made for public access in the website of Transport Committee of Shenzhen Municipality [17]. Most importantly, we release the first big urban data [18], which include the large-scale Shenzhen data including taxi, bus, subway, smartcard, and cellphone data. This is the first time that such comprehensive urban data are released for the benefit of research community.

9 Conclusion

In this work, we design, implement and evaluate a human mobile modeling technique called coMobile based on context-aware tensor decomposition and iterative multi-view learning. Our endeavors offer a few valuable insights: (i) the human mobility modeling based on single-view data introduces biases, which can be partially addressed by using historical data; (ii) to model human mobility, every view itself is incomplete but they are often complementary to each other, and thus it is essential to model the completeness degree of a view before inferring the mobility; (iii) multi-view learning for human mobility requires an iterative optimization process to improve the accuracy of modeling, and thus how to select an objective function and constraint function to ensure the convergence is essential for real-time applications.

10 Acknowledgements

This work was supported in part by the US NSF Grants CNS-1239226, CNS-1446640 and China 973 Program No. 2015CB352400.

11 References

[1] BACKSTROM, L., SUN, E., AND MARLOW, C. Find me if you can: Improving geographical prediction with social and spatial proximity. In *WWW '10*.

[2] BERTSEKAS, D. P. Non-linear programming. In *Athena Scientific* (1999).

[3] BHATTACHARYA, S., PHITHAKKITNUKOON, S., NURMI, P., KLAMI, A., VELOSO, M., AND BENTO, C. Gaussian process-based predictive modeling for bus ridership. UbiComp '13.

[4] CHO, E., MYERS, S. A., AND LESKOVEC, J. Friendship and mobility: User movement in location-based social networks. KDD '11.

[5] DUFKOVÁ, K., LE BOUDEC, J.-Y., KENCL, L., AND BJELICA, M. Predicting user-cell association in cellular networks. MELT'09.

[6] GANTI, R., SRIVATSA, M., RANGANATHAN, A., AND HAN, J. Inferring human mobility patterns from taxicab traces. UbiComp '13.

[7] GIANNOTTI, F., NANNI, M., PEDRESCHI, D., PINELLI, F., RENSO, C., RINZIVILLO, S., AND TRASARTI, R. Unveiling the complexity of human mobility by querying and mining massive trajectory data. *The VLDB Journal*.

[8] GONZALEZ, M. C., HIDALGO, C. A., AND BARABASI, A.-L. Understanding individual human mobility patterns. Nature.

[9] ISAACMAN, S., BECKER, R., CÁCERES, R., MARTONOSI, M., ROWLAND, J., VARSHAVSKY, A., AND WILLINGER, W. Human mobility modeling at metropolitan scales. MobiSys '12.

[10] ISAACMAN, S., BECKER, R. A., CÁCERES, R., KOBOUROV, S. G., MARTONOSI, M., ROWLAND, J., AND VARSHAVSKY, A. Ranges of human mobility in los angeles and new york. In *PerCom Workshops* (2011).

[11] JIANG, S., FIORE, G. A., YANG, Y., FERREIRA, JR., J., FRAZZOLI, E., AND GONZÁLEZ, M. C. A review of urban computing for mobile phone traces: Current methods, challenges and opportunities. UrbComp '13.

[12] KARATZOGLOU, A., AMATRIAIN, X., BALTRUNAS, L., AND OLIVER, N. Multiverse recommendation: N-dimensional tensor factorization for context-aware collaborative filtering. In *Proceedings of the Fourth ACM Conference on Recommender Systems* (New York, NY, USA, 2010), RecSys '10, ACM, pp. 79–86.

[13] KOLDA, T. G., AND BADER, B. W. Tensor decompositions and applications. In *SIAM Review 51* (2009).

[14] LATHIA, N., AND CAPRA, L. How smart is your smartcard?: Measuring travel behaviours, perceptions, and incentives. UbiComp '11.

[15] SIMINI, F., GONZLEZ, M. C., MARITAN, A., AND BARABSI, A.-L. A universal model for mobility and migration patterns. Nature.

[16] SUN, L., LEE, D.-H., ERATH, A., AND HUANG, X. Using smart card data to extract passenger's spatio-temporal density and train's trajectory of mrt system. UrbComp '12.

[17] TRANSPORT COMMITTEE OF SHENZHEN. Statistical data for transportation in shenzhen. http://www.sz.gov.cn/jtj/tjsj/zxtjxx/.

[18] ZHANG, D., HUANG, J., LI, Y., ZHANG, F., XU, C., AND HE, T. Exploring human mobility with multi-source data at extremely large metropolitan scales. MobiCom '14, pp. 201–212.

[19] ZHENG, Y. Trajectory data mining: An overview. *ACM Trans. Intell. Syst. Technol. 6*, 3 (May 2015), 29:1–29:41.

[20] ZHENG, Y., CAPRA, L., WOLFSON, O., AND YANG, H. Urban computing: Concepts, methodologies, and applications. *ACM Trans. Intell. Syst. Technol. 5*, 3 (Sept. 2014), 38:1–38:55.

[21] ZHENG, Y., AND XIE, X. Learning travel recommendations from user-generated gps traces. *ACM Trans. Intell. Syst. Technol.*.