Information Fusion 23 (2015) 25–32

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Human mobility synthesis using matrix and tensor factorizations

Dezhong Yao [a], Chen Yu [a,*], Hai Jin [a], Qiang Ding [b]

[a] Services Computing Technology and System Lab, Cluster and Grid Computing Lab, School of Computer Science and Technology, Huazhong University of Science and Technology, 1037 Luoyu Rd, Wuhan 430074, China
[b] Huawei Technologies Co., Ltd., Beijing 10084, China

ARTICLE INFO

Article history:
Available online 12 June 2014

Keywords:
Multi-target prediction
Mobility model
Tensor decomposition
Human mobility

ABSTRACT

Human mobility prediction is of great advantage in route planning and schedule management. However, mobility data is a high-dimensional dataset in which multi-context prediction is difficult in a single model. Mobility data can usually be expressed as a home event, a work event, a shopping event and a traveling event. Previous works have only been able to learn and predict one type of mobility event and then integrate them. As the tensor model has a strong ability to describe high-dimensional information, we propose an algorithm to predict human mobility in tensors of location context data. Using the tensor decomposition method, we extract human mobility patterns with multiple expressions and then synthesize the future mobility event based on mobility patterns. The experiment is based on real-world location data and the results show that the tensor decomposition method has the highest accuracy in terms of prediction error among the three methods. The results also prove the feasibility of our multi-context prediction model.

1. Introduction

Understanding and predicting human mobility from complex contextual data is an emerging domain relevant to many applications. Mobility features are useful for social science studies and developing context-awareness tools. For example, discovering and predicting traffic jams in an urban city environment is based on knowing how frequently populations move about in their daily lives. Similarly, understanding the spread of a disease is one of the ways that humans themselves move and interact [1]. Other useful applications, like urban planning, routine planning, management scheduling and determining the behavior and habits of individuals are all based on human mobility.

Previous studies have shown that human movement is predictable to a certain extent on different geographic scales [2]. On the basis of this theory, we can predict the next place that will be visited by a user in the next few hours. Some studies are based on probabilistic models, such as Markov models [3–5], Bayes models [6,7], and pattern mining methods [8] have been proposed to continuously predict the next place of individuals' mobility. Many of these prediction algorithms estimate users' future locations using a sequence of history locations visited by the users in the past. However, such models cannot explain the structure of a user's mobility features and are week in predicting long-term mobility [9]. Other studies [9,10] can only predict one kind of location-based event and then calculate which event has the highest probability of occurring at a certain time.

In this paper, we will introduce a tensor-based mobility modeling method and predict future events using a tensor decomposition method. We have three goals to achieve in this research. The first is to use a tensor to model high-dimensional location-based mobility events. The second is to analyze the structure of a user's mobility on different weekdays. Once our description modeling has been examined, our third goal is to predict multiple location events at the same time. Our algorithm has strong ability in dealing with multiple-context prediction problems. It first learns the structure of a user's mobility. Given the beginning location event of a day, our method can synthesize the following location events of the whole day. For long-term prediction, our method can also synthesize a whole day's location events based the features of that day.

We address the mobility structure learning procedure as a component dimensionality reduction process. Some studies [9,10] use the Principal Components Analysis (PCA) method to identify structure in routine. PCA [11] is a good method for learning the most important events (components) from a history record. The tensor-based description has better performance in anglicizing multiple location-based events and it can also maintain the inner relationships among those multiple location-based events. The

* Corresponding author. Tel.: +86 027 8754 3529; fax: +86 027 8755 7354.
E-mail addresses: dyao@hust.edu.cn (D. Yao), yuchen@hust.edu.cn (C. Yu), hjin@hust.edu.cn (H. Jin), q.ding@huawei.com (Q. Ding).

http://dx.doi.org/10.1016/j.inffus.2014.05.005
1566-2535/

prediction method is a tensor data reconstructing process. After learning the mobility structure of a human, we can synthesize a whole day's location events based on the features of that day.

In summary, our contributions are:

- We propose a new mobility prediction algorithm using a tensor model to directly predict mobility events in a multi-type context environment.
- We first construct the tensor model to describe the multiple types of mobility event of a human.
- We propose three different prediction algorithms to compute future locations based on the tensor description model and we compare the performance of each algorithm.

The remainder of this paper is organized as follows. We first review related works on mobility prediction in Section 2. Then we briefly introduce background knowledge about tensor in Section 3. In Section 4, we give the tensor-based location events description method. In Section 5, we present three tensor-based mobility-event prediction models and find an efficient optimization method. The accuracy comparison results of each model are described in Section 6. Finally, we give our conclusions and future plans in Section 7.

2. Related work

Many studies have been undertaken to predict future locations and activities. However, most of the current models can only learn and predict one type of context pattern at a time and mix the weekday pattern and holiday pattern together during learning. As human daily activities are complex data with multiple context information, we cannot learn the patterns separately. Our work is primarily motivated to overcoming these limitations.

Eagle et al. [12] studied the Reality Mining dataset to discovery humans' daily routines. Their basic idea was to use the PCA method on one type of user's mobile data to obtain the top eigenvector which presents the routines of that user. The mobile phones carried by the participants were located by the cell tower. This method models the event of location as an independent context and predicts those contexts separately. Human contexts are not independent of each other and they have influence with each other in prediction. Some works prove that it is more accurate to model the original data together using the tensor model [13]. Farrahi et al. [14,15] applied LDA (Latent Dirichlet Allocation) [16] to the same Reality Mining dataset to discover latent routine topics in people's daily lives. To adapt the framework of LDA, they coupled location and time together, termed activity. However, this choice eliminates the chance of explicitly modeling users' latent behavior patterns, the bridge between spatial and temporal features, making it hard to interpret the result and predict users' future locations using only time information. Another drawback is that this model also requires the pre-labeling of cell tower semantics, which is not practical.

Much effort on descriptive models has been motivated by the desire to extract patterns of human activities, and subsequently leverage them in simulations that accurately mimic the observed general statistics of real trajectories [17]. More recently, Scott et al. used both temporal and spatial features for occupancy prediction [18]. However, all these works focus on aggregate activity and do not address the problem of location prediction, which is the primary focus of this paper. Jeung et al. [8] evaluated a hybrid location model that invokes two different prediction algorithms, one for queries that are temporally close, and the other for predictions further into the future. However, their approach requires selecting a large number of parameters and metrics. Additionally, Jeung et al. experimented with mostly synthetic data. By contrast, we present a unified mobility model and evaluate it on a real-world mobile phone dataset.

More recently, Sadilek and Krumm [9] proposed a probabilistic framework to predict long-term human mobility. They used the PCA model with holiday features. They achieved good performance in predicting location in the next month or the next year. This is because they distinguished the workdays and holidays very well with more features. However, their model still analyzes the location context separately and their model is not suitable for short-term prediction environments.

In other areas, the tensor decomposition method is a very useful technique and it has been used in many applications [19,20]. In recent works [21,22], this method has been used to analyze social network links. Rendle and Schmidt-Thieme [23] built a personalized tag recommendation system using tensor decomposition method. It is very powerful for learning the hidden structure of data, so we used to analyze human mobility data.

There are already some works [24,25] on building models of mobility, both descriptive and predictive. But there is a gap in modeling and prediction multiple-context mobility, which is our contribution, seeing in Table 1. As pointed out in [26,27], assuming relatedness in different type of events and simply learning them separately is detrimental. So it is important to build a model which will generally benefit related events and keep performance when those events are unrelated. There are lots of work on modeling and prediction methods for single-context. The problems on multiple-context modeling and prediction are not fully considered yet. We are the first to use the tensor model to describe the multiple contexts information and propose three prediction algorithms to compute future locations based on the tensor model.

3. Tensor techniques

A *tensor* is a high order generalization of a vector (first order tensor) and a matrix (second order tensor). An N-order tensor is defined as $\chi \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$. In our model, we use a third-order tensor: $I \times J \times K$ to describe the mobility data. The entry (i, j, k) of a 3rd-order tensor is the index of a value in the tensor space. Indexing from an arbitrary direction, we have tube fibers as shown in Fig. 1. A tube fiber at position $i,\ j$ is denoted as $\chi_{ij:}$. When the two modes are fixed, we have a slice tensor, which stands for a two-dimensional section of a tensor. The slice can be denoted as $\chi_{::k}$, which is illustrated in Fig. 1.

3.1. Tucker decomposition

Tensor decomposition techniques [22] can be applied to analyze multi-dimensional data. There are two main decomposition models: the Tucker family of decomposition [28] and the CANDECOMP/PARAFAC (CP) decomposition [29]. Tucker factorizes a tensor into the product of a core tensor $\mathcal{Z}$ and matrices (*e.g.*, $U_1,\ U_2,\ U_3$) in each dimension. The Tucker decomposition method compresses the original data into the orthogonal space, which can maintain the data's inner structure. However, as CP decomposition cannot maintain the inner relationship of the original data, we chose the Tucker decomposition method in this paper. Given a 3rd-order tensor $\chi \in \mathbb{R}^{I \times J \times K}$ and using the Tucker decomposition method, the tensor can be decomposed as:

Table 1
The context of our contributions.

	Single-context	Multiple-context
Descriptive	Previous work	Proposed work
Predictive	Previous work	Proposed work