

Geospatial Area Embedding Based on the Movement Purpose Hypothesis Using Large-Scale Mobility Data from Smart Card

Masanao Ochi¹, Yuko Nakashio², Matthew Ruttley³, Junichiro Mori¹, Ichiro Sakata^{1,2}

¹Graduate School of Engineering, The University of Tokyo, Tokyo, Japan

²Faculty of Engineering, The University of Tokyo, Tokyo, Japan

³Hearst, New York, USA

Email: masanao.oochi@gmail.com

How to cite this paper: Ochi, M., Nakashio, Y., Ruttley, M., Mori, J. and Sakata, I. (2016) Geospatial Area Embedding Based on the Movement Purpose Hypothesis Using Large-Scale Mobility Data from Smart Card. *Int. J. Communications, Network and System Sciences*, 9, 519-534.

<http://dx.doi.org/10.4236/ijcns.2016.911041>

Received: September 26, 2016

Accepted: November 21, 2016

Published: November 24, 2016

Copyright © 2016 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution-NonCommercial International License (CC BY-NC 4.0).

<http://creativecommons.org/licenses/by-nc/4.0/>



Open Access

Abstract

With the deployment of modern infrastructure for public transportation, several studies have analyzed movement patterns of people using smart card data and have characterized different areas. In this paper, we propose the “movement purpose hypothesis” that each movement occurs from two causes: where the person is and what the person wants to do at a given moment. We formulate this hypothesis to a synthesis model in which two network graphs generate a movement network graph. Then we develop two novel-embedding models to assess the hypothesis, and demonstrate that the models obtain a vector representation of a geospatial area using movement patterns of people from large-scale smart card data. We conducted an experiment using smart card data for a large network of railroads in the Kansai region of Japan. We obtained a vector representation of each railroad station and each purpose using the developed embedding models. Results show that network embedding methods are suitable for a large-scale movement of data, and the developed models perform better than existing embedding methods in the task of multi-label classification for train stations on the purpose of use data set. Our proposed models can contribute to the prediction of people flows by discovering underlying representations of geospatial areas from mobility data.

Keywords

Network Embedding, Auto Fare Collection, Geographic Information System, Trajectory Data Mining, Spatial Databases

1. Introduction

As location-based sensor devices and networks have been widely spread, a large amount of mobility data of users, which can be potentially used for several research purposes, has been accumulated [1] [2]. In addition to such sensor devices, the deployment of recent infrastructure for public transit such as automated fare collection (AFC) systems with smart cards has supported the collection of large volumes of mobility data including people's activities with detailed time and space information [3].

Researchers have used such large amount of mobility data for the purpose of location-based recommendation such as personalization point of interest (POI) [4]. More recently, several studies have used the mobility data for regional development [5], urban planning [6], and policymaking [7]. One of key questions in those studies is how to model and predict people flow in a specific area where the mobility data have been collected.

Modeling and predicting people flow in a specific area results in understanding the characteristics or roles of the area by combining activity patterns of people with external information about the area [8]. Recent studies have analyzed transition patterns of people from one area to another using smart card data and characterized the areas or identified the segmentation of the areas [9] [10]. These studies solely assume that an area falls into some pre-defined demographics based on people flow in the area. However, if we regard massive transition patterns of people on an area as the context of its area, we can notice that the characteristics or roles of the area are dynamically changing according to its context of how people move on the area and for what purpose people visit the area. In other words, if two areas have similar characteristics or roles, they should have common underlying representation of areas that can be defined by such context. If we can obtain such latent representation of areas, it contributes to modeling and predicting people flow with massive mobility data more effectively and precisely.

The basic notion of representation learning [11] is that two entities are semantically similar if they are sharing common contexts; this is known as a distributional hypothesis in linguistics, which states that words that occur in similar contexts tend to have similar meanings [12]. That idea of representation learning has been recently expanded to a network embedding method [13] [14] [15] that tries to solve the problem of embedding networks into low-dimensional vector spaces by assuming that two nodes are similar if they are closely connected in a network. In the case of finding latent representation of geographical areas, if we consider areas as nodes and transition patterns of people between areas as links, we can formalize the problem of finding such representation as an extension of studying the embedding of a network. Intuitively, transition patterns of people in a business district are different from those in a residential district. Therefore, we can distinguish those different types of geospatial areas by embedding a people transition network in our low-dimensional vector spaces.

In this research, we aim to find latent representation of geographical areas using the representation learning technique. Such representation can be used for urban planning and regional development by revealing potential roles of geographical areas and their relations, which cannot be always observed from superficial information in mobility

data. We can employ the notion of existing network embedding methods to find such representation from massive people flow data. However, one cannot simply apply existing embedding methods to our problem of embedding geospatial areas. For people movement in a large network of transportation systems such as railroads, several geographical constraints exist on their movement. For example, I, who live in Tokyo, do not go to Osaka to shop for daily necessities; I always buy daily necessities nearby and I don't go all the way to far away with trivial things. Therefore we can assume that people usually tend to minimize their movements depending on their activities, given some available means of transportation at their current location. We define such geographical constraints as the “**movement purpose hypothesis.**” If we consider geospatial areas as a network connected with links of people with movement patterns between areas, and if we then try to embed the network in a low-dimensional vector space to obtain representations of areas, we have to consider such geographical constraints on movement of people in the real world.

In this paper, we propose a novel embedding method to obtain a vector representation of a geospatial area using movement patterns of people from large-scale smart card data. Our proposed method consists of two embedding models, which are the “**concatenating model**” and the “**internally dividing model,**” based on the movement purpose hypothesis. We conducted an experiment using massive smart card data in a large network of railroads in the Kansai region of Japan. We obtained a vector representation of each railroad station using the proposed embedding models and evaluate it in the task of multi-label classification for railroad stations. We demonstrate that our proposed models work well on actual massive mobility data from smart cards of the railroads. Our proposed method can identify stations in a large network of railroads, which are geographically distributed but share similar characteristics or roles in the region. Therefore, we can support a city planner, a marketer, and a policy maker to design their strategies or implement their policies for regional development by providing potential characteristics of geographical areas and their relations.

Our contributions in this paper are four-fold:

- 1) We propose the movement purpose hypothesis and develop novel-embedding models to obtain a vector representation of a geospatial area using movement patterns of people.
- 2) We demonstrate that our developed models work well using actual large-scale mobility data from smart cards of the railroads in Japan.
- 3) We also demonstrate that our proposed models can successfully identify stations, which are geographically distributed but share similar characteristics or roles.
- 4) According to the results of parameter estimation of our proposed embedding model, we find that the purpose of visit for a station is 1.1 times more important than the geographical distance between stations for people movement in a large network of railroads.

2. Related Works

Our work is mainly related to mobility data analysis and network-embedding learning.

In this section, we discuss our research position and novelty in relation to existing related works.

2.1. Modeling Characteristics of Geographical Areas Using Mobility Data

Recent sensor networks and infrastructures for public transit such as automated fare collection (AFC) systems with smart cards have supported the collection of large volumes of mobility data including people's activities with detailed time and space information. In particular, mobility data from the AFC systems are currently used for several purposes such as visualization [16], disaster prevention [17], and service management [18].

Moreover, aiming at several applications for location-based services including a personalized point of interest (POI) recommendation for users [4], regional development [5], urban planning [6], and policymaking [7], several studies have addressed a question of how to model people flow in a specific area and understand the characteristics of the area with such large amount of mobility data.

Recent studies have analyzed movement patterns of people from one area to another using smart card data and have characterized the areas or have enabled segmentation of the areas [9] [19]. These studies solely assume that an area falls into some pre-defined demographics based on people flow in the area. However, if we regard massive transition patterns of people on an area as the context of its area, we can notice that the characteristics or roles of the area are dynamically changing according to its context of how people move on the area and for what purpose people visit the area. In this paper, we aim at obtaining common underlying representation of areas that can be defined by such context using an embedding method. When understanding the characteristics or roles of the area, previous studies require pre-defined demographics of areas like shopping area or business district [6] [20]. On the other hand, our proposed models can learn such information from a few tagged areas using a semi-supervised learning.

2.2. Embedding of Network Data

In this research, we aim to find latent representation of geographical areas using the representation learning technique. We can employ the notion of existing network embedding methods to find such representation from massive people flow data. The network embedding method comes from graph theory and linguistics word embedding methods. In the context of the graph theory, adjacency matrix factorization techniques like singular value decomposition (SVD) and non-negative matrix factorization (NMF) are the prototype [21]. On the other hand, word embedding methods have been recently advanced. The basic notion of word embedding is that two entities are semantically similar if they share common contexts; this is known as distributional hypothesis in linguistics, which states that words that occur in similar contexts tend to have similar meanings [12] [22]. Some works tried embedding network graph structure directly [13]

[14] [15]. These network-embedding methods are useful in many tasks such as visualization, node classification, and link prediction [23].

Network embedding has been further developed for time series analysis [24] and for heterogeneous network [25]. The predictive text embedding (PTE) method for heterogeneous network embedding, which can embed words, documents, and labels to a low-dimensional vector space. The PTE method embeds these three different heterogeneous networks to a same vector space and obtain vectors with a semi-supervised learning style. Our proposed models also use three heterogeneous networks, but it embeds them to two different vector spaces, the geospatial vector space and the role vector space. We describe this in more detail in the following section.

3. Method

This section first describes the “**Movement Purpose Hypothesis**” which the people flow is caused from geolocation and purpose. Next we explain how the network form from massive people flow and the necessity of label propagation on the network. We extend propagating labeled network embedding model for massive people flow data. Finally, we propose models based on the hypothesis and explain precisely.

3.1. Movement Purpose Hypothesis

We propose the “movement purpose hypothesis,” as shown in **Figure 1** for people flow data such as the GPS data, the cell phone base station data, the train travel data, etc. We apply the train travel records as the people flow data on the Japan Kansai region extracted from the smart card system in this paper. So, we represent an area as a station in the figure. This hypothesis presumes that a person moves somewhere to accomplish a purpose that the person cannot accomplish there. In other words, the movements of people (People flow) are represented as the sum of the geographical proximities between areas (Geographical constraints) and the role of the area (Purpose proximity). This model describes that people move to a nearby location, which means a destination to realize their purposes, from the present location. As accumulating thus people’s location and desire, the people flow network is shaped (**Figure 1** right). On the other hand, we propose that the people flow regards as the sum of the amount of geolocation data and the amount of purpose data. In **Figure 1**, we illustrate that two networks (the geographical constraints network and the purpose network) generate the massive people flow network. And we think that thus three networks’ relationship depends on the distance on the latent vector representation.

There are three graphs that do not mutually share their vectors, the people flow graph (G_{ss}), the geographical constraints graph (G_{sc}), and the purpose proximity graph (G_{sr}). More specifically, a vector representation in each graph is the following: \vec{u}_i^{ss} for all vertices $v_i \in G_{ss}$, \vec{u}_i^{sc} for all vertices $v_i \in G_{sc}$, and \vec{u}_i^{sr} for all vertices $v_i \in G_{sr}$. The model shown in the **Figure 1** hypothesis leads to the following equation, which is established among vectors.

$$\vec{u}_i^{ss} = \vec{u}_i^{sc} + \vec{u}_i^{sr} \quad (1)$$

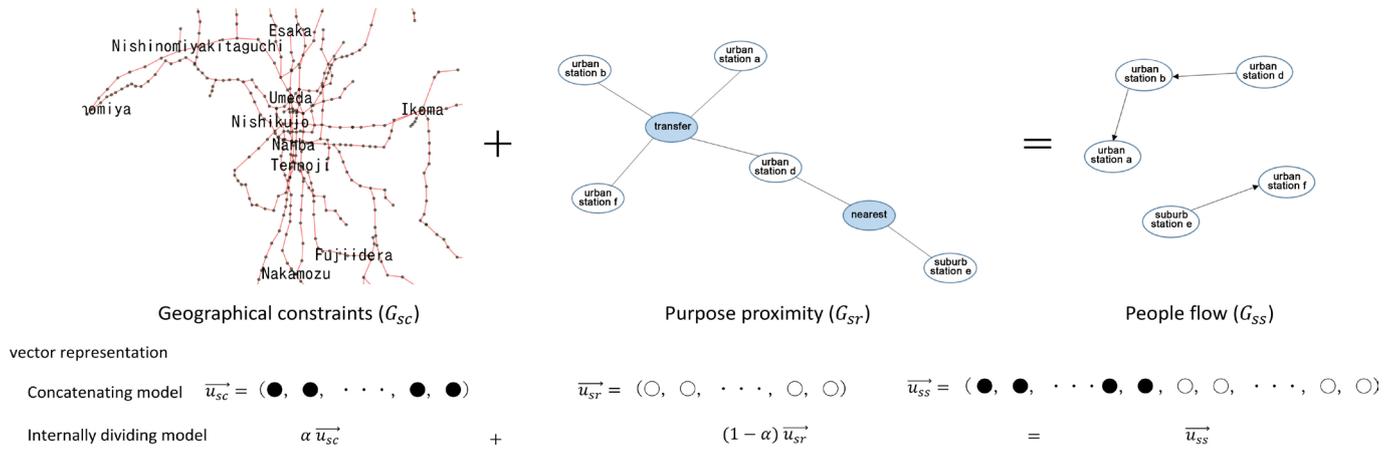


Figure 1. Schematic showing the “movement purpose hypothesis” and proposed models.

We interpret this equation as two types: “**concatenating model**” and “**internally dividing model.**” For the “concatenating model,” we interpret the operator “+” as connecting two vectors and producing a new vector with dimensions that are twice as numerous as the number of dimensions of each vector, not that we add each element in the two vectors (Figure 1, concatenating model). Furthermore, for the “internally dividing model,” we interpret the Equation (1) as the people flow graph node (\vec{u}_i^{ss}) locates between the geographical constraints graph node (\vec{u}_i^{sc}) and the purpose proximity graph node (\vec{u}_i^{sr}) (Figure 1, internally dividing model). We explain these two models more in the following subsections.

3.2. Concatenating Model

Based on the concatenating model, vector representations are acquired by the learning algorithm shown in Table 1. This algorithm needs the geographical constraints network (G_{sc}), the purpose proximity network (G_{sr}), the people flow network (G_{ss}), the number of sampling (T), the initial learning rate (ρ_0), the number of negative sampling (K), and the dimension of the embedding (d) as input. We apply the network embedding model called the “LINE (2nd) model” proposed by Tang *et al.* [15]. This model approximates second-order proximity between two vertices, optimizing each representation vector. The objective function is as follows:

$$O = - \sum_{(i,j) \in E} w_{ij} \ln p(v_j | v_i) \tag{2}$$

In this equation, w_{ij} indicates the empirical edge weight from the vertex v_i to the vertex v_j . $p(v_j | v_i)$ which is the transition probability from v_i to v_j is estimated using the embedding vector \vec{u}_i of the vertex v_i and the context vector \vec{u}'_j of the vertex v_j as following:

$$p(v_j | v_i) = \frac{\exp(\vec{u}'_j{}^T \cdot \vec{u}_i)}{\sum_{k' \in |V|} \exp(\vec{u}'_{k'}{}^T \cdot \vec{u}_i)} \tag{3}$$

We set this objective function for three networks individually and derive update

Table 1. Learning algorithm of the concatenating model.

1:	Input: $G_{sc}, G_{sr}, G_{ss}, T, \rho_0, K, d$.
2:	Output: $\vec{u}^{sc}, \vec{u}^{sr}, \vec{u}^{ss}$.
3:	Initialize each vector $\vec{u}^{sc}, \vec{u}^{sr}, \vec{u}^{ss}, \vec{u}'^{sc}, \vec{u}'^{sr}, \vec{u}'^{ss}$.
4:	for $t=1$ to T
5:	Sample an edge e_{ij}^{sc} from G_{sc} .
6:	Load \vec{u}_i^{sc} and \vec{u}_j^{sc} from the corresponding part of \vec{u}_i^{ss} and \vec{u}_j^{ss} .
7:	Update \vec{u}_i^{sc} and \vec{u}_j^{sc} using the objective function O_{sc} .
8:	Overwrite the corresponding part of \vec{u}_i^{ss} and \vec{u}_j^{ss} with \vec{u}_i^{sc} and \vec{u}_j^{sc} .
9:	Sample an edge e_{ij}^{sr} from G_{sr} .
10:	Load $\vec{u}_i^{sr}, \vec{u}_j^{sr}$ from the corresponding part of \vec{u}_i^{ss} and \vec{u}_j^{ss} .
11:	Update \vec{u}_i^{sr} and \vec{u}_j^{sr} using the objective function O_{sr} .
12:	Overwrite the corresponding part of $\vec{u}_i^{ss}, \vec{u}_j^{ss}$ with $\vec{u}_i^{sr}, \vec{u}_j^{sr}$.
13:	Sample an edge e_{ij}^{ss} from G_{ss} .
14:	Update \vec{u}_i^{ss} and \vec{u}_j^{ss} using the objective function O_{ss} .
15:	END

equations by differentiating them with respect to the each vertex vector (\vec{u}_i) and the each context vector (\vec{u}_i^l). We acquire vertex vector sequentially (Lines 7, 11 and 14) based on the concatenating model (Lines 8 and 12) using this SGD style learning algorithm.

3.3. Internally Dividing Model

For the “internally dividing model,” the node vector in the people flow graph (\vec{u}_i^{ss}) locates between the geographical constraints graph node vector (\vec{u}_i^{sc}) and the purpose proximity graph node vector (\vec{u}_i^{sr}) as the following equation.

$$\vec{u}_i^{ss} = \alpha \vec{u}_i^{sc} + (1 - \alpha) \vec{u}_i^{sr} \quad (4)$$

This equation models that people decide the destination place in consideration of both the physical place relation and the purpose they want to accomplish there.

We set the objective function as in the Equation (2) for each graph. However, when updating the vector \vec{u}_i^{ss} in G_{ss} , the vector \vec{u}_i^{ss} depends on the vector \vec{u}_i^{sc} and the vector \vec{u}_i^{sr} through the Equation (4). Therefore, it is necessary to derive new update rules for the vectors and the parameter α . For the objective function O_{ss} with the people flow graph (G_{ss}), we carefully differentiate all dependent variables ($\vec{u}^{sc}, \vec{u}^{sr}, \vec{u}^{sc}, \vec{u}^{sr}$) and parameter (α). We can derive the following update rules for the people flow graph. Due to the lack of space, we show the updating equations about vertex vectors and α .

$$\frac{\partial O_{ij}}{\partial \vec{u}_i^{sc}} = (1 - \sigma_{ji}) \alpha \vec{u}_i^{ss} - \sum_{k=1}^K \sigma_{ki} \alpha \vec{u}_i^{ss} \quad (5)$$

$$\frac{\partial O_{ij}}{\partial \vec{u}_i^{sr}} = (1 - \sigma_{ji})(1 - \alpha)\vec{u}_i^{ss} - \sum_{k=1}^K \sigma_{k'i}(1 - \alpha)\vec{u}_i^{ss} \tag{6}$$

$$\begin{aligned} \frac{\partial O_{ij}}{\partial \alpha} = & (1 - \sigma_{ji}) \left\{ (\vec{u}_j^{sc} - \vec{u}_j^{sr})^T \cdot \vec{u}_i^{ss} - \vec{u}_j^{ssT} \cdot (\vec{u}_i^{sc} - \vec{u}_i^{sr}) \right\} \\ & - \sum_{k=1}^K \sigma_{k'i} \left\{ (\vec{u}_k^{sc} - \vec{u}_k^{sr})^T \cdot \vec{u}_i^{ss} - \vec{u}_k^{ssT} \cdot (\vec{u}_i^{sc} - \vec{u}_i^{sr}) \right\} \end{aligned} \tag{7}$$

In these equations, $\sigma_{ji} = 1 / (1 + \exp(\vec{u}_j^{tT} \cdot \vec{u}_i))$. We also apply the joint training style same as the PTE (joint) method [25].

4. Data Description and Input Arrangement

As described in this paper, our proposed models need three networks: the people flow network, the geographical constraints network and the purpose proximity network. To arrange these three networks as input, we apply three datasets for the experiment. In this section, we explain these three datasets and the arrangement.

Getting on and off dataset for the people flow network: This dataset includes massive smart card data for the Japan Kansai region (southwestern half of Japan, including Osaka). This dataset has passenger’s smart card log provided by six railway companies. The providers have anonymized this dataset. The dataset contents mainly consist of six elements: each user of the gender, age, getting on and off date and time, and boarding and destination station. The summary of this dataset is shown in **Table 2**. We make the people flow network using this dataset, which is people getting on and off a train between two stations. This is a directed graph and the weight of each edge is $P(\text{destination station} | \text{boarding station})$. We select only weekday morning movement data from 7 AM to 10 AM in April, 2013 to capture the purpose of going to work in the morning.

Train route map dataset for the geographical constraint: This time, we apply the train network information as geographic proximity information obtained through the Japan train line API¹. We construct the train route map through this. The graph is undirected and the weights of all edges are equal and the route map is shown in **Figure 1** left (geographical constraints).

Purpose of use dataset for the purpose proximity network: This paper is intended to estimate each station’s role. As described herein, we produce a dataset using the results of the person trip survey. In Japan, the Ministry of Land, Infrastructure and Transport takes a nationwide survey through questionnaire from many persons every decade. We apply the 2010 results² to our experiment, which includes how much people come to each station for what purpose. The purposes of the getting off each station are “commuting to work”, “commuting to school”, “going home”, “on business”, and “others”. A summary of this dataset is shown in **Table 3**. We make the station-purpose graph as the purpose proximity network from this dataset which presents a probability distribution of purposes to go to a station. This graph is undirected. When making the

¹<http://www.ekidata.jp/>.

²https://www.kkr.mlit.go.jp/plan/pt/data/pt_h22/index.html.

Table 2. Overview of the getting on and off dataset.

Starting date	April 01, 2013
Ending date	April 30, 2013
Total number of records	68,763,457
Number of unique users	3,679,251
Number of station varieties	672
Number of railway companies	6

Table 3. Summary of the purpose of use dataset.

Purpose	Number of users
Commuting to work	1,278,288
Commuting to school	349,234
Going home	2,192,826
On business	358,891
Others	1,313,767
Total	5,493,006
Number of stations	599

people flow network, we select only weekday morning data. So, we do not use the a “going home” purpose in this dataset and use the remaining four purposes, because we think that most people do not return home in the morning.

5. Experiment and Results

In this section, we evaluate the effectiveness of the developed models for geospatial data. For this purpose, we compare various algorithms and conduct an experiment. As reported below, we describe the results.

5.1. Experimental Procedure

As described in this paper, we conducted a multi-label classification experiment because the purposes of dropping off passengers at a station are plural. To be exact, purposes will differ from person to person. We regard a station as a probability distribution of some purposes and estimate it in the experiment.

The experimental procedure is the following. First, obtaining the vector representation using the listed methods in Section 5.2. Second, the training classifier for each experiment using training labeled data set made from a part of the purpose of use dataset. Finally, we conduct a prediction evaluation using test data produced from the rest of the dataset and evaluate the obtained result using some measurements.

For multi-label classification, we use a multiclass logistic regression classifier. We use the LIBLINEAR package³ as the classifier. We use three measurements for the mul-

³<https://www.csie.ntu.edu.tw/~cjlin/liblinear/>.

ti-label classifications, which are the “KL divergence”, the “Mean Reciprocal Rank” (MRR), and the “Mean Average Precision” (MAP). For this experiment, the number of classes is four described in Section 4. We evaluate the method accuracy using two cross-validations randomized five times repeatedly. In other words, we use all the getting on and off data for obtaining vector representations, but we use only half of the stations in the purpose of use dataset for obtaining vector representations and classifier training. We use the rest to evaluate the classifier accuracy. We repeat this experiment procedure five times by randomizing the purpose of use dataset.

Finally, we evaluate geographical locations around each purpose vector. The evaluation metric is the average value of the standard deviation of the actual geolocation of stations near the purpose label vector. Because, when the average of the standard deviation of the nearby station of the purpose label is large, the station group is extracted for the purpose of the station without geographical constraints.

5.2. Compared Algorithms

We use the following methods to compare algorithms.

1) Weighted random: random sampling from a discrete probability distribution. In advance, we calculate each purpose distribution from a training dataset. When predicting the purpose of dropping off at a station in test data, the method predicts it by selecting a purpose randomly according to the arbitrary distribution.

2) Word2Vec [12]: Word2Vec is an efficient word embedding model that learns the representation of each word in a large corpus. We simply use the Skip-gram model in this experiment.

3) GloVe [26]: GloVe is another efficient word embedding model. The method uses global word-word co-occurrence statistics from a corpus to learn word representation vectors.

4) DeepWalk [14]: DeepWalk is the first network embedding method which can learn the representation of networks. This model only works for an unweighted graph. For each vertex, truncated random walk is used to translate the graph structure into linear sequences.

5) LINE [15]: LINE is the other network embedding method. LINE defines the first proximity and the second proximity between vertices using edge weight information. It obtains the representation by approximating the inner product value between the vertex and context vector to each proximity (LINE(1st) and LINE (2nd)). The LINE will achieve the best performance when concatenating the representation the first proximity and the second proximity (LINE(concat)).

6) PTE [25]: PTE is the network embedding method for a heterogeneous network. This method applies to three different networks which are the word-word, word-document, and word-label networks. They propose two learning styles, which are “pre-train” and “joint” learning style. We select “joint” learning style, which is slightly better than pre-train learning style in their report (PTE(joint)). This method can embed vertices in three network graphs to same vector spaces. The same node in different graphs

has the same vector representation among all graphs.

7) Proposed: Our proposed models are all for learning geospatial area embedding through large-scale mobility data from smart cards. We offer two models based on the “Movement Purpose Hypothesis” described in the Section 3.1, which is the concatenating model (“concat”) and the internally dividing model (“divide”). Our proposed models can embed vertices in three network graphs to different vector spaces. A single node in different graphs has different vector representations with each graph $(\vec{u}_{sc}, \vec{u}_{sr}, \vec{u}_{ss})$.

Word2Vec and GloVe are necessary for sentences as input information because of word embedding methods. We regard the sequence of stations which is history of each user getting on and off as a sentence. Word2Vec, GloVe, DeepWalk, and LINE methods are unsupervised style learning. Therefore, we merely apply user information related to getting on and off at different stations for training. PTE and our proposed models are semi-supervised style learning. We set the people flow network as the word-word network, the geographical constraints network as the word-document network, and the purpose proximity network as the word-label network. On all method, the dimension of the node vector is set as 200, but in the proposed concatenating model, $\vec{u}_i^{ss}, \vec{u}_i^{tss}$ has twice the number of dimensions: 400.

5.3. Results

This section presents the performance and characteristics of our proposed models.

5.3.1. Performance of Multi-Label Classification

Table 4 shows the performance of multi-label classification. One can start with a comparison of weighted random with others. Except for weighted random, all other methods are embedding word or node to vector space. In the KL divergence metric, all other methods are superior to the weighted random method. In other metrics, all other method results are equal to or better than the weighted random method. Therefore, applying embedding method to the people flow data is reasonable and efficient to extract the purpose distribution for each station.

Next, we compare the performance of GloVe with others. GloVe indicates the best result at the KL divergence metrics because only GloVe uses global co-occurrence information in the dataset. The effect of long context co-occurrence information also shows the result between LINE (1st) and LINE (2nd). Although LINE (1st) directly approximates the edge weight between two nodes, LINE (2nd) approximates two hops sharing node proximity. This effect appears in the KL divergence and the MRR result. These results indicate that using the global graph structure is good for multi-label classification.

We compare our proposed models with the PTE (joint) method. Particularly, the proposed model of (divide \vec{u}_{ss}) is superior to the PTE method in point of the KL divergence and the MRR metrics. Moreover, proposed (concat \vec{u}_{ss}) is also superior in terms of the KL divergence and the MAP metrics. These results indicate that our proposed models use labeled information more efficiently than PTE (joint) method on the people flow data.

Finally, we make a comparison of our proposed models. The proposed (divide \vec{u}_{ss}) is superior to the proposed (concat \vec{u}_{ss}) concerning the KL divergence and the MRR metrics. This difference derives from the update style difference between the two models. The concatenating model is updating half of vector element once, but the internally dividing model is updating all vector elements. This difference appears to learned vector representations. We describe this difference in greater detail in the following sections. In both models, \vec{u}_{ss} results are superior to \vec{u}_{sc} and \vec{u}_{sr} models results, which shows that it is necessary to both the geospatial information and the purpose information for the multi-label classification.

Regarding α estimation result in the proposed (divide) model, we show the result in **Table 5**. This result indicates that $\vec{u}_{ss} = 0.468\vec{u}_{sc} + 0.532\vec{u}_{sr}$, which means that \vec{u}_{sr} is more important than \vec{u}_{sc} .

5.3.2. Geographical Locations around Each Purpose Vector

Next, it is necessary to unveil the obtained purpose vector characteristics. Therefore, we inspect stations around each purpose vector. As described in this paper, we attempt to extract purposes of a station to go accurately and the purposes of a station are irrelevant to the geospatial location of the station. If so, our proposed method will gather distant stations one after another around a purpose vector. We evaluate this hypothesis to

Table 4. Results of multi-label classification. The KL divergence is better if it is a smaller value. Other metrics are all better if larger values.

Method	KL div.	MRR	MAP
Weighted random	40.734e-2	45.000e-2	74.341e-2
Word2Vec	38.810e-2	44.973e-2	74.313e-2
GloVe	36.187e-2	47.802e-2	73.608e-2
DeepWalk	39.496e-2	45.192e-2	74.176e-2
LINE(1st)	40.006e-2	45.000e-2	74.341e-2
LINE(2nd)	37.796e-2	51.209e-2	73.553e-2
LINE(concat)	37.560e-2	51.071e-2	73.343e-2
PTE(joint)	39.417e-2	45.192e-2	74.368e-2
Proposed (concat \vec{u}_{sc})	40.425e-2	45.000e-2	74.341e-2
Proposed (concat \vec{u}_{sr})	38.766e-2	45.082e-2	74.341e-2
Proposed (concat \vec{u}_{ss})	38.933e-2	45.137e-2	74.368e-2
Proposed (divide \vec{u}_{sc})	38.606e-2	45.000e-2	74.341e-2
Proposed (divide \vec{u}_{sr})	39.051e-2	45.852e-2	74.167e-2
Proposed (divide \vec{u}_{ss})	37.216e-2	51.511e-2	72.610e-2

Table 5. Estimation result of α parameter.

Variable	Average	Std dev.
α	0.468	0.0153

confirm the standard deviation of station geolocation. The result is presented in **Table 6**. This table is the standard deviation of station geolocation around each purpose vector in 10 nearest stations.

Comparison of the proposed (divide \vec{u}_{ss}) model with the PTE(joint) method shows that the proposed (divide \vec{u}_{ss}) standard deviations are larger than the PTE(joint) one for all purposes. This shows that the proposed (divide \vec{u}_{ss}) model can gather distant stations around each purpose node. The proposed (concat) results have a smaller standard deviation than other methods have. In the following visualization result section, we consider the results in greater detail.

5.3.3. Visualization of Vector Representation

Finally, we present an illustrative visualization of each method. We present a visualization in **Figure 2**. Because of space limitations, we select six methods of visualization. In the figure, each point represents a station or a purpose. Then they are colored by six train companies. In the figure, (a) and (b) are past works and (c)-(f) are our proposed method visualizations.

The (a) DeepWalk vector forms clusters gathering at each company. In this people flow dataset, we found from statistically results that most people usually move in a small area and they do not transfer so much. Therefore, the DeepWalk visualization result is reasonable because it captures local context information. This result also shows (e) proposed (divide \vec{u}_{sc}). Linearly aligned stations are apparent in the figure, showing that these stations are along the same train line.

In (f) proposed (divide \vec{u}_{sr}), the visualization result is mixed with six companies around purpose vectors. And (d) proposed (divide \vec{u}_{ss}) is the sum of (e) proposed (divide \vec{u}_{sc}) and (f) proposed (divide \vec{u}_{sr}). By this representation, this method gathers distant stations around each purpose vector and achieves a useful purpose estimation result in MAP metric on **Table 4**.

6. Discussion and Summary

As described in Section 5.3, our proposed models achieve better results than the PTE

Table 6. Standard deviation of station geolocation around each purpose vector (Nearest@10).

Method	"On business"		"Others"		"To work"		"To school"		Average	
	long SD	lat SD	long SD	lat SD	long SD	lat SD	long SD	lat SD	long SD	lat SD
PTE (joint)	4.052e-2	6.884e-2	6.283e-2	8.801e-2	4.991e-2	8.412e-2	12.866e-2	9.020e-2	7.048e-2	8.280e-2
Proposed (concat \vec{u}_{sc})	6.516e-2	8.739e-2	4.804e-2	7.941e-2	5.530e-2	8.124e-2	6.399e-2	7.714e-2	5.812e-2	8.129e-2
Proposed (concat \vec{u}_{sr})	6.126e-2	7.224e-2	7.586e-2	7.782e-2	6.912e-2	7.102e-2	10.655e-2	8.183e-2	7.820e-2	7.573e-2
Proposed (concat \vec{u}_{ss})	3.584e-2	7.897e-2	4.025e-2	7.259e-2	3.498e-2	7.174e-2	7.048e-2	7.869e-2	4.539e-2	7.550e-2
Proposed (divide \vec{u}_{sc})	10.596e-2	9.604e-2	9.551e-2	9.603e-2	9.593e-2	10.133e-2	8.789e-2	8.831e-2	9.632e-2	9.543e-2
Proposed (divide \vec{u}_{sr})	11.701e-2	9.416e-2	14.268e-2	11.357e-2	11.731e-2	11.242e-2	14.934e-2	15.349e-2	13.159e-2	11.841e-2
Proposed (divide \vec{u}_{ss})	11.701e-2	9.416e-2	14.268e-2	11.357e-2	11.731e-2	11.242e-2	14.934e-2	15.349e-2	13.159e-2	11.841e-2

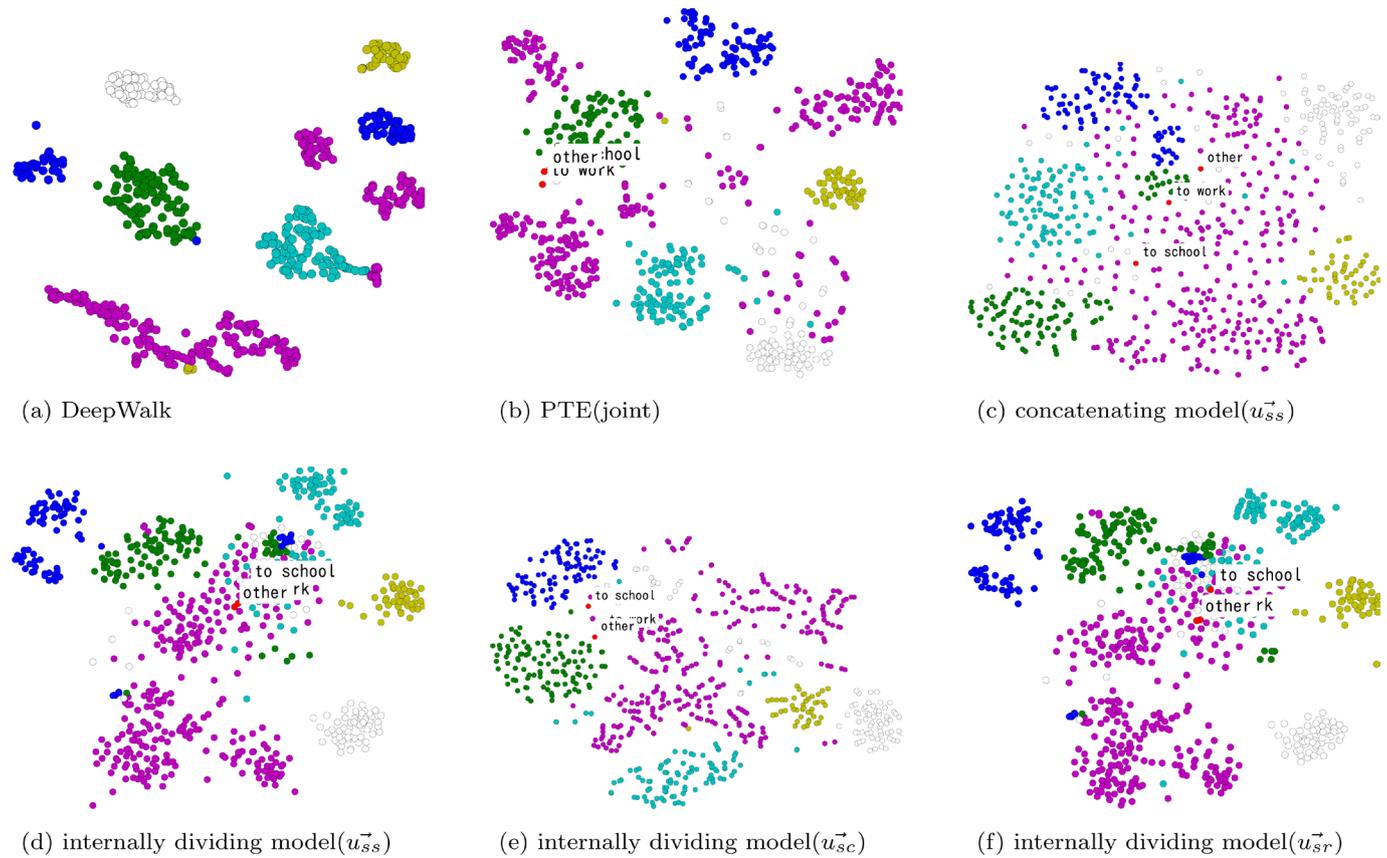


Figure 2. The obtained vector visualization by t-SNE [27] toolkit. It is noteworthy that each point represents a station or a purpose. They are colored according to train companies.

method. These results indicate that, for large-scale movement data, which have spatial dependence, the proposed models capture the characteristics of the purpose of each area better than the PTE method does. People’s moving areas are usually small. They live in defined areas. In light of this constraint, our proposed models work better than the PTE method. For the multi-label classification task, our proposed models (concat \vec{u}_{ss} and divide \vec{u}_{ss}) show good results in **Table 4**. This result underscores the correctness our proposed “Moving Purpose Hypothesis.” Especially for vector visualization results (**Figure 2**), the proposed (divide) models decompose each area to geolocation dependency vectors and purpose vectors. Finally, the α parameter estimation result is impressive. This result means that the purpose is 1.1 times more important than the distance. Therefore, people move to distant places when they have a purpose that they actively want to complete.

However, the currently proposed models’ performance is slightly better than unsupervised embedding methods because our proposed models use only two-hop proximity, and they do not capture the global network structure. As the next step, we should consider a graph global structure with the heterogeneous network and how to apply the labeled network more efficiently. The graph global structure can be captured by the GraRep [13] or GloVe [26]. It is necessary to refer to such approaches for extracting

purposes to go to an area.

We believe that there is considerable research room to representation learning for the geospatial network.

References

- [1] Zheng, Y. (2015) Trajectory Data Mining: An Overview. *ACM Transactions on Intelligent Systems and Technology*, **6**, 29:1-29:41. <http://dx.doi.org/10.1145/2743025>
- [2] Feng, Z. and Zhu, Y. (2016) A Survey on Trajectory Data Mining: Techniques and Applications. *IEEE Access*, **4**, 2056-2067. <http://dx.doi.org/10.1109/ACCESS.2016.2553681>
- [3] Sun, L. and Jin, J.G. (2015) Modeling Temporal Flow Assignment in Metro Networks Using Smart Card Data. *18th International Conference on Intelligent Transportation Systems (ITSC)*, September 2015, 836-841. <http://dx.doi.org/10.1109/itsc.2015.141>
- [4] Yuan, Q., Cong, G. and Sun, A. (2014) Graph-Based Point-of-Interest Recommendation with Geographical and Temporal Influences. *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, New York, 659-668. <http://dx.doi.org/10.1145/2661829.2661983>
- [5] Liu, Y., Kang, C., Gao, S., Xiao, Y. and Tian, Y. (2012) Understanding Intra-Urban Trip Patterns from Taxi Trajectory Data. *Journal of Geographical Systems*, **14**, 463-483. <http://dx.doi.org/10.1007/s10109-012-0166-z>
- [6] Zheng, Y., Liu, Y., Yuan, J. and Xie, X. (2011) Urban Computing with Taxicabs. *Proceedings of the 13th International Conference on Ubiquitous Computing*, New York, 89-98. <http://dx.doi.org/10.1145/2030112.2030126>
- [7] Kim, D., Sarker, M. and Vyas, P. (2016) Role of Spatial Tools in Public Health Policymaking of Bangladesh: Opportunities and Challenges. *Journal of Health, Population and Nutrition*, **35**, 1-5. <http://dx.doi.org/10.1186/s41043-016-0045-1>
- [8] Phithakkitnukoon, S., Horanont, T., Di Lorenzo, G., Shibasaki, R. and Ratti, C. (2010) Activity-Aware Map: Identifying Human Daily Activity Pattern Using Mobile Phone Data. *Proceedings of the First International Conference on Human Behavior Understanding*, Berlin, 14-25. http://dx.doi.org/10.1007/978-3-642-14715-9_3
- [9] Long, Y. and Shen, Z. (2015) Geospatial Analysis to Support Urban Planning in Beijing. *Discovering Functional Zones Using Bus Smart Card Data and Points of Interest in Beijing*, Springer International Publishing, Cham, 193-217. <http://dx.doi.org/10.1007/978-3-319-19342-7>
- [10] Zhang, F., Yuan, N.J., Wang, Y. and Xie, X. (2015) Reconstructing Individual Mobility from Smart Card Transactions: A Collaborative Space Alignment Approach. *Knowledge and Information Systems*, **44**, 299-323. <http://dx.doi.org/10.1007/s10115-014-0763-x>
- [11] Bengio, Y., Courville, A. and Vincent, P. (2013) Representation Learning: A Review and New Perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **35**, 1798-1828. <http://dx.doi.org/10.1109/TPAMI.2013.50>
- [12] Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S. and Dean, J. (2013) Distributed Representations of Words and Phrases and Their Compositionality. In: Burges, C.J.C., Bottou, L., Welling, M., Ghahramani, Z. and Weinberger, K.Q., Eds., *Advances in Neural Information Processing Systems* 26, Curran Associates, Inc., 3111-3119.
- [13] Cao, S., Lu, W. and Xu, Q. (2015) Grarep: Learning Graph Representations with Global Structural Information. *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, New York, 891-900. <http://dx.doi.org/10.1145/2806416.2806512>

- [14] Perozzi, B., Al-Rfou, R. and Skiena, S. (2014) Deepwalk: Online Learning of Social Representations. *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, 701-710. <http://dx.doi.org/10.1145/2623330.2623732>
- [15] Tang, J., Qu, M., Wang, M., Zhang, M., Yan, J. and Mei, Q. (2015) Line: Large-Scale Information Network Embedding. *Proceedings of the 24th International Conference on World Wide Web*, New York, 1067-1077. <http://dx.doi.org/10.1145/2736277.2741093>
- [16] Itoh, M., Yokoyama, D., Toyoda, M., Tomita, Y., Kawamura, S. and Kitsuregawa, M. (2014) Visual Fusion of Mega-City Big Data: An Application to Traffic and Tweets Data Analysis of Metro Passengers. *IEEE International Conference on Big Data (Big Data)*, October 2014, 431-440. <http://dx.doi.org/10.1109/bigdata.2014.7004260>
- [17] Yokoyama, D., Itoh, M., Toyoda, M., Tomita, Y., Kawamura, S. and Kitsuregawa, M. (2014) A Framework for Large-Scale Train Trip Record Analysis and Its Application to Passengers' Flow Prediction after Train Accidents. *Advances in Knowledge Discovery and Data Mining*, Springer, 533-544. http://dx.doi.org/10.1007/978-3-319-06608-0_44
- [18] Lathia, N. and Capra, L. (2011) How Smart Is Your Smartcard? Measuring Travel Behaviours, Perceptions, and Incentives. *Proceedings of the 13th International Conference on Ubiquitous Computing*, New York, 291-300. <http://dx.doi.org/10.1145/2030112.2030152>
- [19] Zhang, F., Zhao, J., Tian, C., Xu, C., Liu, X. and Rao, L. (2016) Spatiotemporal Segmentation of Metro Trips Using Smart Card Data. *IEEE Transactions on Vehicular Technology*, **65**, 1137-1149. <http://dx.doi.org/10.1109/TVT.2015.2409815>
- [20] Roth, C., Kang, S.M., Batty, M. and Barthélemy, M. (2011) Structure of Urban Movements: Polycentric Activity and Entangled Hierarchical Flows. *PLoS ONE*, **6**, e15923. <http://dx.doi.org/10.1371/journal.pone.0015923>
- [21] Dhillon, I.S. and Sra, S. (2005) Generalized Nonnegative Matrix Approximations with Bregman Divergences. *Neural Information Processing Systems (NIPS)*.
- [22] Levy, O. and Goldberg, Y. (2014) Neural Word Embedding as Implicit Matrix Factorization. In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D. and Weinberger, K.Q., Eds., *Advances in Neural Information Processing Systems 27*, Curran Associates, Inc., 2177-2185.
- [23] Zhao, Y., Liu, Z. and Sun, M. (2015) Representation Learning for Measuring Entity Relatedness with Rich Information. *Proceedings of the 24th International Conference on Artificial Intelligence (IJCAI'15)*, Buenos Aires, 25-31 July, 2015, 1412-1418.
- [24] Kulkarni, V., Al-Rfou, R., Perozzi, B. and Skiena, S. (2015) Statistically Significant Detection of Linguistic Change. *Proceedings of the 24th International Conference on World Wide Web*, Florence, 18-22 May 2015, 625-635. <http://dx.doi.org/10.1145/2736277.2741627>
- [25] Tang, J., Qu, M. and Mei, Q. (2015) PTE: Predictive Text Embedding through Large-Scale Heterogeneous Text Networks. *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'15)*, Sydney, 10-13 August 2015, 1165-1174. <http://dx.doi.org/10.1145/2783258.2783307>
- [26] Pennington, J., Socher, R. and Manning, C.D. (2014) GloVe: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, 25-29 October 2014, 1532-1543.
- [27] Van der Maaten, L. and Hinton, G. (2008) Visualizing Data Using t-SNE. *Journal of Machine Learning Research*, **1**, 1-48.



Submit or recommend next manuscript to SCIRP and we will provide best service for you:

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.

A wide selection of journals (inclusive of 9 subjects, more than 200 journals)

Providing 24-hour high-quality service

User-friendly online submission system

Fair and swift peer-review system

Efficient typesetting and proofreading procedure

Display of the result of downloads and visits, as well as the number of cited articles

Maximum dissemination of your research work

Submit your manuscript at: <http://papersubmission.scirp.org/>

Or contact ijcns@scirp.org