City2City: Translating Place Representations across Cities

Takahiro Yabe
Lyles School of Civil Engineering
Purdue University, USA
tyabe@purdue.edu

Kota Tsubouchi
Yahoo Japan Corporation
Tokyo, Japan
ktsubouc@yahoo-corp.jp

Toru Shimizu
Yahoo Japan Corporation
Tokyo, Japan
toshimiz@yahoo-corp.jp

Yoshihide Sekimoto
Institute of Industrial Science
University of Tokyo, Japan
sekimoto@iis.u-tokyo.ac.jp

Satish V. Ukkusuri
Lyles School of Civil Engineering
Purdue University, USA
sukkusur@purdue.edu

ABSTRACT

Large mobility datasets collected from various sources have allowed us to observe, analyze, predict and solve a wide range of important urban challenges. In particular, studies have generated place representations (or embeddings) from mobility patterns in a similar manner to word embeddings to better understand the functionality of different places within a city. However, studies have been limited to generating such representations of cities in an individual manner and has lacked an inter-city perspective, which has made it difficult to transfer the insights gained from the place representations across different cities. In this study, we attempt to bridge this research gap by treating *cities* and *languages* analogously. We apply methods developed for unsupervised machine language translation tasks to translate place representations across different cities. Real world mobility data collected from mobile phone users in 2 cities in Japan are used to test our place representation translation methods. Translated place representations are validated using landuse data, and results show that our methods were able to accurately translate place representations from one city to another.

CCS CONCEPTS

• **Human-centered computing** → Ubiquitous and mobile devices; • **Computing methodologies** → Knowledge representation and reasoning.

KEYWORDS

machine translation, place representations, urban functions, human mobility, mobile phone data

ACM Reference Format:

Takahiro Yabe, Kota Tsubouchi, Toru Shimizu, Yoshihide Sekimoto, and Satish V. Ukkusuri. 2019. City2City: Translating Place Representations across Cities. In *27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (SIGSPATIAL '19)*, November 5–8, 2019, Chicago, IL, USA. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3347146.3359063>

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
SIGSPATIAL '19, November 5–8, 2019, Chicago, IL, USA
© 2019 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-6909-1/19/11.
<https://doi.org/10.1145/3347146.3359063>

Figure 1: Illustration of our problem setting. Given representations of places in 2 cities (ϕ, ψ) generated from mobility patterns, our problem is to translate the place representations of city ϕ to the vector space of city ψ , so that similar places from both cities are mapped closely in the common vector space, as shown in the bottom right panel.

1 INTRODUCTION

Recent studies have made significant progress on understanding the semantics of places by generating vector representations from large scale mobility data [5, 8, 11, 13, 16]. Many of these methods produce place representations in a similar manner to generating word representations in the natural language processing field (e.g. word2vec [9]), by treating words and places, sentences and trip sequences analogously. These place representations successfully characterize the functionality of places within cities, and have been applied in various tasks related to urban planning [12].

However, such studies have been limited to understanding the place representations of cities in an individual manner and has lacked an inter-city perspective. Because the representations of different cities were not generated in a common vector space, it has been difficult to transfer insights based on place representations from one city to another, let alone transferring various phenomena (e.g. evacuation after disasters) across cities. If we could map place representations learned in one city to another, in the same way we translate words to words across different languages, we would be able to utilize knowledge accumulated in other cities to perform better analyses and predictions.

In this study, we attempt to bridge these gaps by treating *cities* and *languages* analogously, which extends the analogies made by

previous studies (“places and words” and “trip sequences and sentences”). We propose models that extend the methods developed in the natural language processing field for unsupervised machine language translation tasks [1, 4, 15]. Figure 1 shows an illustration of our problem setting and approach. Given representations of places X_ϕ , X_ψ in cities ϕ and ψ , directly overlaying X_ϕ on X_ψ would be uninformative, since the vector spaces are not aligned with each other. Using methods to translate representations, we obtain $\tilde{X}_\phi = f(X_\phi)$ which is aligned to the space of city ψ , allowing us to compare representations of places in different cities for further analyses and predictions. The model performances are tested using real world data collected from mobile phones in 2 cities in Japan, and are validated using landuse data. Results show that our methods are able to accurately translate place representations from one city to another.

The main contributions of this paper are as follows:

- We propose and test methods to translate place representations across cities, which can map places from different cities with similar functions closely together.
- We verify that our method can successfully translate place representations, using real mobility world data from 2 cities.
- We make the translated place representations publicly available for researchers and practitioners.

2 METHODOLOGY

2.1 Extracting Human Mobility Patterns

We first extract human mobility pattern datasets for each city, from the location data observed from mobile phones. Each observation of the location data contains the user ID, timestamp, longitude and latitude. To cope with noisy location observations (e.g. spatial errors in GPS data), we perform mean shift clustering to estimate the true location for each observation, as described in previous studies (e.g. [3]). For each user, we read their location data in time order, and search for locations where the user has stayed within the distance defined by the spatial threshold (1000m) for a duration longer than the time defined by the temporal threshold (30 mins).

2.2 Generating Place Representations

To obtain the representations of places in a city, we solve a self-supervised task in which an LSTM RNN model is trained to predict the next staypoint of a user using mobility data, which is analogous to language models which are trained to predict the next word in a sentence. After training an LSTM RNN model using staypoint sequences of a city c , we extract and stack the embedding layer’s parameters of the size $n_c \times d$, and define it as the matrix of place representations $\mathbf{X}_c$. We refer to this place representation learning architecture as “**MobLSTM**” in the following sections.

2.3 Translating Place Representations

2.3.1 Joint Learning Approach. In the first approach, we merge the mobility datasets of two cities into one, train the model over the merged data, and use the transposed embedding layer matrix of the size $d \times (n_\phi + n_\psi)$ as the representation matrix. The rationale behind this approach is that, representations of places with similar functions will be visited in a similar manner (e.g. time of day, day

of week, after and before certain places) regardless of the city the places belong to. A previous study shows that this approach is effective in translating embeddings of one language to another in an unsupervised manner [10]. We refer to this translation method as “**Joint-MobLSTM**”.

2.3.2 Procrustes Transformation Approach. Given place representations $\mathbf{X}_\phi$ and $\mathbf{X}_\psi$, and a dictionary of pairs of places that are expected to be mapped close to each other, orthogonal Procrustes is applied to align them together into a common vector space. The solution gives the best rotational alignment of the two vector spaces. Although our original problem is in the unsupervised setting, we generate synthetic representation pairs by pairing up the top $N = 500$ visited places from both cities. We refer to this translation method as “**MobLSTM-P**”.

2.3.3 Adversarial Training Approach. A similar approach as Conneau et al. [4] is taken here, where a model is trained the discriminate between representations randomly selected from $R\mathbf{X}_\phi$ and $\mathbf{X}_\psi$. R is then trained to prevent the discriminator from making accurate predictions [6]. The standard training procedure of deep adversarial networks is used for train the adversarial model [7]. We refer to this translation method as “**MobLSTM-Adv**”.

3 DATA

Yahoo Japan Corporation collects location information of mobile phone app users. The users in this study have accepted to provide their location information. The data are anonymized so that individuals cannot be specified, and personal information such as gender, age and occupation are unknown. Each GPS record consists of a user’s unique ID, timestamp, longitude, and latitude. The data has a sample rate of approximately 2% of the population. There are around 100,000 unique active users from the two cities (Kumamoto and Okayama).

To validate whether the generated and translated place representations correctly reflect the functionality of the places, we use the Urban Area Land Use Mesh Data in the National Land Numerical Information Database provided by the Ministry of Infrastructure, Land, and Transport and Tourism of Japan. The dataset divides all urban areas of the entire country into $100m \times 100m$ grid cells, and assigns one category to each grid cell out of 17 options.

4 EXPERIMENT

4.1 Model Hyperparameter Settings

To conduct the representation learning of places described, we setup the model and input data as follows. The model consists of the embedding layer, LSTM RNN block, readout layer, and the output layer. While the main input of the model is a sequence of staypoints representing a user’s movement, we added two supplementary values, which are the timestamp of when the user had entered that place and the duration time of the stay, to incorporate time-dependency of the users’ behavior. The embeddings of staypoints were set to 96-dimensional vectors. The timestamp and stay duration were converted to 8-dimensional and 4-dimensional vectors respectively, and the three vectors at each step were concatenated into a 108-dimensional vector. The LSTM RNN block scanning over the embedding sequence consists of two layers of

the size 128, and the hidden vectors of both layers were fed into the readout layer of the size 96, which were then read by the output layer producing the probability distribution over staypoints for the next place prediction. The parameter matrix of the staypoint embedding was reused as the output layer’s matrix to reduce the total number of parameters and make the training data usage more efficient. We applied dropout with the keep probability 0.7 to the embedding layer, readout layer, and output layer. We continued the training for 20 epochs, evaluated performance on the validation data at the end of each epoch, and used the embedding matrix of the best model for subsequent processing.

4.2 Evaluation Metrics

Given two sets of places, we measure the average mutual norm distance and average mutual cosine similarity of the place representations of those places.

Average Mutual Norm Distance (AMND). Given 2 place representations $x_i, x_j \in \mathbb{R}^d$, norm distance is defined as $\text{Norm-Dist}(i, j) = \|x_i - x_j\|$. The average mutual norm distance between sets of places I and J is defined by the following:

$$s_{nd}(I, J) = \frac{1}{Z(I, J)} \sum_{i \in I} \sum_{j \in J; j \neq i} \text{Norm-Dist}(i, j) \quad (1)$$

where $Z(I, J)$ is the number of unique combinations between the places in sets I and J .

Average Mutual Cosine Similarity (AMCS). Cosine similarity is a commonly used metric to measure the similarity between 2 place representations $x_i, x_j \in \mathbb{R}^d$, and is defined as $\text{cos-sim}(i, j) = \frac{x_i \cdot x_j}{\|x_i\| \|x_j\|}$. We use the average mutual cosine similarity to measure the similarity between representations of two sets of places. Similar to the average mutual norm distance, average mutual cosine similarity between sets of places I and J is defined by the following:

$$s_{cs}(I, J) = \frac{1}{Z(I, J)} \sum_{i \in I} \sum_{j \in J; j \neq i} \text{cos-sim}(i, j) \quad (2)$$

4.3 Results

4.3.1 Intra-city Validation of Place Representations. To validate whether the place representations were correctly generated (i.e. representations of places with the same functionality are mapped closely), we measured the AMND and AMCS between places with the same landuse labels (business districts, shopping malls, residential areas, and farmland areas). We refer to this validation scheme as “intra-city validation”. Figure 2 shows the validation results for both cities. Note that for norm distance (A), lower is better, and for cosine similarity (B), higher is better. All error bars (vertical lines) show the standard deviation of the results of 10 iterations. Results show that both models were able to generate accurate representations, and embedded places with same landuse labels closer to each other than randomly selected pairs of places. MobLSTM and Joint-MobLSTM had comparable performances for generating place representations, however we see that MobLSTM had slightly better performances (lower norm distances and higher cosine similarity) for many of the landuse types. This result agrees with our intuition, because MobLSTM is able to allocate more dimensions in the parameter space to encode information related to places in their own city, whereas

Figure 2: Intra-city validation results in both cities. Both models generated representations of all landuse types with statistically significant intra-city similarity.

Figure 3: Inter-city translation results from Kumamoto to Okayama. MobLSTM-P is able to translate place representations across Kumamoto and Okayama for business, shopping, and residential areas, but not for farmland areas where little human mobility patterns are observed.

Joint-MobLSTM shares the parameter space across different cities, having less dimensions to encode the representations for each city.

4.3.2 Inter-city Translation of Place Representations. To quantitatively validate the performance of translating place representations across cities, we measured the AMND and AMCS between places with same landuse types across different cities (e.g. similarity between representations of places with shopping malls in Kumamoto and representations of places with shopping malls in Okayama). We refer to this validation scheme as “inter-city translation”. Figure

3 shows the translation accuracy of all tested methods. The poor performance of MobLSTM confirms the negative example illustrated in Figure 1, and that a translation operation is indeed needed to compare place representations from different cities.

The rest of the models compare the performances of different translation methods. For business, shopping, and residential areas, the AMND seems to be lower than random for all methods, which implies that all of these types of places are clustered together in the vector space, away from the farmland areas. The AMCS metric allows us to measure more specific differences in the place representations, by normalizing the vectors by their lengths. AMCS results show that MobLSTM-P is able to translate representations of business, shopping and residential places successfully (statistically significantly). Even though Joint-MobLSTM and MobLSTM-Adv approaches are shown to succeed in language translation tasks, they fail to do so in place translation tasks. For Joint-MobLSTM to perform better, further searching for the appropriate model architecture may be effective, however is not cost effective considering the vast model space. MobLSTM-Adv failing to align the two vector spaces indicates that the probability distribution of the place representations of two cities are completely different, in contrary to word vector spaces. Even with MobLSTM-P, representations of farmlands were not successfully translated across cities. This is because we are not using many farmland areas as anchor points in our dictionary for solving the optimization problem, due to the lack of observed human mobility patterns in such areas. Overall, results in Figure 3 confirm that MobLSTM-P is successful in translating representations of places visited by people across cities.

5 RELATED WORKS

Models such as SkipGram and POI2vec applied ideas similar to word2vec [9] on social media check-in data, where sequences of POIs are treated as sentences [5, 8]. Place2vec uses the physical proximity between POIs and the number of visit counts to increase training data so that such pairs of POIs would have similar embeddings [13]. ZE-Mob proposes a origin-destination coupled embedding model, where the assumptions are that origin and destinations of the same trip should have similar representations [14].

Unsupervised machine translation is a popular topic in the representation learning literature, due to the difficulty of collecting large scale cross-lingual training data [2]. This setting is similar to our problem, since we are not given any training data to learn which places should be closely embedded. Studies take different approaches to unsupervised word translation tasks; Zhang et al. [15] uses an adversarial training approach, Conneau et al. [4] learns a linear transformation matrix to map words in one language to another, Artetxe et al. [1] applies better initial estimates using probability density functions of distances to other word embeddings, and Wada and Iwata [10] apply a shared LSTM model to jointly embed two languages.

6 DISCUSSIONS AND CONCLUSION

Despite the popularity in understanding urban functions through place representations generated from mobility patterns, no works so far have attempted to translate the representations across different cities. To bridge this gap, we applied and tested methods inspired by

unsupervised machine language translation. Experimental results showed that we were able to translate place representations, so that functionally similar places from different cities were mapped closely together in the common vector space.

For future works, we will further investigate whether this method works between a more diverse set of cities, such as Tokyo (Japan) and Indianapolis (USA), where various aspects differ more than between Kumamoto and Okayama. Applying this technique to real world problems such as selecting appropriate locations to open new stores is also of future research interest.

ACKNOWLEDGMENTS

T.Y. is partly supported by the Doctoral Fellowship provided by the Purdue Systems Collaboratory. The work of T.Y. and S. V. U. is partly funded by NSF Grant CMMI-1638311. This study was supported by Joint Research Program No. 878 at CSIS, University of Tokyo.

REFERENCES

- [1] Mikel Artetxe, Gorka Labaka, and Eneko Agirre. 2018. A robust self-learning method for fully unsupervised cross-lingual mappings of word embeddings. *arXiv preprint arXiv:1805.06297* (2018).
- [2] Mikel Artetxe, Gorka Labaka, Eneko Agirre, and Kyunghyun Cho. 2017. Unsupervised neural machine translation. *arXiv preprint arXiv:1710.11041* (2017).
- [3] Daniel Ashbrook and Thad Starner. 2003. Using GPS to learn significant locations and predict movement across multiple users. *Personal and Ubiquitous computing* 7, 5 (2003), 275–286.
- [4] Alexis Conneau, Guillaume Lample, Marc'Aurelio Ranzato, Ludovic Denoyer, and Hervé Jégou. 2017. Word translation without parallel data. *arXiv preprint arXiv:1710.04087* (2017).
- [5] Shanshan Feng, Gao Cong, Bo An, and Yeow Meng Chee. 2017. Poi2vec: Geographical latent representation for predicting future visitors. In *Thirty-First AAAI Conference on Artificial Intelligence*.
- [6] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. 2016. Domain-adversarial training of neural networks. *The Journal of Machine Learning Research* 17, 1 (2016), 2096–2030.
- [7] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In *Advances in neural information processing systems*. 2672–2680.
- [8] Xin Liu, Yong Liu, and Xiaoli Li. 2016. Exploring the Context of Locations for Personalized Location Recommendations. In *IJCAI*. 1188–1194.
- [9] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*. 3111–3119.
- [10] Takashi Wada and Tomoharu Iwata. 2018. Unsupervised cross-lingual word embedding by multilingual neural language models. *arXiv preprint arXiv:1809.02306* (2018).
- [11] Hongjian Wang and Zhenhui Li. 2017. Region representation learning via mobility flow. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. ACM, 237–246.
- [12] Mengwen Xu, Tianyi Wang, Zhengwei Wu, Jingbo Zhou, Jian Li, and Haishan Wu. 2016. Demand Driven Store Site Selection via Multiple Spatial-temporal Data. In *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (SIGSPATIAL '16)*. ACM, New York, NY, USA, Article 40, 10 pages. <https://doi.org/10.1145/2996913.2996996>
- [13] Bo Yan, Krzysztof Janowicz, Gengchen Mai, and Song Gao. 2017. From ITDL to Place2Vec: reasoning about place type similarity and relatedness by learning embeddings from augmented spatial contexts. In *Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 35.
- [14] Zijun Yao, Yanjie Fu, Bin Liu, Wangsu Hu, and Hui Xiong. 2018. Representing Urban Functions through Zone Embedding with Human Mobility Patterns. In *IJCAI*. 3919–3925.
- [15] Meng Zhang, Yang Liu, Huanbo Luan, and Maosong Sun. 2017. Adversarial training for unsupervised bilingual lexicon induction. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 1959–1970.
- [16] Shenglin Zhao, Tong Zhao, Irwin King, and Michael R Lyu. 2017. Geo-teaser: Geo-temporal sequential embedding rank for point-of-interest recommendation. In *Proceedings of the 26th international conference on world wide web companion*. International World Wide Web Conferences Steering Committee, 153–162.