Automated Feature-Topic Pairing: Aligning Semantic and Embedding Spaces in Spatial Representation Learning

Dongjie Wang  
University of Central Florida  
Orlando, Florida, United States

Kunpeng Liu  
University of Central Florida  
Orlando, Florida, United States

David Mohaisen  
University of Central Florida  
Orlando, Florida, United States

Pengyang Wang  
University of Macau  
Macau, China

Chang-Tien Lu  
Virginia Tech  
Virginia, United States

Yanjie Fu*  
University of Central Florida  
Orlando, Florida, United States

ABSTRACT
Automated characterization of spatial data is a kind of critical geographical intelligence. As an emerging technique for characterization, Spatial Representation Learning (SRL) uses deep neural networks (DNNs) to learn non-linear embedded features of spatial data for characterization. However, SRL extracts features by internal layers of DNNs, and thus suffers from lacking semantic labels. Texts of spatial entities, on the other hand, provide semantic understanding of latent feature labels, but is insensible to deep SRL models. How can we teach a SRL model to discover appropriate topic labels in texts and pair learned features with the labels? This paper formulates a new problem: feature-topic pairing, and proposes a novel Particle Swarm Optimization (PSO) based deep learning framework. Specifically, we formulate the feature-topic pairing problem into an automated alignment task between 1) a latent embedding feature space and 2) a textual semantic topic space. We decompose the alignment of the two spaces into: 1) point-wise alignment, denoting the correlation between a topic distribution and an embedding vector; 2) pair-wise alignment, denoting the consistency between a feature-feature similarity matrix and a topic-topic similarity matrix. We design a PSO based solver to simultaneously select an optimal set of topics and learn corresponding features based on the selected topics. We develop a closed loop algorithm to iterate between 1) minimizing losses of representation reconstruction and feature-topic alignment and 2) searching the best topics. Finally, we present extensive experiments to demonstrate the enhanced performance of our method.

CCS CONCEPTS
• Computing methodologies → Learning latent representations.

KEYWORDS
spatial representation learning, multiple spaces alignment

Permission to make digital or hard copies of all or part 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 components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

SIGSPATIAL ’21, November 2–5, 2021, Beijing, China  
© 2021 Association for Computing Machinery.  
ACM ISBN 978-1-4503-8664-7/21/11. . . $15.00  
https://doi.org/10.1145/3474717.3484212

1 INTRODUCTION
Spatial representation learning (SRL) refers to exploiting representation learning techniques to learn features of spatial network data, which has been successfully applied in many real-world scenarios, such as transportation networks, power networks, social networks, water supply networks [21]. In reality, many practical applications need to understand not just which features are effective, but also what these effective features stand for. This issue relates to two tasks: 1) deep representation learning; 2) label generation and matching for latent embedded features. Although there has been a rich body of work in SRL, including node embedding, autoencoder, random walk, adversarial learning, generative learning based methods with spatial data [14–17], research in unifying the two tasks is still in its early stage.

In response, we formulate the problem as a task of feature-topic pairing (Figure 1), which is to align a latent embedding feature space, consisting of multiple latent features, and a textual semantic topic space, consisting of multiple topic labels during SRL. The basic idea is to teach a machine to extract topic labels from texts, and then pair the labels with learned features. To that end, we propose to develop a novel deep learning framework to unify feature learning, topic selection, feature-topic matching. There are three unique challenges in addressing the problem: (1) Label Generation Challenge, in which a textual semantic topic space is difficult to construct due to the unstructured spatial texts; (2) Measurement Challenge, in which a promising measurement is highly desired to evaluate the alignment or quantify the matching score between the topic label space and the embedding feature space; (3) Optimization Challenge, in which a deep optimization framework is needed for to jointly and simultaneously unify the three tasks of feature learning, topic label selection, and feature-topic pairing.
Embedding of Spatial Entities. We construct a graph to capture the spatial autocorrelation between spatial entities. Specifically, we describe a spatial entity in terms of its POIs, by building two graphs. (i) POI-POI distance graph: denoted by $G^d$, where POI categories are nodes and the average distances between POI categories are edge weights. (ii) POI-POI mobility graph: denoted by $G^m$, where nodes are POI categories, and edge weights are human mobility connectivity, which is extracted by the method in [16]. We then apply Graph Auto Encoder (GAE) [7] as the spatial representation learner to learn spatial embeddings over these two constructed graphs respectively. Finally, we aggregate the embeddings of these two graphs by avaraging, so as to construct the unified spatial embedding of the entity, denoted by $r^n \in \mathbb{R}^K$.

### 2.3 PSO Based Feature-Topic Pairing

#### 2.4.1 Measuring the Alignment of Embedding and Semantic Spaces.

To pair features with topics, we conduct space alignment from the point-wise and pair-wise perspectives, with considering the alignment of the coordinate system and information contents respectively. To be convenient, we take the n-th entity as an example to explain the calculation process.

1) **Point-wise Alignment Loss:** $L_p$. Intuitively, the embedding feature of the spatial entity and corresponding topic should reach a consensus on describing an spatial entity, thus correlations are expected to be maximized between them. Therefore, we first select $K$ values from the topic vector $t_n$ as the vector $\hat{t}_n \in \mathbb{R}^K$, which contains the most representative semantics in the semantic space. Then, we maximize the correlation between $\hat{t}_n$ and the spatial embedding $r_n$, which is equal to minimize the negative correlation between the two vectors. The formula of the minimizing process as follows:

$$L_p = -\sum_{n=1}^{N} \text{cov}(\hat{t}_n, r_n) / \delta(\hat{t}_n)\delta(r_n),$$

where $\text{cov}(\cdot)$ denotes the covariance calculation; $\delta(\cdot)$ denotes the standard deviation.

2) **Pair-wise Alignment Loss:** $L_C$. On the other hand, the embedding feature and the corresponding topic should show consistency on the pair-wise similarity in each space to reflect the pair-wise alignment. Therefore, we minimize the difference between the pair-wise similarity between these two spaces. Specifically, we first construct the topic-topic similarity matrix $S$ and the feature-feature similarity matrix $S'$. Specifically, for $S \in \mathbb{R}^{K \times K}$, we calculate the similarity between any two topics. For $S' \in \mathbb{R}^{N \times N}$, we calculate the similarity between two features of spatial embeddings. We keep the pair-wise consistency between $S$ and $S'$ by minimizing the Frobenius norm, as follows:

$$L_C = ||S - S'||_F.$$
objective function includes the losses of graph reconstruction, semantic alignment, and the regression estimator in the downstream task, is trained to learn spatial representations, using each selected topic subset. As an application, we use the embedding of spatial entities (residential communities) to predict their real estate prices, and the loss of the regression model \( L_{\text{Reg}} \) is:

\[
L_{\text{Reg}} = \frac{1}{N} \sum_{n=1}^{N} (c_n - c_n^*)^2.
\]

where \( c_n \) is the golden standard real estate price and \( c_n^* \) is the predicted price. Next, we calculate the fitness of each particle according to the total loss of the deep model. The fitness can be calculated by:

\[
\text{Fitness} = L_C + L_P + L_R + L_{\text{Reg}}.
\]

Then, we utilize the fitness to inform all particles how far they are from the best solution. Next, each particle moves forward to the solution based on not only its current status but also all particles’ movement. After the fitness value of PSO converges, PSO identifies the best topic subset. Finally, the semantically-rich embeddings of spatial entities, given by: \( \mathbf{R} = \{ \mathbf{R}_i \}_{i=1}^{N} \).

3 EXPERIMENTAL RESULTS

3.1 Evaluation Task

In this work, we apply the proposed AutoFTP to the price prediction of real-estate as the evaluation task. Specifically, we first apply AutoFTP to learn a series of representations of spatial entities based on their geographical structural information and related text descriptions. Then, we build up a deep neural network (DNN) model for predicting average real estate price of each spatial entity according to its corresponding representation. We use RMSE, MAE, MAPE and MSLSE as the evaluation metric.

3.2 Data Description

Table 2 shows the statistics of five data sources used in the experiments. Specifically, the taxi traces data describes the GPS trajectory of taxis in Beijing in three months; the residential regions, texts, and real estate price data sources are crawled from www.fang.com; and the POIs information are extracted from www.dianping.com.

3.2.1 Baseline Algorithms. We compared our proposed method with seven baseline algorithms: AttentionWalk [1], ProNE [22], GatNE [2], GAE [7], DeepWalk [11], Node2Vec [3], and Struc2Vec [12]. Besides, regarding the are four losses in AutoFTP: reconstruction loss \( L_R \), point-wise alignment loss \( L_P \), pair-wise alignment loss \( L_C \), and regression loss \( L_{\text{Reg}} \), we also derive four variants: (ii) AutoFTP\(^{(R+P)}\), which keeps \( L_R \) and \( L_P \) of AutoFTP; (iii) AutoFTP\(^{(R+C)}\), which keeps \( L_R \) and \( L_C \) of AutoFTP; (iv) AutoFTP\(^{(R+P+C)}\), which keeps \( L_R \), \( L_P \), and \( L_C \) of AutoFTP.

3.3 Overall Performance

Table 1 shows the comparison of all the 11 models. As can be seen, AutoFTP, in overall, outperforms the baseline algorithms in terms of RMSE, MAE, MAPE and MSLSE. A possible reason for this observation is that compared with other baseline algorithms, AutoFTP not just captures geographical structural information but also preserves rich semantics of spatial entity. Besides, the regression estimator (the downstream task) of AutoFTP provides a clear learning direction (accuracy) for spatial representation learning. Thus, in the downstream predictive task, the spatial embedding features learned by AutoFTP beats all baselines.

4 RELATED WORK

Graph Representation Learning with Latent Semantics. Graph representation learning refers to techniques that preserve the structural information of a graph into a low-dimensional vector [13, 18]. However, owing to traditional graph representation learning models are implemented by deep neural networks, the learned embeddings lack interpretability. Recently, to overcome this limitation, researchers leveraged the texts related to graphs to learn semantically rich representations [9, 19].

Topic Models in Spatio-temporal Domain. Topic models aim to automatically cluster words and expressions patterns for characterizing documents [8, 20]. Recently, to understand the hidden semantics of spatial entities, many researchers applied topic models...
in the spatio-temporal data mining domain [5, 6]. Thus, in this paper, we employ a pre-trained language model to get the embeddings of keywords and utilize Gaussian Mixture Model to extract topic distribution based on the embeddings.

5 CONCLUSION

We presented a novel spatial representation learning (SRL) framework, namely AutoFTP. The spatial embeddings produced by traditional SRL models lack semantic meaning. To overcome this limitation, we formulated the feature-topic paring problem. We proposed a novel deep learning framework to unify representation learning, topic label selection, and feature-topic pairing through a PSO-based optimization algorithm. Extensive experiments demonstrated the effectiveness of AutoFTP by comparing it with other baseline models. For future work, we plan to extend our approach from geospatial networks to other applications that consist of graphs and texts, such as social media and software code safety.

ACKNOWLEDGMENTS

This research was partially supported by the National Science Foundation (NSF) via the grant numbers: 1755946, 2040950, 2006889, 2045567, 2141095.

REFERENCES

[1] Sami Abu-El-Haija, Bryan Perozzi, Rami Al-Rfou, and Alexander A Alemi. 2018. Watch Your Step: Learning Node Embeddings via Graph Attention. Advances in Neural Information Processing Systems 31 (2018), 9180–9190.
[2] Yukiyo Cen, Xu Zou, Jianwei Zhang, Hongxia Yang, Jingren Zhou, and Jie Tang. 2019. Representation learning for attributed multiplex heterogeneous network. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 1358–1368.
[3] Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable feature learning for networks. In Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining. 855–864.
[4] Han He. 2014. HanLP: Han Language Processing. https://github.com/hankcs/HanLP
[5] Lian Huang, Yuanqiao Wen, Wei Guo, Xinyan Zhu, Chunhui Zhou, Fan Zhang, and Man Zhu. 2020. Mobility pattern analysis of ship trajectories based on semantic transformation and topic model. Ocean Engineering 201 (2020), 107092.
[6] Qionghao Huang, Changjin Huang, Jin Huang, and Hamadu Fujita. 2019. Adaptive resource prefetching with spatial-temporal and topic information for educational cloud storage systems. Knowledge-Based Systems 181 (2019), 104791.
[7] Thomas N Kipf and Max Welling. 2016. Variational graph auto-encoders. arXiv preprint arXiv:1611.07308 (2016).
[8] Hakyeon Lee and Pilsung Kang. 2018. Identifying core topics in technology and innovation management studies: A topic model approach. The Journal of Technology Transfer 43, 5 (2018), 1291–1317.
[9] Pengyang Wang, Yanjie Fu, Yuanchun Zhou, Kunpeng Liu, Xiaolin Li, and Kien A. Bui. 2020. Exploiting Mutual Information for Substructure-aware Graph Representation Learning. In IJCAI. 3415–3421.

Table 1: Overall Performance with respect to RMSE, MAE, MAPE and MSLE. (The smaller value is the better performance is)

| Method     | RMSE   | MAE   | MAPE   | MSLE |
|------------|--------|-------|--------|------|
| AutoFTP    | 18.646 | -     | 16.192 | 58.851 |
| AttentionWalk | 21.418 | +14.9%| 19.712 | 68.590 |
| ProNE      | 21.830 | +17.1%| 19.929 | 69.188 |
| GatNE      | 21.229 | +13.9%| 19.288 | 67.043 |
| GAE        | 21.338 | +14.4%| 19.676 | 68.579 |
| DeepWalk   | 23.561 | +26.4%| 21.987 | 76.038 |
| Node2Vec   | 22.688 | +21.7%| 21.084 | 73.135 |
| Struc2Vec  | 21.589 | +15.8%| 19.937 | 69.423 |
| AutoFTP(R) | 21.965 | +17.8%| 20.283 | 70.991 |
| AutoFTP(R+C) | 20.509 | +9.99%| 18.921 | 66.477 |
| AutoFTP(R+P) | 21.014 | +12.7%| 19.413 | 67.920 |
| AutoFTP(R+P+Ca) | 20.211 | +8.39%| 18.676 | 65.685 |

Table 2: Statistics of the Experimental Data

| Data Sources | Properties         | Statistics |
|--------------|--------------------|------------|
| Taxi Traces  | Number of taxi     | 13,597     |
|              | Time period        | Apr. - Aug. 2012 |
| Residential Regions | Number of residential regions | 2,990 |
|              | Time period of transactions | 04/2011 - 09/2012 |
| POIs         | Number of POIs     | 32,868     |
|              | Number of POI categories | 20 |
| Texts        | Number of textual descriptions | 2,990 |
|              | Time Period        | 04/2011 - 09/2012 |
| Real Estate Prices | Number of real estate prices | 41,753 |
|              | Time Period        | 12/2011 - 06/2012 |

[10] Rada Mihalcea and Paul Tarau. 2004. Textrank: Bringing order into text. In Proceedings of the 21st ACM SIGIR conference on Research and development in information retrieval. 154–161.
[11] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. Deepwalk: Online learning of social representations. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 701–710.
[12] Leonardo FR Ribeiro, Pedro HP Saverese, and Daniel R Figueiredo. 2017. struc2vec: Learning node representations from structural identity. In Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining. 385–394.
[13] Daixun Wang, Peng Cui, and Wenwu Zhu. 2016. Structural deep network embedding. In Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 1235–1244.
[14] Dongjie Wang, Pengyang Wang, Jingbo Zhou, LeiLei Sun, Bowen Du, and Yanjie Fu. 2020. Defending Water Treatment Networks: Exploiting Spatio-temporal Effects for Cyber Attack Detection. In 2020 IEEE International Conference on Data Mining (ICDM). IEEE, 32–41.
[15] Rada Mihalcea and Paul Tarau. 2004. Textrank: Bringing order into text. In Proceedings of the 21st ACM SIGIR conference on Research and development in information retrieval. 154–161.
[16] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. Deepwalk: Online learning of social representations. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 701–710.
[17] Pengyang Wang, Yanjie Fu, Jiawei Zhang, Xinyan Zhu, and Churu Aggarwal. 2018. You are how you drive: Peer and temporal-aware representation learning for knowledge graph embedding with text descriptions. In Thirty-First IJCAI: 2457–2466.
[18] Pengyang Wang, Yanjie Fu, Jiawei Zhang, and Kien A. Bui. 2020. Exploiting Mutual Information for Substructure-aware Graph Representation Learning. In IJCAI. 3415–3421.