

明聚类效果愈差, 样本 i 更应该被分类到其他簇. 本文使用平均轮廓系数以评估整体样本的聚类结果.

3.2 实验结果与分析

3.2.1 空间位置关联分析

基于使用本文提出的向量化表示方法所生成的128维 POI 向量化表示, 计算 POI 之间的相似度 S_{poi} . POI-POI 相似度矩阵如表 2 所示. 具体的, 可发现写字楼与企业园区, 科研机构与高等院校, 住宅区与购物中心的关联度更高, 其向量间的 Cosine 距离也较小. 此外, 由于使用多源交通数据构建网络, 也能使得向量含有居民行为模式特征, 例如机场的 POI 向量与写字楼、高等院校的 POI 向量的距离较小, 而与超市、公

园的 POI 向量的距离较大, 表明其关联程度较低. POI 关联性结果可验证本文提出的向量生成方法符合 POI 之间的关联程度愈高, 其间的 Cosine 距离愈小这一规律, 表明使用本文方法能够得到有效且可信度高的向量化表示.

为对比验证本文提出的向量方法的有效性, 将本文方法生成的 POI 向量与其他的已有方法进行对比. 对比方法分别包括 DeepMove^[10], Node2Vec^[17], 其 POI 的相关性关联热力图如图 5 所示. 对比发现, 使用本文方法得到的向量能够关联 POI 的空间位置特征与居民的行为模式特征; POI 向量间的相似度基本符合先验知识, 并且区分度高, 优于已有方法.

表 2 POI-POI 相关性 (Cosine 距离) 矩阵

POI类型	写字楼	企业园区	银行	科研机构	高等院校	住宅区	购物中心	超市	公园	图书馆	投资理财	机场	度假村	停车场	服务区
写字楼	0.0000	0.0023	0.0014	0.0108	0.0045	0.0007	0.0025	0.0049	0.0057	0.0756	0.0930	0.1393	0.3239	0.5386	0.6786
企业园区	0.0023	0.0000	0.0022	0.0091	0.0069	0.0021	0.0051	0.0093	0.0080	0.0672	0.1027	0.1525	0.3409	0.5300	0.6883
银行	0.0014	0.0022	0.0000	0.0104	0.0065	0.0009	0.0024	0.0055	0.0056	0.0719	0.1008	0.1444	0.3326	0.5324	0.6880
科研机构	0.0107	0.0089	0.0102	0.0000	0.0098	0.0089	0.0110	0.0113	0.0131	0.0983	0.1003	0.1535	0.3135	0.5395	0.6711
高等院校	0.0045	0.0067	0.0064	0.0098	0.0000	0.0043	0.0060	0.0072	0.0093	0.0855	0.0898	0.1456	0.2962	0.5636	0.6760
住宅区	0.0007	0.0021	0.0009	0.0091	0.0043	0.0000	0.0020	0.0043	0.0050	0.0759	0.0944	0.1410	0.3218	0.5415	0.6856
购物中心	0.0025	0.0050	0.0023	0.0110	0.0060	0.0020	0.0000	0.0037	0.0075	0.0814	0.0942	0.1411	0.3230	0.5403	0.6743
超市	0.0049	0.0092	0.0055	0.0114	0.0073	0.0043	0.0038	0.0000	0.0076	0.1015	0.0867	0.1407	0.2959	0.5575	0.6823
公园	0.0060	0.0083	0.0058	0.0139	0.0098	0.0051	0.0079	0.0079	0.0000	0.0904	0.1018	0.1470	0.3150	0.5407	0.7112
图书馆	0.0870	0.0762	0.0815	0.1144	0.0987	0.0865	0.0942	0.1162	0.0992	0.0000	0.2363	0.2366	0.4872	0.5250	0.7807
投资理财	0.1023	0.1114	0.1094	0.1116	0.0992	0.1029	0.1043	0.0949	0.1069	0.2260	0.0000	0.2131	0.2634	0.6961	0.7047
机场	0.1476	0.1593	0.1508	0.1644	0.1548	0.1479	0.1504	0.1483	0.1486	0.2178	0.2051	0.0000	0.3509	0.6700	0.5825
度假村	0.5120	0.5312	0.5185	0.5010	0.4700	0.5035	0.5139	0.4652	0.4752	0.6694	0.3784	0.5238	0.0000	1.0000	0.8628
停车场	0.8514	0.8259	0.8302	0.8624	0.8943	0.8473	0.8596	0.8765	0.8156	0.7213	1.0000	1.0000	1.0000	0.0000	1.0000
服务区	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9437	0.8105	0.8043	0.9322	0.0000

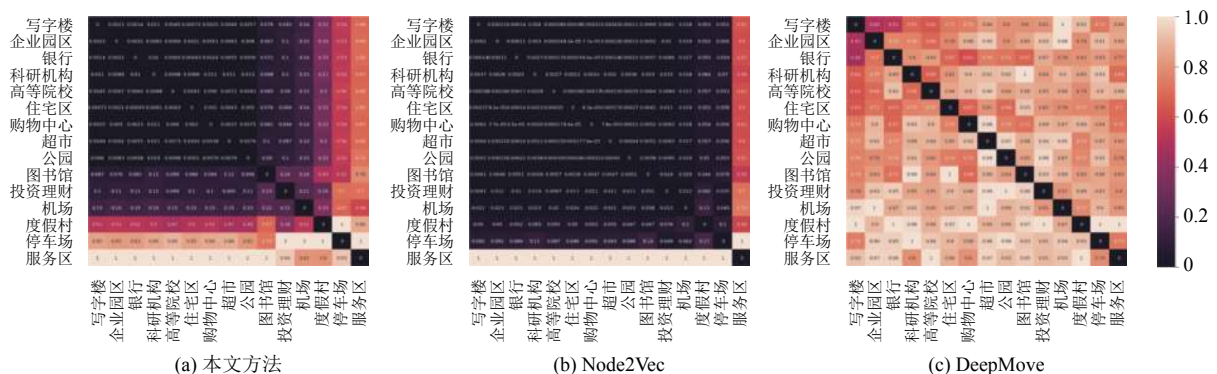


图 5 分别使用本文提出的向量化表示方法、Node2Vec 与 DeepMove 方法生成 POI 向量的相关性热力图

为验证本文提出的多源数据构建大规模交通网络方法与交通网络采样方法的有效性, 设置多组对比实验分别对模型、交通网络数据源、采样方法加以评估. 基于 K-means 聚类方法对 POI 向量进行聚

类操作, 其超参数 K 为聚类类目, 可表征为 POI 的类目. 在已有基于兴趣点 (POI) 大数据的研究综述中表明沈阳市有 21 个 POI 类目^[18], 由于 POI 类目受多种人为主观分类因素影响, 因此可认为 K 值应在 5~20

范围内浮动. 为避免聚类类目 K 值选取不当对实验结果造成的干扰, 本文将 K 分别设定为 5、10、15、20 以评估不同 POI 聚类类目下的聚类效果, 其中 K 值越大, 类别越细致. 并且每个 K 值分别进行 5 次实验取平均值. 使用平均轮廓系数对 POI 向量聚类效果进行评估. 表 3 给出了使用多种对比模型的实验结果, 其中涉及的对比方法均使用融合了长、短距离轨迹的交通网络. 实验结果表明, 当 POI 聚类类目 $K=15$ 、 $K=20$ 时, 由于本文提出方法相较于对比方法, 既能够聚合了邻居节点的信息, 又能够利用空间位置自身的特征, 改善 DeepMove^[10] 与 Node2Vec^[17] 方法仅使用网络中的节点序列特征这一问题, 因此当 POI 分类类目较细致时, 使用本文提出方法生成的向量能够保留更加细粒度的特征, 其聚类效果优于对比方法. 表 4 对比了不同采样方法的实验结果, 实验结果表明在多数 K 值下, 使用本文提出的路网采样方式生成的向量, 其聚类效果都优于使用图神经网络中随机采样方式. 表 5 对比了不同数据源的实验结果, 实验结果表明, 当 K 值较小时, 使用代表长距离出行的公共交通线路所构建的交通网络, 生成的向量聚类效果更好, 但当 K 值逐步增大时, 使用长、短距离融合的交通网络所生成的向量, 其聚类效果比较稳定, 当 $K=10$ 、 $K=15$ 时, 优于仅使用长距离出行轨迹或短距离出行轨迹构建的交通网络. 因此可推断得出, 当 POI 分类较为粗泛时, 长距离轨迹能区分不同功能区域, 因此长距离轨迹构建的网络效果较好, 当 POI 分类较为细致时, 融合长、短距离

的出行轨迹能够捕捉到更加细粒度并且相对完整的出行模式, 因此构建的交通网络所涵盖的信息更加丰富, 效果更优.

表 3 不同向量化方法的实验结果对比 (轮廓系数)

模型	$K=5$	$K=10$	$K=15$	$K=20$
融合模型	0.5137	0.3254	0.4984	0.3035
Node2Vec	0.3466	0.1579	0.193	0.265
DeepMove	0.7991	0.4988	0.4525	0.2698

表 4 不同采样方法的实验结果对比 (轮廓系数)

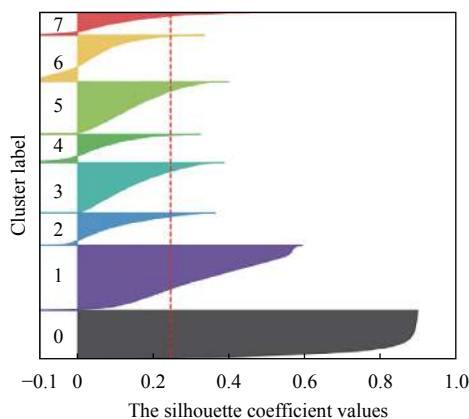
采样方法	$K=5$	$K=10$	$K=15$	$K=20$
路网采样	0.5137	0.3254	0.4984	0.3035
随机采样	0.4701	0.4369	0.3923	0.2853

表 5 不同数据源的实验结果对比 (轮廓系数)

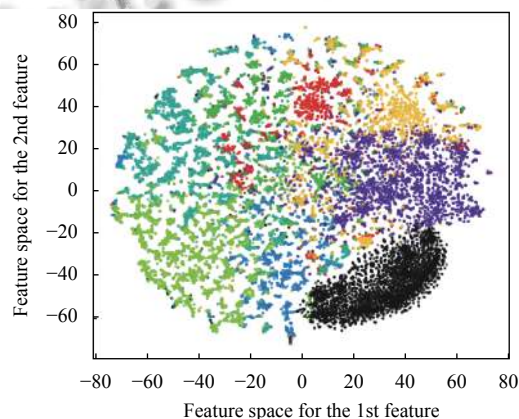
交通网络数据源	$K=5$	$K=10$	$K=15$	$K=20$
多源数据-长、短距离融合	0.5137	0.3254	0.4984	0.3035
共享单车轨迹-短距离出行	0.5139	0.4122	0.4013	0.2811
公共交通线路-长距离出行	0.6601	0.6317	0.4488	0.2543

3.2.2 空间位置聚类分析

首先, 利用 K-means ($K=8$) 对空间位置向量进行聚类分析, 其中每个类别的轮廓系数如图 6(a) 所示. 为了更加直观的反映聚类结果, 使用 t-SNE^[19] 方法将空间位置向量从 128 维空间降维至低维空间, 降维后的空间向量化表示的聚类结果如图 6(b) 所示, 不同的颜色表明不同的类. 本文结合低维空间向量化表示的可视化映射图作为一种直观的评价方式. 通过观察图 6 可发现, 使用本文提出的空间向量化表示方法得到的空间位置向量, 映射在低维空间后, 有较为明显的类间边界, 能够验证本文提出的空间向量化表示方法的有效性.



(a) 空间位置向量聚类的轮廓系数分布图



(b) 经过 t-SNE 降维后的空间向量化表示聚类可视化视图

图 6 聚类分析结果

4 结论与展望

本文提出了一种基于图神经网络的空间向量化表

示方法. 基于共享单车轨迹数据与公共交通线路数据, 将长、短距离出行轨迹进行匹配连接, 构建大规模交

通网络,该交通网络能够覆盖多种出行模式.提出了融合POI与轨迹信息的空间向量化表示方法,综合位置自身的空间特征与其邻域的特征,并优化节点采样方法,提高了空间向量化表示的表达能力.以北京市的共享单车轨迹数据与公共交通路网数据为实例,经验证本文提出的空间向量化表示方法能够综合空间特征、邻域特征与居民出行模式,该向量可作为空间特征用于交流流量预测,交通调度与管理,地理画像,位置推荐等实际应用中.

在未来的工作中,将进一步研究融合多源数据,例如出租车,网约车数据,以构建大规模的交通网络.以及当网络规模增大时,如何提升模型性能,使其能够处理更大规模的交通网络.

参考文献

- 1 Liu ZD, Li ZJ, Li M, *et al.* Mining road network correlation for traffic estimation via compressive sensing. *IEEE Transactions on Intelligent Transportation Systems*, 2016, 17(7): 1880–1893. [doi: [10.1109/TITS.2016.2514519](https://doi.org/10.1109/TITS.2016.2514519)]
- 2 Yan B, Janowicz K, Mai GC, *et al.* From ITDL to Place2Vec: Reasoning about place type similarity and relatedness by learning embeddings from augmented spatial contexts. *Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. New York, NY, USA. 2017. 1–10.
- 3 Jin JQ, Xiao ZJ, Qiu Q, *et al.* A GeoHash based Place2Vec model. *IGARSS 2019–2019 IEEE International Geoscience and Remote Sensing Symposium*. Yokohama, Japan. 2019. 3344–3347.
- 4 Ying R, He RN, Chen KF, *et al.* Graph convolutional neural networks for web-scale recommender systems. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. New York, NY, USA. 2018. 974–983.
- 5 Cai HY, Zheng VW, Chang KCC. A comprehensive survey of graph embedding: Problems, techniques, and applications. *IEEE Transactions on Knowledge and Data Engineering*, 2018, 30(9): 1616–1637. [doi: [10.1109/TKDE.2018.2807452](https://doi.org/10.1109/TKDE.2018.2807452)]
- 6 Goyal P, Ferrara E. Graph embedding techniques, applications, and performance: A survey. *Knowledge-Based Systems*, 2018, 151: 78–94. [doi: [10.1016/j.knsys.2018.03.022](https://doi.org/10.1016/j.knsys.2018.03.022)]
- 7 Chai D, Wang LY, Yang Q. Bike flow prediction with multi-graph convolutional networks. *Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. New York, NY, USA. 2018. 397–400.
- 8 Diao ZL, Wang X, Zhang DF, *et al.* Dynamic spatial-temporal graph convolutional neural networks for traffic forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*. Honolulu, HI, USA. 2019. 890–897.
- 9 Wang YD, Yin HZ, Chen HX, *et al.* Origin-destination matrix prediction via graph convolution: A new perspective of passenger demand modeling. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. New York, NY, USA. 2019. 1227–1235.
- 10 Zhou Y, Huang Y. DeepMove: Learning place representations through large scale movement data. *Proceedings of 2018 IEEE International Conference on Big Data*. Seattle, WA, USA. 2018.
- 11 Balkić Z, Šoštarić D, Horvat G. GeoHash and UUID identifier for multi-agent systems. *Proceedings of the 6th KES international conference on Agent and Multi-Agent Systems*. Berlin, Germany. 2012. 290–298.
- 12 Hamilton W, Ying ZT, Leskovec J. Inductive representation learning on large graphs. *Advances in Neural Information Processing Systems*. Long Beach, CA, USA. 2017. 1024–1034.
- 13 Mikolov T, Sutskever I, Chen K, *et al.* Distributed representations of words and phrases and their compositionality. *Proceedings of the 26th International Conference on Neural Information Processing Systems*. Red Hook, NY, USA. 2013. 3111–3119.
- 14 Ramos J. Using TF-IDF to determine word relevance in document queries. *Proceedings of the First Instructional Conference on Machine Learning*. Honolulu, HI, USA. 2003. 890–897.
- 15 Pennington J, Socher R, Manning C. Glove: Global vectors for word representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar. 2014. 1532–1543.
- 16 Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 1987, 20: 53–65. [doi: [10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)]
- 17 Grover A, Leskovec J. Node2Vec: Scalable feature learning for networks. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA. 2016. 855–864.
- 18 薛冰, 李京忠, 肖骁, 等. 基于兴趣点(POI)大数据的人地关系研究综述: 理论、方法与应用. *地理与地理信息科学*, 2019, 35(6): 51–60. [doi: [10.3969/j.issn.1672-0504.2019.06.009](https://doi.org/10.3969/j.issn.1672-0504.2019.06.009)]
- 19 Van Der Maaten L, Hinton G. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 2008, 9(11): 2579–2605.