A Related Works

A.1 Mobility Data Mining

Location-based services have given rise to a new and promising research topic known as mobility data mining, which has led to the emergence of three significant tasks that contribute to enhancing the quality of services: next location prediction (LP), next time prediction (TP), and trajectory user link (TUL). Recent studies have confirmed that deep learning techniques, specifically recurrent neural networks (RNNs) and attention mechanisms, are highly effective in capturing sequential and periodic patterns of human mobility. By combining deep learning techniques, researchers have made significant advancements in capturing both the sequential and periodic patterns of human mobility. The core of these models is the modeling of check-in sequences, which leads to improved accuracy in location prediction and trajectory analysis.

LP aims to anticipate a user's future location based on their historical movement. Several notable models have emerged as prominent approaches in LP. DeepMove [9] leverages RNNs and attention mechanisms to capture the spatial-temporal intentions in users' location data and predict their next destination. STAN [29] introduces a spatial-temporal attention network that incorporates spatial and temporal contexts for accurate prediction. LSTPM [39] focuses on long and short-term patterns in user trajectory using an attention-based LSTM [18] model. SERM [47] utilizes an encoder-decoder architecture with a spatial-temporal residual network to capture user preferences and predict future locations. PLSPL [43] trains two LSTM models for location- and category-based sequences to capture the user's preference. LightMove [20] designs neural ordinary differential equations to enhance robustness against sparse or incorrect inputs. HMT-GRN [22] alleviates the data sparsity problem by learning different User-Region matrices of lower sparsities in a multitask setting. Graph-Flashback [35] constructs a spatial-temporal knowledge graph to enhance the representation of POIs, having a great advantage when dealing with the historical sequence input of the same length. GETNext [46] introduces a user-agnostic global trajectory flow map as a means to leverage the abundant collaborative signals.

TUL is a significant task that focuses on establishing connections between different trajectories, facilitating the analysis of user movement patterns, and uncovering valuable insights about their behavior. Notable models have been specifically developed to address the challenge of predicting trajectory links. TULER [12] takes advantage of advanced algorithms to establish links between trajectories, allowing for a comprehensive understanding of user movement patterns. TULVAE [55] uses latent variables to model the variability in trajectories, capturing hierarchical and structural semantics and improving the identification and linking performance. MoveSim [10] simulates human mobility using a generative adversarial framework that incorporates attention mechanisms to capture complex spatial-temporal transitions in human mobility. DeepTUL [30] utilizes deep learning techniques to extract representations from trajectory data and facilitate the prediction of trajectory links. S2TUL [6] utilizes graph convolutional networks and sequential neural networks to capture trajectory relationships and intra-trajectory information. GNNTUL [53] employs graph neural networks for human mobility and associates the traces with users on social networks.

TP focuses on estimating the time at which a user is likely to visit their next location. To accomplish this, it is common practice to use intensity functions to represent the rate or density of event occurrences, various models have been developed to model the intensity function and make accurate time predictions effectively. Modeling the intensity function using RNNs or attention mechanisms is a common approach for predicting the occurrence of events. IFLTPP [37] approximates any distribution of inter-event times using normalizing flows and mixture distributions. RMTPP [8] utilizes RNNs to model the intensity function. SAHP [52] combines the Hawkes process with self-attention mechanisms to capture the temporal dependencies and spatial influences in event sequences. THP [57] combines the Hawkes process with transformer-based architectures to capture temporal dependencies in event sequences. NSTPP [15] utilizes neural ODEs to model discrete events in continuous time and space, enabling the learning of complex distributions in spatial and temporal domains. IMTPP [16] models the generative processes of observed and missing events and utilizes unsupervised modeling and inference methods for time prediction. DSTPP [48] proposes a novel parameterization framework that uses diffusion models to learn complex joint distributions.

In recent years, many sequence representation models have become a hot topic, including those for natural language sequences and spatial-temporal trajectory sequences. For natural language sequences, SimCSE [13] is a classical sequence representation language model that utilizes a straightforward contrastive learning framework. It employs standard dropout as noise, making it easy to implement. However, it cannot effectively distinguish between hard negative samples or semantically similar sequences. Thus, VaSCL [49] leverages neighborhood information to generate virtual augmentations, improving the quality of data transformations compared to traditional methods. Inspired by models for natural language sequence representation, researchers in the spatial-temporal domain have also conducted similar studies on spatial-temporal sequences. SML [54] first employs self-supervised contrastive learning to effectively manage sparse and noisy trajectory data, enhancing trajectory representation through spatial-temporal data augmentation. However, the spatial-temporal augmentation in SML needs to be manually specified, which is relatively cumbersome. Therefore, ReMVC [50] effectively addresses this issue by adopting a cross-view contrastive approach. It uses contrastive learning to extract meaningful region representations, improving intra-view and inter-view learning. And CACSR [14] effectively adds adversarial perturbations and automated data augmentation, enhancing contrastive training processing. While it lacks interpretability and does not understand the semantic information of time and space from the actual semantics of human activities.

In summary, it can be seen that whether it is end-to-end models designed for specific tasks or spatiotemporal sequence representation learning models, their understanding of human activities is not yet deep enough, and they lack a profound understanding of the essence and spatial-temporal patterns behind human activities.

References

[1] Marah Abdin, Sam Ade Jacobs, Ammar Ahmad Awan, Jyoti Aneja, Ahmed Awadallah, Hany Awadalla, Nguyen Bach, Amit Bahree, Arash Bakhtiari, Harkirat Behl, et al. Phi-3 technical report: A highly capable language model locally on your phone. *arXiv preprint arXiv:2404.14219*, 2024.

[2] Stella Biderman, Hailey Schoelkopf, Quentin Gregory Anthony, Herbie Bradley, Kyle O'Brien, Eric Hallahan, Mohammad Aflah Khan, Shivanshu Purohit, USVSN Sai Prashanth, Edward Raff, et al. Pythia: A suite for analyzing large language models across training and scaling. In *International Conference on Machine Learning*, pages 2397–2430. PMLR, 2023.

[3] Defu Cao, Furong Jia, Sercan O Arik, Tomas Pfister, Yixiang Zheng, Wen Ye, and Yan Liu. Tempo: Prompt-based generative pre-trained transformer for time series forecasting. *arXiv preprint arXiv:2310.04948*, 2023.

[4] Ching Chang, Wen-Chih Peng, and Tien-Fu Chen. Llm4ts: Two-stage fine-tuning for time-series forecasting with pre-trained llms. *arXiv preprint arXiv:2308.08469*, 2023.

[5] Tri Dao et al. Tinyllama: An open-source small language model. *arXiv preprint arXiv:2401.02385*, 2024.

[6] Liwei Deng, Hao Sun, Yan Zhao, Shuncheng Liu, and Kai Zheng. S2tul: A semi-supervised framework for trajectory-user linking. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*, WSDM '23, page 375–383. Association for Computing Machinery, 2023.

[7] Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, and Luke Zettlemoyer. Qlora: Efficient finetuning of quantized llms. *Advances in Neural Information Processing Systems*, 2023.

[8] Nan Du, Hanjun Dai, Rakshit Trivedi, Utkarsh Upadhyay, Manuel Gomez-Rodriguez, and Le Song. Recurrent marked temporal point processes: Embedding event history to vector. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, page 1555–1564. Association for Computing Machinery, 2016.

[9] Jie Feng, Yong Li, Chao Zhang, Funing Sun, Fanchao Meng, Ang Guo, and Depeng Jin. Deepmove: Predicting human mobility with attentional recurrent networks. In *Proceedings of the 2018 world wide web conference*, pages 1459–1468, 2018.

[10] Jie Feng, Zeyu Yang, Fengli Xu, Haisu Yu, Mudan Wang, and Yong Li. Learning to simulate human mobility. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 3426–3433, 2020.

[11] Xue-Yong Fu, Md Tahmid Rahman Laskar, Elena Khasanova, Cheng Chen, and Shashi Bhushan TN. Tiny titans: Can smaller large language models punch above their weight in the real world for meeting summarization? *arXiv preprint arXiv:2402.00841*, 2024.

[12] Qiang Gao, Fan Zhou, Kunpeng Zhang, Goce Trajcevski, Xucheng Luo, and Fengli Zhang. Identifying human mobility via trajectory embeddings. In *IJCAI*, volume 17, pages 1689–1695, 2017.

[13] Tianyu Gao, Xingcheng Yao, and Danqi Chen. Simcse: Simple contrastive learning of sentence embeddings. *arXiv preprint arXiv:2104.08821*, 2021.

[14] Letian Gong, Youfang Lin, Shengnan Guo, Yan Lin, Tianyi Wang, Erwen Zheng, Zeyu Zhou, and Huaiyu Wan. Contrastive pre-training with adversarial perturbations for check-in sequence representation learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(4):4276–4283, Jun. 2023.

[15] Vinayak Gupta, Srikanta Bedathur, and Abir De. Learning temporal point processes for efficient retrieval of continuous time event sequences. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 4005–4013, 2022.

[16] Vinayak Gupta, Abir De, Sourangshu Bhattacharya, and Srikanta Bedathur. Learning temporal point processes with intermittent observations. In *Proc. of the 24th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2021.

[17] Xiaoxin He, Xavier Bresson, Thomas Laurent, and Bryan Hooi. Explanations as features: Llm-based features for text-attributed graphs. *arXiv preprint arXiv:2305.19523*, 2023.

[18] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[19] Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. In *ICLR*, 2022.

[20] Jinsung Jeon, Soyoung Kang, Minju Jo, Seunghyeon Cho, Noseong Park, Seonghoon Kim, and Chiyoung Song. Lightmove: A lightweight next-poi recommendation fortaxicab rooftop advertising. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, CIKM '21, page 3857–3866. Association for Computing Machinery, 2021.

[21] Ming Jin, Shiyu Wang, Lintao Ma, Zhixuan Chu, James Y Zhang, Xiaoming Shi, Pin-Yu Chen, Yuxuan Liang, Yuan-Fang Li, Shirui Pan, et al. Time-llm: Time series forecasting by reprogramming large language models. *arXiv preprint arXiv:2310.01728*, 2023.

[22] Nicholas Lim, Bryan Hooi, See-Kiong Ng, Yong Liang Goh, Renrong Weng, and Rui Tan. Hierarchical multi-task graph recurrent network for next poi recommendation. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, page 1133–1143, 2022.

[23] Shengchao Liu, Weili Nie, Chengpeng Wang, Jiarui Lu, Zhuoran Qiao, Ling Liu, Jian Tang, Chaowei Xiao, and Animashree Anandkumar. Multi-modal molecule structure–text model for text-based retrieval and editing. *Nature Machine Intelligence*, 5(12):1447–1457, 2023.

[24] Xin Liu, Yong Liu, Karl Aberer, and Chunyan Miao. Personalized point-of-interest recommendation by mining users' preference transition. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, pages 733–738, 2013.

[25] Xin Liu, Daniel McDuff, Geza Kovacs, Isaac Galatzer-Levy, Jacob Sunshine, Jiening Zhan, Ming-Zher Poh, Shun Liao, Paolo Di Achille, and Shwetak Patel. Large language models are few-shot health learners. *arXiv preprint arXiv:2305.15525*, 2023.

[26] Yong Liu, Guo Qin, Xiangdong Huang, Jianmin Wang, and Mingsheng Long. Autotimes: Autoregressive time series forecasters via large language models. *arXiv preprint arXiv:2402.02370*, 2024.

[27] Yong Liu, Wei Wei, Aixin Sun, and Chunyan Miao. Exploiting geographical neighborhood characteristics for location recommendation. In *Proceedings of the 23rd ACM international conference on conference on information and knowledge management*, pages 739–748, 2014.

[28] Kevin Lu, Aditya Grover, Pieter Abbeel, and Igor Mordatch. Frozen pretrained transformers as universal computation engines. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pages 7628–7636, 2022.

[29] Yingtao Luo, Qiang Liu, and Zhaocheng Liu. Stan: Spatio-temporal attention network for next location recommendation. In *Proceedings of the Web Conference 2021*, pages 2177–2185, 2021.

[30] Congcong Miao, Jilong Wang, Heng Yu, Weichen Zhang, and Yinyao Qi. Trajectory-user linking with attentive recurrent network. In *Proceedings of the 19th international conference on autonomous agents and multiagent systems*, pages 878–886, 2020.

[31] OpenAI. Gpt-4 technical report, 2023.

[32] Ziqi Pang, Ziyang Xie, Yunze Man, and Yu-Xiong Wang. Frozen transformers in language models are effective visual encoder layers. *arXiv preprint arXiv:2310.12973*, 2023.

[33] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*, pages 8748–8763. PMLR, 2021.

[34] Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, et al. Improving language understanding by generative pre-training. 2018.

[35] Xuan Rao, Lisi Chen, Yong Liu, Shuo Shang, Bin Yao, and Peng Han. Graph-flashback network for next location recommendation. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '22, page 1463–1471, New York, NY, USA, 2022. Association for Computing Machinery.

[36] Xubin Ren, Wei Wei, Lianghao Xia, Lixin Su, Suqi Cheng, Junfeng Wang, Dawei Yin, and Chao Huang. Representation learning with large language models for recommendation. *CoRR*, abs/2310.15950, 2023.

[37] Oleksandr Shchur, Marin Biloš, and Stephan Günnemann. Intensity-free learning of temporal point processes. *International Conference on Learning Representations (ICLR)*, 2020.

[38] Chenxi Sun, Yaliang Li, Hongyan Li, and Shenda Hong. Test: Text prototype aligned embedding to activate llm's ability for time series. *arXiv preprint arXiv:2308.08241*, 2023.

[39] Ke Sun, Tieyun Qian, Tong Chen, Yile Liang, Quoc Viet Hung Nguyen, and Hongzhi Yin. Where to go next: Modeling long-and short-term user preferences for point-of-interest recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 214–221, 2020.

[40] Zhou Tian, Niu Peisong, Wang Xue, Liang Sun, Rong Jin, and Patchtst Timesnet. One fits all: Power general time series analysis by pretrained lm. *CoRR*, abs/2302.11939, 2023.

[41] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.

[42] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, pages 5998–6008, 2017.

[43] Yuxia Wu, Ke Li, Guoshuai Zhao, and QIAN Xueming. Personalized long-and short-term preference learning for next poi recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 2020.

[44] Dingqi Yang, Daqing Zhang, Zhiyong Yu, and Zhu Wang. A sentiment-enhanced personalized location recommendation system. In *Proceedings of the 24th ACM Conference on Hypertext and Social Media*, pages 119–128. ACM, 2013.

[45] Dingqi Yang, Daqing Zhang, Zhiyong Yu, and Zhiwen Yu. Fine-grained preference-aware location search leveraging crowdsourced digital footprints from lbsns. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, pages 479–488. ACM, 2013.

[46] Song Yang, Jiamou Liu, and Kaiqi Zhao. Getnext: trajectory flow map enhanced transformer for next poi recommendation. In *Proceedings of the 45th International ACM SIGIR Conference on research and development in information retrieval*, pages 1144–1153, 2022.

[47] Di Yao, Chao Zhang, Jianhui Huang, and Jingping Bi. Serm: A recurrent model for next location prediction in semantic trajectories. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 2411–2414, 2017.

[48] Yuan Yuan, Jingtao Ding, Chenyang Shao, Depeng Jin, and Yong Li. Spatio-temporal diffusion point processes. *arXiv preprint arXiv:2305.12403*, 2023.

[49] Dejiao Zhang, Wei Xiao, Henghui Zhu, Xiaofei Ma, and Andrew Arnold. Virtual augmentation supported contrastive learning of sentence representations. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 864–876, 2022.

[50] Liang Zhang, Cheng Long, and Gao Cong. Region embedding with intra and inter-view contrastive learning. *IEEE Transactions on Knowledge and Data Engineering*, 2022.

[51] Peiyuan Zhang, Guangtao Zeng, Tianduo Wang, and Wei Lu. Tinyllama: An open-source small language model, 2024.

[52] Qiang Zhang, Aldo Lipani, Omer Kirnap, and Emine Yilmaz. Self-attentive Hawkes process. In Hal Daumé III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 11183–11193. PMLR, 13–18 Jul 2020.

[53] Fan Zhou, Shupei Chen, Jin Wu, Chengtai Cao, and Shengming Zhang. Trajectory-user linking via graph neural network. In *ICC 2021 - IEEE International Conference on Communications*, pages 1–6, 2021.

[54] Fan Zhou, Yurou Dai, Qiang Gao, Pengyu Wang, and Ting Zhong. Self-supervised human mobility learning for next location prediction and trajectory classification. *Knowledge-Based Systems*, 228:107214, 2021.

[55] Fan Zhou, Qiang Gao, Goce Trajcevski, Kunpeng Zhang, Ting Zhong, and Fengli Zhang. Trajectory-user linking via variational autoencoder. In *IJCAI*, pages 3212–3218, 2018.

[56] Tian Zhou, Peisong Niu, Liang Sun, Rong Jin, et al. One fits all: Power general time series analysis by pretrained lm. *Advances in neural information processing systems*, 36, 2024.

[57] Simiao Zuo, Haoming Jiang, Zichong Li, Tuo Zhao, and Hongyuan Zha. Transformer hawkes process. In *Proceedings of the 37th International Conference on Machine Learning*, ICML'20. JMLR.org, 2020.