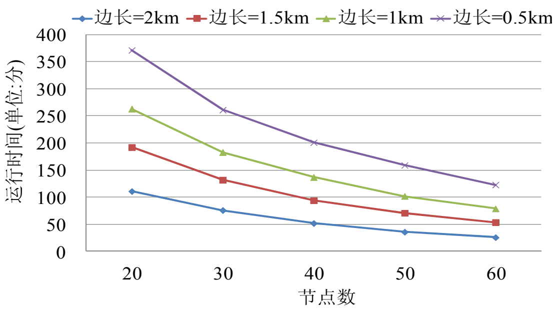


Fig.6 Running time of GPMA
图 6 不同网格粒度下的时间性能

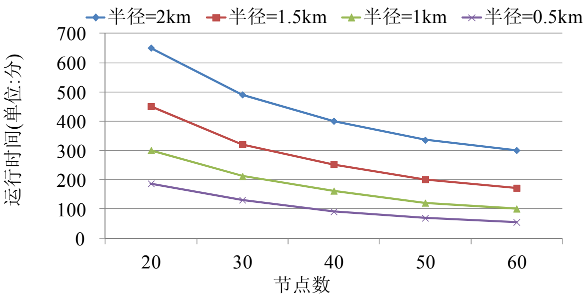


Fig.7 Running time of SPMA
图 7 不同半径下的时间性能

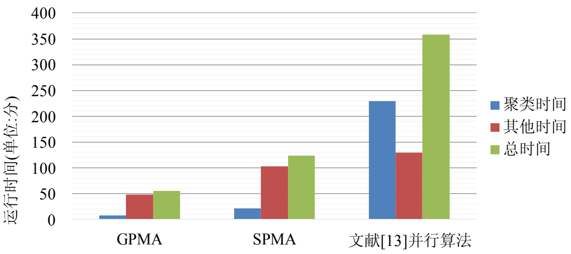


Fig.8 Comparison of running time
图 8 算法运行时间对比图

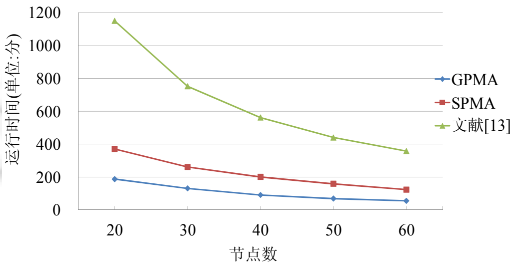


Fig.9 Scalability of proposed algorithms
图 9 算法扩展性

6 结束语

本文介绍了从海量低质的手机轨迹数据中挖掘用户重要位置的通用框架,以及基于该框架的两种算法:GPMA 和 SPMA,进一步从 3 个方面修正所挖掘出的重要位置.本文所介绍的两种算法均是先从用户的活动记录中找出可能包含重要位置的区域,再对区域进行聚类,最后从聚类结果中分析出用户的重要位置.两者的区别是,GPMA 使用网格区域来表示可能包含重要位置的区域,而 SPMA 使用基站覆盖范围来表示.相比已有工作,本文围绕重要位置挖掘这一核心点,考虑了如何提高低质轨迹数据的可用性.同时,还融合其他数据以提高结果的精度,并针对两类特殊人群,设计了退休等无工作地人群的发现算法和夜间工作人群挖掘算法,以提高结果的合理性.实验结果表明,本文算法相比已有算法,其结果的 F_1 值更高,位置结果更精确.同时,本文所提两种算法中 GPMA 的时间性能更好,SPMA 的准确度和精确度更高.

References:

[1] Gonzalez MC, Hidalgo CA, Barabasi AL. Understanding individual human mobility patterns. Nature, 2008,453(7196):779−782. [doi: 10.1038/nature06958]

[2] Song C, Qu Z, Blumm N, Barabasi AL. Limits of predictability in human mobility. Science, 2010,327(5968):1018−1021. [doi: 10.1126/science.1177170]

[3] Song C, Koren T, Wang P, Barabasi AL. Modelling the scaling properties of human mobility. Nature Physics, 2010,6(10):818−823. [doi: 10.1038/nphys1760]

[4] Li Z, Ding B, Han J, Roland K, Peter N. Mining periodic behaviors for moving objects. In: Proc. of the 16th Int'l Conf. on Knowledge Discovery and Data Mining (SIGKDD 2010). New York: ACM Press, 2010. 1099−1108. [doi: 10.1145/1835804.1835942]

[5] Ashbrook D, Starner T. Using GPS to learn significant locations and predict movement across multiple users. Personal and Ubiquitous Computing, 2003,7(5):275−286. [doi: 10.1007/s00779-003-0240-0]

[6] Sadilek A, Kautz H, Bigham JP. Finding your friends and following them to where you are. In: Proc. of the 5th ACM Int'l Conf. on Web Search and Data Mining. New York: ACM Press, 2012. 723−732. [doi: 10.1145/2124295.2124380]

[7] Cho E, Myers SA, Leskovec J. Friendship and mobility: User movement in location-based social networks. In: Proc. of the 17th Int'l Conf. on Knowledge Discovery and Data Mining (SIGKDD 2011). New York: ACM Press, 2011. 1082−1090. [doi: 10.1145/2020408.2020579]

[8] Bao J, Zheng Y, Mokbel MF. Location-Based and preference-aware recommendation using sparse geo-social networking data. In: Proc. of the 20th Int'l Conf. on Advances in Geographic Information Systems (SIGSPATIAL 2012). New York: ACM Press, 2012. 199−208. [doi: 10.1145/2424321.2424348]

[9] Yuan NJ, Wang Y, Zhang F, Xie X. Reconstructing individual mobility from smart card transactions: A space alignment approach. In: Proc. of the 13th Int'l Conf. on Data Mining (ICDM 2013). IEEE, 2013. 877−886. [doi: 10.1109/ICDM.2013.37]

[10] Wu WQ. 本市人均拥有手机 1.4 部.2014. http://www.jfdaily.com/shanghai/bw/201406/t20140625_481869.html

[11] Isaacman S, Becker R, Cáceres R, Kobourov S, Martonosi M, Rowland J, Varshavsky A. Identifying important places in people's lives from cellular network data. In: Pervasive Computing. Berlin, Heidelberg: Springer-Verlag, 2011. 133−151. [doi: 10.1007/978-3-642-21726-5_9]

[12] Li JZ, Liu XM. An important aspect of big data: Data usability. Journal of Computer Research and Development, 2013, 50(6):1147−1162 (in Chinese with English abstract).

[13] Cheng R, Chen J, Xie X. Cleaning uncertain data with quality guarantees. In: Proc. of the Very Large Data Bases Endowment (VLDB 2008). 2008,1(1):722−735. [doi: 10.14778/1453856.1453935]

[14] Scellato S, Noulas A, Lambiotte R, Mascolo C. Socio-Spatial properties of online location-based socialnetworks. In: Proc. of the 5th Int'l Conf. on Weblogs and Social Media. Palo Alto: AAAI Press, 2011,11:329−336. http://www.icwsm.org/2011/index.php

[15] Chen J, Hu B, Zuo X, Yue Y. Personal profile mining based on mobile phone location data. Geomatics and Information Science of Wuhan University, 2014,39(6):734−738.

[16] Dean J, Ghemawat S. MapReduce: Simplified data processing on large clusters. Communications of the ACM, 2008,51(1):107−113. [doi: 10.1145/1327452.1327492]

[17] Ester M, Kriegel HP, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. In: Simoudis E, Han JW, Fayyad U, eds. Proc. of the 2nd Int'l Conf. on Knowledge Discovery and Data Mining (SIGKDD 1996). Palo Alto: AAAI Press, 1996,96:226−231. http://www.aaai.org/Press/Proceedings/kdd96.php

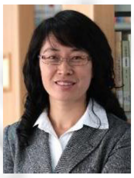
附中文参考文献:

[10] 吴卫群.本市人均拥有手机 1.4 部.2014. http://www.jfdaily.com/shanghai/bw/201406/t20140625_481869.html

[12] 李建中,刘显敏.大数据的一个重要方面:数据可用性.计算机研究与发展,2013,50(6):1147−1162.

[15] 陈佳,胡波,左小清,乐阳.利用手机定位数据的用户特征挖掘.武汉大学学报•信息科学版,2014,39(6):734−738.

章志刚(1988-),男,江苏海安人,博士,主要研究领域为分布式计算,基于位置的服务.



王晓玲(1975-),女,教授,博士生导师,CCF 会员,主要研究领域为半结构化数据管理,数据服务,隐私保护.

金澈清(1977-),男,博士,教授,博士生导师,CCF 会员,主要研究领域为海量数据管理,包括基于位置的服务,数据质量,不确定数据管理.

周傲英(1965-),男,博士,教授,博士生导师,CCF 杰出会员,主要研究领域为 Web 数据管理,数据密集型计算,内存集群计算,大数据基准测试和性能优化.