以看出一致性处理算法的效果，Cons-QT 算法的表现明显优于 Noisy-QT 算法. 随着查询面积的增大，一致性处理算法降低的相对误差越来越小. 随着 Size(Q)的增加，相对误差逐渐减少. 这是由于查询面积越大，其包含的噪音数据的比例就越小. 从实验结果中可以看出，本文提出的方法在不同的查询面积上是稳定的. 按照图 7 的数据趋势来看，随着 Size(Q)的增加，当其大于 50%时，相对误差值会更小.

5.2.2 噪音 R-树的数据可用性实验

本实验在隐私预算 ε 取不同值的情况下测试算法的相对误差. 通过改变查询路段的长度随机生成了 10000 个查询，查询路段的长度 Length(L)设为 5 到 20 个结点不等. 根据 Length(L)的大小将查询分成 4 个不同的组. 每个相对误差值都是执行 10 次查询的平均值. 隐私预算 ε 取值为 0.5 到 1.5 时. 实验结果如图 8 所示.

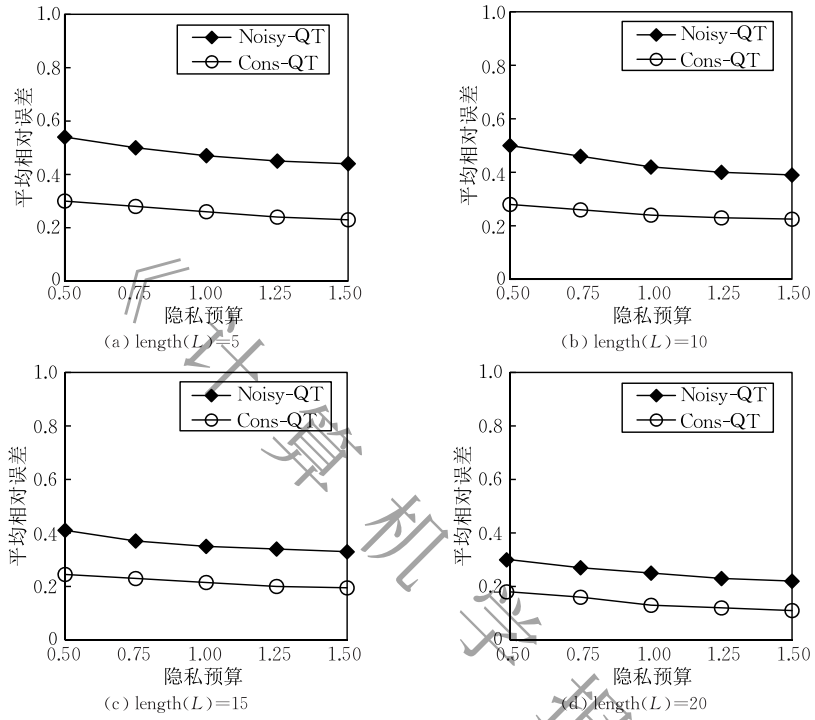

图 8 噪音 R-树算法的平均相对误差

图 8 中，x-轴表示不同的隐私预算值，y-轴表示平均相对误差值. 查询长度 length(Q)分别设为 5，10，15，20 个路网结点. 平均相对误差随着隐私预算 ε 的增大而减小，与前述实验相同：根据差分隐私的性质，隐私预算 ε 越大，构建噪音 R-树时添加的噪音量越小，查询结果越精确. 从图 6 中可以看出数据一致性处理算法的作用，其明显提高了数据的可用性. 随着 length(Q)的增加，相对误差也在减小，这是由于查询范围越大，包含的噪音比例相对较小. 从实验结果中可以看出，本文提出的方法在数据可用性上是稳定的.

5.3 算法运行效率

本实验测试了 4 个算法的运行效率. 算法的运行时间长短与数据集大小 $|D|$ 相关，因此实验测试在不同数据集大小时的运行时间. 图 9 中，x-轴表示数据集大小，y-轴表示算法的运行时间. 本实验的结果是在隐私预算 $\varepsilon=1.0$ 时的运行时间. 图 9(a)的数据集是在 Gaussian 数据集中随机截取大小不同的子集作为数据集进行查询，以测试算法的运行时间，查询区域 Size(Q)的大小取值为总数据集的 15%；图 9(b)的数据集是从 OLDEN 中随机截取相应大小的子集，查询长度 length(Q)=10. 图 9 中的结果是多次运行算法得到的平均运行时间.

从图 9 的实验结果可以看出，算法的运行时间随着数据集的增加近似成线性增长. 经过一致性处理的算法比简单的噪音树算法运行时间长，但运行时间增加的幅度有限. 这也说明使用经过一致性处理的算法有良好的效率和数据可用性. 实验结果表明，运行时间与数据集大小近似成正比. 因此，两个算法可以运行在较大规模的数据集上，算法有较好的可扩展性.

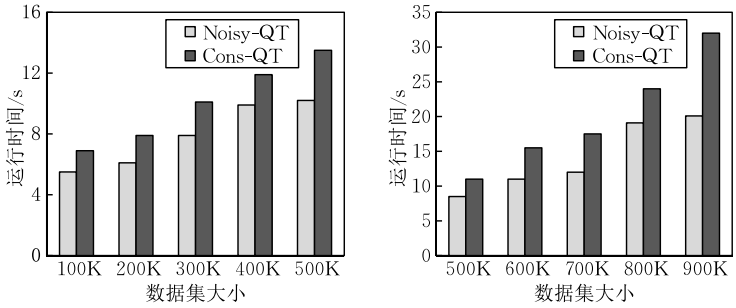

图 9　算法运行时间测试

6　总结与展望

本文提出了一种满足差分隐私的轨迹数据发布方法.针对自由空间移动对象的运行特点,提出了支持空间范围计数查询的噪音四分树构建方法,及基于最大运行速度限制的一致性处理算法;在路网空间提出了基于路段计数查询的噪音 R-树构建方法,及在路网限制条件下的一致性处理算法.本文通过理论证明了上述两个算法满足 ε-差分隐私,且生成的数据在空间范围查询上可达到(ε,δ)-可用性.最后,本文在模拟数据集上进行了数据可用性和算法可扩展性实验,实验结果证明本文提出的方法具有较低的相对查询误差率及良好的可扩展性,算法可用在较大规模数据集上.

本文的研究工作是基于数据收集者是"可信"的假设之上的.然而,越来越多的实际案例表明:所谓的"可信"第三方数据收集者并不可靠,许多数据收集者有意或无意地泄露用户的原始数据,造成大量用户的个人隐私数据泄露.在今后的研究中,可以考虑采用本地化差分隐私技术(Local differential privacy)进行位置或其它个人敏感数据的收集和发布[22],使得第三方收集者也无法获取原始数据,从源头上遏制由个人数据泄露造成的用户隐私泄露.此外,在大数据环境中,个人位置数据和其它个人数据关联起来造成的个人隐私泄露风险量化和攻击模型建模也是值得研究和关注的问题.

参　考　文　献

[1] Dwork C. Differential privacy//Proceedings of the 33rd International Colloquium on Automata, Languages and Programming(ICALP'06). Venice, Italy, 2006: 1-12

[2] Zhang Xiao-Jian, Meng Xiao-Feng. Differential privacy in data publication and analysis. Chinese Journal of Computers, 2014, 37(4): 927-949(in Chinese)
(张啸剑,孟小峰. 面向数据发布和分析的差分隐私保护. 计算机学报, 2014, 37(4): 927-949)

[3] Cormode G, Procopiuc C, Srivastava D, et al. Differentially private spatial decompositions//Proceedings of the IEEE 28th International Conference on Data Engineering (ICDE'12). Washington, USA, 2012: 20-31

[4] Xiao Y, Xiong L, Yuan C. Differentially private data release through multidimensional partitioning//Proceedings of the 7th VLDB Workshop on Secure Data Management. Singapore, 2010: 150-168

[5] Chen R, Fung B C M, Desai B C, Sossou N M. Differentially private transit data publication: A case study on the Montreal transportation system//Proceedings of the 18th ACM SIGKDD Conference on Knowledge Discovery and Data Mining(KDD'12). Beijing, China, 2012: 213-221

[6] Hua J, Gao Y, Zhong S. Differentially private publication of general time-serial trajectory data//Proceedings of the 2015 IEEE Conference on Computer Communications(INFOCOM). Hong Kong, China, 2015: 549-557

[7] Xiao Y, Xiong L. Protecting locations with differential privacy under temporal correlations//Proceedings of the ACM SIGSAC Conference. Denver, USA, 2015: 1298-1309

[8] Yigitoglu E, Damiani L M, Abul O, Silvestri C. Privacy-preserving sharing of sensitive semantic locations under road-network constraints//Proceedings of the 13th IEEE International Conference on Mobile Data Management (MDM'12). Bengaluru, India, 2012: 186-195

[9] Silvestri C, Yigitoglu E, Damiani L M, Abul O. SAWLnet: Sensitivity aware location cloaking on road-networks//Proceedings of the 13th IEEE International Conference on Mobile Data Management (MDM'12). Bengaluru, India, 2012: 336-339

[10] Peng T, Liu Q, Meng D, Wang G. Collaborative trajectory privacy preserving scheme in location-based services. Information Science, 2017, 387: 165-179

[11] Theodorakopoulos G, Shokri R, Troncoso C, et al. Prolonging the hide-and-seek game: Optimal trajectory privacy for location-based services//Proceedings of the 13th Workshop on Privacy in the Electronic Society (WPES 2014). Scottsdale, USA, 2014: 73-82

[12] Dwork C, McSherry F, Nissim K, Smith A. Calibrating noise to sensitivity in private data analysis//Proceedings of

the 3rd Theory of Cryptography Conference(TCC'06). New York, USA, 2006: 265-284

[13] McSherry F, Talwar K. Mechanism design via differential privacy//Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07). Providence, Rhode Island, 2007: 94-103

[14] McSherry F. Privacy integrated queries: An extensible platform for privacy-preserving data analysis//Proceedings of the 2009 ACM SIGMOD/PODS Conference(SIGMOD'09). Providence, Rhode Island, 2009: 19-30

[15] Chen R, Mohammed N, Fung B C M, et al. Publishing set-valued data via differential privacy. Proceedings of the VLDB Endowment, 2011, 4(11): 1087-1098

[16] Xiao X, Wang G, Gehrke J. Differential privacy via wavelet transforms//Proceedings of the 26th IEEE International Conference on Data Engineering (ICDE'10). Long Beach, USA, 2010: 225-236

[17] Finkel A R, Bentley L J. Quad trees: A data structure for retrieval on composite keys. Acta Informatica, 1974, 4: 1-9

[18] Guttman A. R-Trees: A dynamic index structure for spatial searching//Proceedings of the 1984 ACM SIGMOD Conference (SIGMOD'84). Boston, USA, 1984: 47-57

[19] Pan X, Xu J, Meng X. Protecting location privacy against location-dependent attacks in mobile services. IEEE Transactions on Knowledge and Data Engineering, 2012, 24(8): 1506-1519

[20] Hay M, Rastogi V, Miklau G, Suciu D. Boosting the accuracy of differentially private histograms through consistency. Proceedings of the VLDB Endowment, 2010, 3(1): 1021-1032

[21] Brinkhoff T. A framework for generating network-based moving objects. GeoInformatica, 2002, 6(2): 153-180

[22] Chen R, Li H, Qin A K, et al. Private spatial data aggregation in the local setting//Proceedings of the IEEE International Conference on Data Engineering(ICDE'16). Helsinki, Finland, 2016: 289-300

HUO Zheng, born in 1982, Ph. D., lecturer. Her research interests include privacy-preserving techniques, mobile data management, etc.

MENG Xiao-Feng, born in 1964, Ph. D., professor, Ph. D. supervisor. His research interests include big data management, mobile data management, privacy-preserving techniques, etc.

Background

This research is partially supported by the Grants from the Natural Science Foundation of China (Nos. 91646203, 61532010), the Natural Science Foundation of Hebei Province (Nos. F2015207009, D2015207008) and the Young Prominent Talent Project of Hebei Province Higher School (Nos. BJ2016019, BJ2014021).

Recent years, with the development of location-aware devices, more and more locations and traces of moving objects are collected and then published for mobile-related applications. Analyzing trajectories of passengers in an area may help people making commercial decisions, such as where to build a restaurant. It also can be seen in traffic control systems, analyzing trajectories of vehicles in a city may help government to optimize traffic control strategy. Although publishing trajectories is beneficial for mobility-related decision making processes, it may represent serious threats to individual privacy, since trajectories contain rich spatio-temporal information, which may reveal individual's habits, health condition, social customs, etc. In order to protect trajectory privacy, more and more attentions have been given to this area. In the past years, researchers proposed several techniques based on k-anonymity, however, they could not provide sufficient privacy guarantee since they are highly depend on the background knowledge, the attackers hold. This paper proposes a method under differential privacy, which is a strong privacy model, as we know.

WAMDM lab have studied location privacy-preserving since 2006, several key problems are solved, Related research findings are published in CIKM, ACM SIGSPATIAL GIS and IEEE TKDE, we have studied trajectory privacy-preserving techniques since 2010, related research findings are published in DASFAA, Chinese Journal of Computers, etc. We have studied differential privacy since 2012, publications include SDM, DASFAA, Frontiers of Computers, etc.