International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 3, Issue 8, pp 48-51, August-2019

Efficient Statistics Estimation on Points of Interests

Vuppari Krishnamaraju Sai

Assistant Professor, Department of Computer Science And Engineering, Malla Reddy College of Engineering for Women, Hyderabad -500100, Telangana, India

Abstract - In recent times one of the key attention for business like hotels and restaurants startup by providing valuable information by analyzing the popularity on a certain location and competitors. These details are gathered using PoI (Point of Interest) for a marketing research. Though we do not have direct access of PoI database we propose sample statistics method such as sum and aggregate average with the use of very few queries. As a result of our experiment using real datasets we end up in the same accuracy as such in state-of-the-art method but with six times less queries than the usual.

INTRODUCTION

Aggregate statistics such as sum, aggregate average and distribution of points of interests (PoIs), e.g., Google maps [2] and Foursquare [3] services used in major restaurants and hotels, present important information through applications for a marketing decision. The knowledge in PoI rating allows us to understand the ranking of some PoI's comparative service quality. Furthermore, a restaurant start-up can collect the information of people food preferences in a geo- graphic area by comparing the fame of restaurant. PoIs provide information on serving of diverse cuisines in the area of interest [4], with this information in parallel; it can also give a fair estimation of its market size based on statistics like the PoI's of checked in Foursquare users within the area. Similarly, to launch a hotel it can utilize PoIs' related to hotels. PoIs' like ratings and reviews will allow us to understand its current scenario of markets.

To calculate the above aggregate statistics in exact, it requires all PoIs within the area of interest to be retrived. However most of the service providers do not provide an access to the database where PoIs present in nor full access, so on other option we can only have a public map APIs to discover and collect user needed PoIs. Furthermore, public APIs has its own limitations, they provide PoIs based on the need; as a result it is more of cost to gather PoIs of a wide area.

For an example, foursquare map API [5] returns up to 50 PoIs per query and it allows 500 queries per hour per account. As a sample, to collect PoIs of 14 cities in Foursquare, Lietal [6] spent almost two months, to explore more in the above challenge, we are in need of samples.

That is, to calculate PoI statistics small portion of PoI samples are required. Unless one has a direct full access to get information from PoI databases, no one can directly sample over PoIs, so it is difficult to sample PoIs regularly. After sampling a few of PoIs using both these methods, one has no guarantees whether the PoI statistics obtained directly are to be trusted. To handle this problem, Dalvi et al. [7] suggested a method to exact the sampling bias. However the suggested method is costly because it requires a huge number of queries to collect each sample PoI (e.g., on average 55 queries are used in their paper).The method in [8] samples PoIs with unknown bias stands difficult to eliminate its sampling bias.

In this paper we propose a new method called random region zoom-in (RRZI) to get rid of the estimation bias. The idea works behind RRZI is to sample a random set of sub-regions from a certain area of interest and then collect PoIs within sampled regions. However,an unknown sampling bias is introduced when we query a sampled sub- region with a large number of PoIs, if we only collect PoIs returned.

Otherwise, to comprehensively collect PoIs within we need to divide it further the sampled sub- region within it which requires a huge number of queries. To meet this challenge, we further divide the area of interest into fully accessible sub-regions without any overlapping, if the region includes PoIs is comparatively less than the maximum number of PoIs returned for a query then the region is known as a fully accessible region.

This works so efficient in collecting PoIs within a sampled sub- region, this requires just a query. We demonstrate that RRZI is efficient, and which requires only few queries for a fully accessible region as a sample. Besides its efficiency, the sampling bias of RRZI is easy to be corrected, which requires no extra queries in comparison with the existing methods [7], [8]. To further reduce the number of queries, we propose a mix method RRZI URS, which first works on a small sub-region from the area of interest at random and then samples PoIs within the sub- region using RRZI.

Moreover, for map services such as Google maps providing the total number of PoIs within an input search region, we propose a method to improve the accuracy of RRZI by utilizing this meta information. of queries required to achieve the same level of accuracy in state-of-the-art methods.

EXISTING SYSTEMS

The existing sampling methods have been shown that PoIs are not evenly distributed, but are clustered into sparse and dense areas. The lack of the PoI geographical distribution makes it challenging to accurately estimate PoI statistics, since it is hard to sample PoIs from are uniformly.

Besides its efficiency, the sampling bias requires no extra queries in comparison with the existing methods.

DISADVANTAGES

While such a sampling methods search interface is often sufficient for an individual user looking for the nearest shops or restaurants, data analysts and researchers interested in an LBS service often desire a more comprehensive view of its underlying data. For example, an analyst of the fast-food industry may be interested in obtaining a list of all McDonald's restaurants in the world, so as to analyze their geographic coverage, correlation with income levels reported in Census, etc. Our objective in this paper is to enable the crawling of an LBS database by issuing a small number of queries through its publicly available kNN web search interface, so that afterwards a data analyst can simply treat the crawled data as an offline database and perform whatever analytics operations desired.

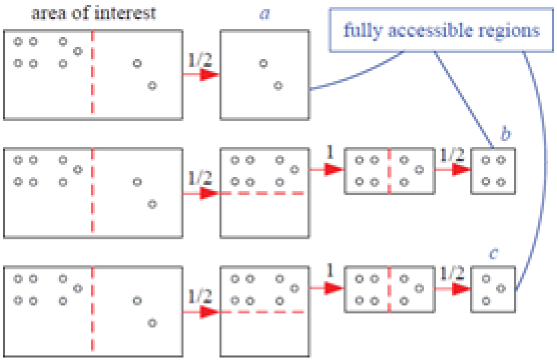
PROPOSED SYSTEMS

Propose sampling methods estimate PoI statistics om a specified area, RRZI divides the current queried region into two sub-regions without overlapping, and then randomly selects a non-empty sub-region as the next region to query. It repeats this process until it observes a fully accessible region. However the method is costly because it requires a large number of queries for each sampled PoI. The method in samples PoIs with unknown bias, it is difficult to remove its sampling bias. We propose a new method random region zoom-in (RRZI) to eliminate the estimation bias. collect PoIs within sampled regions. However, when we query a sampled sub-region including a large number of PoIs, an unknown sampling bias is introduced if we only collect PoIs returned. Otherwise, we need to further divide.

DVANTAGES

A small fraction of PoIs are collected for a sample and helped to calculate PoI statistics. Due to the lack of a direct and full access to PoI databases, one cannot directly sample over PoIs, so it is difficult to sample PoIs regularly.

SYSTEM ARCHITECTURE



IMPLEMENTATION

Implementation is the critical stage where the working system runs successfully to achieve a confidence of the users in using this system and becomes effective.

The execution of this stage needs careful planning, more investigation of the existing model and it's disadvantage and limitations in changeover method and also its evaluation.

Modules Description:

In our Project, We describe three modules

i) Points Of Interests,

ii) Sampling,

iii) Measurement

Points of Interests:

For marketing decision PoIs' play a vital role in providing the need especially in launching restaurants and hotels.

Example, the knowledge in PoI rating helps us to evaluate the service quality ranking of PoI. Moreover, the PoIs' of food infer in a particular region helps for the restaurant start-up which helps in serving various cuisines based on the area of interest.

Sampling:

Sampling Methods estimate PoI statistics in accurate , we show that mix methods also reduce the number of queries required to sample a fully accessible region in comparison with RRZI and RRZIC. We first introduce a uniform region sampling (URS) method, which is used to sample sub-regions uniformly. A parameter to control the size of sub-regions sampled by URS.

Algorithms:

i).Random Region Zoom-in ii).Random Region Zoom-in Count

Random Region Zoom-in:

Random Region Zoom-in (RRZI) method is a proposed method Public map APIs might impose a limit on the size of input regions. For example, Foursquare returns an error message "Your geographic boundary is too big. Please search a smaller area."; this method does not require any extra . Moreover, we show that our mix methods also reduce the number of queries required to sample a fully accessible region in comparison with RRZI and RRZIC. using RRZI method by calculating PoIs within the sub-region.

Random Zoom-in Count for a region:

CONCLUSION

Our methods are proposed to sample PoIs on maps, and give consistent data of PoI with aggregate statistics. We also have shown that the mix method RRZI_URS is more accurate than RRZI under the same number of queries used. RRZIC_MHWRS utilizing this information is more accurate than RRZI_URS. Based on a different real datasets the experimental results they sharply reduce the number of queries required to achieve the same estimation such of state-of- the-art methods.

REFERENCES

[1] Pinghui Wang, MOE Key Laboratory for Intelligent Networks and Network Security, Xi'an Jiaotong University, Shaanxi, China.

[2] Wenbo He, , School of Computer Science, McGill University, QC, Canada.

[3] Xue Liu, , School of Computer Science, McGill University, QC, Canada.

[4] Y. Zhu, J. Huang, Z. Zhang, Q. Zhang, T. Zhou, Y. Ahn, "Geography and similarity of regional cuisines in china", arXiv preprint arXiv:1307.3185, 2013.

[5] Y. Li, M. Steiner, L. Wang, Z.-L. Zhang, J. Bao, "Exploring venue popularity in foursquare", Proc. 5th IEEE Int. Workshop Netw. Sci. Commun. Netw., pp. 1- 6, 2013.

[6] N. Dalvi, R. Kumar, A. Machanavajjhala, V. Rastogi, "Sampling hidden objects using Nearest-neighbor oracles", Proc. ACM SIGKDD, pp. 1325-1333, Dec. 2011.

[7] Y. Li, M. Steiner, L. Wang, Z.-L. Zhang, J. Bao, "Dissecting foursquare venue popularity via random region sampling", Proc. ACM Conf. CoNEXT Student Workshop, pp. 21-22, 2012.

[8] S. Chib, E. Greenberg, "Understanding the Metropolis-hastings algorithm", The Am. Statist., vol. 49, no. 4, pp. 327-335, Nov. 1995.

[9] W. K. Hastings, "Monte carlo sampling methods using Markov chains and their applications", Biometrika, vol. 57, no. 1, pp. 97-109, Apr. 1970.

[10] A.H. Teller, E. Teller, "Equations of state calculations by fast computing machines", IEEE J. Sel. Areas Commun., vol. 21, no. 6, pp. 1087-1092, Jun. 2011.

[11] Z. Bar-Yossef, M. Gurevich, "Efficient search engine measurements", Proc. WWW, pp. 401-410, 2007.

[12] Zhang, N. Zhang, G. Das, "Mining a search engine’s corpus: Efficient yet unbiased sampling and aggregate estimation", Proc. ACM SIGMOD Int. Conf. Manage. Data, pp. 793-804, 2011.

****