Benchmarking Big Data for Trip Recommendation

Kuien Liu
Institute of Software,
Chinese Academy of Sciences
Beijing 100190, China
kuien@iscas.ac.cn

Yaguang Li
Institute of Software,
Chinese Academy of Sciences
Beijing 100190, China
yaguang@nfs.iscas.ac.cn

Zhiming Ding
Institute of Software,
Chinese Academy of Sciences
Beijing 100190, China
zhiming@iscas.ac.cn

Shuo Shang
Department of Software Engineering,
China University of Petroleum
Beijing 102249, China
jedi.shang@gmail.com

Kai Zheng
School of Information Technology and Electrical
Engineering, The University of Queensland
Brisbane QLD 4072, Australia
kevinz@itee.uq.edu.au

***Abstract*—The availability of massive trajectory data collected from GPS devices has received significant attentions in recent years. A hot topic is trip recommendation, which focuses on searching trajectories that connect (or are close to) a set of query locations, e.g., several sightseeing places specified by a traveller, from a collection of historic trajectories made by other travellers. However, if we know little about the sample coverage of trajectory data when developing an application of trip recommendation, it is difficult for us to answer many practical questions, such as 1) how many (future) queries can be supported with a given set of raw trajectories? 2) how many trajectories are required to achieve a good-enough result? 3) how frequent the update operations need to be performed on trajectory data to keep it long-term effective? In this paper, we focus on studying the overall quality of trajectory data from both spatial and temporal domains and evaluate proposed methods with a real big trajectory dataset. Our results should be useful for both the development of trip recommendation systems and the improvement of trajectory-searching algorithms.**

Keywords—Big Data; Trip Recommendation; Spatio-temporal Trajectory Data; Benchmark

I. INTRODUCTION

With the drastically increasing size of trajectory data generated by location-based services and applications which are collected from inexpensive GPS-enabled devices, the availability of such massive trajectory data has received significant attentions in recent years and spawned various novel applications, such as trajectory search and recommendation [1-8], which is designed to retrieve from a database the raw trajectories that best connect (or are close to) a few selected locations (e.g., a set of user specified geographical locations on map). As exemplified in Fig. 1, this service can benefit travellers when they are planning a trip to multiple places of interest in an unfamiliar city by providing similar routes travelled by other people for reference. Cyclopath [9] is a vivid example, which tends to find bike routes that match the personalized cycling demand and share personal cycling knowledge with the community by utilizing an unique rating system. Moreover, trajectory search and recommendation can benefit users in many important aspects including route planning, carpooling, friend recommendation, traffic analysis, urban computing, novel location based services, etc.

Unlike the conventional navigation services, such as Google Maps, which performs path-finding algorithms (e.g., $A*$), trip recommendation, which is based on massive historical trajectories, prefers to find the results by data searching rather than intensive computing. In this sense, proposed trip recommendation is a typical application of **big data** scenarios. One of its advantages is that it tends to find out the "best" routes for users. In many real application scenarios, users may not be fully satisfied with the "fastest" or "shortest" routes recommended by conventional navigation services. For example, a tourist/travel agency may favour the longer routes around a national park for the sake of not missing important scenic spots.

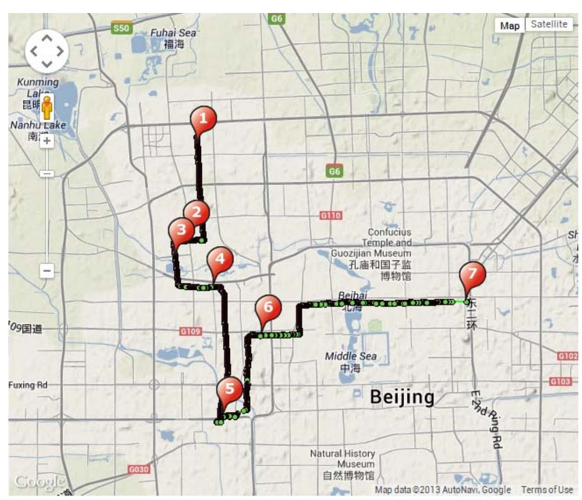

Fig. 1. An Example of Trip Recommendation using Big Data [8].

978-1-4799-3572-7/14/$31.00 ©2014 IEEE

In this sense, trip recommendation based on massive trajectory data is a kind of empiricist approaches which is widely adopted in the age of big data. It highly depends on the deep understanding of historical datasets. In other words, if we know little about the sample coverage of trajectory data when developing an application of trip recommendation, it is difficult for us to answer many practical questions, e.g.,

- How many (future) queries can be answered properly with a given set of raw trajectories?
- How many trajectories are required to achieve a good-enough result?
- How frequent the update operations needed to perform on trajectory data to keep it long-term effective?

In this paper, we focus on studying the sample coverage of trajectory data from both spatial and temporal domains and evaluating the latest methods of trip recommendation [7] which are implemented as an embedded component on a light-weight platform [8] that has been obtaining large-scale practical trajectory data (i.e., hundreds of gigabytes of data per month) of over forty thousand taxis in Beijing. Moreover, we further analyse the experiment results from a practical perspective and give suggestions on above questions.

Benchmarks on spatial-temporal data have attracted many researchers, such as those works listed by Düntgen et.al. [10], but they can neither state the sample coverage of spatio-temporal trajectory data nor answer these three questions above qualitatively. Some papers, e.g., [14], in a sense address a similar problem to us, but they have a different focus and cannot be applicable to trip recommendation directly.

Our results should be useful for both the development of trip recommendation systems themselves and the improvement of trajectory-searching algorithms. To demonstrate how to utilize the benchmarks in trip recommendation, we propose a novel data sampling schema which can significantly reduce the size of trajectory data used for recommendation with only slight decrease in terms of effectiveness.

The remainder of the paper is structured as follows. Section II gives a brief overview of evaluation methodology. Section III analyses the results of our benchmarks. Section **Error! Reference source not found.** provides an example to demonstrate how to further improve trip recommendation based on the results. Section V concludes the paper.

II. Evaluation Methodology

In this section we first define the scenarios of trip recommendation we want to evaluate, and then we briefly introduce the experimental platform and metrics used in the evaluation process. TABLE I. summarizes the notations used through the paper.

A. Problem Statement

Given a query (with several specified locations) and a trajectory database, a trip recommendation system may serve a traveller with an appropriate route, **but this is not always true**. As illustrated by Fig. 2, some locations (e.g., the Hotel) may be far from the recommended trip t, i.e., $d_t >> d_r$, which inevitably reduces the feasibility of applications. We call this phenomenon as the ***Best but Not Good enough*** (BNG) problem. That is because the key idea behind trajectory searching and recommendation is to retrieve the most similar (i.e., the best) trajectories to a sequence of places according to their spatial (and/or temporal) shapes [1-4]. Although some works have made some improvements by considering information in textual domains [5-8], they are not helpful to this problem. We need to study the quality of trajectory data used for recommendation. This is our primary motivation to propose and carry out this work.

TABLE I. Summary of notions

Notation	Definition
G	The graph of road network $G=<V,E>$
D	A trajectory dataset
α	The sampling ratio of total queries.
β	The sampling ratio of the trajectory dataset
d_t	The distance between the recommended trip and the specified location
d_r	The distance between the nearest road and the specified location
Tr	A trajectory, $Tr \in D$
Q	A set of query locations
q	A query location in Q
$D_r(G,q)$	The minimum distance from location q to the nearest road on G
$D_t(Tr,q)$	The minimum distance from q to the Tr
$TR(Q,D,k)$	The k most similar trajectories recommended from D to Q, k can be set to 1.

Trip Recommendation (TR) Given a trajectory set D, a query Q, a positive integer k, and trajectory similarity search component TS returns k distinct trajectories from D that have minimum match distances with respect to Q.

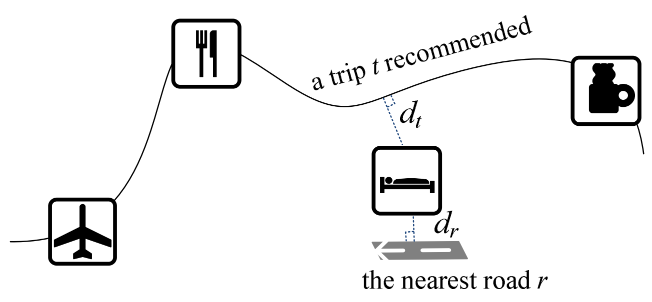

Fig. 2. The *Best but Not Good enough* (BNG) Problem

"Best but Not Good enough" (BNG) Detection Given a recommended trajectory Tr, a positive floating-point number f, Tr is a BNG choice if and only if there exists at least one location q in Q whose minimum distance to Tr is at least f times farther than that to the nearest roads, i.e.,

$$\exists q \in Q, D_t(Tr,q) \geq \Delta, \Delta = f \cdot D_r(G,q)$$

Ideally, a trajectory recommendation system (*TR*) will return more than one (k>1) non-*BNG* candidates to users. To evaluate the overall coverage of a given trajectory data respect to a set of given queries, a benchmark is required to estimate the quality of service.

B. Platform

Fig. 3 gives an overview of proposed benchmarking platform which consists of four major modules: query constructor, query executor, metric adjustor, and *BNG* detector, where the first two modules reproduce the scenarios of trip recommendation and the latter two modules deals with the evaluation tasks with metrics explained in subsection II.C.

The query constructor chooses α parts of trajectories, and then transforms them into (potential) queries according to certain semantic features, e.g., long-term residence time, points of interest (POIs) or vacant status (for taxis in our datasets), in an offline manner. The query executor employs the trajectory searching engine, which has been implemented in our previous works [7, 8], to recommend trips based on β part of the trajectories, where $\alpha+\beta \leq 1$. As the process of recommendation is quite time-consuming, we first partition the dataset into blocks and dispatch each block together with a subprogram to each machine, finally collect and analysis the results in a collaborative strategies. The metric adjustor generates parameters of instructions according to user specification, while the *BNG* detector evaluates the number of queries that are consistent with specified policies, and finally returns the report in a visual way.

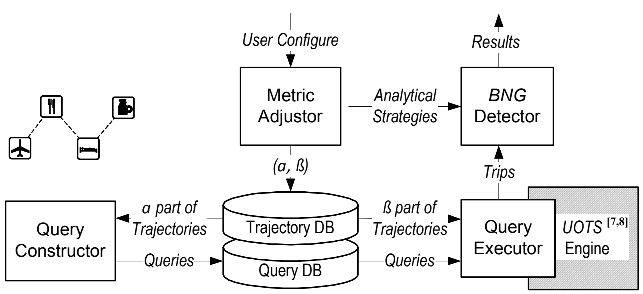

Fig. 3. Overview of our Platform

C. Metrics

Given a trajectory set, the metrics which reflect the overall rate that all (potential) queries that can be satisfied is referred to as the sample coverage of trajectory data. We evaluate the sample coverage of trajectory data for trip recommendation from both spatial and temporal domains, and abstract three metrics, i.e., temporal coverage (*TC*), spatial coverage (*SC*) and the corresponding metric confidence (*MC*), whose measurements are shown below.

- *TC* is the overall rate of the case in which there exists at least one non-*BNG* answers among the returned values of all systems, i.e., $TC = 1 - N_{BNG}/N_{query}$, $k \geq 1$.
- *SC* is the distribution of *BNG* queries whose start and/or end locations are not satisfied in specified spatial area.
- *MC* is confidence that trips candidates recommended are non *BNG*, i.e., the distribution of result candidates, k.

To achieve an intuitive understanding of the metrics, especially the *SC*, we draw the *BNG* results on maps directly.

Data sizes of historical trajectories and queries are adjusted according to different strategies, e.g., partially impacted by *(α, β)* pair raised in Fig. 3. We evaluate these metrics according to the practical questions listed in Section I, i.e., a couple of *how many* and *how frequent*.

III. Results and Analysis

In this section, we first present the experimental settings, then we evaluate a trajectory dataset according to three metrics used in the proposed platform, and finally we give a short summary of the experimental results.

A. Dataset and Experimental Setup

In our experiments, we mainly use the following two datasets. The first dataset contains the trajectory data collected from more than 40,000 taxis during a period of one month in the year 2012. This dataset is collected from an on-going project Moir [11] and contains more than 1,000,000,000 trajectory points. We use α percent of the data to generate queries and β percent of the data as the source data. Each query is a trip consisting of one starting location and one ending point, i.e., <*start*, *end*>, each of which can be inferred from the vacant status of the trajectory records.

Another dataset is the smart card dataset [12] which contains the check-in information of bus and railway travellers. The time period of this dataset is also one month. This dataset is dedicated as a source of queries. It contains 10,000 queries, and each of them also consists of a start location and an end location. ***Remark.*** The smart card dataset provides us with chances to construct complicated trips with more than two locations in future work.

As the queries are generated from taxi trips, all the start points and the end points are on the road network, i.e., $D_r(G, q) \leftarrow 0$. In order to calculate trajectory coverage in a general way, we consider the maximum deviation $\Delta = f \cdot D_r(G, q)$ as a constant, and a recommended trajectory is regarded as good if the distance between the trajectory and the query point, i.e. $D_r(Tr, q)$, is smaller than the maximum deviation distance Δ.

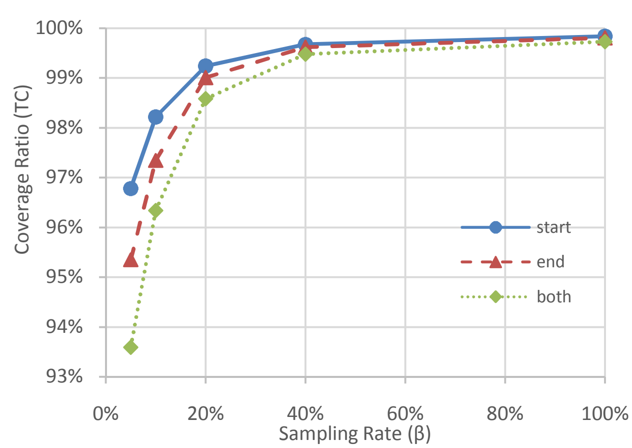

Fig. 4. Effectiveness of Sampling Rate on Taxi Dataset

B. Coverage Ratios with Various Sampling (α,β)

Fig. 4 shows the relationship between the sampling rate of trajectory data (corresponding to β), and trajectory coverage TC. A query point (*start* or *end*) is considered satisfied if it has at least one candidate within the deviation distance of Δ, and a query is considered as satisfied if both of its start and end points are satisfied. The solid blue line denotes the ratio of satisfied query start points; the dashed orange line denotes the ratio of satisfied query end points; and the dotted grey line denotes the ratio of satisfied queries, i.e. TC.

In this experiment, we set the maximum allowed deviation distance $\Delta=100m$, and the queries are generated using the trajectory data of another adjacent day. As indicated in Fig. 4, 5% of the trajectory data is big enough to satisfy 93% percent of the queries. Therefore, we choose $\beta=5\%$ as the default sampling rate for the following experiments.

TABLE II. SAMPLING OF QUERY DATA (Δ)

Sampling ratio (α)	**TC**	**Time(s)**
5%	98.98%	7.4
100%	99.03%	152.8

As shown in TABLE II. , 5% samples can well represent the distribution of the total queries while being much faster. Therefore, we will use $\alpha=5\%, \beta=5\%$ as the default settings.

C. Effect of Maximum Allowed Deviation (Δ)

In this section, we further study the effect of maximum allowed deviation, i.e., Δ, on the coverage ratio. As Fig. 5 shows, with the increase of the maximum allowed deviation, the satisfied ratio grows rapidly. When deviation Δ reaches 200m, most queries can be satisfied. Larger Δ means further distance the user needs to go before he can take the trip, so its value should be limited based on actual service use. For the following experiments, we choose $\Delta=100m$ as the default maximum deviation.

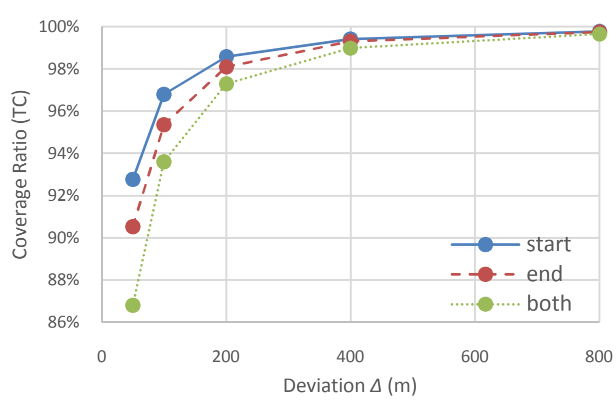

Fig. 5. Effect of Maximum Allowed Deviation

D. Long-term Effectiveness of Historical Data

We use 5% of the trajectory data collected in 09/30/2012 (*Sep.* for short) and 12/01/2012 (*Dec.* for short) respectively as the trajectory data source to answer the queries ranging from 12/02/2012 to 12/31/2012.

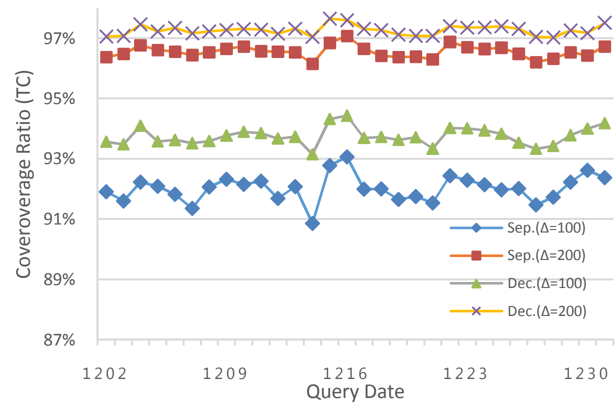

Fig. 6. Effectiveness for the Future

As shown in Fig. 6, TC does not change too much within the month. This may because that the trajectory patterns stay roughly the same with the period of one month. On the other hand, the prediction result of 12/01/2012 is better than that of 09/30/2012, this may be caused by the changes of weather (from autumn to winter) during the interval of two months. Another observation is that the results have periodic trends on a weekly basis, which motivates us to regroup the historical data by weekday/weekend and keep fewer data from each group.

E. Distribution of BNG queries

Fig. 7 demonstrates the distribution of queries (i.e., SC) built upon taxi data with parameters $\alpha=5\%, \beta=5\%, \Delta=100m$. The green points denote satisfied start points while the red points denoted unsatified end points. As indicated in Fig. 7, most of the end points of BNG queries are located in suburb, which means that trips whose destinations are suburb are less likely to be satifsified. This is also consistent with the common sense.

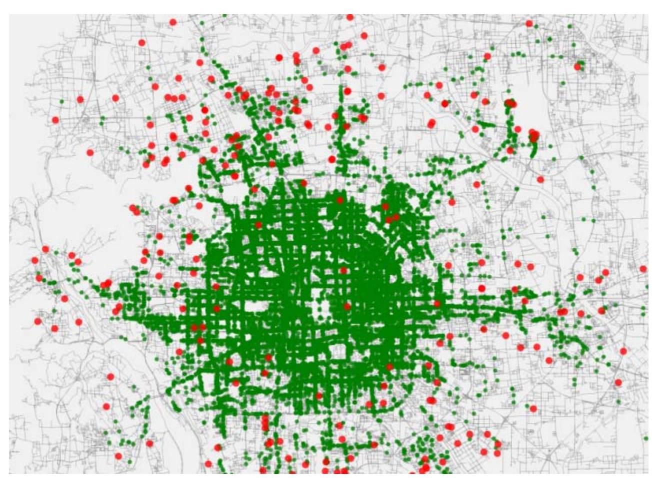

Fig. 7. Distribution of (Partial) *BNG*-queries

F. *Distribution of Result Candidates (K)*

As a metric of recommendation confidence (i.e., MC), variable k denotes the number of candidate trajectories of the query result. As Fig. 8 shows, the number of candidates increases when the maximum allowed distance Δ grows larger. To better understand Fig. 8**,** we transfer it into the accumulated percentage as shown in Fig. 9. It is easy to see that more than 50% (80%) queries have more than 30 (10) candidates while $\beta=5\%$. With this observation, we give an enlightenment on how to further optimize the sampling strategies of historical trajectories.

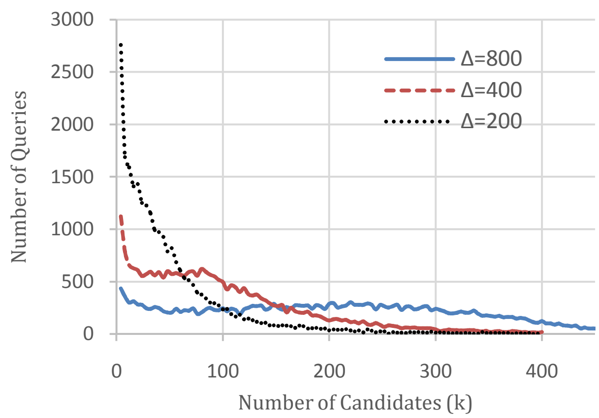

Fig. 8. Distribution of the Number of Candidates

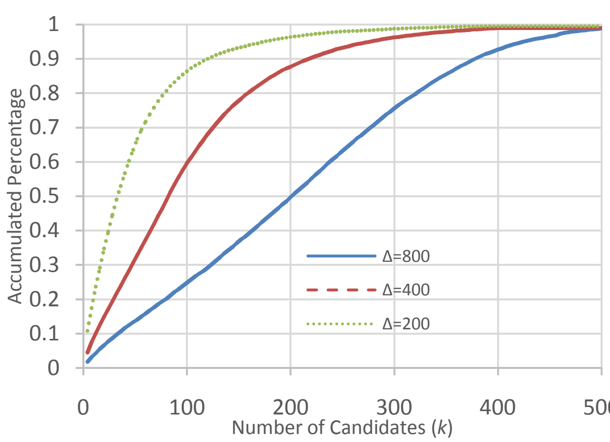

Fig. 9. Accumulated Percentage of Candidates

G. *Effectiveness on Smart Card Dataset*

Furthermore, the smart card dataset is used to experiment with the effectiveness of trip recommendation based on taxi dataset, which is positively confirmed by Fig. 10. In the experiments, 10,000 queries are extracted from the check-in records of bus and railway travelers with the most advanced methods [12] and performed on the trajectory recommendation platform [8].

One observation in Fig. 10 is that the coverage ratio is heavily affected by the maximum deviation Δ. To find out the reasons, we further study the distribution of BNG queries on smart card dataset. As Fig. 11 demonstrates, most of the BNG queries are near the subway stations. This is because a part of queries are trips by buses, which have a transportation network that does not exactly match the subway system, the passengers need to walk hundreds of meters to take transfer between bus and subway systems.

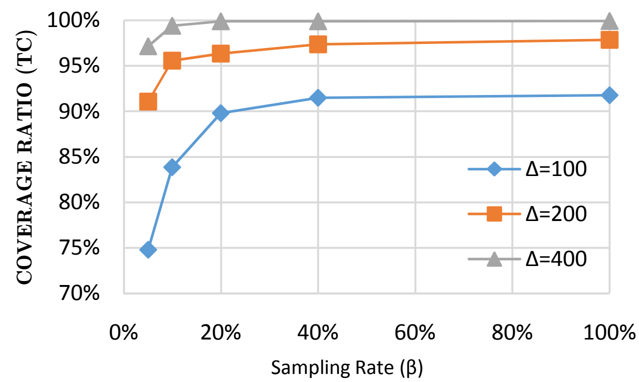

Fig. 10. Effectiveness of Sampling Rate on Smart Card Dataset

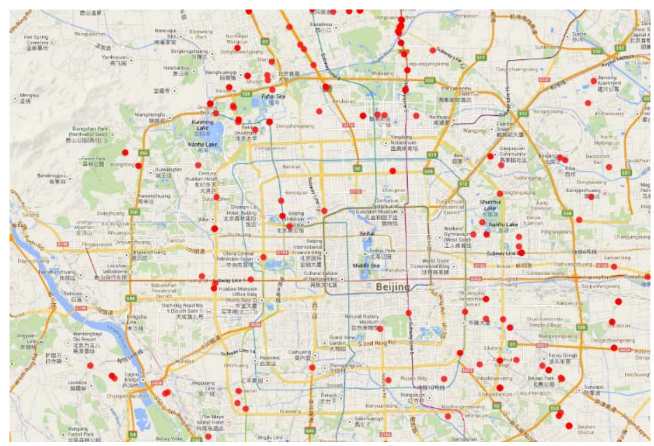

Fig. 11. Distribution of the *BNG* queries on Smart Card Dataset.

H. *Summary of Experiments*

With the experimental results, we can find the answers to three questions raised in Introduction in quantitative ways:

- 5% historical data of one day is large enough to cover 95% of the queries built from next 30 days.
- We only need to sample and store the data from dozens of days to get relatively good recommendation results for the whole year, and the size of data to store is no more than that of original data of a week.
- Update the trajectory every month is frequent enough to get satisfactory results.

IV. How to Benefit from the Results

In this section we give several examples on how to benefit from the evaluation results and start a discussion on further improvement of the trip recommendation system.

A. *Example 1: Fine-grained Data Sampling*

The enormous volumes of trajectory data can easily overwhelm existing trajectory searching and recommendation applications. This brings new challenges in storing, transmitting and processing these data, which, at the same time, highlights the need for data sampling technique for trajectories. Ideally, we call for data sampling methods that can discard as many trajectories as possible while guaranteeing the success rate of trip recommendation.

The evaluation results of this paper have inspired us that it is possible to achieve a fine grained data sampling by tailoring the data according to its spatial and temporal coverage. For example, we can re-organize the data according to each weekday (or weekend) and store only *h* days of historical trajectories from each set. The time intervals between these *h* days are not even-distributed, e.g., following the exponential attenuation when stretching back. Besides, we can further sample the data of each day according to the result of spatial coverage, e.g., leaving less trajectories that traverse downtown area while keeping more for the surrounding suburbs.

B. *Example 2: Adjustable Route Joining*

From the results we find that raw trajectories themselves may be difficult to satisfy all the queries when the data are sparse. However, we may find a reachable route which consists of several trajectories that join and connect in nature. Unfortunately, there exists so many candidate combination plans among these trajectories, it is a challenging problem to determine which one is the best. At this point, we can refer to the metrics used for evaluation, e.g., MC which represents the frequency of movement behaviors. We may choose the trajectory with many convoys, i.e., higher MC, and construct a trip by combining several trajectories with the biggest accumulated confidence value.

C. *Example 3: Multi-modal Routing*

One interesting observation, by comparing Fig. 7 and Fig. 11, is that datasets collected from different modes of transportation have distinguished space coverage. It is pretty easy to understand: urban public transport such as bus, tram, MTR, ferry and MRT travels along fixed (or seldom changed) routes, while the movements of taxi and bicycle are very diverse (bicycle are commonly used in many countries, e.g., China and the Netherlands, and keep their attraction in future as a green low carbon transport). A practical trip recommendation system is better to support multi-modal routing. Techniques proposed in this paper are also applicable to this scenario by providing suggestions on what proportion of different data sets to sample and how to combine them to achieve an acceptable coverage rate.

Furthermore, some empirical evidences in this paper have certain reference values to more applications scenarios. For example, the long-term effectiveness of historical data (corresponding to Fig. 6) can be used to improve the effectiveness of large scale trajectory compression [13] up to 20+%, by partitioning and reorganizing trajectories to fit the constrained memory used for compression according to the similarity of trajectories across spatial and temporal domains. We do not give more examples due to space limitations.

V. CONCLUSION

In this paper we study the practical problem of *Best but Not Good enough (BNG)* in trip recommendation applications. We evaluate a trajectory dataset according to three metrics used in the proposed platform, and further exemplify how to benefit from the results. Our work should be useful for both the development of trip recommendation systems themselves and the improvement of trajectory-searching algorithms.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (Nos. 61202064), the National High Technology Research and Development Program of China (863 Program) (No. 2013AA01A603), and the Strategic Priority Research Program of the Chinese Academy of Sciences (No. XDA06010600).

REFERENCES

[1] Zaiben Chen, Hengtao Shen, Xiaofang Zhou, Yu Zheng, and Xing Xie. Searching trajectories by locations: an efficiency study. In *SIGMOD*, pages 255–266, 2010.

[2] Yu Zheng, Lizhu Zhang, Zhengxin Ma, Xing Xie, and Wei-Ying Ma. Recommending friends and locations based on individual location history. *ACM T Web* 5(1):1-44, Feb. 2011.

[3] Lu An Tang, Yu Zheng, Xing Xie, Jing Yuan, Xiao Yu, Jiawei Han. Retrieving k-Nearest Neighboring Trajectories by a Set of Point Locations.In *SSTD*, pages 223-241, 2011.

[4] Haibo Wang and Kuien Liu. User oriented trajectory similarity search. In *SIGKDD UrbComp*, pages 103–110, 2012.

[5] Shuo Shang, Ruogu Ding, Bo Yuan, Kexin Xie, Kai Zheng and Panos Kalnis. User Oriented Trajectory Search for Trip Recommendation. In *EDBT*, pages 156-167, 2012.

[6] Kai Zheng, Shuo Shang, Jing Yuan and Yi Yang. Towards Efficient Search for Activity Trajectories. In *ICDE*, pages 230-241, 2013.

[7] Shuo Shang, Ruogu Ding, Kai Zheng, Christian S. Jensen, Panos Kalnis and Xiaofang Zhou. Personalized Trajectory Matching in Spatial Networks. The *VLDB J.*, online 2013.

[8] Kuien Liu, Bin Yang, Shuo Shang, Yaguang Li and Zhiming Ding. Moir/UOTS: Trip Recommendation with User Oriented Trajectory Search. In *MDM*, pages 335–337, 2013.

[9] Katherine Panciera, Reid Priedhorsky, Thomas Erickson, and Loren Terveen. Lurking? cyclopaths?: a quantitative lifecycle analysis of user behavior in a geowiki. In *SIGCHI*, pages 1917-1926, 2010.

[10] Christian Düntgen,Thomas Behr,Ralf Hartmut Güting. BerlinMOD: a benchmark for moving object databases. *VLDB Journal*, 18(6): 1335-1368, Dec. 2009.

[11] Kuien Liu, Ke Deng, Zhiming Ding, Mingshu Li, and Xiaofang Zhou. Moir/MT: Monitoring large-scale road network traffic in real-time. In *VLDB*, pages 1538–1541, 2009.

[12] Nicholas Jing Yuan, Yingzi Wang, Fuzheng Zhang, et al, Reconstructing Individual Mobility from Smart Card Transactions: A Space Alignment Approach, In *ICDM*, pages 877-886, 2013.

[13] Kuien Liu, Yaguang Li, Jian Dai, Shuo Shang, and Kai Zheng. Compressing large scale urban trajectory data. In *CloudDP*, April 2014.

[14] Yanhua Li, Moritz Steiner, Jie Bao, Limin Wang, and Ting Zhu. Region Sampling and Estimation of GeoSocial Data with Dynamic Range Calibration. In *ICDE*, April 2014.