2017 IEEE 18th International Conference on Mobile Data Management

Assessing the Accuracy Benefits of On-The-Fly Trajectory Selection in Fine-Grained Travel-Time Estimation

Robert Waury, Jilin Hu, Bin Yang, Christian S. Jensen
Department of Computer Science
Aalborg University
Email: {rwaury,hujilin,byang,csj}@cs.aau.dk

***Abstract*—Today's one-size-fits-all approach to travel-time computation in spatial networks proceeds in two steps. In a preparatory off-line step, a set of distributions, e.g., one per hour of the day, is computed for each network segment. Then, when a path and a departure time are provided, a distribution for the path is computed on-line from pertinent pre-computed distributions. Motivated by the availability of massive trajectory data from vehicles, we propose a completely on-line approach, where distributions are computed from trajectories on-the-fly, i.e., when a query arrives. This new approach makes it possible to use arbitrary sets of underlying trajectories for a query. Specifically, we study the potential for accuracy improvements over the one-size-fits-all approach that can be obtained using the on-the-fly approach and report findings from an empirical study that suggest that the on-the-fly approach is able to improve accuracy significantly and has the potential to replace the current one-size-fits-all approach.**

1. Introduction

A range of services related to road networks rely on the computation of travel-times of paths [1]. For example, this applies to vehicle routing services that compute fastest paths and to services that pre-compute payments for transportation based on estimated travel-times. The travel-time of a path is time-varying due to congestion and external conditions such as precipitation and road-surface conditions; and travel-time also varies across drivers.

Due to the proliferation of GPS data from vehicles, travel-time estimation is increasingly being based on vehicle trajectory data. As more and more such data becomes available, we believe that it is becoming feasible to estimate travel-times in a much more fine-grained manner than what is possible with today's state-of-the-art approach, which is intuitively a one-size-fits-all approach.

The one-size-fits-all approach pre-computes histograms off-line for each road network segment based on travel-times extracted from trajectories that traversed the segment. Specifically, this approach typically partitions a day into intervals, e.g., 15-minute, 30-minute, or 1-hour intervals, and computes a histogram for each interval based on trajectories that occurred during that interval. It is also possible to pre-compute histograms for consecutive segments, i.e., paths [2] [3]. When a query in the form of a path and a departure time are provided, this approach produces a travel-time histogram for the traversal by combining pertinent pre-computed histograms. The off-line computation of histograms offers efficiency, but it also restricts flexibility, as all queries must use the pre-computed histograms.

Instead, we propose an on-line approach, where histograms are computed on-the-fly at query time. While this may reduce running time efficiency, it is much more flexible: depending on the query, any subset of relevant trajectories can be used to form a histogram for segments and paths. We hypothesize that this approach has the potential to improve the accuracy of travel-time distribution estimates substantially over the one-size-fits-all approach. Put differently, we believe that if we choose the "right" trajectories from which to compute histograms, we get much more accurate histograms than is possible with the one-size-fits-all approach.

To motivate this study, consider the road network of Denmark, which consists of around 1.6 million edges. Precomputing histograms for all potentially interesting subsets of trajectories, e.g., one set for every driver in personalized routing [4] [5], is infeasible as exponentially many such subsets may exist.

The paper aims to provide insight into the potential of on-the-fly travel-time histogram computation. To do this, we define several specific procedures for histogram computation that exemplify both the one-size-fits-all approach and the on-the-fly approach. We then detail a methodology for comparing the different procedures empirically using GPS data. In particular, we build histograms for a road network using some of the data. Then we use trajectories from the remaining data for testing. Intuitively, we provide means of determining which histograms best predict the travel-times seen in the test trajectories. Finally, we apply the methodology to a substantial GPS dataset and report findings that suggest that on-the-fly histogram computation indeed has potential to improve the accuracy of travel-time distributions in road networks.

The paper's contributions are as follows: (i) we motivate and identify a new paradigm for travel-time distribution computation; (ii) we define specific procedures for travel-time histogram computation; (iii) we identify a methodology for empirical evaluation of the accuracy of histograms; and (iv) we report findings from an empirical study with a

2375-0324/17 $31.00 © 2017 IEEE
DOI 10.1109/MDM.2017.40

substantial GPS dataset.

The remainder of the paper is structured as follows. Section 2 describes the paper's setting and trajectory selection criteria and defines the histogram computation techniques to be compared, Section 3 details the empirical evaluation methodology, and Section 4 reports on the empirical study. Finally, Section 5 concludes.

2. Travel-Time Estimation Methods

This section provides the setting of our study and describes four histogram computation methods.

2.1. Trajectories in Spatial Networks

A spatial network, as shown in Figure 1, is a weighted directed graph $G = (V, E, l)$, where V is a vertex set, $E \subseteq V \times V$ is an edge set, and $l : E \rightarrow \mathbb{R}$ is a length function. Every edge $e \in E$ represents a road segment.

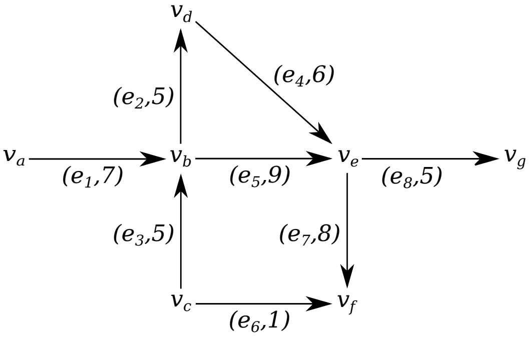

Figure 1. Example Road Network Graph

A trajectory tr in a spatial network is represented as a tuple.

$$tr = (a, s),$$

where $a = (a_1, .., a_k)$ is a tuple of trajectory-specific information like driver and vehicle ID and s is a sequence of triples (t, e, p), where t is a timestamp, $e \in E$ is a segment in the spatial network G, and p is the vehicle's position on segment e, given as a distance from the start of the segment, meaning that $\forall tr\ \forall (t, e, p) \in tr.s\ (p \leq l(e))$. Table 1 provides five example trajectories.

TABLE 1. Example Trajectories

ID	user	s
tr_1	A	$\langle(1, e_1, 3), (2, e_1, 6), (3, e_5, 4), (4, e_8, 5)\rangle$
tr_2	A	$\langle(21, e_1, 4), (22, e_5, 3), (23, e_8, 4)\rangle$
tr_3	A	$\langle(36, e_1, 4), (37, e_2, 3), (38, e_4, 3), (39, e_4, 6)\rangle$
tr_4	B	$\langle(3, e_1, 2), (4, e_1, 4), (5, e_1, 6), (6, e_5, 3), (7, e_8, 3)\rangle$
tr_5	B	$\langle(16, e_1, 1), (17, e_1, 4), (18, e_1, 7), (19, e_3, 3),$ $(20, e_6, 1)\rangle$

A traversable sequence of neighboring segments is called a path. Examples of paths include $\langle e_1, e_5, e_8\rangle$ and $\langle e_1, e_2, e_4\rangle$. Path-based traffic information can provide more detailed insights into travel-time, e.g., if a path contains left turns where right of way has to be observed, path-based histograms can pick up those path-specific delays, whereas single-segment histograms would only give an average based on all directions.

TABLE 2. A Summary of the four Methods

Method	Temporal Features	Path Cost	Predicate
LM	Static, Off-line	No	No
M2	Static, Off-line	Yes	No
M3	Dynamic, On-line	Yes	No
M4	Dynamic, On-line	Yes	Yes

2.2. Histogram Computation Techniques

Given a path $P = \langle e_1, e_2, \ldots, e_X\rangle$ and a departure time t, we apply four different methods to compute a travel-time distribution of using path P at t. We represent a distribution as a histogram and use $H_{P,t}$ to denote the travel-time histogram of traversing path P at t. A summary of the four methods is given in Table 2.

Legacy Method (LM) [4] [6] [7]: In LM we partition every day into multiple equal-sized intervals, e.g., 96 15-min intervals or 24 1-hour intervals. For each segment e in the road network and for each interval I, we identify a set of trajectories $T_{e,I}$ that occurred on segment e during interval I. If the cardinality of set $T_{e,I}$ exceeds threshold β, we derive a histogram $H^{LM}_{e,I}$ from the trajectories in $T_{e,I}$, which represents the travel-time distribution of traversing e during interval I.

This procedure is conducted *off-line*, meaning that all histograms are pre-computed once and for all.

To compute the travel-time distribution of P at t using LM, we first identify the interval I such that $t \in I$. Next, we are able to obtain relevant histograms for all segments in P, i.e., $H^{LM}_{e_1,I}$, $H^{LM}_{e_2,I}$, …, $H^{LM}_{e_X,I}$. Finally, we convolve the segments' histograms to derive the histogram for path P at t:

$$H^{LM}_{P,t} = \odot^{X}_{i=1} H^{LM}_{e_i,I},$$

where $\odot$ is the convolution operator.

For example, for $P = \langle e_1, e_2, e_3, e_4\rangle$, the travel-time distribution at t is computed as follows.

$$H^{LM}_{P,t} = H^{LM}_{e_1,I} \odot H^{LM}_{e_2,I} \odot H^{LM}_{e_3,I} \odot H^{LM}_{e_4,I}$$

Convolution of histograms has been used in multiple earlier approaches [8] [9] [10] and assumes independence between different segments, which often does not hold for all segments [6].

In LM, the intervals of the histograms that are use in convolution are pre-defined and not dependent on the departure time t. Thus LM is a *static* approach. Next, LM considers all trajectories and does not apply any predicates to further filter trajectories. Thus, LM is a *non-specific* approach. Finally, LM only assigns histograms to *segments*, but not to paths. This explains the "Path Cost" column in Table 2.

Method 2 (M2) [2] [3]: We consider a static, ***path-based***, and non-specific method, M2.

M2 differs from LM in that it computes histograms not only for all segments, but also for some paths.

Specifically, if more than β trajectories occurred on path P_k during interval I, we maintain a histogram $H^{M2}_{P_k,I}$ for path P_k. The cardinality of any path P_k is typically small. The larger the cardinality of a path P_k, the less likely the path is to have more than β trajectories.

Similar to LM, the procedure is also conducted *off-line*.

To estimate the travel cost distribution of P at t, we first identify the interval I such that $t \in I$. Next, we choose the coarsest combination of available histograms of segments and subsequences of P and convolve them to obtain $H^{M2}_{P,t}$.

$$H^{M2}_{P,t} = \odot_{P_k \in CoarsestSet} H^{M2}_{P_k,I},$$

For example, let $P = \langle e_1, e_2, e_3, e_4 \rangle$. Assume that we have, in addition to the histograms of all segments, histograms $H^{M2}_{\langle e_1,e_2 \rangle,I}$, $H^{M2}_{\langle e_2,e_3 \rangle,I}$, $H^{M2}_{\langle e_1,e_2,e_3 \rangle,I}$. Then, we have $CoarsestSet = \{\langle e_1, e_2, e_3 \rangle, e_4\}$ since they together cover path P and $\langle e_1, e_2, e_3 \rangle$ is coarser than paths $\langle e_1, e_2 \rangle$ and $\langle e_2, e_3 \rangle$. Thus, we have the following:

$$H^{M2}_{P,t} = H^{M2}_{\langle e_1,e_2,e_3 \rangle,I} \odot H^{M2}_{e_4,I}.$$

Consider another scenario where we also have histogram $H^{M2}_{\langle e_1,e_2,e_3,e_4 \rangle,I}$. Then $CoarsestSet = \{\langle e_1, e_2, e_3, e_4 \rangle\}$, and thus the histogram can be returned directly, and no convolution is needed.

Method 3 (M3): We consider a ***dynamic***, path-based, and non-specific method, M3.

In M3, we consider dynamically constructed intervals that are dependent on the given departure time t. Specifically, we construct interval $I' = [t - \frac{\alpha}{2}, t + \frac{\alpha}{2},]$, where α is an interval parameter. For example, if $\alpha = 60$ minutes, we construct interval I' = [9:06, 10:06] if the departure time t is 9:36. We only consider the trajectories that occurred during I' to derive histograms for segments or paths.

In addition, when we consider building histograms for the segments and paths that do not contain the first segment in the given path P, we use a *shift-and-enlarge* procedure [3] that shifts interval I' by the minimum traversal time and enlarges I' by the maximum traversal time of the proceeding segments. For instance, assume that traversing $\langle e_1, e_2, e_3 \rangle$ takes at least 2 minutes and at most 5 minutes. When retrieving the trajectories for building a histogram for segment e_4, we should then use a shifted-and-enlarged interval I' = [9:08,10:11].

This procedure is conducted *on-line*, since the dynamically constructed interval I' is based on the departure time t, meaning that the histograms cannot be pre-computed.

Finally, we compute the path cost distribution for path P at t as follows.

$$H^{M3}_{P,t} = \odot^{X}_{P_k \in CoarsestSet} H^{M3}_{P_k,I'}.$$

Method 4 (M4): We consider a dynamic, path-based, and ***specific*** method, M4.

In M4, we only consider the trajectories further that satisfy particular predicates during the dynamically constructed intervals. Such predicates help us choose trajectories under specific conditions and may enable more accurate travel-time estimation when such specific conditions apply.

For example, a predicate defined on drivers enables us to consider trajectories from only a specific driver. Alternatively, a predicate defined on weather enables us to only consider trajectories from different weather conditions, e.g., rainy and icy conditions. This procedure is also on-line. We use $H^{M4}_{P_k,I',f}$ to denote the histogram of path P_k that is derived from trajectories that satisfy predicate f during interval I'. Then, we have the following:

$$H^{M4}_{P,t,f} = \odot^{X}_{P_k \in CoarsestSet} H^{M4}_{P_k,I',f}$$

For our example in Figure 1 and Table 1, the histogram $H^{M4}_{P,t,f}$ with $P = \langle e_1, e_5, e_8 \rangle$, $t = 4$, $\alpha = 6$, and $f =$ "$user = B$" would be based on the trajectory set $\{tr_4\}$, and $H^{M3}_{P,t}$ would be based on $\{tr_1, tr_4\}$.

3. Experimental Design

This section explains how we evaluate the accuracy of the histograms computation techniques from Section 2.2. The objective of the study is to assess the potential for higher travel-time distribution prediction accuracy of histograms constructed using the on-the-fly approach. In order to assess the accuracy benefits of on-the-fly approach, we utilize the concept of likelihood.

3.1. Experimental Setup

Since our method is expected to increase predictive performance, we consider trajectories from a real-world dataset (cf. Section 4.1). To evaluate our results, we first store all trajectories from the first year (until December 2007) in a temporal index. All trajectories for the following year are initially held back and are then inserted in order. From those, k trajectories tr_q with a duration of at, least 120 seconds are randomly selected, and a query $Q = (P, t, \alpha, R, f)$ is generated from them before they are inserted. P is the path for which the histograms are computed, t is the center of an interval I, α is the interval parameter, R is the recurrence of the interval, and f is a predicate. The query is run with $t = t_1$, where t_1 is the first timestamp of each tr_q, P is the sequence of segments traversed in $tr_q.s$, and f is the predicate used for M4 $user \in tr_q.a$. For α and R, multiple different values are considered.

We compare all methods introduced in Section 2.2. In addition to the above parameters, the parameters $d \geq 1$ and $\beta \geq 1$ are specified. Parameter d denotes the maximum cardinality of paths for which traversal times are collected and β denotes the minimum sample size. The histogram collection therefore returns up to d histograms for every segment in S. If the number of trajectories traversing them is below β, they yield empty histograms.

3.2. Likelihood Ratio

Since our method aims to improve travel-time predictions, we evaluate the method introduced in Section 2.2 by considering the likelihood the retrieved histograms for our sample trajectory tr_q would have predicted an actual traversal time TT_e of each segment or path in $tr_q.s$.

To achieve this, we first define a discrete probability density function p_H derived from a histogram H:

$$p_H(TT_e) = \frac{m(TT_e)}{n},$$

where $m : \mathbb{R} \rightarrow \mathbb{N}$ is a function that maps a traversal time to the number of trajectories that traversed the segment or path of the histogram within the time range of the bin of TT_e. Further, $n = \sum m_i$ is the number of trajectories sampled in all bins of the histogram.

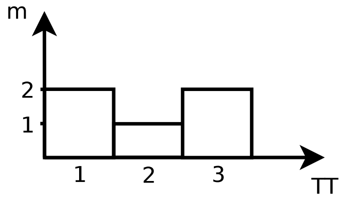

Figure 2. Histogram for e_1 in G

The histogram for e_1 from our example trajectories in Table 1 is shown in Figure 2, where $m(1) = 2$, $m(2) = 1$, $m(3) = 2$, and $n = 5$. It follows that $p_H(1) = \frac{2}{5}$, $p_H(2) = \frac{1}{5}$, and $p_H(3) = \frac{2}{5}$.

Next, from the probability density function p_H, the likelihood $\mathcal{L}(H|TT_e) = p_H(TT_e)$ of the traversal time TT_e is computed. If likelihood for the M3 dynamic histograms $\mathcal{L}_{M3} = \mathcal{L}(H^{M3}|TT_{e_i})$ consistently improves compared to the LM static approaches $\mathcal{L}_{LM} = \mathcal{L}(H^{LM}|TT_{e_i})$, the on-the-fly approach can provide more accurate travel-time estimations.

The likelihood ratio $\rho = \frac{\mathcal{L}_{M3}}{\mathcal{L}_{LM}}$, where $\mathcal{L}_{LM}$ is the likelihood derived from the convolved static histogram, and $\mathcal{L}_{M3}$ is the likelihood derived from the dynamic one, is then computed from the respective histograms.

If $\frac{\mathcal{L}_{M3}}{\mathcal{L}_{LM}} > 1$ then the likelihood for the dynamic approach is higher and therefore is a better predictor of travel-time. The same comparison is performed for the likelihood of the M4 dynamic user histogram $\mathcal{L}_{M4}$, and for the M2 non-convolved static histogram $\mathcal{L}_{M2}$.

4. Empirical Study

We evaluate our results by examining the accuracy of travel-time estimation based on a real-life trajectory dataset. In the study, we exclude traversal times for the first and last segments of trajectories to avoid any outliers that may exist for such segments. In the experiments, we only consider paths with cardinalities up to 3 since the share of empty results for M4 exceeds 30% in our dataset after that. We use a sample size of $k = 1,000$ trajectories, and we use equi-width histograms with bin width $h = 1s$.

4.1. Young Drivers Dataset

The Young Drivers dataset covers ca. 670,000 trajectories within Aalborg and the surrounding area during the period from December 2006 to December 2008 and comprises over 100 million map-matched GPS records [11] sampled at 1 Hertz [12]. The data was collected from private cars, so all trajectories with the same vehicle ID are assumed to be generated by the same driver. This allows us to use the vehicle ID as a predicate in Method 4. Apart from timestamp, trajectory ID, and vehicle ID, the records also contain a segment ID and the vehicle's position on the segment. These records are grouped by trajectory ID, sorted by timestamp, and then aggregated to form a trajectory set TR as defined in Section 2.1.

4.2. Histogram Accuracy

We first consider the accuracy of the advanced methods when compared to the baseline LM. Specifically, we measure the likelihood ratios $\rho_{M2} = \frac{\mathcal{L}_{M2}}{\mathcal{L}_{LM}}$, $\rho_{M3} = \frac{\mathcal{L}_{M3}}{\mathcal{L}_{LM}}$, and $\rho_{M4} = \frac{\mathcal{L}_{M4}}{\mathcal{L}_{LM}}$ of the histograms computed by the different methods. When the fraction of likelihood ratios above 1 exceeds 0.5 substantially, this suggests that the method represented by the nominator is an improvement over the method represented by the denominator.

Figures 3–5 show how often the likelihood of M2, M3, and M4 histograms improves upon the likelihood of the LM histograms for single segments and paths of length 2 and 3 when selecting trajectories in I for every day, with α set to either 60 or 120 minutes and when varying β from 5 to 100 trajectories.

Similarly, Figures 6–8 show the fraction of likelihood ratios above 1 when using trajectories from all weekdays or all weekends. Specifically, if departure time t falls on a weekday, all trajectories within I on weekdays are considered, and if it falls on a Saturday or Sunday, all trajectories within I on weekends are considered. For the single segment results reported in Figures 3 and 6, the difference between convolved (LM) and non-convolved (M2) static is omitted since no convolution occurs.

From the figures, we see that the M3 histograms, despite considering a larger time range due to the *shift-and-enlarge* procedure, provide no noticeable improvement on the LM histograms. The same observation holds for the M2 histograms. However, the M4 histograms exhibit a consistent improvement, especially for $\beta \geq 10$. We also see that for the non-specific histograms, $\alpha = 60$ performs better than the larger time range in most cases. For the M4 histograms, no such pattern can be observed. While outperforming the LM baseline for single segment histograms, the other two non-specific methods are only slightly better than the baseline method for two-segment histograms (cf. Figures 4 and 7), and both methods are slightly worse than the baseline for three-segment histograms (cf. Figures 5 and 8). With single segments as well as paths, the M4 histograms consistently outperform their non-specific counterparts; and for paths,

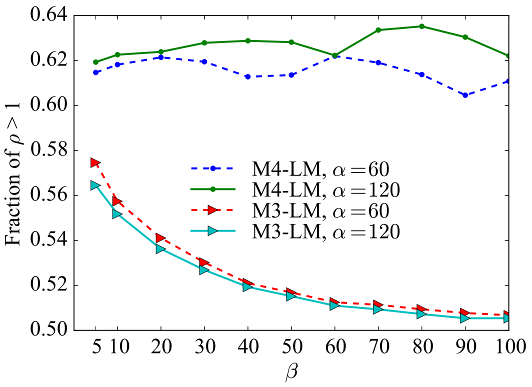

Figure 3. Single Segment, Daily

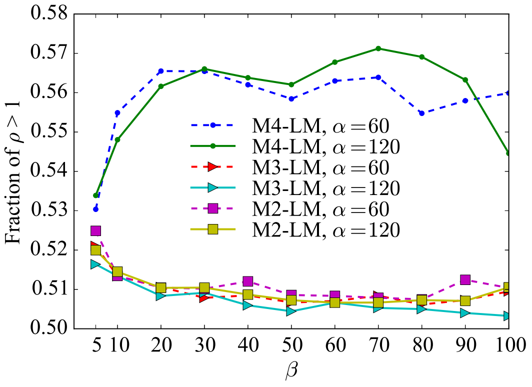

Figure 4. Two Segments, Daily

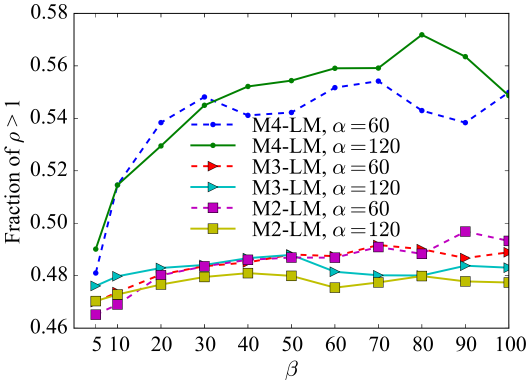

Figure 5. Three Segments, Daily

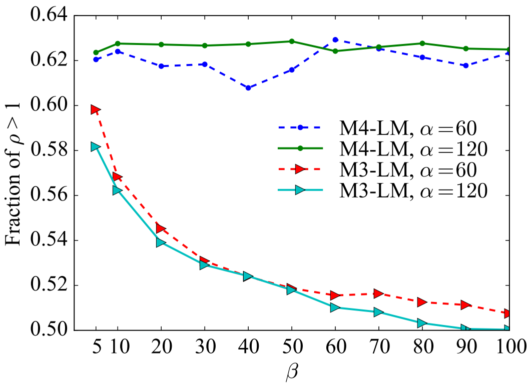

Figure 6. Single Segment, Weekend/Weekday

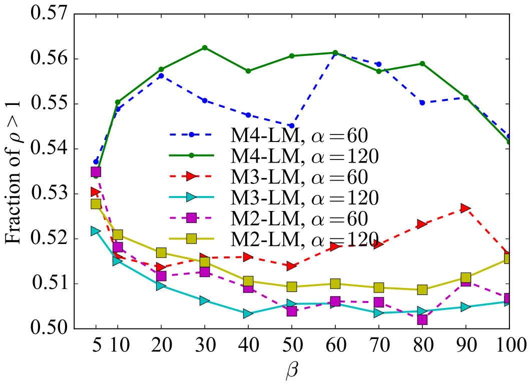

Figure 7. Two Segments, Weekend/Weekday

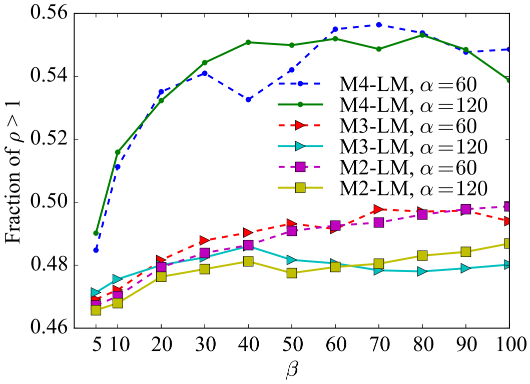

Figure 8. Three Segments, Weekend/Weekday

they are the only methods that manage to improve significantly on the baseline. The improvement remains consistent even for larger sample sizes for the single segment histogram methods as well as for the path-based methods.

4.3. Fractions of Non-Empty Results

Next, we evaluate how significantly filtering by user reduces the sample that is considered for histogram construction. This is to understand the trade-off between higher sample size and "fit" of the trajectories that is made when using the different methods. To evaluate the impact, we look at how often queries return non-empty histograms for segments in S, i.e., when the number of trajectories $n \geq \beta$. Figures 9–11 show how often queries with $d = 1$ and a daily interval return non-empty results with our dataset for different segments given different time ranges α and minimum sample sizes β. The fractions of non-empty returns of the M3 method (cf. Figure 10) are slightly higher than those of the static methods (cf. Figure 9), which is to be expected due to the *shift-and-enlarge* procedure that increases the time range for every subsequent segment.

The fraction of non-empty returns for M4, shown in Figure 11, is significantly lower than for the two other

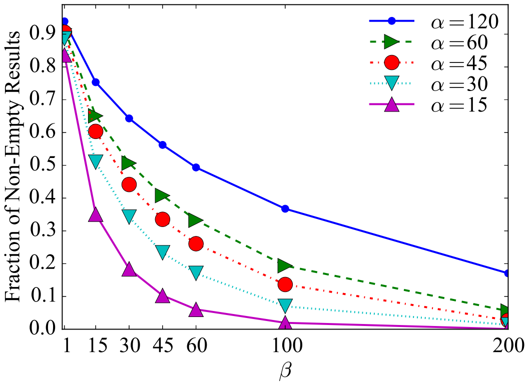

Figure 9. LM/M2 Non-Empty Result Fraction

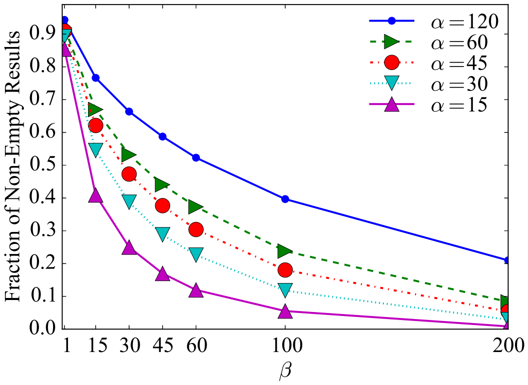

Figure 10. M3 Non-Empty Result Fraction

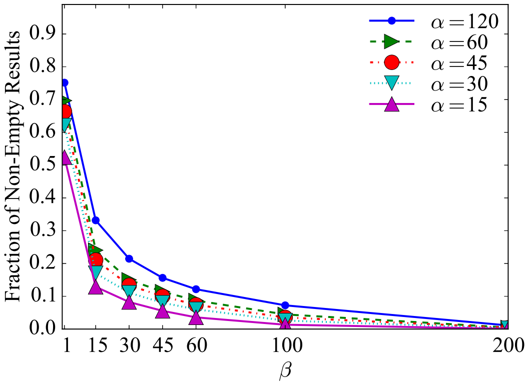

Figure 11. M4 Non-Empty Result Fraction

histogram types, since single drivers conduct considerably fewer trips than the whole population of drivers. Datasets as sparse as the Young Drivers dataset would therefore require the histograms computed with the other methods as fall-back from M4 to provide sufficient coverage of segments and paths.

4.4. Computational Efficiency

While computational efficiency is not the focus of this paper, but is rather a different subject to be considered after the improved accuracy of the on-the-fly approach is established, current results are encouraging.

With our in-memory temporal index, the retrieval of all M3 histograms over a period between one and two years, with $d = 3$ took between 1.4 and 2.5 ms per trajectory for the daily and weekend/weekday interval, and between 0.2 and 0.5 ms for the weekly interval. For the M4 histograms, the retrieval times fell from 0.9 to 1.5 ms and from 0.15 to 0.25 ms, respectively.

5. Conclusion

The empirical study indicates that on-the-fly histogram construction is able to provide considerable accuracy improvements over the state-of-the-art, even for sparser datasets. The evaluation also shows that the improvements are consistent across smaller as well as larger sample sizes. With GPS data becoming more and more abundant, on-the-fly histogram construction will become even more attractive in the future.

Acknowledgments This research was supported in part by a grant from the Obel Family Foundation and by the DiCyPS project.

References

[1] C. Guo, C. S. Jensen, and B. Yang, "Towards total traffic awareness," *SIGMOD Record*, vol. 43, no. 3, pp. 18–23, 2014.

[2] J. Hu, B. Yang, C. S. Jensen, and Y. Ma, "Enabling time-dependent uncertain eco-weights for road networks," *GeoInformatica*, vol. 21, no. 1, pp. 57–88, 2017.

[3] J. Dai, B. Yang, C. Guo, C. S. Jensen, and J. Hu, "Path cost distribution estimation using trajectory data," *PVLDB*, vol. 10, no. 3, pp. 85–96, 2016.

[4] B. Yang, C. Guo, Y. Ma, and C. S. Jensen, "Toward personalized, context-aware routing," *The VLDB Journal*, vol. 24, no. 2, pp. 297–318, 2015.

[5] J. Dai, B. Yang, C. Guo, and Z. Ding, "Personalized route recommendation using big trajectory data," in *ICDE*, 2015, pp. 543–554.

[6] B. Yang, C. Guo, C. S. Jensen, M. Kaul, and S. Shang, "Stochastic skyline route planning under time-varying uncertainty," in *ICDE*, 2014, pp. 136–147.

[7] C. Guo, B. Yang, O. Andersen, C. S. Jensen, and K. Torp, "Ecosky: Reducing vehicular environmental impact through eco-routing," in *ICDE*, 2015, pp. 1412–1415.

[8] Z. Ji, "Path finding under uncertainty," *Journal of Advanced Transportation*, vol. 39, no. 1, pp. 19–37, 2005.

[9] E. Nikolova, M. Brand, and D. R. Karger, "Optimal route planning under uncertainty," in *ICAPS*, 2006, pp. 131–141.

[10] S. Lim, C. Sommer, E. Nikolova, and D. Rus, "Practical route planning under delay uncertainty: Stochastic shortest path queries," in *Robotics: Science and Systems*, 2013, pp. 249–256.

[11] F. C. Pereira, H. Costa, and N. M. Pereira, "An off-line map-matching algorithm for incomplete map databases," *European Transport Research Review*, vol. 1, no. 3, pp. 107–124, 2009.

[12] B. Yang, M. Kaul, and C. S. Jensen, "Using incomplete information for complete weight annotation of road networks," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 5, pp. 1267–1279, 2014.