F-Trail: Finding Patterns in Taxi Trajectories

Yasuko Matsubara[1], Lei Li[2], Evangelos Papalexakis[3], David Lo[4], Yasushi Sakurai[1], and Christos Faloutsos[3]

[1] NTT Communication Science Labs,
[2] University of California, Berkeley,
[3] Carnegie Mellon University,
[4] Singapore Management University

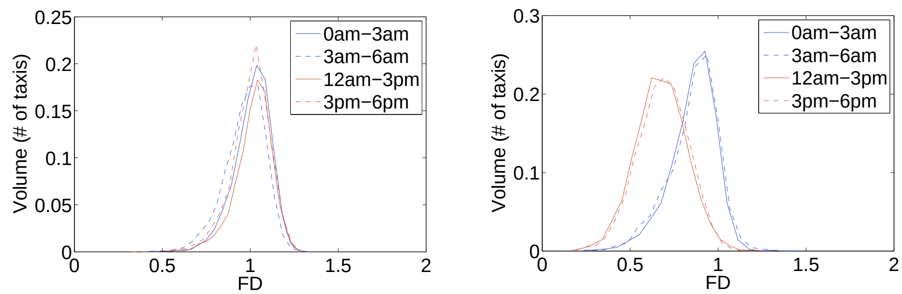
Abstract. Given a large number of taxi trajectories, we would like to find interesting and unexpected patterns from the data. How can we summarize the major trends, and how can we spot anomalies? The analysis of trajectories has been an issue of considerable interest with many applications such as tracking trails of migrating animals and predicting the path of hurricanes. Several recent works propose methods on clustering and indexing trajectories data. However, these approaches are not especially well suited to pattern discovery with respect to the dynamics of social and economic behavior. To further analyze a huge collection of taxi trajectories, we develop a novel method, called F-Trail, which allows us to find meaningful patterns and anomalies. Our approach has the following advantages: (a) it is fast, and scales linearly on the input size, (b) it is effective, leading to novel discoveries, and surprising outliers. We demonstrate the effectiveness of our approach, by performing experiments on real taxi trajectories. In fact, F-Trail does produce concise, informative and interesting patterns.

1 Introduction

What patterns can we derive, using trajectory data from a whole fleet of taxis? What is the normal pattern of activity, and which (if any) outliers exist? We seek to discover such patterns, so that we can spot anomalies, and help the taxi operating company understand general trends, with the ultimate goals of maximizing fuel efficiency, profits, and passenger satisfaction. Trajectory analysis has attracted a lot of interest, including trajectories of (migrating) animals [21, 8, 7, 10], of hurricane paths [4, 2], as well as from an indexing point of view [6, 16]. The latter studied indexing, but not pattern discovery; among the former, the emphasis was on clustering and distance functions on trajectories.

Motivation and challenges Uncovering rules governing collective taxi behavior is a challenging task because of the myriad factors that influence an individual's decision to take a particular action. In this work, we study 10,000 trajectories generated by anonymous taxi drivers, with the aim of measuring social and economic activity. Intuitively, the task of this paper is as follows:

Given a set of GPS coordinates for every taxi, every few minutes, and its status (i.e., 'full' or 'empty'), find the general trends of the fleet of taxis, and



(a) Full taxis (b) Empty taxis

Fig. 1. Different behavior of empty taxis, at night: PDF of fractal dimension (FD) of all trips, for each segment of a day (0-3am, 3-6am etc). (a) Full taxis (b) empty ones. Notice that trajectories of full taxis are linear-like (fractal dimension FD≈1), for any time slot; empty taxis have a wandering behavior (FD $<$ 1) during the day, but linear behavior during the night.

spot outliers.

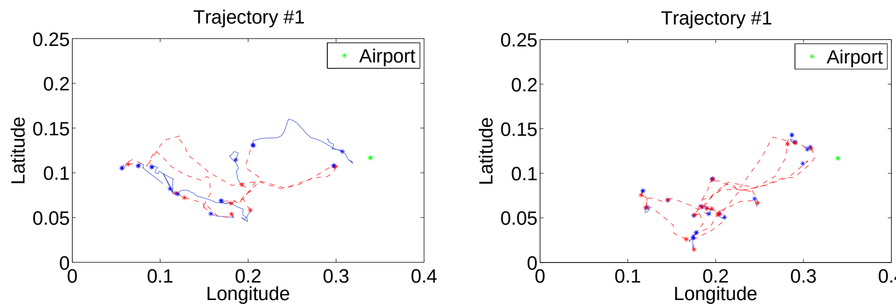
Here, we present a novel method, F-Trail, for finding meaningful patterns and anomalies in a huge number of trajectories. Our approach can scalably and automatically identify typical patterns of taxi behavior, and actually "see" such patterns from the point of view of topology. More specifically, we propose using the *fractal dimension* as a characteristic for trend analysis and extreme detection.

Figure 1 shows a snapshot of our discoveries: Namely, taxi drivers follow linear-like paths when 'full'; and convoluted, 'wandering' paths when empty during the day, but clearly different behavior when 'empty' at night. More specifically Figure 1(a) is the probability density function (PDF) of the 'fractal dimension' (FD) of all trajectories with a passenger (say, "full taxi trails"). Since $FD \approx 1$, for any time of the day, we conclude that, when 'full', taxi drivers follow linear, piece-wise linear, and in general, smooth paths. Conversely, Figure 1(b) shows that, when 'empty', they follow short, wandering paths, creating a bursty-like pattern (see blue dots in Figure 2(b)), with a much lower FD.

The only exception is during the night (solid blue and dotted blue curves in Figure 1(b)): the drivers abandon their 'wandering' behavior, since they don't expect to find nearby customers, and instead go almost straight to taxi plazas or the airport (green dot in Figures 2(a-b)). Notice the linear-like blue paths ('empty trails') in Figure 2(a). We present several more observations, later.

Contributions In this paper we propose a new approach, namely F-Trail, which includes fast and effective techniques that can learn the key trends of a large collection of taxi trails. The contributions of this paper are as follows:

1. *Effective*: We apply F-Trail to a real trajectory dataset, which allows us to identify major trends in taxi behavior, and spot outliers.
2. *Adaptive:* F-Trail describes the common behavior and anomalies of taxi trajectories from the point of view of an individual taxi.



(a) Taxi #1 - night (0am-6am) (b) Taxi #1 - day (12am-6pm)

Fig. 2. Night vs day trajectories of the same taxi: (a) during the night, the blue (='empty') trail is linear-like, probably going to the airport (green dot), or to taxi plazas; (b) during the daytime, the blue (empty) trail is point-like, due to wandering behavior, to find nearby passengers. In all cases, the red (='full') trails are linear-like. Note that we anonymize latitude and longitude information due to privacy concerns.

3. *Scalable*: The careful design of F-TRAIL makes it linear on the number of input size in terms of wall clock time.

Outline The rest of the paper is organized as follows: Section 2 discusses related work. Section 3 introduces our approach and describes how to analyze individual taxi behavior. We conclude in Section 4.

2 Related work

The previous work on mining trajectory data can be grouped into the following categories: (1) design of distance and similarity scores and (2) indexing methods for spatial-temporal databases. For trajectory similarity functions, Vlachos et al. [21] use the longest common sub-sequence, while [8, 7] use minimal description language. Given a similarity score, [21] study the trajectories of marine mammals, while [8, 7] use it to find clusters and outlier trajectories. Gaffney et al. [4, 2] use generative models to group the trajectories of moving objects such as hurricanes. Giannotti et al. [5] study the trajectory pattern mining problem. Very recently, Yuan et al. [24, 23, 11] study a large number of taxi trajectories in Beijing, and present new and sophisticated models to find the different functional regions, and optimal driving route.

Remotely related is the work on indexing and searching moving objects: The work in [16] builds index on moving both spatial and temporal dimensions with pre-aggregation to support OLAP operations. In related work [6, 3] also propose various solutions for answering region retrieval queries, predicting the past and future positions. Similarity search and pattern discovery in time sequences have also attracted huge interest [13, 18, 15].

One of the contributions of F-TRAIL is that it uses fractal concepts to spot patterns in trajectories. Fractals and self-similarity have been used in numerous settings, like financial time series analysis [12], modeling ethernet traffic [22], social network analysis [14, 9], and numerous other settings (see, e.g., [19, 17]).

3 F-Trail

Given several thousands of trajectories, we need to find commonalities and extremes. What do the trips have in common? Can we extract features from such trajectories, to help us understand the dataset? We propose extracting the *fractal dimension* of every such trip. The fractal dimension has several desirable properties: (a) it is invariant to affine transformations, (b) it is fast to compute, and (c) it captures the complexity of the trajectory. Next we give brief background and the intuition behind our proposed solution.

3.1 Preliminaries

(Hausdorff) fractal dimension There exist many fractal dimensions (Hausdorff, Minkowski, correlation, information, etc). Among them, we use only the first one, which is formally defined as follows:

The *(Hausdorff) fractal dimension*, or simply the *fractal dimension* for a range of scales (r_1, r_2) and for a given, self-similar, point-set in an n-dimensional address space, is defined as the exponent of the law [19]:

$$N(r) \;=\; C\, r^{-FD} \quad (r_1 < r < r_2) \tag{1}$$

where (r_1, r_2) is a suitable range of scales, and $N(r)$ is the number of non-empty cells, when we impose a grid of side r, on our dataset. Intuitively, if we plot $N(r)$ vs r in log-log scales, the plot will be linear for the range of scales (r_1, r_2). We refer to such plots as *Hausdorff plots*, and we report the slope and constant (= intercept), for each taxi trajectory.

Computational Complexity. Linear. More specifically, we have:

Lemma 1. *The computation time for the fractal dimension is $O(N)$, that is, linear on the number of points N.*

Proof: Using the so-called *box-counting* method [19], we only need to go over the data points a few times. **QED**

Basic properties for trajectory data Our dataset consists of 10,000 trajectories taken from anonymized taxi drivers in the large city, where each trajectory corresponds to the trail of each taxi for an entire day. The dataset has the following attributes: GPS coordinates (i.e., longitude, latitude), a timestamp, and the passenger status (i.e., 'full' or 'empty').

Figure 3 shows an example of a taxi trajectory of an entire day. The horizontal and vertical axes show the longitude and latitude of the GPS points respectively, where red and blue lines show the trails of the taxi for each status (full and empty). Note that we anonymize latitude and longitude information due to privacy concerns. The red and blue dots indicate the locations of 'pick-up' and 'drop-off' points. The green dot at (0.3391, 0.1168) is the location of the international airport, and (0.15 - 0.25, 0.06) is a downtown area. In this figure, we can see that most trajectory lines are between the airport and downtown via highways.

Here we define some terminology. We will refer to each set of trail points with passengers (shown as red lines in Figure 3), as a *'full' trail*, and refer to a set of points without passengers (i.e., empty) as an *"empty" trail*. That is, the entire trajectory consists of a combination of 'full' and 'empty' trails.

3.2 Intuition - Fractal dimension as a feature

Our goal is to analyze the trajectories, and specifically to characterize the underlying behavior of taxi drivers, and gain insight into how and why the observed characteristics arise.

The taxi drivers have social and economic priorities and follow their own strategies for success. We want to extract detailed information on their behavior, especially regarding their mobility patterns. There exist numerous time-series analysis methods, including FFT and wavelets, but they depend strongly on the locations of trajectories, which makes it hard to find the economic strategies and social behavior. How can we characterize the shape of trajectories, while ignoring their location?

Approach 1 *We propose to use the fractal dimension of each taxi trajectory, as a feature for finding patterns and groups.*

Several real datasets are self-similar, and thus have an intrinsic, or *fractal* dimension: the peripheries of clouds and rain patches ($D = 1.3$), coast-lines ($D = 1.1$ to 1.58 for Norway), and many more [1]. Are our taxi trajectories self-similar? The answer is 'yes', for the vast majority of them. The ones that don't, are either too short, or deserve further examination, being the exceptions. Next we give the intuition and necessary definitions.

Figure 4 shows some trajectories as well as the tools to measure their fractal dimension: the odd columns (i.e., Figure (a,c,e)) are synthetic, and the even ones (i.e., Figure (b,d)) are real trajectories. Intuitively, a set of points (like our taxis' (x,y) coordinates per unit time) is a fractal if it exhibits self-similarity over all scales. The way to interpret the value of the fractal dimension is as follows:

- $FD = 1$: This happens when the trajectory has iso-spaced points, along a line or a smooth curve, (see synthetic dataset Figure 4(c)).
- $FD > 1$: This happens when the taxi does twists and turns, like the real trajectory of Figure 4(d).
- $FD < 1$, This happens when the taxi does many stops, like the real trajectory of Figure 4(b).
- $FD = 0$, if the taxi is completely static, in which case the trajectory reduces to a (multi)point.

The two extremes in Figure 4(a,e) correspond to the so-called 'Cantor dust', and the 'Hilbert curve'. The former is derived from a line segment, by recursively

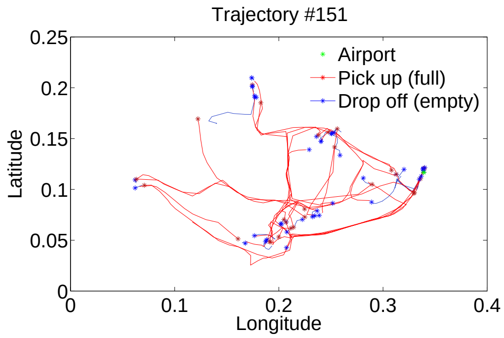
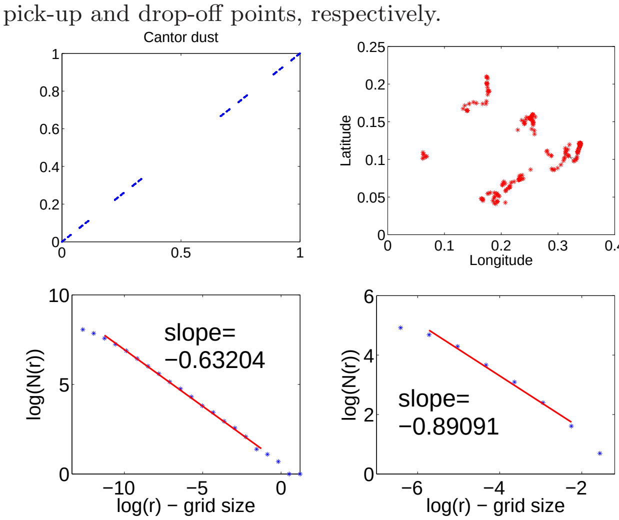
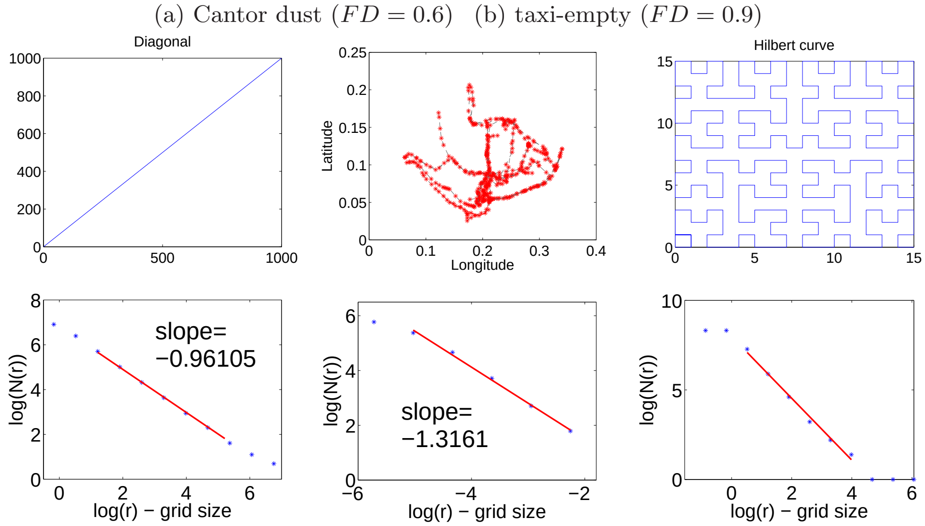


Fig. 3. Plotting conventions for a taxi trajectory: the red/blue lines indicate trails of the taxi with/without passengers, respectively. The red and blue dots indicate the locations of pick-up and drop-off points, respectively.



(a) Cantor dust ($FD = 0.6$) (b) taxi-empty ($FD = 0.9$)



(c) line ($FD = 1$) (d) taxi-full ($FD = 1.3$) (e) Hilbert curve ($FD = 2.0$)

Fig. 4. Intuition behind trajectories and their fractal dimension(s). Top row: trajectories (sets of (x,y) points). Bottom row: the 'Hausdorff' plots, in log-log scales - the slope is the fractal dimension. Columns have trajectories of increasing fractal dimension ($\approx$ complexity), from $FD = 0.6$ up to 2.

deleting the middle third, and has fractal dimension $\log(2)/\log(3) \approx 0.63$. The latter is a *space filling curve*, with fractal dimension = 2, covering the whole 2-d space, in the limit.

Thus far we have introduced the fractal dimension for individual trajectory analysis in order to understand the taxi behavior. However, the behavior of each taxi could vary in a month, a week, or even in a single day since the dynamic strategy typically beats the static strategy. Actually, each empty taxi exhibits distinct behavior in different time ranges according to the distribution of passengers (see Figure 1).

Approach 2 *We propose to apply a short-window approach to the fractal dimension, which is more flexible for trajectory analysis.*

Instead of handling the entire trajectory, we locally analyze the fractal dimension of each snapshots to obtain a better understanding of time-varying social behavior.

3.3 F-Trail analysis

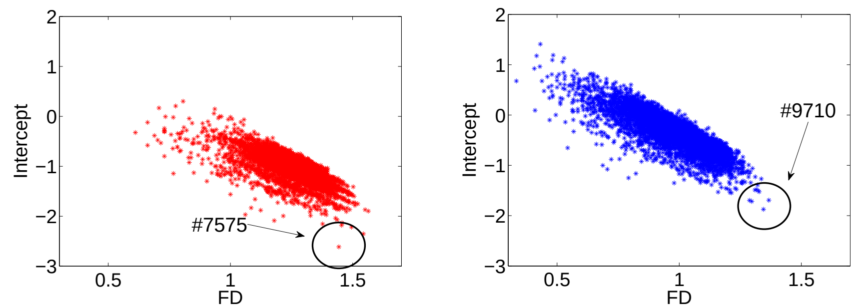
We now introduce our approach and describe how to analyze individual taxi behavior.

Fractal trajectories - are there clusters? For a few trajectories, a human could eye-ball them, and derive the above patterns. But, how can we accomplish this automatically for thousands of trajectories? Our first idea is to compute the fractal dimensions for individual trajectory analysis. We begin by investigating the sociability of taxi movement by measuring the fractal dimensions of trajectories of two statuses (i.e., 'full' and 'empty' trails). We compute the fractal dimension of each trail, which help us to understand how the taxi drivers find the passengers and how they pick them up.

Do taxi drivers take their passengers over direct paths? Are their trajectories different, when they are empty, looking for passengers? It turns out that F-Trail can answer both questions, and the answers are 'yes', with a few twists. Let's see the details.

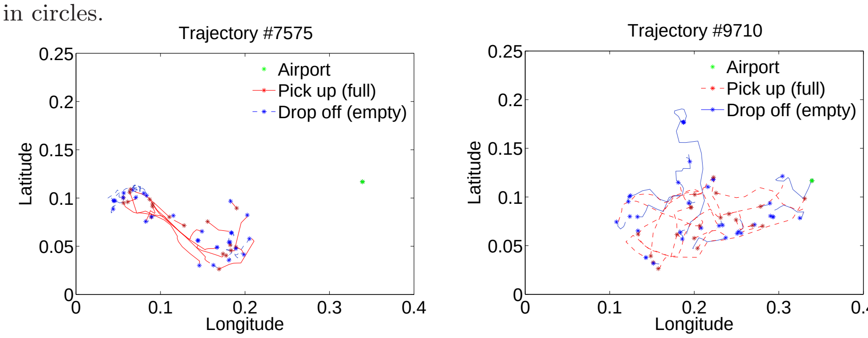
Figure 5 shows the fractal dimension (FD) versus the intercept of full and empty trails. For most 'full' trails, the fractal dimension is between $1.1 - 1.3$, while for 'empty' trails, it is between $0.8 - 1.1$. Here, $FD = 0$ is the burstiest (i.e., static taxi), $FD = 1$ corresponds to a taxi moving uniformly on a line or smooth curve, and $FD = 2$ (maximum) would be for a taxi that is uniformly distributed over the whole address space. This figure shows remarkable differences in the behaviors of each status. Thus, we have:

Observation 1 (Typical behavior) *Typical taxi behavior over an entire day is to have fractal dimensions between* $FD = 1.1 - 1.3$ *for full trails and* $FD = 0.8 - 1.1$ *for empty trails.*



(a) FD (full trails)
(b) FD (empty trails)

Fig. 5. Pattern and extremes: FD vs intercept plots of the trajectories: Each dot corresponds to a trail of (a) full taxi and (b) empty taxi. Full taxis tend to have higher fractal dimension. Also note the extremes, like 'full' taxi #7575 and 'free' taxi #9710, in circles.

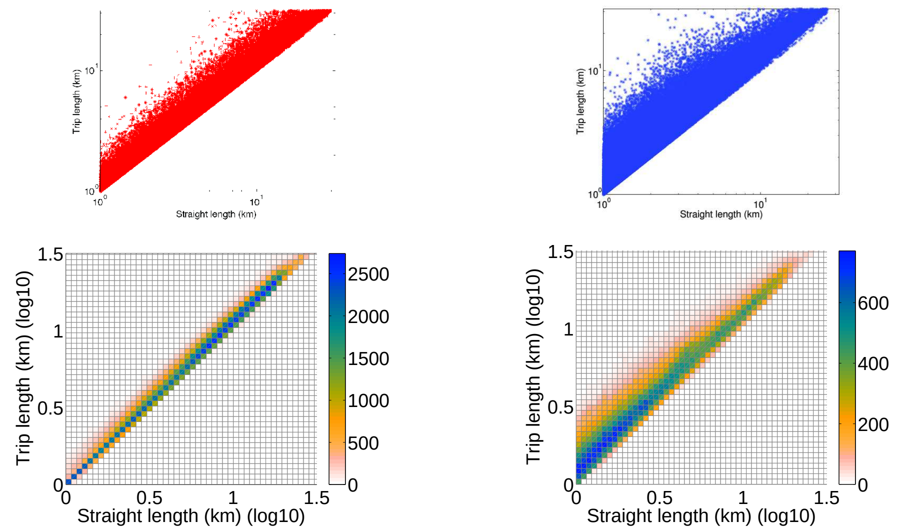


(a) Extreme (in the "full" status)
(b) Extreme (in the "empty" status)

Fig. 6. Extremes of full and empty trails: (a) full trail of taxi # 7575 (red line) has high FD, but the intercept ($\approx$ total length) is the lowest: a lot of short, straight-like rides. (b) Empty trail of taxi # 9710, having the highest FD: it's blue rides are long, and mainly straight like, which means no local 'wandering' for its next passenger.

Notice that the above observation is invariant to affine transformations. Moreover, it helps us spot a clear distinction between 'full' and 'empty' trail sets (see Figure 5, FD vs. intercept plot). This is because, unlike many full taxis that head straight for their destinations, empty taxis frequently do many stops and turns, to find a new passenger, which leads to low fractal dimension.

A further observation that we can derive from the fractal dimensions of trajectories shown in the same figure is that there are several extremes for each of the full and empty trail sets. For instance, the trajectory #7575 is the one extreme, due to its low intercept, which implies that the particular trajectory covers a very small area. In fact, the taxi driver adopted a different strategy and focused only on a highly-populated area (see Figure 6(a)). The trajectory #9710 is an extreme example of the empty status and shows a high fractal dimension. Actually the trajectory includes many line-like trips (long blue segments), for his/her long 'passenger search' (see Figure 6(b)), in contrast to the majority of taxi drivers, who wander locally, looking for their next passenger.



(a) Full taxis (slope=1.04, intercept=0.03) (b) Free taxis (slope=0.97, intercept=0.16)

Fig. 7. Straight distance vs. trip length of each ride (log-log scales): scatter plots (top) and density plots (bottom). Full taxis pick the shortest path (intercept $\approx$0); empty taxis drive about 40% more length than necessary (intercept=0.16, $10^{0.16} \approx 1.4$).

Thus we have almost answered the first question: taxis seem take their passengers to smooth, line-like curves. The next natural question is: are these the shortest paths? We address this question next.

Trip length vs. crow's flight - any waste? So far we have examined the dynamics of the taxi behaviors at an individual level, and proposed a simple model capturing the fractal dimensions. To check whether 'full' taxis actually use the shortest path, we do another scatter-plot: For every trip, we plot the "crow's flight" length (= Euclidean distance between pick-up point and drop-off point) and the reported length (= sum of lengths between successive location snapshots). Figure 7 shows these scatter plots, where every dot is a trip. Of course, nothing is below the diagonal, and there is heavy over-plotting. The bottom row contains the density plots.

Notice that the 'full' trips are almost on the diagonal (slope=1, intercept$\approx$0), which means that the taxi drivers are efficient, in the shortest path sense.

Observation 2 (Taxi driver efficiency) *Most taxi drivers take their passengers via the shortest path (or very close to it).*

By contrast, when 'empty', the slope is still 1, but the intercept is higher (0.16, in log-log scales), which means that the drivers 'wander', with many turns and returns, until they find a new passenger. Mathematically: $l = 10^{0.16} * s^1$ or

$$l \approx 1.4 * s \quad (2)$$

where l is the reported length of the trip, and s is the straight ("crow's flight") length.

Observation 3 (40 percent wandering) *Drivers on 'empty' status have more convoluted trajectories than on 'full' status, driving about* 40% *longer than necessary.*

This observation agrees with the intuition: Drop-offs are typically in residential areas, with many, narrow roads, and the taxi drivers have to turn and loop, until they find another passenger. Thus, the trajectory is more convoluted. In contrast, 'full' taxis typically go on highways, which are straight or with a few smooth turns and thus the trajectories are simpler and more efficient.

Fractal dimension - any changes with the hour of the day? So far we have seen that most empty taxis are likely to pick up new passengers geographically close to the last drop-off point, and thus we would expect that this strategy to minimize their effort (e.g., maximize fuel efficiency, profits). However, as described in the introduction, this is not always the case. We thus introduce a *short-window* approach to the fractal dimension (FD), which enable us to characterize the taxi behavior for each time interval (i.e., snapshot). Specifically, instead of analyzing all the (x,y) snapshots of, say, taxi i, we study separately the snapshots at 0-3am, 3-6am, etc.

As we have seen in Figure 1, the behavior of 'full' taxis is the same for all hours of the day, but the 'empty' ones vary in a very interesting way:

Observation 4 ('empty' at night: differ) *There is a clearly distinct pattern of empty taxis at night time: instead of convoluted, 'wandering' paths, drivers choose line-like paths.*

In retrospect, the observation above makes sense: During the night (0-6am), the drop-off place is probably at a residential area, and there is slim chance to find another passenger nearby. Thus, taxi drivers choose to drive straight to places with higher chances of demand (airport, down-town, etc).

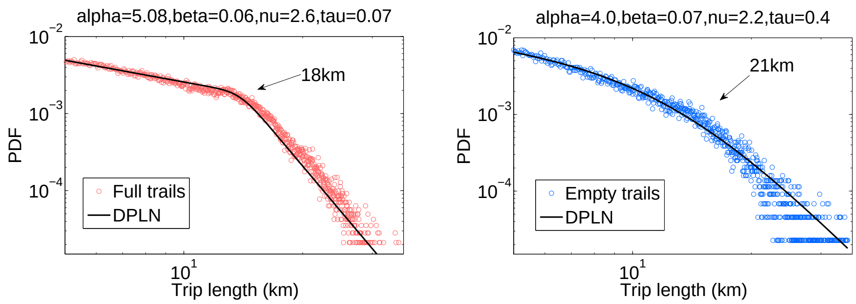
Power law in the trip-length distribution? Whenever there is self-similarity and fractals, we often have power laws and scale-free distributions. Does this hold for the trip-lengths, in our case? Surprisingly, it turns out that the double Pareto lognormal (DPLN) distribution yields good fits to our data.

The DPLN distribution generalizes the power-law and lognormal distributions, and is expressed by the following equation,

$$f(x|\alpha,\beta,\nu,\tau) = \frac{\alpha\beta}{\alpha+\beta}[\exp^{\alpha\nu+\alpha^2\tau^2/2} x^{-\alpha-1}\Phi(\frac{\log x-\nu-\alpha\tau^2}{\tau}) + x^{\beta-1}\exp^{-\beta\nu+\beta^2\tau^2/2}\Phi^c(\frac{\log x-\nu+\beta\tau^2}{\tau})] \quad (3)$$

where Φ and Φ^c are the cdf and complementary cdf of $N(0,1)$, and for further details (e.g., parameters α, β, ν, τ), see [20].

Figure 8 shows the DPLN fitting results of trip-length vs count distribution. The figure shows the PDF for the (a) full (red circles) and (b) empty (blue circles) rides, in log-log scales. There is a power-law behavior up to the 'knee' (at about 20 km), and then a sharp (power-law) drop after that. Notice that the knee is at roughly the radius of the city of study. Consequently, we have:



(a) Full taxis (b) Empty taxis

Fig. 8. PDF of trip length (length vs count, in log-log scales): knees at characteristic scales of ≈20 km, which is roughly the radius of the city.

Observation 5 (Trip length: DPLN behavior) *The trip length distribution is skewed, for both 'full' and 'empty' rides, with a power-law that has a 'knee' at* ≈ 20 *km. This is exactly the so-called "doubly Pareto lognormal" (DPLN).*

From Figure 8 (a), we observe two types of customers: the 'below-city-radius' ones, that take short trips, or airport-to-downtown; and the (much more rare) 'above-city-radius' ones, that maybe hire a taxi for sightseeing, or for several days. Such results (i.e., trip-length DPLN fitting) could be used to analyze the tradeoff between the cost of finding passengers (Figure 8 (b)) and the fares received from them (Figure 8 (a)) to design pricing structure, which would maximize the total revenue.

4 Conclusions

We investigated patterns of human mobility on a large collection of taxi trajectories, and proposed a new method, F-Trail, to find meaningful patterns and anomalies. Our approach is (a) linear on the input size, and (b) able to spot meaningful general trends, as well as outliers. In more detail, our main discoveries are as follows:

- Typical for taxi trajectories is to have fractal dimension between 1.1 to 1.3 for full trails and 0.8 to 1.1 for empty trails, which means that most taxis take their passengers to smooth, line-like curves, while empty taxis have many stops or bursty-like patterns.
- Most surprisingly, we found that most taxi drivers change their strategies in different time ranges according to passenger demand, i.e., there was a very interesting deviation of FD for empty taxi during the night.

Acknowledgement. We would like to thank the taxi company that provides us the data. This work was sponsored by the Defense Threat Reduction Agency and was accomplished under contract No. HDTRA1-10-1-0120 and by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-09-2-0053.

References

1. A. Belussi and C. Faloutsos. Estimating the selectivity of spatial queries using the 'correlation' fractal dimension. In *Proc. of VLDB*, pages 299–310, Zurich, Switzerland, Sept. 1995.
2. D. Chudova, S. Gaffney, E. Mjolsness, and P. Smyth. Translation-invariant mixture models for curve clustering. In *KDD*, pages 79–88, 2003.
3. P. Cudre-Mauroux, E. Wu, and S. Madden. Trajstore: An adaptive storage system for very large trajectory data sets. In *ICDE*, pages 109 –120, march 2010.
4. S. Gaffney and P. Smyth. Trajectory clustering with mixtures of regression models. In *KDD*, pages 63–72, 1999.
5. F. Giannotti, M. Nanni, F. Pinelli, and D. Pedreschi. Trajectory pattern mining. In *KDD*, pages 330–339, 2007.
6. M. Hadjieleftheriou, G. Kollios, V. J. Tsotras, and D. Gunopulos. Indexing spatiotemporal archives. *VLDB J.*, 15(2):143–164, 2006.
7. J.-G. Lee, J. Han, and X. Li. Trajectory outlier detection: A partition-and-detect framework. In *Proceedings of the 2008 IEEE 24th International Conference on Data Engineering*, pages 140–149, Washington, DC, USA, 2008. IEEE Computer Society.
8. J.-G. Lee, J. Han, and K.-Y. Whang. Trajectory clustering: a partition-and-group framework. In *SIGMOD*, pages 593–604, 2007.
9. J. Leskovec, D. Chakrabarti, J. M. Kleinberg, and C. Faloutsos. Realistic, mathematically tractable graph generation and evolution, using kronecker multiplication. *PKDD*, pages 133–145, 2005.
10. Z. Li, B. Ding, J. Han, R. Kays, and P. Nye. Mining periodic behaviors for moving objects. In *KDD*, pages 1099–1108, 2010.
11. W. Liu, Y. Zheng, S. Chawla, J. Yuan, and X. Xing. Discovering spatio-temporal causal interactions in traffic data streams. In *KDD*, pages 1010–1018, 2011.
12. B. Mandelbrot. *Fractal Geometry of Nature*. W.H. Freeman, New York, 1977.
13. Y. Matsubara, Y. Sakurai, C. Faloutsos, T. Iwata, and M. Yoshikawa. Fast mining and forecasting of complex time-stamped events. In *KDD*, pages 271–279, 2012.
14. Y. Matsubara, Y. Sakurai, B. A. Prakash, L. Li, and C. Faloutsos. Rise and fall patterns of information diffusion: model and implications. In *KDD*, pages 6–14, 2012.
15. Y. Matsubara, Y. Sakurai, and M. Yoshikawa. Scalable algorithms for distribution search. In *ICDM*, pages 347–356, 2009.
16. D. Papadias, Y. Tao, J. Zhang, N. Mamoulis, Q. Shen, and J. Sun. Indexing and retrieval of historical aggregate information about moving objects. *IEEE Data Engineering Bulletin*, 2002.
17. H.-O. Peitgen, H. Juergens, and D. Saupe. *Chaos and Fractals: New Frontiers of Science*. Springer-Verlag New York Inc., 1992.
18. Y. Sakurai, C. Faloutsos, and M. Yamamuro. Stream monitoring under the time warping distance. In *Proceedings of ICDE*, pages 1046–1055, Istanbul, Turkey, April 2007.
19. M. Schroeder. *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise*. W.H. Freeman and Company, New York, 1991.
20. M. Seshadri, S. Machiraju, A. Sridharan, J. Bolot, C. Faloutsos, and J. Leskovec. Mobile call graphs: beyond power-law and lognormal distributions. In *KDD*, pages 596–604, Las Vegas, Nevada, USA, 2008.
21. M. Vlachos, G. Kollios, and D. Gunopulos. Discovering similar multidimensional trajectories. In *Data Engineering, 2002. Proceedings. 18th International Conference on*, pages 673–684, 2002.
22. W. Willinger, M. Taqqu, R. Sherman, and D. V. Wilson. Self-similarity through high variability: statistical analysis of ethernet LAN traffic at the source level. *ACM SIGCOMM'95. Computer Communication Review*, 25:100–113, 1995.
23. J. Yuan, Y. Zheng, and X. Xie. Discovering regions of different functions in a city using human mobility and pois. In *KDD*, pages 186–194, 2012.
24. J. Yuan, Y. Zheng, X. Xie, and G. Sun. Driving with knowledge from the physical world. In *KDD*, pages 316–324, 2011.