Driving Analytics: Will it be OBDs or Smartphones?

Rufeng Meng
University of Illinois

Chengfeng Mao
University of Illinois

Romit Roy Choudhury
University of Illinois

ABSTRACT

This paper shows that smartphones are capable of estimating a car's speed in an overwhelming majority of situations. Comparing against OBD data as ground truth, we find that continuous GPS based estimates offer greater than 98% correlation across various road, traffic, and weather conditions. Of course, GPS is energy hungry and may only be relevant for taxi or Uber-like applications where the user can plug in their phones in the car charger. For regular drivers who are not likely to plug in their phones, we find that sensor based techniques, combined with map and/or crowd-sourced data, can achieve greater than 94% correlation. Given that smartphones can offer reliable speed estimates as well as other contextual information (e.g., SMS/email distraction, user's location, or even the hours of sleep the previous night), we believe that the smartphone platform is a superset of OBD. Further sophistication in sensor fusion and machine learning should enable detection of risky driving behavior, such as hard braking, aggressive acceleration, road accidents. Finally, given the rapid hardware and software innovation on the smartphone ecosystem, there is little doubt that it will hold centerstage in the emerging market of driving analytics.

1. INTRODUCTION

With the convergence of personal sensing, computing, and communications, the ability to quantify human behavior in daily lives is rapidly on the rise. One of the areas gearing up for technological disruption is *driving analytics*, also called *vehicular telematics* [1, 2, 8]. Briefly, driving analytics refers to the ability to *measure* the quality of human driving, ultimately assigning a score it. A driver's score, for instance, could to be designed to reflect the riskiness in her driving [14], while a different score could even reflect her fuel efficiency. Car insurance companies are keen on such analytics with the aim of enabling personalized insurance; apps like Uber, that allow for citizens to moonlight as cab drivers, also desire such analytics to be able to attach a "reputation index" to drivers. Needless to say, today's quality and risk assessments are coarse grained, based on parameters such as an individual's age, marital status, years of driving, and residence zip code. The future lies in far finer grained insights – statistics on the jerkiness of a driver, ability to maneuver through traffic, dozing off while driving, time spent on familiar roads, SMSs sent during driving, etc. [2]. If risk assessments in other industries – healthcare, stocks, real estate – are all based on detailed, high dimensional data, so should driving.

A large number of companies (both established and startups) have plunged into this space, performing various forms of vehicular activity recognition and analytics. These companies may be categorized into two distinct classes based on how they source their data – (1) those that are using the on-board diagnostic (OBD) devices, and (2) those that rely only on smartphones.

While OBDs benefit from precise measurements of vehicular speed, accelerator throttle, and various engine conditions, all sourced from the vehicle's internal computer, the smartphones rely on the various sensors that capture information about the cars motion patterns, the driver's activities, and even broader background context (such as the driver being at the pub just prior to driving). Unsurprisingly, a debate is brewing that essentially asks: *can the smartphone be an effective substitute for the OBD?* More specifically, given that a vehicle's speed is a critical input to driving analytics, the debate degenerates to whether smartphone sensors can infer instantaneous speed. If it can, then the smartphone can become a superset of OBDs, i.e., it can match an OBD in understanding the car's motion patterns, and far exceed the OBD in sensing the driver's personal behavior. Since driving score is a function of both behaviors – the car's and the human's – the smartphone may be in the best position to serve the need.

The above question is non-trivial because even though smartphones are equipped with various sensors like accelerometers and gyroscopes, none of them directly measure the speed of a car. Instead they measure the linear and angular acceleration of the object on which they are placed [10]. This is an indirect measure of the vehicle's speed, and also polluted by interfering motions such as the car bouncing on the road, potholes [6], lane changes, turns, and even the user using the phone. Figures 1(a) and (b) show a Samsung Galaxy phone's accelerometer and gyroscope data, respectively, in contrast to the pristine speed data from an OBD in Figure 1(c), when both devices were placed in the same car. Clearly, inferring speed from the (fluctuating) smartphone data is not straightforward. Of course, smartphone GPSs compute speed but its heavy energy consumption [5, 4, 9, 13] renders them unsuitable for continuous usage. This paper is aimed at designing an energy-efficient speed-estimation technique on smartphones, comparable to that of the OBD.

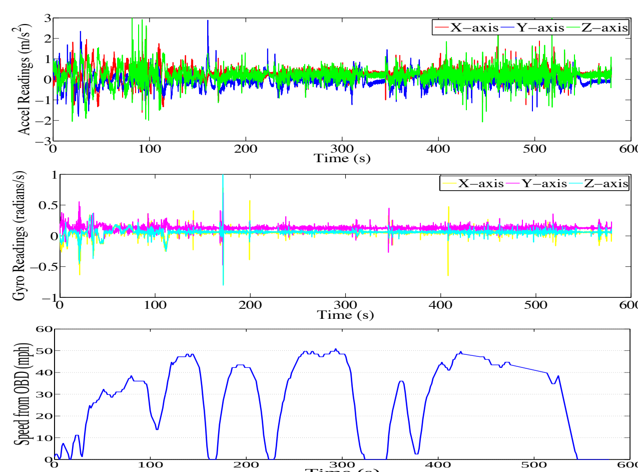

Figure 1: (a) Accelerometer and (b) gyroscope data when phone placed in the cup holder and driving in the university town. (c) OBD data captured from the same car trips.

Figure 2: Eight Samsung (Android) smartphones placed at various positions in the car during the data collection phase. One phone taped to the dashboard to gather data not polluted by the wobbling of the phone. Also, middle figure shows the OBD connected to the car and laptop.

While indirect speed estimation is challenging on smartphones, opportunities exist too. The variety of other sensors on the phone, that apparently are not useful for speed estimation, could be creatively used to aid accelerometers and gyroscopes. Compasses can indicate when a user is taking turns; pressure sensors can suggest the variation in the terrain of the road; GPS can be used infrequently if it significantly cuts back on estimation error. Moreover, the access to the Internet/cloud offers access to maps, speed limits on roads, etc. all of which can be fed to an inference engine. Finally, opportunities arise in cases where the user drives down habitual paths – the repeated sensor data from these roads can be mined to improve estimation.

Main Findings:

This paper explores the landscape of such opportunities and quantifies the gap between OBD and smartphone-estimated speed. Reported results suggest that smartphones can attain speed estimations that correlate 98% to OBD without energy constraints, and around 94.6% to OBD when GPS is completely turned off. These results are reasonably robust to various scenarios, such as in highways and city streets, weather and traffic conditions, and different phone placements. While this is already promising for smartphones, we believe that further improvements are possible through more sophisticated machine learning and signal processing on the sensor data. We hint at some of these techniques in Section 5, but leave a deeper treatment to future work. We end the paper with a summary of how smartphones offer multi-modal capabilities that together make them powerful, agile to the fast changing market of personal computing and analytics.

2. MEASUREMENT SETUP

We begin with a description of our experiment set up, followed by metrics to evaluate our speed estimation schemes.

2.1 Data Collection Methodology

We use Samsung smartphones as our experimentation platform. We employ 8 devices from 3 different models, running Android 4.2+. The phones are placed at 8 different orientations inside the car, including driver's pant pocket, shirt pocket, cup holder, passenger's pockets, back-passenger's hands, etc. One of the phones is taped on the dashboard so that it does not "wobble"as the car moves – the top part of the phone screen is made to point towards the driving direction. The phones log data from a range of sensors, namely, accelerometer, gyroscope, compass, barometer, GPS, microphone, WiFi, and cell tower. The motion sensors are sampled at the highest possible rate; the microphone is sampled at 16 KHz; the GPS is sampled every second to extend battery life.

The phones run a custom-made data sensor-data logging application. To synchronize measurements, the app is activated on all the phones, following which the phones are held together and shaken vehemently. The accelerometer data from the shake synchronizes the phones to the granularity of milliseconds. A remote server gathers all the data and runs visualization and analytics on them. In addition, an OBD-II device is plugged into the car and connected to the laptop – all relevant fields offered by the OBD are logged, including vehicle speed, throttle, etc. The car is driven in the Champaign–Urbana area in Illinois, under various environments, including highways, congested urban roads, empty roads at night. Data is also collected under various weather conditions (especially to understand the effect of GPS), as well as under different speed regimes on these roads. The data collection lasted for 12 different trips, performed over 1 week, resulting in around 500 MB of data.

2.2 Metrics of Interest

Our goal is to characterize the gap between OBD and smartphone-estimated speed. Treating the two data streams as time series signals, we compute the *covariance coefficient* for each trip defined as:

$$\rho = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y}$$

This coefficient ranges from 0 to 1; higher the coefficient, better is the similarity between OBD and the estimated speed.

We also quantify the difference between OBD and smartphone–estimated speed (referred to as *SES* henceforth) by measuring the distribution of the instantaneous differences between the two. Specifically, we will plot the cumulative distribution function (CDF) of $\delta(t)$ defined below.

$$\delta(t) = OBD(t) - SES(t)$$

As the SES moves closer to the OBD values, the CDF curve of $\delta(t)$ is expected to move closer to the $X = 0$ vertical line. We will observe the behavior of $\delta(t)$ for varying parameters, such as speed regimes, type of roads, phone positions, etc.

3. OBD VS. SMARTPHONE GPS

For taxi companies or UberX applications, or even when people drive for long distance, drivers may be able to plug their phones

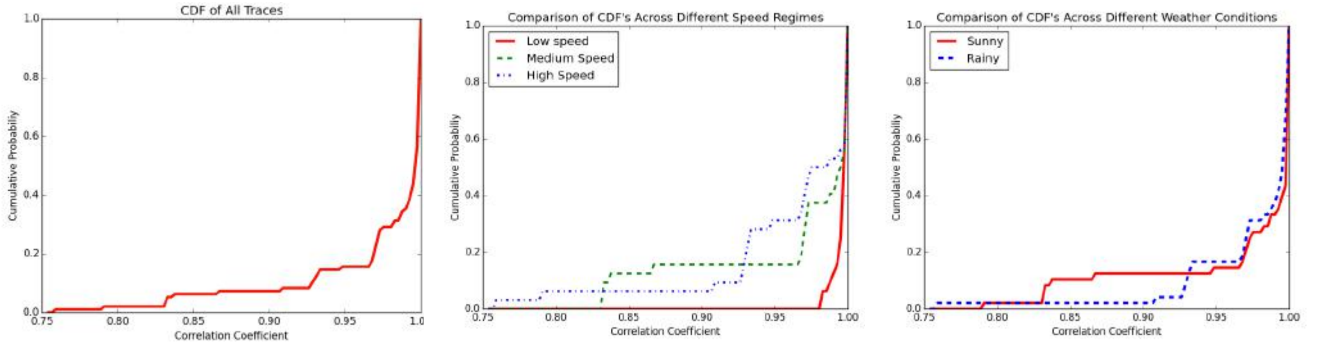

Figure 3: (a) CDF of correlation between GPS and OBD across all traces in all conditions. (b) CDF of correlation for 3 speed regimes. (c) CDF of correlation for sunny and rainy weather, characterizing the impact of clouds on GPS speed estimates.

into car chargers. In such cases, the smartphone GPS can be activated to continuously estimate the car's speed. This section verifies if GPS is indeed a strong approximation for OBD, especially under various phone orientation, weather, and road conditions. If so, then we should be able to use GPS as the ground-truth thereafter, making our experiments easier and scalable.

Overall Results

Figure 3(a) reports the overall CDF of the correlation coefficient between GPS and OBD for all traces in our database. Observe that the median correlation is upward of 0.99. Of course, there is a tail at which the correlation falls to 0.8, however, this is in less than 2% of the cases.

Varying Speed Regimes

To understand the reasons, Figure 3(b) plots the CDF of correlation for three different speed regimes, namely low, medium, and high. The regimes are defined as *low=[0,40], medium=[41, 60], high=[61, 100]* miles per hour. Evidently, the correlation is stronger for lower car speeds, but the difference is small. Even when the car is traveling at a high speed, correlation borders around 0.9. Also, as we see later, the GPS values are sometimes one-off outliers, and simple outlier detection algorithms eliminate those values. The post-processed GPS speeds produce even higher correlations.

Impact of Weather

Figure 3(c) reports on the impact of two different weather conditions on GPS speed estimates. The curve for "sunny" weather is slightly better than "rainy". However, as long as the GPS obtained locks from the satellites, the results seemed comparable.

Different Phone Placements

Figure 4 shows the various percentile of the correlation distribution between OBD and GPS for 8 different phone placements. Correlation is consistently high except in one phone position – the shirt pocket – where a small percentile of the distribution drops down to 0.78. While we are unsure about the reason of this anomaly, we believe this is a less common phone placement among others. The correlation is consistently strong in all others.

4. OBD VS. SMARTPHONE SENSORS

This section turns to the case where smartphone GPS cannot be employed to preserve battery life. We ask if its still possible to estimate speed using sensor data from smartphones, and combining them, if needed, with information from the Internet (e.g., maps, road traffic, historical data)

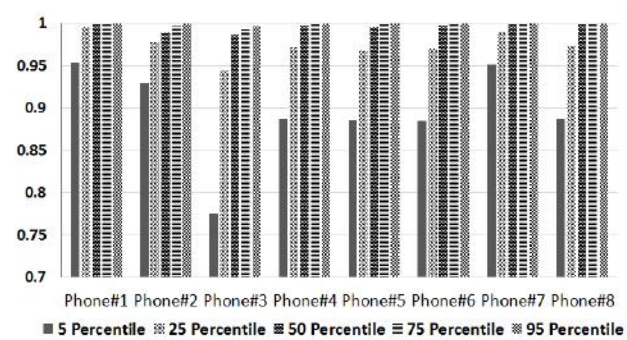

Figure 4: Correlation for 8 phone positions. The phone positions are: 1) Mounted. 2) In Cup-Holder. 3) Driver's shirt pocket. 4) Driver's pants pocket. 5) Co-driver's seat. 6) In bag (bag is in back seat). 7) Front passenger is using a phone. 8) Phone is stuck on the dashboard on the co-driver's side having the same heading as the vehicle.

4.1 Estimating Speed

Our experiment set up is as described earlier, except that users are now use the GPS data as the ground truth. We systematically develop a speed estimation scheme in successive stages. We assume that a driver has expressed her destination to the phone app, and hence, the route on which she is driving is known from a Google/Bing/Mapquest map.

4.1.1 Using Speed Limits

Knowing the route a car is driving on, the phone inside the car can infer the road it is currently moving on. This is possible by matching the real-time trace of the compass data with the expected shape of the route on a map. Figure 5 shows a simple example of the compass trace and how it matches to the route on a map. Inferring the road segment on which the car is located at a given time, it is now possible to learn the speed limit of that road segment. As a first step, we use the speed limit as a naive estimate of the car's speed. Figure 6(a) compares the actual speed of the car (from continuous GPS samples) against the speed-limit based estimate for one of the driving trips. The estimate is obviously poor, and results in a correlation value of 0.092. Figure 6(b) plots the distribution of the "speed difference", defined as the difference in the actual speed minus the estimated speed, computed every second. The left tail of the distribution curve is quite heavy. As the speed estimates improve, the curve should move closer to the $X = 0$ vertical line.

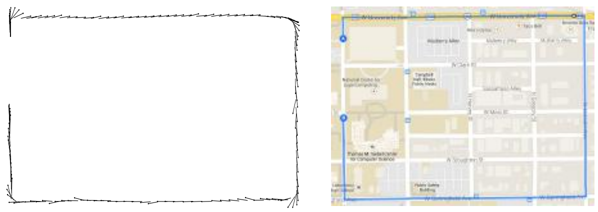

Figure 5: Compass data match well with road traces from maps.

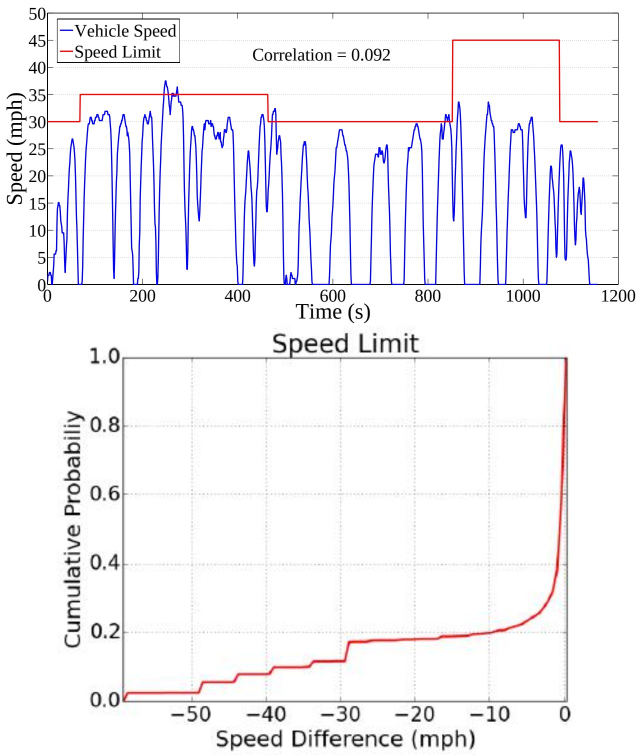

Figure 6: (a) Crude speed estimation based on speed limit alone. (b) CDF of speed difference over all trips and across all phone positions.

Unfortunately, accurate speed limits for all roads may not be easily available. Some APIs exist – openMAPs, wikispeedia – that have collected this information and made it public. However, the coverage is limited in the US and far less in other parts of the world. Where precise information is absent, the APIs offer some hints on the speed limits based on the category of the roads (e.g., urban, suburban, rural). The result is only a crude speed estimate and needs significant improvement.

4.1.2 Using Stops

A car is likely to stop several times during its drive – perhaps at stop signs, traffic lights, traffic jams, at gas stations, or for other reasons. A smartphone should be able to classify such stops from the accelerometer data, and reduce the estimated speed to zero for these durations. Of course, cars don't stop instantaneously – the speed gradually falls to zero. We mimic this behavior by empirically observing the rate of decelerations from extensively measured data. Figure 7(a) shows an example of how the car speed is dropped to zero with a slope that is a function of the actual speed of the car (in this case the speed limit of the road). The slope captures the phenomenon that decelerations are higher when the car is at a higher speed compared to the case where the car is moving slower.

Figure 7(b) shows the CDF of "speed difference" – a dramatic improvement from the prior estimate. The correlation value increases to 0.818. Of course, the left tail is still quite long and at some instances, the actual speed is greater than the estimated speed, resulting in part of the curve moving to the right of the $X = 0$ vertical line. Improvements are still necessary.

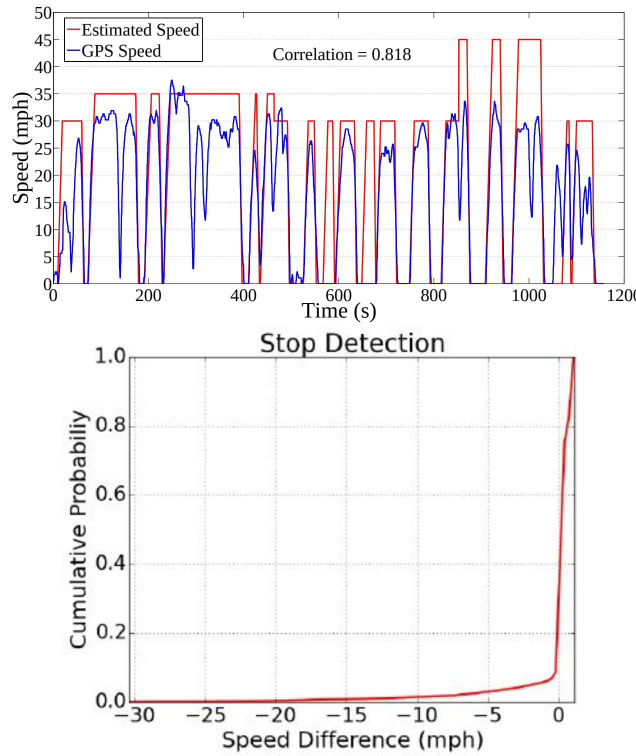

Figure 7: (a) Car speeds brought to zero whenever the accelerometer based classifier detects that the car has stopped. (b) CDF of speed difference when classifying the stop/motion state of the car.

4.1.3 Using Turns

When taking a turn at intersections, a vehicle either stops (as detected above) or slows down (say when there is a green light). The smartphone's compass and gyroscope can together detect turns accurately,[1] since turning angles at intersections are almost always large ($> 60°$). Upon detecting such turns, we also bring down the speed estimate of the car to a value V_{turn}, empirically selected as a fraction of the car's recent speed. Figure 8(a) illustrates the effect – observe how the estimated speed is reduced at around time instants of 40, 280, 430, 825, 1100. This slow down is also not instantaneous, but incorporates the deceleration of the car using the same slope functions we used for stops. Figure 8(b) plots the CDF of the speed difference – the improvement is not visible to the naked eye, but the correlation improves to 0.844.

[1]The compass alone might be adequate to detect such turns, except in cases where the phone is placed too close to the car's engine, which can cause substantial magnetic interferences. To cope with such situations, we utilize the gyroscope as well.

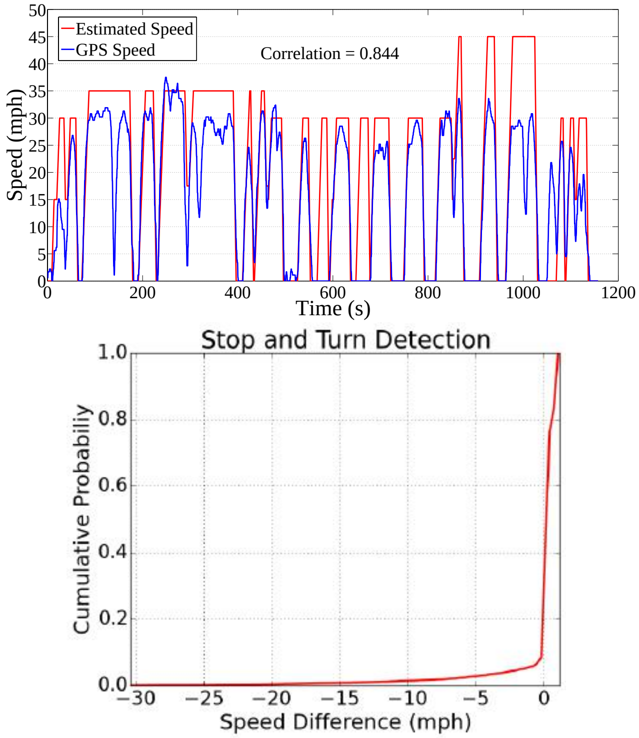

Figure 8: (a) Car speeds estimates reduced whenever compass and gyroscope detect a turn at road intersections. (b) CDF of speed difference.

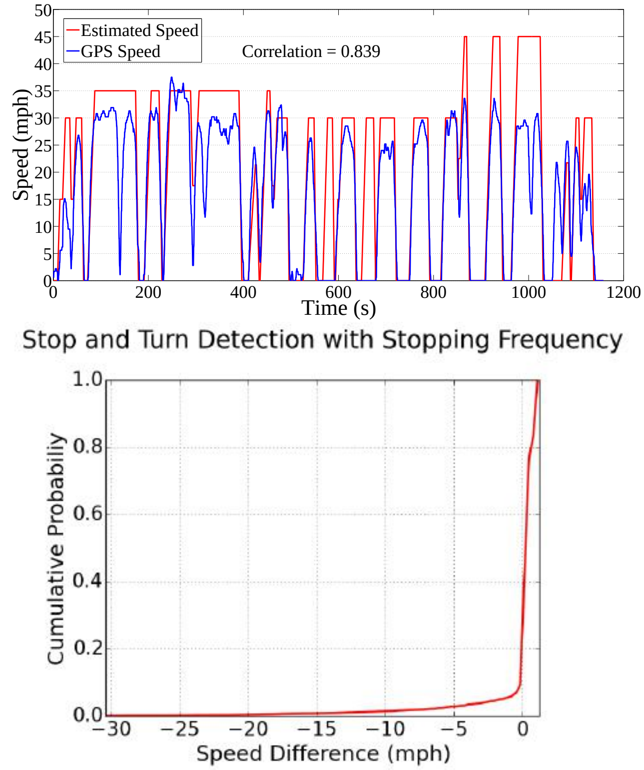

Figure 9: (a) Estimated speed is set at a fraction of the road speed limit when the car stops frequently. (b) CDF of speed difference with stopping frequency.

4.1.4 Using Stopping Frequency

Between stops and turns, the estimated speed is increased back to the speed limit. However, during road congestions for example, a frequently stopping car is not likely to ramp up to the speed limit. We accommodate such scenarios by observing the duration between consecutive stop events. If this duration is small, the estimated speed is assigned to a proportionally small fraction of the speed limit. Figure 9(a) shows a case right after time=400 where this optimization is applicable – unlike Figure 8(a), observe that the speed is not increased to the limit. However, some errors are introduced due to this in scenarios where multiple stop signs are in proximity. In such cases, drivers drive fast between stop signs partly defeating our heuristic. Figure 9(b) shows the CDF of speed difference, with marginal improvement. However, this is also because hardly any trip experienced appreciable traffic congestion.

4.1.5 Using Historical and/or Crowd-sourced Data

At this point, the main problem with speed estimates arises from the difference between actual driving speeds and the speed limits. This is a result of (1) the speed limit data being inaccurate, and (2) cars often not driving at the speed limit, either due to traffic or over-speeding. To mitigate this, we leverage both real-time crowd-sourced data and historical speed data. The former is possible through Google map APIs that offer current/recent speeds for a given road segment. The latter is possible if an app stores the user's historical speed data on a given road, perhaps on the route between a user's home and office [11]. On this typical route, its possible to turn on the GPS for a few times during bootstrap, and utilize this data (i.e., say median speed) as a substitute for the speed limit.

Figure 10(a) shows narrowing down of the gap between the estimated and actual speeds upon applying this optimization. We used the median speed per road segment, measured from 6 traces between a user's home and office. The correlation now jumps to 0.946, appreciably greater than the previous graphs. The estimated speed is now closely mimicking the car's actual GPS speed. Figure 10(b) shows that this accuracy is preserved across all the measurement data – observe that the left tail of the distribution is much shorter. Based on this, we believe it is fair to infer that smartphones, without relying on GPS, can estimate the car's speed in an overwhelming majority of situations. The core opportunities arrive from the ability to opportunistically utilize motion sensors, compass, and (historical) information from the Internet.

5. POINTS OF DISCUSSION

We discuss a few observations relevant in the context of this study.

Phone usage during a drive

The data used in this paper was gathered from normal users who did not use the phone excessively during the driving trips. They occasionally checked emails and moved their phones around, mimicking their own normal behavior. In reality, some drivers may use the phone more frequently, especially if he/she is a co-passenger. Simple techniques can detect such behavior, and the data from these time-windows need to be disregarded for speed estimation. OBDs do not suffer from this issue – they can estimate speed continuously regardless of the driver's phone usage. However, by virtue of this ability, OBDs are also not able to characterize a user's distraction due to SMS, text, phone-calls and other usage patterns. If phone usage is dominant, smart-

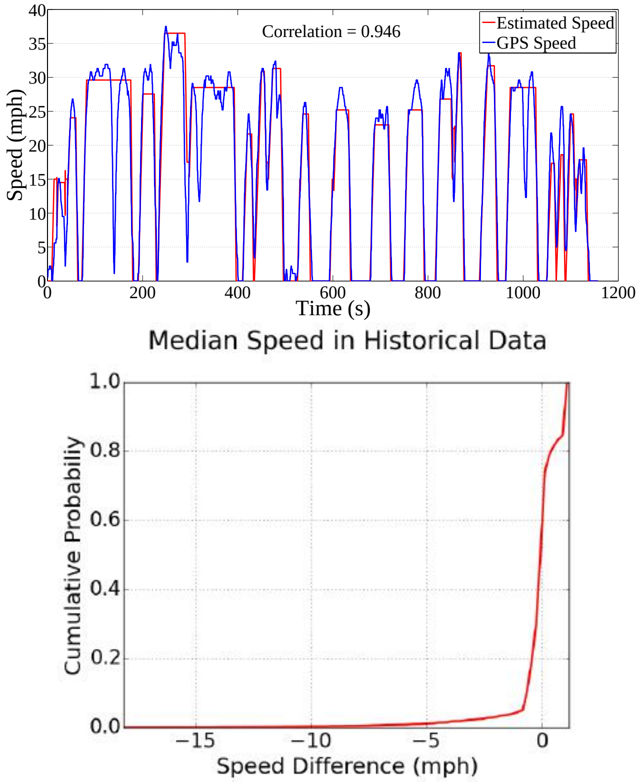

Figure 10: (a) Refining speed estimates based on historical data between a user's home and office. (b) CDF of speed difference.

phones are in the best position to include them in the driver's risk assessment. If phone usage is infrequent, smartphones can estimate car speed for most of the driving duration.

Upper bounds in speed estimation

Needless to say, the techniques proposed in this paper are simple and cannot be claimed to be the best speed estimation possible from smartphone data. Its entirely possible that greater sophistication will extract more accurate estimates. Additional sensors such as barometers and/or doppler shifts in cellular signals could add to the accuracy. The quest for the upper bound in speed estimation from motion sensors is an open question that warrants long term research investigation.

Energy consumption

The use of accelerometers and compasses significantly lowers the energy footprint in contrast to GPS. Gyroscope is more expensive than accelerometers, but was used only to add reliability to the compass – it use can be minimized or even eliminated if needed, without compromising on the quality of estimation. Moreover, if greater accuracy is necessary for speed estimation, GPS can be periodically turned on at strategic times. The tradeoff between accuracy and energy is a continuum, and various opportunities can be employed at increasing costs. We believe our accelerometer-plus-compass based approach operates at a sweet spot in this tradeoff.

6. CONCLUSION

With the personal transportation ecosystem poised for disruption, its unclear if OBDs or smartphones will serve as the platform for technological innovation. This paper is an attempt to compare and contrast their capabilities and how they translate to the end goal of understanding driving analytics. We observe that while smartphone sensors are noisy on one hand, the variety of sensors available to them, in conjunction with their access to the Internet, makes it an extremely agile and powerful platform. Even though no single smartphone sensor can directly estimate speed like the OBD, they can together orchestrate information from various sources to achieve a 96% similarity to OBD. This orchestration involves understanding when a car stops and turns from accelerometers and compasses, utilizing crowd-sourced data from the roads, leveraging the driver's historical patterns, etc. Further research in sensor fusion and machine learning will enable deeper insights into driving behavior, including aggressive braking and acceleration, accidents, dozing off, etc. [14, 8]. Finally, it is possible to execute such data-driven analytics on a low energy budget, implying that a smartphone based system is amenable to public adoption. In light of these advantages, and more that are emerging in the field of mobile sensing [3, 7, 12, 15], the smartphone platform is likely to become a superset of OBDs in a vast majority of driving scenarios.

7. REFERENCES

[1] How big data is transforming the auto insurance industry. http://www.sas.com/resources/whitepaper/wp_56343.pdf.

[2] Insurance telematics: What is it? and why we should care. http://www.verisk.com/visualize/insurance-telematics-what-is-it-and-why-we-should-care.html.

[3] H. Chu, V. Raman, J. Shen, A. Kansal, V. Bahl, and R. R. Choudhury. I am a smartphone and i know my user is driving. In *COMSNETS*, 2014.

[4] I. Constandache, R. R. Choudhury, and I. Rhee. Towards mobile phone localization without war-driving. In *IEEE Infocom*, 2010.

[5] I. Constandache, S. Gaonkar, M. Sayler, R. R. Choudhury, and L. Cox. EnLoc: Energy-efficient localization for mobile phones. In *IEEE INFOCOM Mini Conference*, 2009.

[6] J. Eriksson, L. Girod, B. Hull, R. Newton, S. Madden, and H. Balakrishnan. The pothole patrol: Using a mobile sensor network for road surface monitoring. In *ACM MobiSys*, 2008.

[7] H. Haofu, Y. Jiadi, Z. Hongzi, C. Yingying, Y. Jie, Z. Yanmin, X. Guangtao, and L. Minglu. Senspeed: Sensing driving conditions to estimate vehicle speed in urban environments. In *IEEE Infocom*, 2014.

[8] D. Johnson and M. Trivedi. Driving style recognition using a smartphone as a sensor platform. In *IEEE Intelligent Transportation Systems (ITSC)*, 2011.

[9] M. B. Kjærgaard, J. Langdal, T. Godsk, and T. Toftkjær. Entracked: energy-efficient robust position tracking for mobile devices. In *ACM MobiSys*, 2009.

[10] S.-W. Lee and K. Mase. Activity and location recognition using wearable sensors. *IEEE Pervasive Computing*, 1:24–32, July 2002.

[11] L. Liao, D. Fox, and H. Kautz. Learning and inferring transportation routines. In *Proceedings of the 19th national conference on Artifical intelligence*, AAAI'04, pages 348–353. AAAI Press, 2004.

[12] Y. Wang, J. Yang, H. Liu, Y. Chen, M. Gruteser, and R. P. Martin. Sensing vehicle dynamics for determining driver phone use. In *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*. ACM, 2013.

[13] C. wen You, Y.-C. Chen, J.-R. Chiang, P. Huangb, H. hua Chua, and S.-Y. Lau. Sensor-enhanced mobility prediction for energy-efficient localization. In *Third Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks*, 2006.

[14] J. White, C. Thompson, and H. Turner. Wreckwatch: Automatic traffic accident detection and notification with smartphones. *Mobile Networks Application*, 2011.

[15] J. Yang, S. Sidhom, G. Chandrasekaran, and T. Vu. Detecting driver phone use leveraging car speakers. In *Proceedings of the 17th annual international conference on Mobile computing and networking*, pages 97–108. ACM, 2011.