Muthanna Journal of Engineering and Technology, Vol. (9), Issue (2), Year (2021)

Muthanna Journal of Engineering and Technology

Website: https://muthjet.mu.edu.iq/

Research paper, Short communication, Review, Technical paper

Watching vehicle speed using GPS by using data mining approach

Seror Manea Bahloos[a]

[a]*Al-Muthanna University, Iraq*
**Corresponding author E-mail: seror.manea@mu.edu.iq*

DOI:10.52113/3/eng/mjet/2021-09-02/09-16

Abstract

The suggested effort is an endeavor to regulate the speed of the car using computer software that allows the owner to obtain information about the driver's position, speed, and activities. To do this, the system must be able to send data in real time. The widespread accessibility of GPS-enabled instruments, as well as the enormous quantities of data collected from them, allows us to get a perfect understanding of the condition of traffic and the road network. The current study was prompted through a sample of "T-Drive GPS" trajectory data made public by Microsoft Research in 2010. The final objective was to estimate the average speeds of the road sections using the supplied trajectory data and therefore obtain a speed overview of the road network. The corrected sensor data are used by Driving Sense to detect three types of hazardous behaviors: uncontrolled speed, driving irregularly and shifting the directions. We test the efficacy of our system in real-world scenarios. Driving Sense can identify the convert of directions through driving and anomalous speed control with 93.95 percent accuracy and 90.54 percent recall, correspondingly, according to the findings. Furthermore, the speed estimate mistake is within an acceptable range of less than 2.1 m/s.

Keywords: *GPS, Haversine formula, Longitude, Latitude, Vehicle speed.*

1. Introduction

Our civilization is evolving at a breakneck pace. People nowadays are inundated with information and are always under pressure to be "on time" with some processing of that information. As a result, everything occurs or moves at a quicker rate every day. If we comparing nineteenth century and today's perspective, the contrast is startling. This is also correct for individuals attempting to fulfill their goals through motorized vehicles, which rush in all directions, creating a throng-like scene. Such modes of transmission have grown significantly in recent decades, their performance has improved at a phenomenal pace, and as a result, traffic conditions have deteriorated, particularly in large cities. These changes have a direct impact on city drivers' driving behavior, which has grown more aggressive and incident-prone, decreasing traffic safety. In a broader sense, there's a lot of interest in classifying city drivers' driving styles relied on their attitude through driving, that may abstract using different driving characteristics.

Despite this enthusiasm, the task of extracting "standard behaviors from raw data of actual human drivers has yet to be addressed, and will be a future study topic". Even though field testing with actual drivers in real-world situations were conducted, "the driver's performance in terms of driving style was determined in each test via the subjective assessment of specialists present at the test," which was confirmed by fuel consumption. Many studies highlight the necessity for an objective approach to analyze everyday driving attitude that is derived from driving style [2]. Based on questionnaire surveys, some works attempt to determine the style of driving, which is defined as "the attitude, orientation, and way of thinking for daily driving". Recent work has combined the objective rank approach with frequent learning relied on "Elman's type neural network" to gather factual driving data from human drivers and model human driving behavior, "or classify driving style using a virtual driving simulator to collect realistic driving data from human drivers and model human driving behavior" [3]. There are also in relation with traffic flow modeling, driving course rulings, and emergency driver attitude. However, few studies, particularly in urban traffic, have looked into the designing of personal driving styles of several vehicle drivers depending on several driving parameters.

The goal of such an effort is to categorize drivers depending on their riskiness through improving traffic safety, that is a significant issue throughout the globe. The availability of GPS tracking data" coupled with digitized map data provides a strong foundation for road traffic analysis that may be used for civilian planning, traffic administration, and other applications. The purpose of this study was to analyze road traffic using GPS trajectory data collected for the T-Drive initiative. T-Drive refers to a driving direction service that uses GPS trajectories from most of taxis to assist users determine the quickest route to a location at a certain leaving period [4]. The service was created using real-world taxi trajectory data collected above a three-month period. Since then, the dataset has been utilized in a variety of projects [5, 6]. This research used a part of the dataset that was made publicly available. The purpose was to employ "the T-Drive sample dataset to conduct a speed assessment of the Hyderabad road network."

At first, we preprocessing the trajectory data to eliminate outliers, inconsistencies, and unnecessary information. Following that, we use the map-matching method to link the trajectory points with road sections and compute the rate of speed in all road segment. Consequently, the researchers create a map using the existing road network. The study is performed individually every day and for various time intervals throughout the day for examining the impact of time on road segment speeds. After knowing the speed and location of the following point, the researcher applies a Kalman filter method to obtain a speed estimation for the following point in trajectory data. Then, the researchers recalculate the average road segment speeds based on the findings and compare them to the real-world data.

The accidents happen on roads represent the major reasons of mortality in a lot of nations across the globe as the use of cars has grown over the past century. Approximately 1.3 million people die each year as a result of road traffic crashes [7]. According to a research, human error is responsible for almost 90% of road accidents. Human actions including speeding, driving when inebriated, and employing a cell phone during driving are all significant contributors to driver inattention. Many investigations in different fields have shown that when a driver know he/she is watched, his or her conduct is comparatively safer. Thus, different technologies are invented to identify a driver's condition during driving in order to minimize road accidents [7].

Smartphones with sophisticated processing capabilities and many "sensors such as accelerometers, GPS, magnetometers, and cameras" have seen remarkable growth in recent years. As a consequence, a slew of smart phone augmented reality apps have been suggested, including integrating smart phones with vehicles to provide drivers with assistance. The smart phone-based method has the benefit of avoiding the large investment costs of commercial systems. However, we discovered via our extensive testing in reality that the data given by the smart phone's integrated sensors is probably poorer. The simple integration of this data can be caused from a considerable departure from the reality of the vehicle's statuses, that possesses a major effect on the suggested applications' real-world usefulness.
The current study was prompted through a sample of "T-Drive GPS" trajectory data made public by Microsoft Research in 2010 [1]

2. Data collection and preparation

The road network data was taken from Open Street Maps. The map itself comes in XML format and consists of nodes and ways. Each node is defined by its ID and its geodetic location (latitude and longitude). A way is comprised of nodes and can have multiple tags attached to it (e.g name, what type of road it is, whether it is one-way). In order to speed up XML parsing, the road data was first filtered with Osmosis [8]. All irrelevant information (e.g railways, footways) was removed during filtering. In addition, the map was reduced to a smaller area in the central Hyderabad (bounded by latitude [17.387140] and longitude [78.491684]), since working with the entire map of Hyderabad is quite time consuming in terms of map-matching and road network parsing.
The T-Drive sample dataset, which was utilized for speed estimate research, has 10357 taxi trajectories totalling more than 15 million points. The data was collected in Hyderabad between September 2 and September 8, 2018. Figure 1 shows a sample of the trajectory data.
We eliminate duplicates and data with minimal sample ratios, implausible speeds, or locations that are abroad of limits from GPS trajectory data. The data was subjected to the following filters:

- Repetitive If a trajectory point's timestamp was the same as the preceding point, it was deemed a duplication.
- Long intervals between sampling the trajectory segment was deleted if the delay between two consecutive points was greater than 20 seconds. Because we would have had to utilize a shortest path method to calculate the distance between the spots for speed estimation with low-sampling rates, the objective was to work with dense data.
- Extremely fast a speed of more "than 90 km/h between two" locations was deemed implausible, and the trajectory section was deleted in such instance.
- Out of limits coordinates all locations that were outside of the road network's specified limits were deleted.

```
10,2018-09-02 19:09:32, 17.387140, 82603
10,2018-09-02 19:09:32, 17.387140, 82603
10,2018-09-02 19:09:32, 78.491684, 82603
10,2018-09-02 19:09:32, 78.491684, 82603
```

Fig. 1: Taxi id, timestamp, longitude, and latitude are examples of trajectory data in this format.

3. Map-matching and speed estimation

The sensor data inaccuracy, as we've seen, is caused not just by white noise whereas via a bias too. As a result, "the Kalman filter algorithm cannot be" directly employed. We discover that the data inaccuracy varies from time to time, even while the computer is idle. This implies that whenever we utilize it, we must re measure the data inaccuracy. Authors suggested sensing realistic driving circumstances to establish reference points for measuring speeding mistake and removing collected from biased acceleration in such technique, however, cannot be utilized in a freeway situation with fewer reference points. We present Driving Sense in this article, which may effectively reduce accumulating mistake while cars are driving, allowing for more accurate identification of hazardous driving behaviours.

A version of the "point-to-curve map-matching method" clarified in [9] was used to link trajectory points to road sections. The algorithm's general concept is as follows:

- Find the node in the road network that is closest to the trajectory point by running a query across all of the nodes. A kd-tree [10] data construction was employed to fast the process of searching since the number of trajectory points required to be linked with roadways was in the millions.
- Locate all road segments that intersect with that node.
- Measure the distances between any road section and select the one that is closest to you.

The Haversine formula was used to determine the distance between two GPS locations.

3.1. Haversine formula

$$a = \sin^2(\Delta\varphi/2) + \cos\varphi 1 \cdot \cos\varphi 2 \cdot \sin^2(\Delta\lambda/2) \quad (1)$$
$$c = 2 \cdot \mathrm{atan2}(\sqrt{a}, \sqrt{(1-a)})$$
$$d = R \cdot c$$

Where:
R: denotes the earth's radius
Φ: latitude
λ: longitude

The Haversine formula (equation 1) was mostly utilized during the speed estimation phase, although a basic Euclidean distance was used in other cases (e.g., the nearest node query). The cross-track distance formula was used to determine the distance between a trajectory point and a road segment specified by two nodes in the road network (equation 2):

$$d = a\sin(\sin(\delta_{13}) \cdot \sin(\theta_{13}-\theta_{12}) \cdot R \quad (2)$$

Where:
δ_{13}: represents the distance between the starting point and the 3rd point.
θ^{13}: refers to the starting bearing from the starting point to the 3rd point.
θ_{12}: points out to the starting bearing from the start point to the ending point.

In order to calculate the speed, two consecutive trajectory points have to be connected with the same road segment in the case of speed estimate. The average of whole speeds on the same road was therefore calculated. The procedure was irritated for the total trajectories, and then a total average for each road segment was determined. Java JDK 1.8 was used to do the road speed study repository at [11] The data were displayed using the Java minigeo package once the average speeds for the road sections were calculated.

3.2. GPS

The GPS system consists of a constellation of satellites that receive "signals from the GPS transmitter and transfer data to the receiver". This allows you to monitor the precise position, speed, and events such as door opening and closing, as well as the fuel level of a car equipped with a GPS tracking device. The GPS tracking instrument [transmitter] is installed within the car, where it is difficult to detect by a vehicle thief and therefore difficult to disable. The transmitter transmits signals to the monitoring station on a continual basis. The data/signal reception is handled by the GPS server [12] which stores the data securely and provides it when requested. Finally, when the user transmits data from the mobile, the GPS interface or control system delivers the speed limit data to the speedometer. There are two kinds of GPS:

3.2.1. Passive GPS

It has a receiver that merely "listen" to satellite signals and "records" (saves) them in digital form. You attach the GPS tracker to your PC to check where the vehicle [9] has been driven. Because they are battery operated and have no cables attached to the vehicle, they may be transported from one vehicle to another. A GPS signal is captured once per second (60 times every minute) by the reporter.

3.2.2. Active GPS

It also includes receivers that operate in the same way as passive GPS receivers. An inbuilt cellular device is included in the active units, which makes a phone call every few minutes [10]. The phone call sends satellite readings and other data from the GPS tracker to a website where you can log in and monitor the car in real period to decrease speed.

3.3. Receiving GPS signals

At all times, the satellite broadcast pattern delivers at least four signals to all places on Earth. Signals, however, cannot travel through solid things. When you switch on GPS it requires at least three satellite signals to determine your position. If they don't get strong enough signals, most GPS units will provide a warning. To use a GPS inside the car, you may require an extra antenna.

3.4. Pinpointing your location

Longitude, latitude, and altitude are used by GPS to show your position. To locate your position, use the given co-ordinates on a map. The more precisely you can pinpoint your position, the better. A clock, that utilizes satellite signals to define period in your position, may be included with GPS [13]. GPS gadgets may show you your current speed depending on how far you've travelled in a particular amount of time. A gadget may use GPS to locate itself.

4. Principle of process

The Global Positioning System (GPS) is a satellite-based navigation system. It sends data to the receiver via a digital signal at about 1.5GHz from each satellite. The receiver can then determine the satellite's precise range as well as the satellite's geographic location (GP). The GP [14] is the Earth's position immediately underneath the satellite. A line of position (LOP) ON THE EARTH is established as a result of this.

4.1. Data error analysis

As previously stated, we use the smartphone as a sensing platform in this work to gather vehicle driving data and detect hazardous driving behaviours. However, the data gathered by smartphone sensors is found to be noisy. In this part, we'll look at the effect of sensor data inaccuracy on vehicle driving behaviour estimate.

We begin by conducting an experiment to determine the nature of the sensor data mistake. For doing so, "we place a smartphone on a horizontal plane and hold it still while collecting sensor data". The sample rate has been set to 1 Hz. Each sensor data reading should have a value of zero in an ideal situation. We plot the measurement data using the -acceleration and -gyroscope readings as an example. Measuring data is erratic and differs from the reality. Uncorrected bias errors and white noise are the main sources of error in smartphone sensors, according to our practice test. The data inaccuracy of the accelerometer sensor is higher than that of the gyroscope sensor. As a result, we are primarily concerned with determining the influence of acceleration mistake on the estimation of vehicle driving speed.

Let S: $(T_1, T_2, T_3 \ldots\ldots T_N)$ be a series of gathered data, $T_i=(a_i, P_i)a_i$, refers to the acceleration, and represents the vehicle position. Therefore, the crossing distance through the period span can be expressed as:

$$P = \sum_{k=1}^{N} \nabla P_i = \sum_{k=1}^{N} \left(V_{k-1}\Delta t + \frac{a_k \Delta t^2}{2} \right).$$

With the initial velocity V_0 at the beginning of data collection, the travelling distance P can be computed as

$$P = \sum_{k=1}^{N} \left[\left(V_0 + \sum_{i=1}^{k-1} a_i \Delta t \right) \Delta t + \frac{a_k \Delta t^2}{2} \right] + NV_0 \Delta t + \sum_{k=2}^{N} \sum_{i=1}^{k-1} a_i \Delta t^2 + \sum_{k=1}^{N} \frac{a_k \Delta t^2}{2}.$$

Then, we have the velocity estimation function of V_0 as

$$V_0 = \frac{P - \sum_{k=2}^{N} \sum_{i=1}^{k-1} a_i \Delta t^2 - \sum_{k=1}^{N} \left(a_k \Delta t^2 / 2 \right)}{N \Delta t} = \frac{P}{N\Delta t} - \sum_{k=2}^{N} \sum_{i=1}^{k-1} \frac{a_i \Delta t}{N} - \sum_{k=1}^{N} \frac{a_k \Delta t}{2N}.$$

Thus, the vehicle speed at the time point $\Delta t \cdot N$ can be estimated as

$$V_t = V_0 + \sum_{k=1}^{N} a_k \Delta t = \frac{P}{N\Delta t} - \sum_{k=2}^{N} \sum_{i=1}^{k-1} \frac{a_i \Delta t}{N} - \sum_{k=1}^{N} \frac{a_k \Delta t}{2N} + \sum_{k=1}^{N} a_k \Delta t$$

$$= \frac{P}{N\Delta t} - \sum_{k=2}^{N} \sum_{i=1}^{k-1} \frac{a_i \Delta t}{N} + \sum_{k=1}^{N} \frac{(2N-1)\, a_k \Delta t}{2N} = \frac{P}{N\Delta t} + \sum_{k=1}^{N} \frac{2k-1}{2N} a_k \Delta t.$$

We can see that the velocity of a vehicle is composed of the acceleration and the travelling distance which is acquired by GPS. As we know, the GPS data is inaccurate as well. Even the GPS readings adjusted by WAAS have an inaccuracy of 3 m (standard deviation), not to mention the ones in the region without WAAS. Fortunately, examining the GPS trajectories of various cars, we notice that the GPS inaccuracy is strongly linked over a long driving distance, which is represented by the fact that the vehicle trajectory is closely paralleled with the actual roadways. That is to say, given a sequence of GPS trajectories, they have the same data bias. It is important to mention that we are not the first ones to make such findings; comparable features have previously been found and used by numerous works. Based on this finding, we may infer that the travelling distance calculated via the relative motion distance superposition is trustworthy. Hence, we can figure out the estimation speed mistake of the vehicle as

$$err = V_t^* - V_t = \sum_{k=1}^{N} \frac{2k-1}{2N} (a_k^* - a_k) \Delta t$$

Where:
a_k^*: refers to the ground truth value of the acceleration
We discover that while integrating the accelerometer's data, the estimate inaccuracy accumulates, and the later accelerometer's values have a larger effect on the vehicular speed estimation. Assume "the accelerometer's –axis" is parallel to the vehicle's movement. The speed estimate error for 200 samples is up to 7.48 m/s, which has a significant impact on the identification of vehicular driving behaviour. As a result, it is critical to rectify sensor data errors before utilizing them.

5. Speed estimation results

After pre-processing, there were about 3.8 million trajectory points remaining. After filtering using Osmosis, the road network utilized in speed estimate included 25391 nodes and 7876 road sections. For visualization reasons, the road speeds were divided into three groups (figure 2):

- Fluid- speed < 30 km/ hr
- Dense – speed (20-30) km/ hr
- Congested – speed ≤ 20 km/ hr

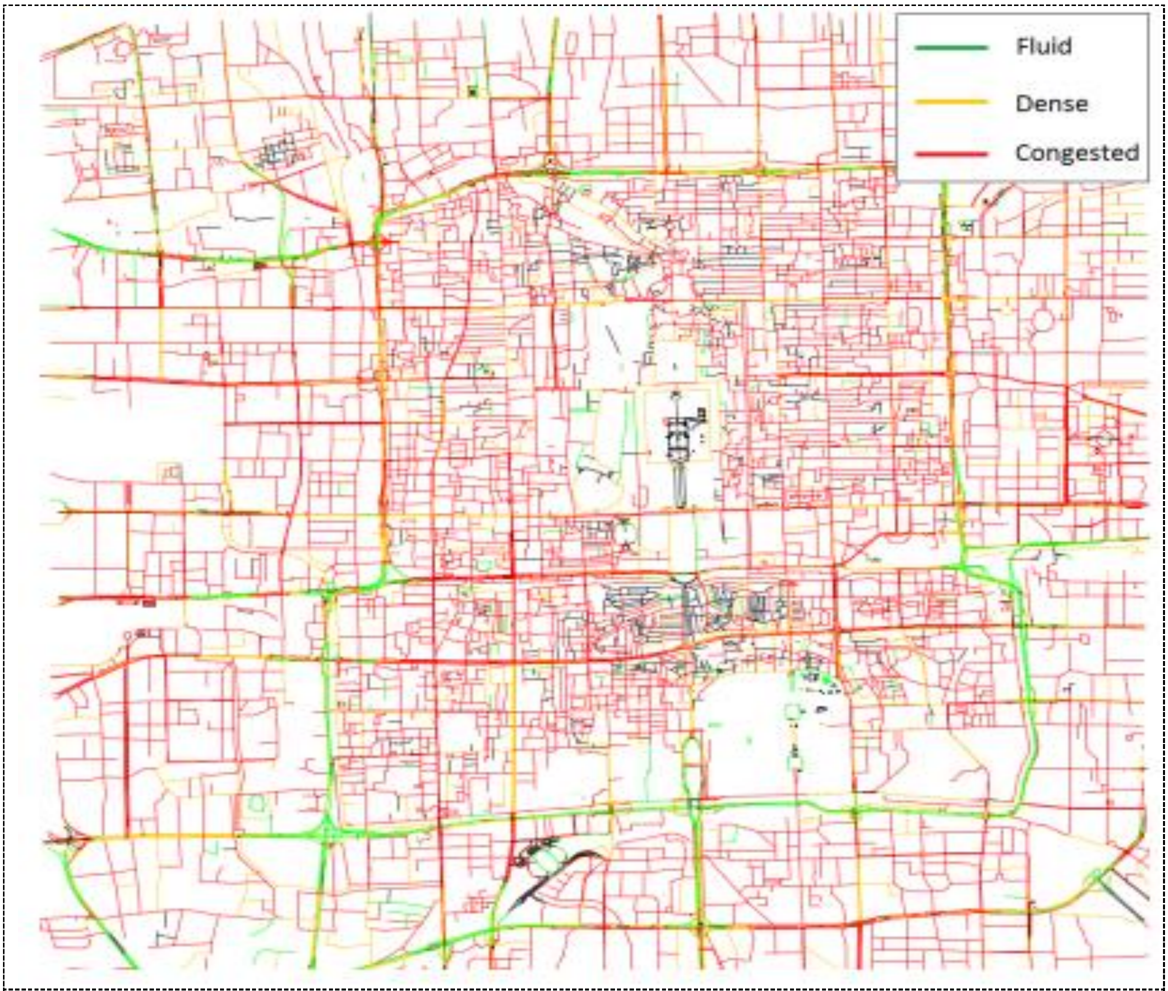

Fig. 2: displays the average road speed estimate findings a spatial representation in Hyderabad.

The average speed was identified for the whole 6402 road sections, with 669 classified "as fluid, 1171 as dense, and 4562 classified as congested the mean speed of the road sections represents 17.313 km/h, with a normaal deviation of 10.351 km/h. Figure 3 depicts the dispersion of estimated speed values.

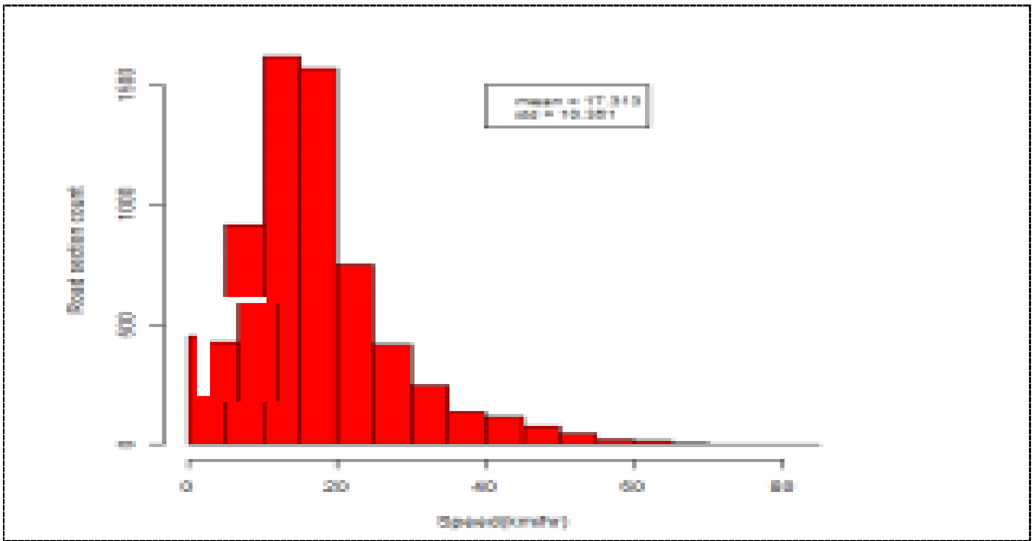

Fig. 3: The central Hyderabad road network's speed distribution in downtown Hyderabad.

The next stage is to look at the road segment speeds for each mode. We may concentrate on each day individually since the data was collected over a 7-day period. In addition, we divided the day's 24-hour chronology into six sections (see figure 4) and performed speed analysis for each component individually.

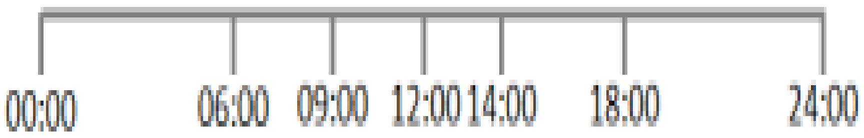

Fig. 4: The speed analysis was performed at several time intervals during the day.

5.1. Kalman filter

The Kalman filter refers to a recursive data filtering technique that takes into account all available data to provide a total best estimate with the lowest Mean Square Error (MSE). The aim is to utilize noisy data to generate more accurate estimations (hopefully). To forecast the measurements in the future state, we utilize all of the available data (e.g. position, velocity, acceleration) in the present state, as well as the system's knowledge. We utilize the actual measurements and the projected values to obtain the best estimate once we get the real data for the next state. The Kalman filter is particularly helpful for real-time data processing. The real Kalman filter formulae and their derivations is the topic of several studies, including [11][12], and have been utilized to some degree in the current study. The objective is to apply the Kalman filter to taxi trajectory data and compute Kalman estimates depending on two trajectory points that are provided. The basic concept is that if we know the speed and position of the present state (i.e. a trajectory point), as well as the position of the following state, we can employ the Kalman filter algorithm to estimate the speed at the next state. The starting "speed is projected to velocities on various axes and the relevant geodetic coordinates are translated to UTM coordinates (UTM stands for "Universal Transverse Mercator) in meters in order to compute a Kalman estimate based on two consecutive trajectory points". Figure 5 depicts the road network in terms of speed after conducting 100 000 Kalman estimates (the total MSE was 1436.6 km h). The starting trajectory points from which the Kalman estimates were computed were selected at random from the filtered trajectory data. Because the visual converts seem to be small, bounding boxes have been used to emphasize some of the discrepancies between the findings with actual estimates (see figure 5).

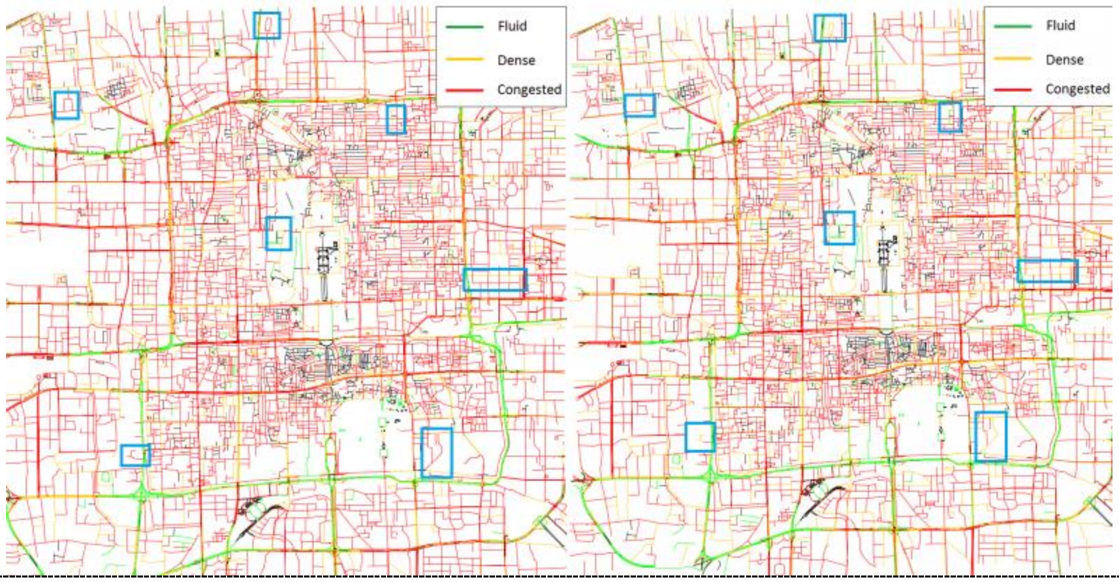

Fig. 5: Variety of estimations in road section speed with real estimates (left image) vs. with 100 000 Kalman estimations (right image).

However, MSE was calculated on a per-road-section basis to obtain a better understanding of the Kalman estimates. Figure 6 shows considerable variations, with some reaching above 4000 (km/ h) 2 in certain instances. Despite this, the MSE remained below 1000 (km/h) 2 in the majority of instances. Furthermore, the Kalman estimate often exceeded the actual estimate, implying that increasing the random number of Kalman estimates would make the entire picture seem considerably more fluid in terms of speed.

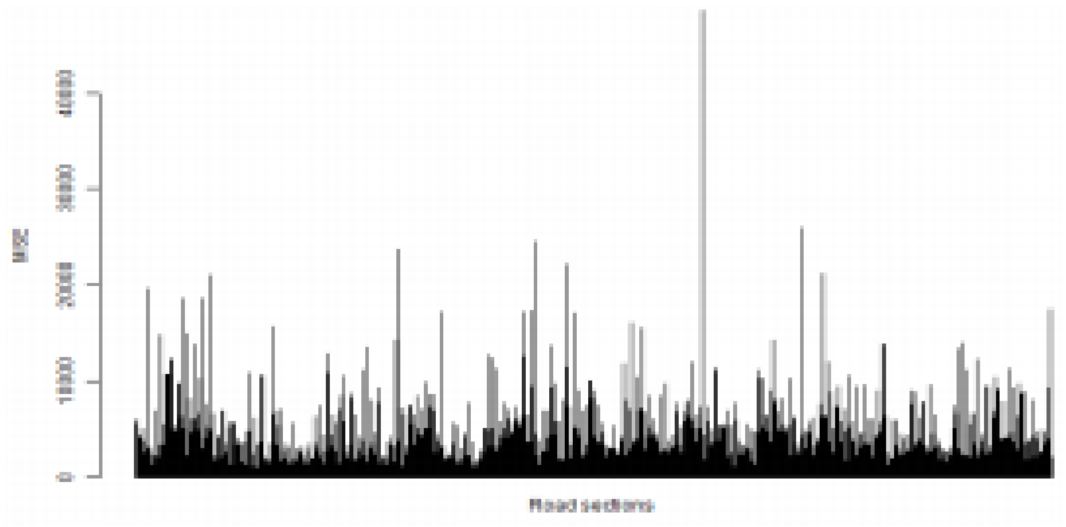

Fig. 6: MSE based on 100 000 Kalman predictions per road segment.

6. Conclusion

The findings of this study may be substantially improved by considering, at the very least, some of these variables, which are readily available (individual characteristics). Other variables, like as the state of mind of the driver, are more difficult to address, but can be done with the appropriate technological solution. Finally, certain problems, such as worry over potential traffic accidents, are difficult to evaluate objectively from an observer's perspective without interviewing or questioning the drivers directly about them. Based on real-world taxi trajectory data, the research conducts a speed analysis of the core Hyderabad road network. We begin by pre-processing the trajectory data and then utilizing a point to-curve map-matching method to associate the points with road sections. We only utilize dense data (sampling interval 20s) for estimating road segment speed. The resultant picture shows that the majority of the routes are classified as crowded. This may be harmed as a result of the present solution's failure to account for traffic lights, which may be particularly noticeable on smaller midtown streets where traffic is heavily controlled by signals. Furthermore, we examine the findings on a per-modality basis by dividing the day into six sections and examining each day individually. The findings indicate that traffic congestion is at its peak between 14:00 and 18:00 and at its lowest between 06:00 and 09:00. In the last section, we generate Kalman estimates at random at various trajectory points and compare them to the actual estimates. The overall picture does not alter substantially as a consequence of the following outcome.

Acknowledgement

I would like to express my special thanks of my parents, my husband, my friends who helped me a lot in finalizing this research paper.

References

[1] Yuan J., Zheng Y., Zhang C. Y., Xie W., Xie, X., Sun, G., Huang, Y. TDrive: Driving Directions Based on Taxi Trajectories. In Proc. ACM SIGSPATIAL GIS, 2010.
[2] Zoran Constantinescu, Driving Style Analysis Using Data Mining Techniques International Journal of Computers, Communications & Control (IJCCC) V:654-663.DOI:10.15837/ijccc.2010.5.2221. December 2010.
[3] Ganti, Raghu K., Fan Ye, and Hui Lei. "Mobile crowdsensing: current state and future challenges." IEEE Communications Magazine 49.11 (2011).
[4] Hiromitsu Hattori , Yuu Nakajima , Toru Ishida Learning From Humans: Agent Modeling With Individual Human Behaviors IEEE Transactions on Systems Man and Cybernetics - Part A Systems and Humans 41(1):1-9 . DOI:10.1109/TSMCA.2010.2055152 . February 2011
[5] Van Buer, Darrel J., et al. "Prediction of vehicle operator destinations." U.S. Patent No. 7,233,861. 19 Jun. 2007.
[6] You, Chuang-Wen, et al. "Carsafe app: Alerting drowsy and distracted drivers using dual cameras on smartphones." Proceeding of the 11th annual international conference on Mobile systems, applications, and services. ACM, 2013.
[7] https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries
[8] Lowrey, Larkin Hill, et al. "Internet-based vehicle-diagnostic system." U.S. Patent No. 6,611,740. 26 Aug. 2003.
[9] Liu, He, et al. "Visual analysis of route diversity." Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on. IEEE, 2011.
[10] Deitel, Paul J. Java how to program. Pearson Education India, 2007.
[11] Tai, Yu-chin, Cheng-wei Chan, and Jane Yung-jen Hsu. "Automatic road anomaly detection using smart mobile device." conference on technologies and applications of artificial intelligence, Hsinchu, Taiwan. 2010.
[12] Vatsavai, Ranga Raju, et al. "Spatiotemporal data mining in the era of big spatial data: algorithms and applications." Proceedings of the 1st ACM SIGSPATIAL international workshop on analytics for big geospatial data. ACM, 2012.
[13] Rakshit, Sarbajit K. "On-demand suggestion for vehicle driving." U.S. Patent No. 9,020,876. 28 Apr. 2015.
[14] Witte, T. H., and A. M. Wilson. "Accuracy of non-differential GPS for the determination of speed over ground." Journal of biomechanics 37.12 (2004): 1891-1898.