CMSC 676: Information Retrieval

Graduate Term Paper

Ms. Kiran Santosh Jambhale

Driver Behavior Analysis Using Automotive Data from Data Acquisition Systems

Abstract:

A driver's behavioral traits can be analyzed from the way he drives his vehicle. There have been studies that show more women prefer an automatic geared vehicle over a geared vehicle. Also, men tend to change the gears more frequently as compared to women. Also, the number of gear changes sheds light upon a few other behavioral traits. This can also be linked to the mindset of the driver at that very point. Data acquisition systems are always fitted in vehicles to gather the driving data. This data can be put to good use to understand the behavioral traits. It has been found out that after having a tiff with their significant others, many people did not focus on their driving or the road. On the other hand, if people knew about this study, they were over careful while driving after a tiff with their significant others. Therefore, data acquisition systems can be used to retrieve substantial data about the drivers which can be put to good use after analysis such as improved safety of the drivers and the passengers. This also helps in improved automobile designs for more comfort, which could help in the health of the drivers in the long run.

1. Introduction:

Driver behavior analysis has its applications in a wide spectrum of fields like automatic vehicles, auto-gear shifting mechanisms, psychological studies, understanding the nature and reason of road accidents, etc. It yields essential information about not just the driver, but also the vehicle. It is known to have benefitted researchers to study the structural dynamics and fatigue of the vehicles. These studies have also known to be useful while curating new traffic laws.

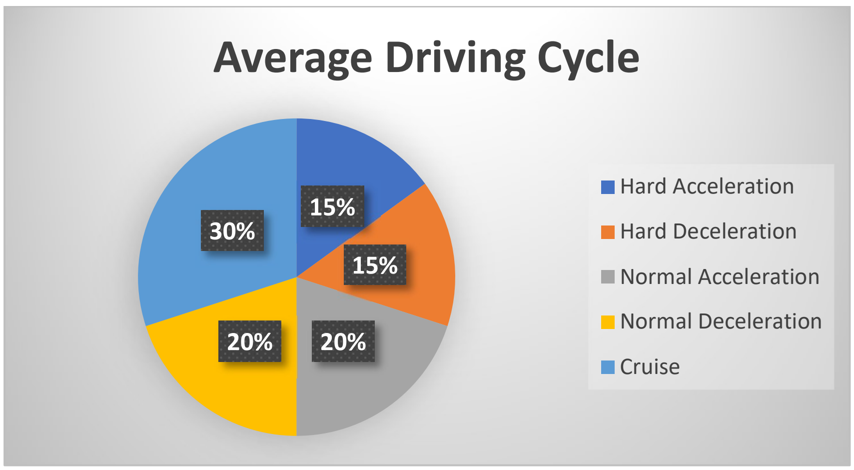

Figure 1: Average Driving Cycle of Individual Drivers

The data for all these studies is obtained from data acquisition systems. Data acquisition systems are installed in various places in the car like the torque wheel, accelerator pedal, brake pedal, clutch, and gearbox (if any). The data is usually stored as a log file connected to these sensors. A few studies use a virtual or smart driving simulator to obtain experiential driving data from drivers and to understand driving tendencies and further classify driving trends by combining evaluation with machine learning techniques based on data generated by smartphone accelerometer sensor. Many studies have suggested various models for virtual traffic flow and decision making by drivers in normal and unusual circumstances such as natural disasters or emergencies. In figure 1, it shows the driving cycle of an individual driver that was captured by the data acquisition systems. It tells how long each of the acceleration and deceleration tasks were performed. These studies have also helped in identifying driver distractions in various scenarios. The data is then processed using information retrieval techniques right from tokenizing and cleaning up to clustering. Machine Learning and neural networks have been known to predict the users' reaction in certain situations and help avoid fatalities. A lot of automatic vehicles, auto-braking systems are designed using these strategies. A corpus of data is studied to identify multimodal features that can be used to distinguish between normal and unusual driving conditions. Furthermore, binary classifiers are trained to differentiate between normal and each of the secondary (unusual) conditions.

2. Data Procurement and Cleaning:

A network of multiple sensors, communication interfaces, cameras, and data acquisition systems are maintained to obtain the automotive data to be analyzed. The data, hence, obtained is stored in a database which could be a log file, a NoSQL database, a SQL database or even a simple excel file. The data obtained contains a lot of noise and unnecessary content which needs to be filtered out. The following sections talk about the data procurement process and the cleaning and filtering process in greater details.

2.1. Data Collection:

For collecting data, noninvasive sensors are installed at various places in a vehicle like the torque wheel, accelerator pedal, braking pedal, clutch, and gearbox (if any). At times, there are also cameras fixed at the front, rear, and sides of the vehicle. These sensors are then corelated using the time stamps to map the events. For instance, if at a moment the torque wheel was used, the time stamps for the same time are checked for the other sensors too to know what other tasks were performed. Also, cameras are monitored to check the surroundings to understand why the specific activity was performed. This data collected is mostly stored in the form of log files and/ or json files. Figure 2 shows the type of data the log files contain.

Quantitative factors	Qualitative factors	
	Driving-related	Driver-related
• Speed - Acceleration • Braking • Orientation • Position • Time range • Mileage • Road type	• Distraction • Attention • Law violation • Maneuvers prediction	• Sensation seeking • Impulsivity • Anger/Vengeance • Narcissism

Figure 2: Sample data collected in log files [4]

This data could contain a lot of noise and hence, preprocessing becomes imperative. This noise could be from a sensor recording movements due to some involuntary actions of the body. Different experiments require different type of sensors. If the aim of the experiment is to study the fatigue or structural dynamics, the sensors are more sensitive to the surroundings. If the recordings required are not too detailed, the sensors could be rustic.

2.2. Data Acquisition Systems and Sensors:

Data Acquisition is the process of obtaining data from the real-world scenarios using actuators and sensors and then, presenting the data in human readable formats for further analysis and study.

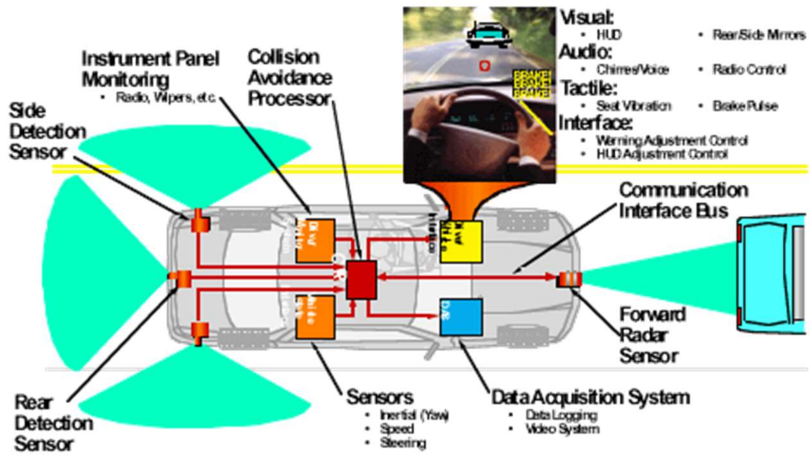

Figure 3: Data Acquisition Sensors in a typical vehicle [9]

The above image (figure 2) shows a diagrammatic representation of the sensors placed in a vehicle. Not all these sensors are always required or used. These sensors collect all the tasks performed by the driver during the journey. The sensors are connected to a data acquisition system that stores the data logged by the sensors in real time. We could call this the black box of the vehicle. The sensors are connected to the data acquisition system using a common communication interface bus or channel. This makes it possible to map the activities of multiple sensors at the time of a particular event.

2.3. Log Files:

The data collected by these sensors, cameras and data acquisition systems is stored in the form of log files. These log files could be json files, excel files, SQL databases, or NoSQL databases. Most commonly, json files and NoSQL databases are seen to be used for storing automotive data, as there are easily scalable and have no fixed format for every entry. This makes it easier for every distinct driving event to be recorded.

2.4. Pre-processing and Cleaning of Data:

The data in these log files consists of a lot of noise. For instance, at times, the brake is simple used while igniting the engine or the clutch is used to brake the vehicle, but there was no requirement for braking, so the clutch was released again. This data can be considered as noise. This data needs to be cleaned so that we have only directly relevant data for analysis. The following figure 3 shows how the workflow for the data pre-processing and cleaning takes place.

Data pre-processing involves a lot of information retrieval strategies. The data obtained is first cleaned to get rid of unwanted content. If the data collected is in a uniform S.I. unit, then the units succeeding all the numeric values can also be removed. This will decrease the volume of the dataset considerably. Later all the terms in these log files are down cased and tokenized. A few log files also need stemming and lemmatization.

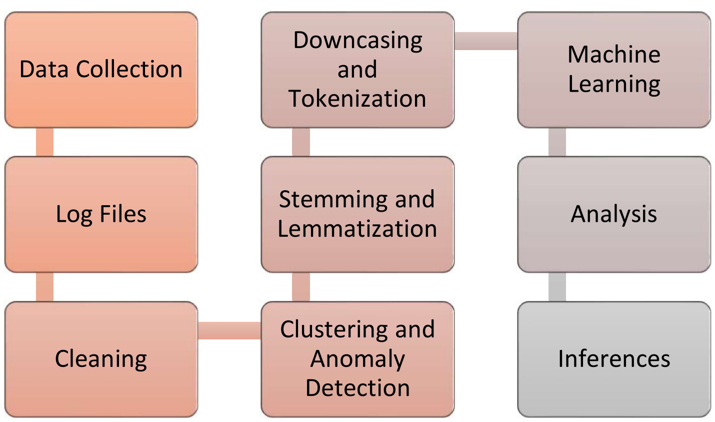

Figure 4: Workflow of Data Pre-Processing

Once the data is tokenized, it can be formed into clusters based on a common characteristic that needs to be studied. This can be done using the term weights that have been generated. For instance, if we need to study acceleration and deceleration trends, we can form clusters for a specific speed interval. This will make it easier to study each cluster and form inferences. This will also help in analysis the driver behavior at every speed interval. Going ahead, the clustering and speed progression will also help in understanding the progression of the drivers' behavioral traits.

3. Literature Survey:

A lot of research has been done in this field for several years. It ranges from studying behavioral traits, studying driving traits, studying driver distractions, etc. This has helped in designing more robust, comfortable, and safe driving experiences and to design laws for traffic.

In a paper by Dilum et al[3], the researchers have studied and analyzed driving tendencies and detected anomalies on that basis. To detect the anomalies, the researchers have clustered the data acquired from data acquisition systems using k-means clustering algorithm. As k-means is an unsupervised learning algorithm, it becomes easy to learn the data. The record whose values keep changing over time are used by this algorithm. If the data records show a regular pattern, it can be said that the data is normal. After the final clusters are formed, if the number of data points in a cluster are unusual or sparse, the cluster is known to show anomaly. Then, the minimum and maximum clusters are calculated. On this basis, a range of values where anomaly is detected is identified. Therefore, if there is a new data value introduced, it will help understand if its anomalous or not.

The paper also talks about Markov models for studying acceleration, engine RPM and speed. Studying the anomalies of a driving pattern of a driver is not based on a single parameter. The authors have used a sliding window approach for an interval of 20 seconds. If a newly introduced event is identified as an anomaly, it is directed to the sliding window that maintains events coming from engine RPM, speed, and acceleration. A certain driving behavior is classified as anomalous, if the anomaly index is more than a set threshold value. A small margin or tolerance is added to reduce the false positive results.

The paper also talks about the usage of Adaboost algorithm. It is a supervised machine learning algorithm which builds an efficient classifier using an ensemble of results of a set of weak classifiers in a linear combination. This model has been used to classify the incoming data values as safe or unsafe driver behavior based on an alpha value.

Turkson et. al [5] studied the driver behavior and the corresponding acceleration intention. The researchers have built a support vector machine classifier on top of the data acquisition system for multiple types of roads. After road tests were performed, fuzzy logic was used to identify driving traits using the mean and standard deviation. The statistics were used as the input to the fuzzy logic controller, that yielded an output which talks about the driving traits. As this study focused on acceleration intention, it yields good insight into fuel economy and fatigue of the vehicle on different types of roads.

4. Compare and Contrast Relevant Work:

The study of behavioral traits of drivers revolves around procuring data using data acquisition systems, information retrieval, and machine learning. Almost all works procure the data using multiple sensors and data acquisition systems. However, how the data is stored varies from application to application. It is observed that most of these log files or recorded data needs to be cleaned and pre-processed to make it suitable for analysis and computations. The choice of tokenizing, stemming, and lemmatizing is varied as per the use case of the experiment and the required input for the machine learning model built.

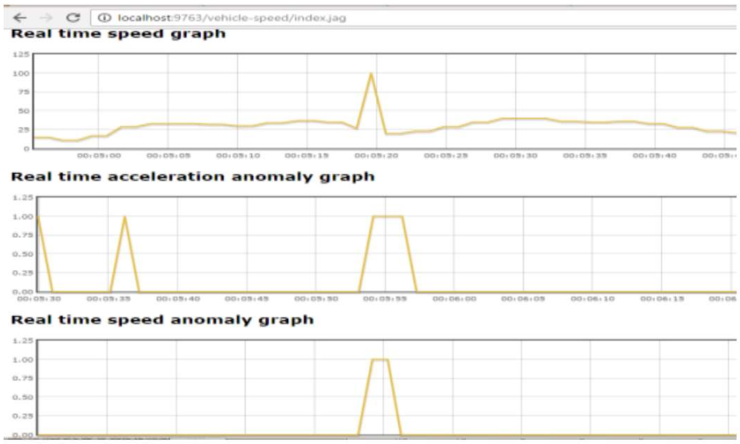

Figure 5:Sample anomaly graphs [3]

The first step after tokenizing and/ or stemming-lemmatization is to cluster the data based on a specific characteristic which could be acceleration/ deceleration, torque speed, parabolic paths, etc. This clustering could be performed using any unsupervised learning algorithm such as k-means. This makes it easy for categorization of data and detecting anomalies. Figure 5 shows a sample of anomaly graphs. These anomalies will help us to determine the exceptional behavioral traits. Besides, as each cluster shows a similarity amongst the data points in consists of, it becomes much convenient to find traits or characteristics specific to those data points. Hence, every driving trend can yield a set of its own characteristics.

Later comes the building of the machine learning model. There are multiple models that could be used such as Naïve Bayes, SVM Classifier, XG-Boost, Adaboost, Vector Space Model, Neural Networks, etc. The classifiers could be used to classify the driving trends as safe or unsafe. This even though a macro level classification, could help in learning if a certain driving task should be considered as negative or positive. With the help of the SVM Classifier, tasks such as sudden braking, smooth acceleration, speed maintenance, general acceleration, and urgent acceleration could be studied. This can also help in understanding the driver distractions. The following figure 6 shows the unsafe driving conditions identified using the Adaboost algorithm for acceleration, engine load, and ratio of relative speed to relative RPM.

Parameter	Unsafe range	Unit
Acceleration	>3.0 and < -3.0	m/s^2
Engine load	<20% and >50%	-
Ratio of relative speed and relative RPM	>1.3 and <0.9	-

Figure 6: Example ranges for unsafe driving conditions [3]

Other trends such as gear changing, use of torque wheel, lane changing can be studied by building a Vector Space Model on each of the cluster that is generated and based on the output generated, a trend for each type of these tasks can be identified effectively.

5. Challenges and Future Work:

The main challenge faced is that the trends vary for different type of roads, vehicle, date of manufacture, driver, geographic location, and many other parameters. Also, one model is not sufficient to study all types of traits. The other challenge is that the sensors introduce a lot of noise and it is difficult to differentiate between the data that can be rendered useful and the data that should be discarded. Also, the anomalies detected could possibly be a potential driving trend to be studied. Therefore, this subjects to change as per the change in the surroundings- internal as well as external.

Future work primarily consists of more robust data acquisition systems and effective models, to deliver yet more efficient results which could help in a variety of fields. A few more parameters could also be used such as the driving surroundings and use of image retrieval and image processing techniques will be helpful in this regard. Also, the voice data has not yet been used. This data could also yield a lot of important and essential information in studying the driving traits.

6. Conclusion:

It could be observed that various conventional and non-conventional information retrieval methods can be used for analysis driver behavior and traits. The most used methods are vector space modelling, fuzzy retrieval, latent semantic indexing, Adaboost, k-means clustering, and neural networks. Many studies have also made use of statistical methods to extract multimodal features to study the reporting from the sensors.

This domain has greatly benefited the automotive design industry, the lawmaking and law enforcement agencies, and the psychologists. Besides, it has also been used in the development of autonomous vehicles, accident detection, road and vehicle condition monitoring, and the study of structural dynamics of the vehicles for automotive improvements.

References:

1. Arumugam, S., Bhargavi, R. A survey on driving behavior analysis in usage-based insurance using big data. *J Big Data* 6, 86 (2019). https://doi.org/10.1186/s40537-019-0249-5

2. Lin, N., Zong, C., Tomizuka, M., Song, P., Zhang, Z., & Li, G. (2014). An Overview on Study of Identification of Driver Behavior Characteristics for Automotive Control. *Mathematical Problems in Engineering*, *2014*, 569109. https://doi.org/10.1155/2014/569109
3. Nirmali, B., Wickramasinghe, S., Munasinghe, T., Amalraj, C. R. J., & Bandara, H. M. N. D. (2017). Vehicular data acquisition and analytics system for real-time driver behavior monitoring and anomaly detection. *2017 IEEE International Conference on Industrial and Information Systems (ICIIS)*, 1–6. https://doi.org/10.1109/ICIINFS.2017.8300417
4. Zinebi, K., Souissi, N., & Tikito, K. (2018, April 1). *Driver Behavior Analysis Methods: Applications oriented study*.
5. Turkson, R. (2016). *The Driving Behavior Data Acquisition and Identifiation Based on Vehicle Bus*. https://doi.org/10.4271/2016-01-1888
6. Aksjonov, A., Nedoma, P., Vodovozov, V., Petlenkov, E., & Herrmann, M. (2018). A Novel Driver Performance Model Based on Machine Learning. *IFAC-PapersOnLine*, *51*(9), 267–272. https://doi.org/10.1016/j.ifacol.2018.07.044
7. Abbadi, M., Abadleh, A., Alja'afreh, S., & Halhouli, Z. (2019). *Machine Learning-Based Approach for Detecting Driver Behavior Using Smartphone Sensors*. *8*(12), 4.
8. Li, N., Jain, J. J., & Busso, C. (2013). Modeling of Driver Behavior in Real World Scenarios Using Multiple Noninvasive Sensors. *IEEE Transactions on Multimedia*, *15*(5), 1213–1225. https://doi.org/10.1109/TMM.2013.2241416
9. *Title: Automotive Collision Avoidance Systems (ACAS) Program Final Report DOT HS 809 080*. (n.d.). Retrieved April 30, 2021, from https://www.google.com/imgres