Mining Spatio-temporal Data for Computing Driver Stress and Observing Its Effects on Driving Behavior

Rohit Verma*, Gyanesha Prajjwal**, Bivas Mitra*, Sandip Chakraborty*
*Indian Institute of Technology Kharagpur, INDIA; **National Institute of Technology Durgapur, INDIA
rohit.verma@iitkgp.ac.in,gp.16u10088@btech.nitdgp.ac.in,{bivas,sandipc}@cse.iitkgp.ac.in

ABSTRACT

With the increase in road fatalities due to various factors like aggressive driving and road rage, quantifying and monitoring the stress level of a driver is an important task for the preparation of driving rosters for the cab companies. Stress monitoring using physiological sensors is a costly and obstructive task, while stress factors impact differently for different individuals based on their personality traits. In this paper, we develop a learning-based model to predict the stress level of a driver and its effect on his driving behavior, solely based on spatio-temporal driving data collected through GPS and inertial sensors. We further establish a correlation between the stress level of a driver and his driving behavior; thus, we develop a complete system to infer stress profiling and its impact on driving behavior based on spatio-temporal driving data. The model has been tested over a publicly available dataset with 6 drivers for 500 minutes of driving data. We observe that the proposed model gives an average prediction accuracy of 79% with low false-positive rates.

CCS CONCEPTS

•Information systems →Spatial-temporal systems; •Computing methodologies →*Neural networks;*

KEYWORDS

Driver Stress, Driving Behavior, Spatio-temporal Driving Data

ACM Reference format:
Rohit Verma*, Gyanesha Prajjwal**, Bivas Mitra*, Sandip Chakraborty*. 2018. Mining Spatio-temporal Data for Computing Driver Stress and Observing Its Effects on Driving Behavior. In *Proceedings of 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Seattle, WA, USA, November 6–9, 2018 (SIGSPATIAL '18),* 4 pages.
DOI: 10.1145/3274895.3274941

1 INTRODUCTION

Road safety is a major problem afflicting people across the globe. The global status report on road safety by World Health Organization (WHO) [11] indicates that the total number of deaths worldwide due to various road accidents has leveled off at 1.3 million per year, with low-income and middle-income countries having higher road traffic fatality rates (over 90% of the world's fatalities). The reasons, among many others, are (i) inadequate adoption and enforcement of traffic laws to prevent major road safety issues like drink-driving and over-speeding, (ii) unplanned road and transport infrastructures, and (iii) lack of concentration during driving due to stress, fatigue, drowsiness or poor mental conditions. Among these, work-stress is one of the major deciding factors that correlate with poor driving behavior [18]. Various recent studies reveal the fact that factors like stress-driving, road rage, driving under fatigue or poor mental conditions etc. are predominant among the ride-sharing, car-pooling, truck and taxi cab drivers across low-income and middle-income countries [19], the primary reasons being hectic driving schedules due to various socioeconomic factors.

Although there have been several studies to understand the relationship between the driver stress and the driving behavior [1, 20], existing works are mostly based on physiological sensors or surveys to understand this relationship. These methods are quite obtrusive and difficult to be used for a long period. Further, physiological sensors are costly for mass-scale deployments over the drivers at low-income and the middle-income countries. On the contrary, spatial and inertial sensor data are widely available – cars are in-general GPS enabled; even a smartphone can provide such data. Therefore, a model to predict the stress level of a driver and its effect on driving behavior, solely based on spatial and inertial sensor data, can have a significant advantage for the development of stress monitoring applications and its large-scale low-cost deployments.

However, development of such a predictive model for correlating the stress profile of a driver and the driving behavior from historical data of driving trajectories is challenging itself. Several direct and indirect factors, like driving schedule, spatial features like road conditions, weather conditions, the mental state of the driver like anxiety etc. contribute to the stress profile of the driver and impacts the driving behavior. All of these factors may not be quantitative always, like the anxiety of the driver due to external factors may not be known apriori.

Against this backdrop, we analyze spatio-temporal driving dataset to develop a system (§4) which predicts the impact of past driving attributes on the stress level of a driver and his driving behavior under various driving and environmental conditions. We employ various driving attributes and spatial features of the driving environment to develop a Neural Network model which computes driver's stress based on the driving schedule roster (§4.1). Following this, we develop a mechanism to score the driver's driving behavior based on his driving style which could be affected by the stress profile (§4.2). Considering the abnormal driving behavior we provide a score to the driver. Finally, we give a quantifiable assertion that the driver stress impacts driving behavior and an increase in stress would deteriorate the driving behavior (§5). We provide a correlation between the two and based on this correlation, we could

SIGSPATIAL '18, Seattle, WA, USA

DOI: 10.1145/3274895.3274941

predict the driver's driving behavior at any later stage from the calculated driver stress by utilizing a simple linear model.

We evaluate our system over UAH-DriveSet data [13]. Experiments performed on these datasets give a 79% accuracy for the stress model and high accuracy and low false-positive rate for the abnormal behavior detection. Moreover, we are able to show a strong correlation between driver stress and driving behavior which can be utilized for the development of various safe driving applications.

2 RELATED WORK

The need to define a relationship between the stress and the driving behavior has been a buzz among the researchers for quite a time in the medical field. In this section, we present a brief survey of the existing literature broadly focusing on three aspects of our work – (a) stress modeling, (b) driver behavior identification and (c) defining a relationship between the stress and the driving behavior.

Stress Modeling: With the increasing number of accidents with the increase of private and public vehicles, researchers have explored physiological data to quantify various stress parameters of drivers using different methodologies, such as correlation analysis [17] and pattern recognition [4] over physiological sensor data, using machine learning techniques over ECG data of the drivers [9], and using machine learning techniques based on driving behavior and road statistics [10]. However, such quantification procedures of stress parameters primarily based on the physiological sensor data or personal survey from the individuals.

Driver Behavior Analysis: The increasing number of vehicles on the road made the problem a generalized one [2]. In recent times there have been works to model and provide a score to the driving behavior for safe driving [21]. Here, the authors have designed machine learning based techniques to identify several abnormal driving behaviors like weaving, sudden brakes, swerving, drifting, etc. and provide a score to the drivers.

Relationship between Stress and Driving Behavior: Various recent works in the literature [20] have explored the reasons for the abnormal driving behavior. In [12], the authors have shown that drowsy driving is one of the major factors behind road fatalities. Scott-Parker *et al.* [15] have attributed stress as a major factor behind the abnormal driving behavior. However, to the best of our knowledge, no existing works have developed a predictive model to correlate driving behavior with the stress profile considering various hidden factors, such as the personality traits and the driving environments. Consequently, in this paper, we devise a non-obtrusive technique for stress prediction based on the historical spatio-temporal driving data and provide a quantifiable relationship between the driver stress and the driving behavior.

3 DATASET DESCRIPTION

In this paper, we conduct the experiments on UAH-DriveSet data [13]. The UAH-DriveSet data [13] is a public dataset captured by the driving monitoring app developed by the authors. The data were collected for five male drivers and one female driver in the age group 20 – 50 over two different routes in Spain – (i) 25 km round-trip in a motorway type of road with 3 lanes in each direction and 120 km/h of maximum allowed speed, and (ii) 16 km round-trip in a secondary road with one lane in each direction and around 90 km/h of maximum allowed speed. The data contains both inertial sensors and GPS data along with a video clip for all the trips. The complete dataset has over 500 minutes of driving data. The data is labeled with three different driving behaviors – normal, drowsy and aggressive.

4 COMPUTING DRIVING STRESS AND DRIVING BEHAVIOR SCORE

The proposed system can be divided into three broad modules (a) Driver stress computation, (b) Driving behavior score calculation, and (c) Prediction of the driving behavior score from the driver stress. In the final developed system, we apply the driver stress computation model to calculate driver stress from the trip data and then use this stress level to predict the driving behavior score. In this section, we describe the model developed to compute driving stress and a method to assign driving behavior score.

4.1 Driver Stress Model

In this section, we develop a driver stress model to compute driver's stress from the trip information. Precisely, we extract the driving schedule, trip time, spatial information like road conditions, traffic conditions, etc. from the collected sensor data and compute the stress of the respective driver. We aim to classify the driver's stress into following three categories – *no stress*, *medium stress*, and *high stress*.

4.1.1 Model Development. In principle, the model should take the driver's trip information as input and compute the driver's stress label. The trip information may include trip time, trip distance, rest time after every trip along with spatial information such as road conditions, traffic congestion, etc. We develop the driver's stress model based on *Neural-Network* (NN) strategy. We implement one NN model to train each driver in isolation.

We next discuss how the ground truth is generated from physiological data and how the model is trained.

4.1.2 Model Training. We leverage on the HCILab dataset [14] to train the NN model. The dataset contains the driving information collected for ten drivers (3 female and 7 male), aged between 23–57 years, with four trips each. The recorded driving data includes acceleration along the three axes, light, and GPS coordinates, speed, altitude, and bearing. We extract 7 base features from the trip details, which are No. of trips (n_T), Trip time covered (t_T), Trip distance covered (d_T), Rest time (r_T), Time of the day (z), Congestion (C) and Road Type (r). Additionally, we leverage on the multiple derived features computed from statistical properties such as mean, variance, kurtosis and 80^{th} percentile calculated on those base features. Finally, a total of 27 features constitute the feature set $\mathbb{F}$ which along with the loss function for softmax regression [6] is used to train our model. Side by side, the dataset provides physiological sensor data (say, electrocardiogram (ECG), skin conductance rate, body temperature, heart rate and the heart rate variability) collected for each driver, which we utilize to compute the driver's stress label (no stress, medium stress and high stress) as the ground truth of the MTL-NN model. The details follow.

Ground Truth Stress Labels:

We implement the model proposed in [9] to calculate the stress of a driver from the available physiological data, which is used for ground truth generation for the HCILab dataset. This model classifies the driver stress into three classes, viz., *no stress*, *medium stress*, and *high stress*. It relies on eight standard ECG features, viz., *average QRS, RR, QQ, SS, QR, RS intervals*, *average heart beats* and *average difference beats* derived from the physiological sensors available in the PhysioNet dataset [5]. Precisely, the PhysioNet dataset contains the recorded data of electrocardiogram, skin conductance and respiration, collected from 17 healthy drivers in Boston area. The dataset was manually annotated with different stress levels for each driver. We implement a sliding window (of window size 15 seconds and shift of 500 ms) to compute the features and assign ground truth labels for each time window[1]. Notably, this short window handles the noisy nature of the physiological signal. Following [9], we implement the decision tree algorithm to classify driver stress labels from physiological data in PhysioNet dataset [5] obtaining 86% accuracy with a train-test split of 90%–10%. We leverage on this (validated) model to generate the ground truth stress label for each driver in the HCILab dataset. For that, we extract the features from physiological data available in the HCILab dataset for each time window and compute the respective stress label.

4.1.3 Model Evaluation. We evaluate the proposed NN model for stress calculation on the HCILab dataset. We calculate the features in each time window (15 seconds and 500 ms slide) and trained the model using ground truth stress labels generated using the PhysioNet dataset. The data is divided into non-overlapping set of 80% and 20% for training and testing respectively. Moreover, the training portion is divided as 60% and 20% of data for training and validation, respectively.

Considering the unbalanced dataset (as in general, most of the data points are in non-stress class), we calculate the results for the *Area Under the Curve* (AUC) metric. We observe that the approach has an an AUC of 0.794.

4.2 Driving Behavior Score

Stress leads to cognitive distraction of a driver which significantly impacts his concentration and affects driving behavior. In this section, we quantify the driving behavior of a driver from the recorded sensor data and assign a driving behavior score. We check how frequently the driver performs dangerous maneuvers [3, 16] such as sudden brakes, sharp turns, side-slipping etc.

4.2.1 Dangerous Maneuvers. We concentrate on the following six types of dangerous maneuvers [21] to compute the driving score – (a) *Weaving*: indicates constantly changing lanes, usually at high speed, creating an 'S' pattern, (b) *Swerving*: abruptly changing direction, (c) *Side-slipping*: driving in straight line, but deviating from the proper driving direction, (d) *Fast U-turn*: suddenly taking a U-turn, (e) *Sharp turn* and (f) *Sudden brake*. We employ the technique proposed in [21] to identify if the driver performs any of the aforesaid dangerous maneuvers. This method extracts unique signatures of these maneuvers from the acceleration and orientation sensor data and then computes a set of features. Following [21], we

[1] Since the window size was not specified in [9], we determine it empirically to obtain results identical to [9].

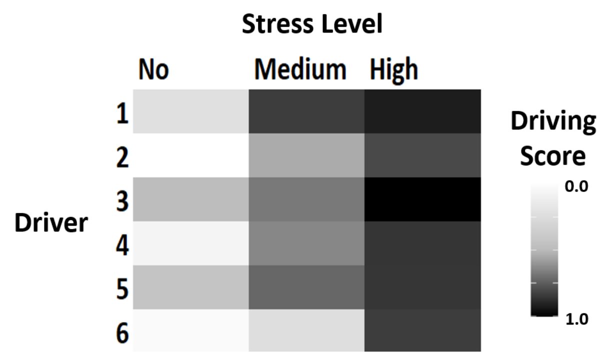

Figure 1: Driving behavior score with respect to the stress value for all the drivers

implement the method using a SVM (with 16 features) to detect if a dangerous maneuver has occurred. This model emits a tuple $\mathcal{M}$ of size six to store the score for each of the detected maneuvers. A score of 1 is assigned if a particular maneuver occurs in a sample window and 0 otherwise.

Finally the overall score is calculated as

$$\mathbb{D} = \frac{1}{6}\sum_{i=1}^{6} \mathcal{M}_i \tag{1}$$

Evidently, we assign a low score for the smooth driving and high score for the aggressive driving behavior.

Evaluation of Driving Score:

Table 1: Accuracy in assigning scores $\mathcal{M}$

Scenario	P	R	AUC
Weaving	0.92	0.95	0.89
Swerving	0.93	0.93	0.90
Side-slipping	0.96	0.97	0.94
Fast U-Turn	0.91	0.91	0.89
Sharp Turn	0.95	0.91	0.85
Sudden Brake	0.96	0.93	0.85

We evaluate the correctness of identifying the dangerous maneuvers. We perform the evaluation experiments on the the UAH-DriveSet data. The UAH-DriveSet provides files enumerating driver over-speeding, weaving, sudden brakes, etc as ground truth. Additionally, we obtain the other information like interaction with PoCs, sharp turns, swerving from the available video data. Table 1 exhibits the evaluation results in terms of precision (P), recall (R) and the AUC metrics for all the three aspects. This is comforting for us to observe that the computed score is able to capture the diverse driving behaviors efficiently.

5 PREDICTING DRIVING BEHAVIOR FROM DRIVER STRESS

We start this section by exhibiting the relationship between a driver's stress ($\mathbb{S}$) and his driving behavior ($\mathbb{D}$) for each driver in the UAH-DriveSet dataset. Fig. 1 shows the average driving score of each driver over the time windows sorted in increasing order, from left to right and displays the corresponding stress level. Evidently, it can be observed that increase in the stress level deteriorates the driving performance. In order to substantiate this empirical observation, we demonstrate that the correlation between $\mathbb{S}$ and $\mathbb{D}$ are statistically significant.

5.1 Correlating Driver Stress and Driving Behavior

We leverage on Kendall's tau coefficient [8] to conduct the correlation analysis of driver stress and driving behavior[2]. In order to calculate the coefficient, we first compute the concordant and discordant pairs. A pair of observations $(\mathbb{S}_i, \mathbb{D}_i)$ and $(\mathbb{S}_j, \mathbb{D}_j)$ are concordant if $\mathbb{S}_i > \mathbb{S}_j$ & $\mathbb{D}_i > \mathbb{D}_j$ or $\mathbb{S}_i < \mathbb{S}_j$ & $\mathbb{D}_i < \mathbb{D}_j$. In the same vein, the pairs are said to be discordant if $\mathbb{S}_i > \mathbb{S}_j$ & $\mathbb{D}_i < \mathbb{D}_j$ or $\mathbb{S}_i < \mathbb{S}_j$ & $\mathbb{D}_i > \mathbb{D}_j$. Kendall's tau coefficient is calculated as $\tau = \frac{n_c - n_d}{n(n-1)/2}$, where, n_c and n_d are the number of concordant and discordant pairs respectively in the dataset, and n is the total number of samples.

We obtain the mean correlation coefficient τ as 0.74 between driver stress $\mathbb{S}$ and driving behavior $\mathbb{D}$ for all the drivers in the dataset (with p-value $2.99x10^{-10}$), which statistically establishes the claim of Fig. 1.

5.2 Model Development: Driving Score Prediction

Finally, we develop a simple machine learning based model to predict the driving behavior score from the driver stress. The model is constructed considering the driver stress level as a single feature. We implement the model using *Simple Linear Regression* (based on the least square technique [7]). We obtain the ground truth driving scores of the drivers from the methodology described in §4.2. During the evaluation, we split the dataset into the ratio of 80% (to train the model) and 20% (for testing); we perform 5-fold cross-validation and report the results in terms of Mean Square Error (MSE), Mean Absolute Error (MAE) and R^2 score. We obtained low *MSE* (0.2763), low *MAE* (0.5237) and high R^2 (0.6021) score which show the elegance of our model.

6 CONCLUSION

With the gaining popularity of the cab-sharing and car-pooling services, a low-cost and easy-deployable solution for driver stress monitoring has become inevitable for the cab companies; driver stress can work as an important deciding factor for the preparation of driving rosters and scheduling trips for cab drivers. In this paper, we have exploited spatio-temporal driving data (GPS and inertial sensor data) to model a driver's stress based on his past driving attributes and trip history; we designed a learning model to quantify the driver stress. We have further established a strong correlation between the driver stress and the driving behavior. Finally, we have developed a simple model to predict the impact of driver stress on the driving behavior. To the best of our knowledge, this is the first work that meticulously quantifies a driver's stress level and measures its effect on driving behavior, solely based on the spatial trajectories of past trips and the inertial sensor logs collected during the driving. The correctness of the stress model has been validated using physiological sensor data as ground truth, and the effectiveness of the prediction model has also been demonstrated using a large set of spatio-temporal driving data.

[2]We use the Kendall's tau coefficient because it has low gross error sensitivity and a low asymptotic variance and hence is suitable for the large sample cases as ours.

REFERENCES

[1] 2018. HAMS: Harnessing AutoMobiles for Safety - Microsoft Research. (2018). https://www.microsoft.com/en-us/research/project/hams/.

[2] Rahel Bieri, Tobias Nef, René M Müri, and Urs P Mosimann. 2015. Development of a novel driving behavior adaptations questionnaire. *International psychogeriatrics* 27, 6 (2015), 1017–1027.

[3] Jerry L Deffenbacher, David M Deffenbacher, Rebekah S Lynch, and Tracy L Richards. 2003. Anger, aggression, and risky behavior: a comparison of high and low anger drivers. *Behaviour research and therapy* 41, 6 (2003), 701–718.

[4] Jennifer Healey and Rosalind Picard. 2000. SmartCar: detecting driver stress. In *Proc. of Pattern Recognition*, Vol. 4. IEEE, 218–221.

[5] Jennifer A Healey and Rosalind W Picard. 2005. Detecting stress during real-world driving tasks using physiological sensors. *IEEE TITS* 6, 2 (2005), 156–166.

[6] David Heckerman and Christopher Meek. 1997. Models and selection criteria for regression and classification. In *Proc. of UAI*. Morgan Kaufmann Publishers Inc., 223–228.

[7] Eric Iksoon Im et al. 1996. *A Note On Derivation of the Least Squares Estimator*. Technical Report.

[8] Maurice G Kendall. 1938. A new measure of rank correlation. *Biometrika* 30, 1/2 (1938), 81–93.

[9] N Keshan, PV Parimi, and Isabelle Bichindaritz. 2015. Machine learning for stress detection from ECG signals in automobile drivers. In *Proc. of Big Data*. IEEE, 2661–2669.

[10] Mario Muñoz-Organero and Victor Corcoba-Magaña. 2017. Predicting Upcoming Values of Stress While Driving. *IEEE Transactions on Intelligent Transportation Systems* 18, 7 (2017), 1802–1811.

[11] World Health Organization. 2015. *Global status report on road safety 2015*. World Health Organization.

[12] Hanbit Park, Seungwon Oh, and Minsoo Hahn. 2009. Drowsy driving detection based on human pulse wave by photoplethysmography signal processing. In *Proc. of IUCS*. ACM, 89–92.

[13] Eduardo Romera, Luis M Bergasa, and Roberto Arroyo. 2016. Need data for driver behaviour analysis? Presenting the public UAH-DriveSet. In *Proc. of ITSC*. IEEE, 387–392.

[14] Stefan Schneegass, Bastian Pfleging, Nora Broy, Frederik Heinrich, and Albrecht Schmidt. 2013. A data set of real world driving to assess driver workload. In *Proc. of AutomotiveUI*. ACM, 150–157.

[15] B Scott-Parker, CM Jones, K Rune, and J Tucker. 2018. A qualitative exploration of driving stress and driving discourtesy. *Accident Analysis & Prevention* 118 (2018), 38–53.

[16] Lipaz Shamoa-Nir and Meni Koslowsky. 2010. Aggression on the road as a function of stress, coping strategies and driver style. *Psychology* 1, 01 (2010), 35.

[17] Mandeep Singh and Abdullah Bin Queyam. 2013. A novel method of stress detection using physiological measurements of automobile drivers. *International Journal of Electronics Engineering* 5, 2 (2013), 13–20.

[18] Predrag Stanojević, Timo Lajunen, Dragan Jovanović, Paul Sârbescu, and Svilen Kostadinov. 2018. The driver behaviour questionnaire in South-East Europe countries: Bulgaria, Romania and Serbia. *Transportation Research Part F: Traffic Psychology and Behaviour* 53 (2018), 24–33.

[19] Yonggang Wang, Linchao Li, and Carlo G Prato. 2018. The relation between working conditions, aberrant driving behaviour and crash propensity among taxi drivers in China. *Accident Analysis & Prevention* (2018).

[20] Yan Wang, Jie Yang, Hongbo Liu, Yingying Chen, Marco Gruteser, and Richard P Martin. 2013. Sensing vehicle dynamics for determining driver phone use. In *Proc. of MobiSys*. ACM, 41–54.

[21] Jiadi Yu, Zhongyang Chen, Yanmin Zhu, Yingying Jennifer Chen, Linghe Kong, and Minglu Li. 2017. Fine-grained abnormal driving behaviors detection and identification with smartphones. *IEEE TMC* 16, 8 (2017), 2198–2212.