Using Context to Identify Difficult Driving Situations in Unstructured Environments

Kevin R. Dixon[1], Justin D. Basilico[1], Chris Forsythe[1], and Wilhelm E. Kincses[2]

[1] Sandia National Laboratories, Albuquerque, NM, 87185 USA
{krdixon,jdbasil,jcforsy}@sandia.gov
[2] Daimler AG Group Research, 71059 Sildelfingen, Germany
wilhelm.kincses@daimler.com

Abstract. We present a context-based machine-learning approach for identifying difficult driving situations using sensor data that is readily available in commercial vehicles. The goal of this system is improve vehicle safety by alerting drivers to potentially dangerous situations. The context-based approach is a two-step learning process by first performing unsupervised learning to discover meaningful regularities, or "contexts," in the vehicle data and then performing supervised learning, mapping the current context to a measure of driving difficulty. To validate the benefit of this approach, we collected driving data from a set of experiments involving both on-road and off-road driving tasks in unstructured environments. We demonstrate that context recognition greatly improves the performance of identifying difficult driving situations and show that the driving-difficulty system achieves a human level of performance on cross-validation data.

1 Introduction

Cars are an essential means of transportation for much of the world. However, the widespread use of automobiles exacts a large toll in the form of property damage, injury, and death. The United States National Highway Traffic Safety Administration reports that "In 2005, there were an estimated 6,159,000 police-reported traffic crashes, in which 43,443 people were killed and 2,699,000 people were injured;" it is the leading cause of death of people aged 3 through 33 [1]. Naturalistic driving studies have shown that having a passenger in the vehicle reduces the odds-ratio of having a crash by 50% [2]. The goal of this research is not to automate driving, but to identify and mitigate potentially dangerous situations for the driver, similar to a "backseat driver," improving safety. To this end, we have conducted a series of experiments in both on-road and off-road driving in unstructured environments. In these experiments, we have shown that our system identifies difficult driving situations with performance similar to that of a human backseat driver, and see significant improvements in the performance of drivers during the experimental conditions. Our driving-difficulty classifier system operates in real time in unstructured environments without human intervention, using sensors that are readily available on commercial vehicles without additional instrumentation.

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 20–29, 2009.

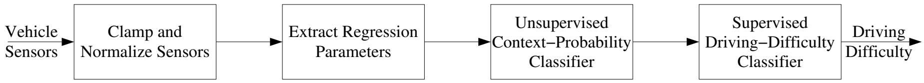

Fig. 1. Data-flow diagram in the context-based difficulty classifier system

We create the driving-difficulty detector using a two-step semi-supervised machine-learning approach [3]. The first step takes unlabeled data from the vehicle's data bus and automatically extracts the context by automatically identifying statistical regularities in the vehicle data. Our hypothesis is that the driver performing the underlying physical task - driving in the given conditions - induces observable regularities in the vehicle data and identifying these regularities, or "contexts," is crucial in achieving a human-level of performance. For example, entering a high-speed roadway tends to result in a driver pressing down the accelerator pedal, entering a period of relatively high lateral acceleration, turning on a lane-change signal, and achieving a fast speed. In this example, the underlying physical task induces regularities in how the driver interacts with the vehicle. We are interested in automatically extracting contexts to determine when the driver is entering a potentially difficult situation. With the contexts identified, the system then maps these contexts onto a difficulty score using a supervised-learning machine-learning algorithm (Fig. 1). To validate the system, we compare the performance of an actual human backseat driver with our automated system, both with and without context recognition, in identifying potentially dangerous driving conditions.

2 Related Work

For over twenty years, there has been interest in developing autonomous driving systems, with an early example being the NAVLAB project [4] and research is ongoing [5]. Autonomous driving systems have recently gained widespread attention in the research community and mainstream media, due in large part to the DARPA Grand Challenge [6] and the follow-on DARPA Urban Grand Challenge. While computer systems and robots may one day replace humans as the main users of the world's highways, it is likely that humans will continue to be the primary drivers of motor vehicles for the near future. This will continue the trend of over 40,000 fatalities per year in the United States alone, coupled with incalculable related damages [1]. The 100-car naturalistic driving study [2] recorded almost 10,000 crashes, near crashes, and "crash-relevant conflicts" over the course of about one year. This averages to about seven incidents per subject per month. One bright spot is that the same study showed that having a passenger in the vehicle reduces the odds-ratio of having a crash by 50% [2]. In some sense, the goal of this research is to have the same crash-reducing effect that passengers had in the naturalistic driving study. There has been substantial research into driver-assistance systems. Many systems focus on placing additional sensors on the vehicle, particularly visible-light cameras [7, 8], to identify previously undetectable situations. Other groups have focused on developing models of human drivers to focus attention [9]. While these are very promising avenues to pursue, we feel that we can offer powerful driver-assistance tools by intelligently analyzing readily available sensors on commercial vehicles to determine how the current situation can impact driver performance. Unsupervised learning has been used as a basis for context recognition for mobile devices [10] and for improving image classification [11]

The work presented in this paper extends the previous work in driving-difficulty systems of [12], which trained a classification system to identify potentially dangerous driving conditions using predefined situations. This system identified eight high-level situations with high accuracy: 1) Approaching or Waiting at Intersection, 2) Leaving Intersection, 3) Entering On-ramp or High-Speed Roadway, 4) Being Overtaken, 5) High Acceleration or Dynamic State of Vehicle, 6) Approaching Slow-Moving Vehicle, 7) Preparing to Change Lanes, and 8) Changing Lanes. However, this system was based purely on supervised-learning classifications on predefined categories. The primary limitation is that predefined categories are inherently limited by the cleverness of the developers to identify all relevant situations, while ignoring irrelevant ones. This also means that the system must have numerous examples of each situation against which to train the classifier. Out of the 24 hours of data collected, the rarest situation, "Entering On-ramp or High-Speed Roadway," was present for less than 1% of the data and it is very challenging for any machine-learning classifier to identify rare events [13]. Building on this previous work, our system uses a two-stage approach to identifying potentially dangerous driving conditions.

3 Algorithms

The central component of our approach is the automated unsupervised learning of context. Because we typically have a much larger amount of unlabeled data than labeled data, we take a semi-supervised approach to learning. The creation of contexts using unsupervised clustering algorithms makes use of all data recorded from an experimental vehicle. The supervised learning of driving difficulty makes use of the smaller amount of labeled data. This allows the driving-difficulty classifier to make productive use of all the unlabeled and labeled data.

3.1 Data Representation

The input to the system is a discrete-time temporal signal, which is extracted from sensors aboard an experimental vehicle from its standard Controller Area Network (CAN) bus (Section 4.1). Because we are interested in the change of the sensor values over time, we extract the rate-of-change and current-value information from each signal over a fixed time window. This feature-extraction process converts temporal signals into a vector-based representation. In terms of the features to use in the driving-context recognition, we feel that:

1. The magnitude of a signal is important. For example, knowing the speed of the vehicle or brake-pedal force can help to disambiguate similar contexts.
2. The general trend of a signal is also important. For example, knowing how sensors are changing can differentiate otherwise identical contexts.

With this in mind, at each time step for each input sensor, we construct a window over some predefined length into the past (typically 5 seconds) and compute the first-order linear-regression slope-intercept coefficients $\{m,b\}$ for that time window.

Converting a windowed temporal signal into a vector using the linear-regression coefficients creates two coordinates; the regression slope (*m*) and the regression intercept (*b*). Consequently, if there are 5 input signals, the result will be a 10-dimensional vector. Our unsupervised-clustering algorithms search for driving contexts in this vector space.

3.2 Unsupervised Context Learning

At each time step, the input to the unsupervised-learning context classifier is the collection of vectors with the slope-intercept regression parameters for each sensor. The unsupervised context-learning algorithm is a reductionist version of the prevalent *k*-means clustering algorithm [3]. To determine vector similarity, we use the Mahalanobis distance and compute the sample mean and full covariance matrices belonging to each cluster. We make an assumption that each regression-coefficient vector is generated independently of all others. With this assumption, the number of data points assigned to a *particular* cluster is a binomial random variable, and we remove a cluster if its corresponding probability is too low. By evaluating the binomial cumulative distribution function, we can determine if a cluster is not significant, in a statistical sense, and should be removed. If we have *k* clusters and *N* data-points, then the expectation is that each cluster contains N/k data-points. From this perspective, we can set a removal threshold based on the fraction of data-points of the expectation. For example, a threshold of 0.5 means that we will remove any clusters containing less than $0.5N/k$ data-points. In practice on our experimental data, this reductionist clustering approach yields relatively stable numbers of clusters from random initializations ($E\{k\}$=53.5,±1.92, $p<0.05$ for a removal threshold of 0.5). We also find the reductionist clustering approach to less sensitive to the initial parameter *k* because if the value of *k* is initially set too high, the algorithm will compensate by removing spurious clusters. Thus, to set *k* we can initially choose a relatively high value and then let the algorithm iteratively remove clusters to find a stable value.

3.3 Supervised Learning of Driving Difficulty

Up to this point, the system has mapped temporal vehicle sensors to a *k*-dimensional vector of context probabilities (cf. Fig. 1). We use supervised learning to map this context-probability vector to a difficulty score. As we describe later in more detail in Section 4.1.2, we collected labels of driving difficulty for a subset of the experimental data, by either backseat observation or *post hoc* video analysis. We use these scalar 1-100 value labels as ground-truth outputs for a supervised-learning algorithm. Because the values are continuous, this difficulty classifier can be stated as a standard regression formulation. Not surprisingly, driving difficulty does not change dramatically from second to second and the ground-truth difficulty labels are highly auto-correlated (R=0.89 at 5-second lag).

4 Experimental Description

We have conducted a series of driving experiments in unstructured environments over the past several years. The first studies were a proof of concept that we could infer

Fig. 2. Frontal camera view from the Camp Pendleton experiments used for *post hoc* labeling and analysis

difficult driving situation from readily available sensors from a commercial vehicle in naturalistic on-road driving conditions [12]. The set of experiments covered by this analysis involved driving in off-road conditions, on semi-improved and unimproved paths, at the United States Marine Corps Base Camp Pendleton. These experiments tested the ability for our system to identify high-difficulty driving conditions without the presence of human-made regularities, such as traffic lights, lanes, and signage. Drivers were instructed to drive on a predefined road circuit, but we did not attempt to alter the roadway and or control external conditions in any of the experiments. As such, we have encountered snow, rain, fog, traffic jams, road construction, mechanical problems, armed guards, artillery howitzers, lost vehicles, and even flocks of sheep (complete with over-protective herding dogs). Through the evolution of these experiments and the knowledge gained, we have learned that identifying driving *context* is crucial in achieving human-level accuracy with a driving-classification system. By context, we mean those regularities that are caused by the human operator (the driver) making the vehicle behave in a constrained manner.

4.1 Data Collection

Before each experiment, the subjects familiarize themselves with the test vehicle and drive on a sample course. Additionally, before the main experiment, we conduct a calibration study where we collect data from a small number of subjects with which we train our difficulty-classification system. The purpose of the calibration study was to duplicate the experimental conditions and gain insight into the phenomena that would be helpful in identifying high-difficulty situations. The use of calibration data also meant that a general driving-difficulty model is used, rather than a unique model for each driver. After the calibration study is complete, the main pool of subjects performs the driving study, as in [14].

Vehicle Data. To obtain information about the state of the vehicle and how the driver is interacting with it, we interfaced through the Controller Area Network (CAN) bus

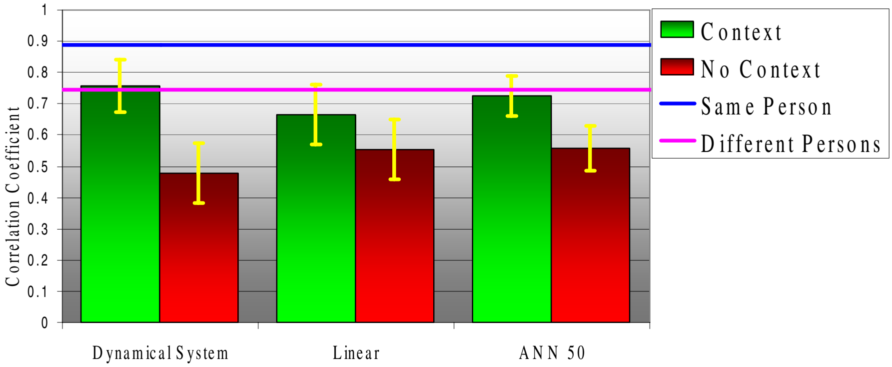

Fig. 3. Performance of the difficulty classifier with and without context recognition in terms of normalized correlation coefficient (ρ). The error bars are the cross-subject 95% confidence intervals. Same persons have an average of ρ=0.95 for subsequent scoring of the same round. Different persons difficulty scores have an average of ρ=0.74 agreement.

of the vehicle to record sensors that are readily available on many commercial vehicles. In all experiments, we sampled data from the CAN bus at 4Hz. In our experience, sampling rates above 4Hz did not improve performance, and sampling slower than 2Hz could result in missed events. Our difficulty-classification system incorporated two types of sensors: sensors that directly measure how the driver is interacting with the vehicle and sensors that measure secondary interactions or vehicle state. From the control-surface state, we made use of steering-wheel position, force applied to the brake pedal, and accelerator-pedal deflection. From the physical state of the vehicle, we made use of wheel speeds, adaptive cruise-control radar, and current gear number. There are many driver-assistance systems that require special-purpose instrumentation [8] and these provide valuable insight into the cost-benefit analysis of additional instrumentation to vehicles. However, our driving-difficulty classifier does not require any experiment-specific instrumentation of the driver or vehicle, meaning that this system is deployable on currently available commercial vehicles.

Difficulty Labels. To generate the ground-truth labels, the difficulty of the current driving situation were scored on 100-point scale (1 to 100) entered by a human labeler with an external dial or a software slider bar. A value of 1 means that the driving is very easy, while a value of 100 means that there is imminent danger. Furthermore, the labelers were instructed that a score of 50 or above indicated a judgment that it would be a bad time to burden the driver with additional tasks, such as a mobile-phone call. Allowing the labelers to input a continuous value on a 100-point scale, instead of a binary difficulty decision, makes it possible to create more accurate machine-learning classifiers. A human labeler can generate difficulty scores in two ways: sitting in the back seat of the vehicle during the experiment or a graphical user interface for *post hoc* analysis. For *post hoc* labeling, we constructed a user interface that displays a video recording taken out the front window of the vehicle, such as Fig. 2, and controls that allowed the labeler to move forward and backward in time so that users may

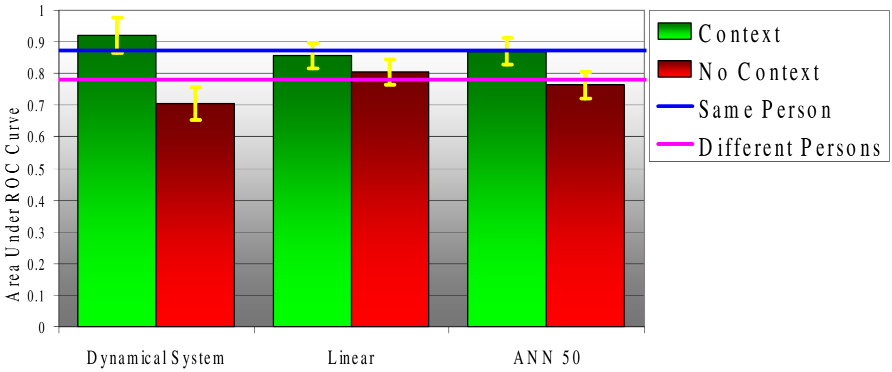

Fig. 4. Performance of the difficulty classifier with and without context recognition in terms of the receiver operator characteristic (ROC) area under curve (AUC). The error bars are the 95% cross-subject confidence intervals. Same persons average an AUC of 0.87 for subsequent scoring of the same round. Different persons agree with each other with an average AUC of 0.78.

adjust difficulty labels to ensure their accuracy. While the human labeler may use the video to generate difficulty labels, the classifier system did not process the images in keeping with the requirement that the system only use sensors currently available on commercial vehicles.

4.2 Off-Road Experiments in Camp Pendleton

We conducted a series of experiments at the United States Marine Corps Base Camp Pendleton, where the experimental platform was a Mercedes-Benz G-class 500 SUV. In these experiments, subjects drove on a mixed semi-improved and off-road circuit four times at 30 km/hour, with each circuit lasting about half an hour. We collected data from nineteen drivers, resulting in 42 hours (609,744 samples) of data. As in our previous experiments, we had to contend with unforeseen events, such as vehicle traffic, road guards, and other equipment. The results described in this paper will be based on the data collected from these experiments (Section 5).

5 Results

To evaluate the results of our driving-difficulty classifier, we compared the context-based difficulty recognizers to those without context recognition. For the results without context recognition, we mapped directly from the regression-coefficients (cf. Fig. 1) to the difficulty labels[1]. In all cases, we tried several regression architectures, including a linear dynamical system, a linear mapping, and a feedforward artificial neural network (ANN).

The linear dynamical system was trained using an iterative one-step optimal Expectation-Maximization routine using least-squares pseudoinversion of the feedforward and feedback matrices. The linear mapping was trained using the closed-form

[1] Mapping from the sensors to driving difficulty did not produce results better than random.

optimal least-squares pseudoinverse. The ANN had *arctan* node activation with 50 hidden units[2], trained with the quasi-Newton Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm with Fletcher-type line search [15]. The unsupervised context-learning algorithm was given 7.5 hours (107,313 samples) of unlabeled driving data from which to extract driving context. The supervised-learning algorithms were trained with 4.6 hours (65,678 samples) of labeled driving-difficulty data. The hold-out cross-validation set was 2.4 hours (34,581 samples) of labeled data from subjects not contained in the supervised-training or context-learning sets.

To compare the performance of the different approaches, we use the correlation coefficient between the estimated driving difficulty (the scalar 1-100 values) and the ground-truth driving difficulty generated by the human labelers. To baseline the results, we also asked human labelers to generate difficulty labels for the same round on subsequent days, and asked different human labelers to generate difficulty labels, and compared their results to other labelers. This yields a correlation for how consistent humans are with themselves, and how consistent different persons are with each other. The results are summarized in Fig. 3. In terms of correlation coefficients, all context-based difficulty classifiers outperform those that do not use context recognition. The context-based linear dynamical systems (ρ=0.76) and the context-based ANN (ρ=0.73) perform to the consistency level of different persons with each other (ρ=0.74). The best non-context-based classifier, the ANN, achieved a statistically significantly worse correlation of ρ=0.56.

Another measure of performance is the receiver operating characteristic (ROC) area under curve (AUC) measure [3]. In our case, this measures the probability that a difficulty estimate will agree with a ground-truth label that the situation is "too difficult," cf. Section 4.1.2. The results are summarized in Fig. 4. Once again, all context-based classifiers outperform those that do not use context recognition. The best performer was the context-based linear dynamical system (AUC=0.92), which performed as well as the self-consistency of human labelers (AUC=0.87). The best non-context-based classifier, the linear mapping, achieved a statistically significantly worse result of AUC=0.80. Thus, the best context-based classifier reduces the AUC error rate by almost 60% over those classifiers that do not use context recognition, achieving human levels of performance on both correlation and AUC measures.

6 Conclusions and Future Work

We have presented a context-based semi-supervised machine-learning approach to identify difficult driving situations. We showed that context-based classifiers outperform those that do not use context recognition and that a context-based linear dynamical system can achieve human-like performance on real-world experimental data. In future work, we plan to look at techniques for automatically adapting the generalized contexts to the behavior of a new driver. This will create contexts that are representative of the actual person-specific driving style. In addition, because we have much

[2] An ANN with 50 hidden units performed better than other hidden-layer sizes on cross-validation data, which is, incidentally, close to the number of contexts discovered by our unsupervised context-learning algorithm on this data set.

more unlabeled data than labeled data, we want to look at bootstrapping techniques for the difficulty scorer.

In the experiments so far, we have applied this technique within the realm of driver overload. In the future, we plan to change our focus to look at the more common condition of driver underload. By underload, we mean those situations that become potentially dangerous because the driver is distracted, inattentive, drowsy, or bored. We plan to extend the context-based approach to unsupervised learning approach in order to identify unusual, potentially dangerous driving situations due to underload.

Acknowledgment

This work was funded by the Office of Naval Research Code 30 and DARPA/DSO/IPTO. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under contract DE-AC04-94AL85000.

References

1. NHTSA: Traffic safety facts: 2006 data. Technical Report DOT HS 810 807, U.S. Department of Transportation, National Highway Traffic Safety Administration (2007)
2. Klauer, S., Dingus, T., Neale, V., Sudweeks, J., Ramsey, D.: The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data. Technical Report DOT-HS-810-594, National Highway Traffic Safety Administration (2006)
3. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification, 2nd edn. Wiley-Interscience, New York (2001)
4. Thorpe, C., Hebert, M., Kanade, T., Shafer, S.: Vision and navigation for the Carnegie-Mellon Navlab. IEEE Transactions on Pattern Analysis and Machine Intelligence 10(3), 362–373 (1988)
5. McDowell, K., Nunez, P., Hutchins, S., Metcalfe, J.S.: Secure mobility and the autonomous driver. IEEE Transactions on Robotics 24(3), 688–697 (2008)
6. Montemerlo, M., Thrun, S., Dahlkamp, H., Stavens, D., Strohband, S.: Winning the DARPA Grand Challenge with an AI robot. In: Proceedings of the AAAI National Conference on Artificial Intelligence (2006)
7. Franke, U., Heinrich, S.: Fast obstacle detection for urban traffic situations. IEEE Transactions on Intelligent Transportation Systems 3(3), 173–181 (2002)
8. Fletcher, L., Apostoloff, N., Petersson, L., Zelinsky, A.: Driver assistance systems based on vision in and out of vehicles. In: Proceedings of the IEEE Intelligent Vehicles Symposium (2003)
9. Miller, B.W., Hwang, C.H., Torkkola, K., Massey, N.: An architecture for an intelligent driver assistance system. In: Proceesings of the IEEE Intelligent Vehicles Symposium, pp. 639–644 (2003)
10. Himberg, J., Korpiaho, K., Mannila, H., Tikanmaki, J., Toivonen, H.T.: Time series segmentation for context recognition in mobile devices. In: IEEE International Conference on Data Mining (2001)

11. Boutell, M., Luo, J.: Incorporating temporal context with content for classifying image collections. In: Proceedings of the 17th International Conference on Pattern Recognition, vol. 2, pp. 947–950 (2004)
12. Dixon, K.R., Lippitt, C.E., Forsythe, J.C.: Supervised machine learning for modeling human recognition of vehicle-driving situations. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (2005)
13. Kubat, M., Holte, R.C., Matwin, S.: Machine learning for the detection of oil spills in satellite radar images. Machine Learning 30(2-3), 195–215 (1998)
14. Kohlmorgen, J., Dornhege, G., Braun, M., Blankertz, B., Muller, K.R., Curio, G., Hagemann, K., Bruns, A., Schrauf, M., Kincses, W.: Improving human performance in a real operating environment through real-time mental workload detection. In: Toward Brain-Computer Interfacing, pp. 409–422. MIT Press, Cambridge (2007)
15. Fletcher, R.: Practical Methods of Optimization, 2nd edn. Wiley Interscience, New York (1987)