DETECTING SLEEPINESS IN TRUCK DRIVERS

David Sandberg*, **Mattias Wahde**
Department of Applied Mechanics
Chalmers University of Technology, 412 96 Göteborg, Sweden
Tel.: +46 31 772 3696
e-mail: {david.sandberg,mattias.wahde}@chalmers.se

ABSTRACT

This paper deals with an important component of driver inattention, namely driver sleepiness, in the particular case of truck drivers. Analysis of data obtained from driving sessions involving ten test subjects shows that, contrary to some results found in the literature, the standard deviation of lateral position is unable to predict driver sleepiness.

By contrast, using a generic variability indicator, with different parameter settings during daytime and nighttime driving, respectively, a performance score (after optimization) of 0.80 (on a scale from 0 to 1) was obtained (compared to 0.54 for the standard deviation of the lateral position). Furthermore, an optimized sleepiness detection system combining several indicators achieves a score of 0.88, showing that different types of indicators can work together to achieve a better prediction of sleepiness than any indicator alone.

KEYWORDS

Sleepiness detection, stochastic optimization, sleepiness in truck drivers.

INTRODUCTION

Inattention to the road ahead while driving is a primary cause for vehicle accidents [15]. An important reason for driver inattention is eye closure due to driver sleepiness [12]. In fact, driver sleepiness has been estimated to be a contributing factor in around 20 % of all vehicle crashes [10, 13, 20]. Truck drivers, in particular, suffer a high risk of having a fatal crash due to sleepiness; it has been suggested that as many as 30 % of all fatal truck crashes are related to sleepiness [3].

The present paper describes research aimed at devising driver support systems capable of robust and reliable sleepiness detection in truck drivers. Such systems, in turn, typically consist of several sleepiness indicators. The research problem thus involves formulating and optimizing sleepiness indicators, as well as finding (also through optimization) useful combinations of such indicators.

DATA

The data used in the present analysis were collected during a study performed in 2008 in eastern Sweden, involving ten test subjects who drove a truck from the city of Linköping to the town of Målilla and then back, along the road Riksväg 34, for a combined period of around 80 hours.

Study design

A within-subject design was used where each test subject performed two driving sessions: one day session from 12:00 to 16:00 and one night session from 00:00 to 04:00. The test subjects were instructed to have normal nights of sleep prior to the study and were monitored during the days before the study as well as on the day of driving. The goal of the study was twofold: to evaluate a driver sleepiness warning system developed within the framework of a Swedish research project[1] and to collect data for further research. Hence, during parts of the four-hour driving sessions a sleepiness warning system was active and may have affected the driving behavior. Thus, only data collected during the parts where that system was inactive have been used in the research presented here. The warning system was active on the way back to Linköping from Målilla on both the day and night sessions. Hence, the remaining data, used here, was about half of the total amount of data collected.

Driving behavior signals

The truck was equipped with a lateral position sensor[2] (sampled at 16 Hz) providing the lateral position of the truck and the heading angle (i.e. the heading of the vehicle relative to the tangent of the driving lane). Furthermore, the steering wheel angle and the speed of the vehicle (sampled at 50 Hz), were obtained from the controller-area network (CAN).

Since the lateral position sensor was mounted at the lateral center of the vehicle, the measurements obtained from this sensor represent the distance from the center of the vehicle to either driving lane boundary; in the work presented here, the lateral position has been taken as the distance to the right lane boundary. The lateral position sensor also provides a confidence signal, taking values in the range $[0, 100]$, such that each sample represents the quality (0 being the lowest quality and 100 the highest) of the corresponding samples in the other signals obtained from the lateral position sensor.

Eyelid recordings

A Driver State Sensor (DSS) system[3] was used throughout the study to record eyelid movements. The DSS also produced a confidence signal, indicating the reliability of the recorded eyelid movement signals.

[1] The study was conducted within the Drowsi research project, in which several industrial, academic and Swedish government agency partners collaborate to further the research on driver sleepiness. See `http://www.ivss.se/drowsi` for more information.

[2] The lateral position sensor used in the study is a camera, mounted between the rear view mirror and the windshield. Machine vision techniques are applied in order to identify the driving lane boundaries, from which the position of the vehicle (within the driving lane) can be computed.

[3] The DSS is manufactured by Seeing Machines (`http://seeingmachines.com`).

Subjective measurements of sleep

Prior to driving, the test subjects had been instructed to estimate their level of sleepiness according to the Karolinska sleepiness scale (KSS) [2] which has been shown to correlate well both with the physiological level of sleepiness [2] and with a deterioration in driving performance [17]. While driving, the test subjects were required to provide a KSS estimate every five minutes, taking the preceding five minutes into account. The KSS estimates obtained from each test subject (TS) were taken as the ground truth, and were used for distinguishing between alert driving (KSS ≤ 6) and sleepy driving (KSS ≥ 7). The problem of estimating driver sleepiness was then cast as a binary classification problem, with the two classes *alert* (C_1) and *sleepy* (C_2) containing driver behavior data from samples with KSS ≤ 6 and KSS ≥ 7, respectively.

Two of the test subjects, namely TS3 and TS10, reported no estimates with KSS ≥ 7, implying that they provided no data to class C_2. Thus, the data obtained from TS3 and TS10 were discarded altogether, leaving data from a total of eight test subjects for use in the analysis.

DETECTING SLEEPY DRIVING

Several indicators of driver sleepiness, and sleep in general, can be found in the literature; some are formed using driving behavior signals (such as the lateral position of the vehicle or the steering wheel angle), while others are based on eyelid movements and physiological signals (such as brain waves, i.e. EEG). Furthermore, it is also possible to combine several indicators of driver sleepiness to form a system for driver sleepiness detection. For example, in [7], several measures of driving performance were combined using linear regression techniques; in [8] an indicator based on blink behavior was combined with a lane drifting measure using a lookup table.

The indicators of driver sleepiness considered in this paper are based either on (i) driving behavior, (ii) eyelid behavior or (iii) a mathematical model of sleepiness. Several indicators of type (i) are studied whereas only one indicator of type (ii), namely Perclos, and one indicator of type (iii), namely the Sleep/Wake Predictor, are considered.

Indicators based on driving behavior

Quite a few indicators of driver sleepiness have been investigated and proposed in the literature. Perhaps the most common indicator is the *standard deviation of the lateral position* which simply measures the average lateral deviation. The rationale behind this indicator is that a high degree of lateral deviation is supposedly indicative of driver sleepiness, as indeed has been demonstrated in [23, 21, 4, 18].

Of course, standard deviations can be computed for other signals as well; here, in addition to the standard deviation of the lateral position, the standard deviations of the heading angle and the steering wheel angle have been considered.

Also quite common in the literature on driver sleepiness is to study the steering wheel behavior, often measured using the *steering wheel reversal rate* [14], defined as the number of reversals[4] r, during a given period, that fall within a certain range $c_1 < r < c_2$. Three gap sizes for reversals, proposed in the literature, have been considered in the present analysis: small reversals ($c_1 = 0.0087$ and $c_2 = 0.0873$) [16], medium reversals ($c_1 = 0.1396$ and $c_2 = 0.2618$) [23] and large reversals ($c_1 = 0.2618$ and $c_2 = \infty$) [23]. Of course, the values of c_1

[4] A reversal is given by two local optima in the steering wheel angle signal, with the size of the reversal being the difference in angle between the two local optima.

and c_2 affect the performance of the steering wheel reversal indicator. These parameters can therefore be optimized in order to improve the indicator, something that has been done in the present study.

Generic variability indicator

A general measure of variability, called the generic variability indicator (GVI), was introduced in [19] with the aim of providing a general and optimizable indicator of driver sleepiness. The GVI was designed to include many of the indicators of driver sleepiness proposed in the literature as special cases, depending on the parameter settings used. The GVI is defined as

$$G = \frac{1}{n}\sum_{i=1}^{n} w(z_i)|z_i|^{\kappa}, \tag{1}$$

where n is the number of samples and

$$z_i \;=\; x_i - (\delta\bar{x} + (1-\delta)p)). \tag{2}$$

δ and p are two tunable parameters. δ is constrained to the interval $[0,1]$, whereas p may, in principle, take any value. x_i is the i^{th} sample of the time series. The weight function $w(z_i)$ is defined by

$$w(z_i) = \frac{c_L}{1+\mathrm{e}^{-\alpha_\mathrm{L}(\mathrm{z_i}-\beta_\mathrm{L})}} + \frac{c_R}{1+\mathrm{e}^{-\alpha_\mathrm{R}(\mathrm{z_i}-\beta_\mathrm{R})}}. \tag{3}$$

As mentioned above, some indicators considered in the literature for driver sleepiness are special cases of this generic variability indicator. By setting the parameters appropriately, the following indicators of driver sleepiness (based on driving behavior, measured using the lateral position) can be obtained: the number of lane exits (abbreviated Lanex); the mean squared error from a given point p in the driving lane; the standard devation and, finally, the average of the lateral position (provided that the origin has been set such that the signal only contains positive values). For example, by setting $\beta_L = \beta_R = 0, \alpha_L = \alpha_R = 0$, $c_L = c_R = 1$, $\delta = 1$, $\kappa = 2$ the (square of the) standard deviation is obtained.

Perclos

The DSS mentioned above was used to record *Perclos* [23], which is a measure of slow eye-closures. Perclos is defined as the proportion of a given period of time (here set to 60 seconds) for which the eyes are at least 80% covered by the eyelids. This measure has been found to correlate well with driver sleepiness [6].

Sleep/Wake Predictor

In addition to the indicators described above, which are based on driving behavior signals and eyelid movement signals, a *model of sleepiness*, namely the *Sleep/wake Predictor* (SWP) [1], has also been considered as an indicator of driver sleepiness. The SWP takes as input the time of day and information on previous sleeping periods and, based on this information, outputs an estimated level of sleepiness. In this work, only time of day, the time since awakening and the length of the most recent sleeping period have been used as input to the model.

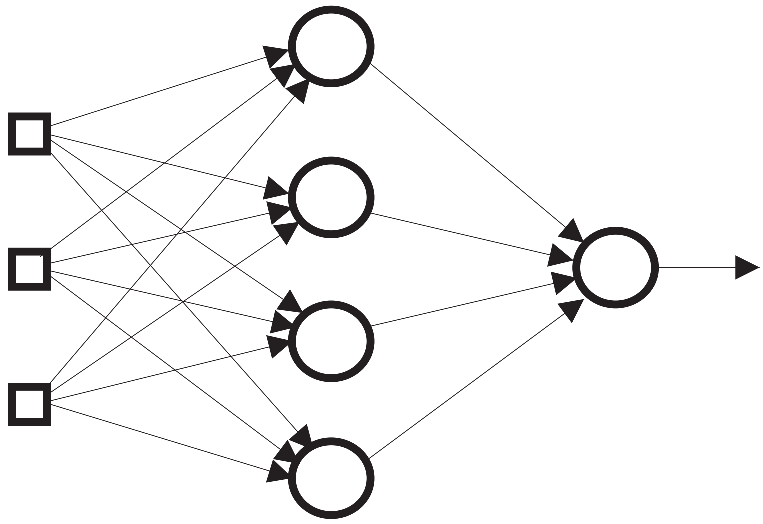

Figure 1: *An example of an artificial neural network with three input signals and one output signal. The neurons are shown as circles. The squares are input elements that simply transfer the input signals without carrying out any computation. The network weights are indicated by lines connecting (from left to right) input elements to the hidden layer of neurons that, in turn, is connected to the output layer (which, in this case, contains a single neuron). The biases are not shown.*

Combinations of indicators

A combination of several indicators may provide a better estimate of sleepiness than any individual indicator alone. Indicators can be combined in several ways, ranging from simple linear combinations of indicator values to much more complex functional forms. In order to capture as many functional forms as possible, artificial neural networks (ANNs) have been used here (as in earlier work; see [18, 19]).

An ANN consists of a number of elementary computational elements referred to as *neurons*. Each neuron takes inputs x_i and generates an output y according to

$$y = \sigma\left(\sum_{i=1}^{m} w_i x_i + b\right), \tag{4}$$

where w_i are the connection weights, b is the bias, and m is the number of inputs to the neuron. σ is a non-linear squashing function that, in addition, keeps the output values in a given range (e.g. $[0,1]$). One can show that an ANN with two layers of weights can, in principle, represent any continuous function [5]. A common special case is the two-layered feedforward neural network, an example of which is shown in Fig. 1. As can be seen in the figure, in addition to the inputs, there are two layers of neurons: a hidden layer and an output layer. The tunable parameters, which determine the computation carried out by the ANN, are the weights and biases. In this paper, particle swarm optimization (PSO) (see below) has been used for tuning the network parameters.

METHOD

Clearly, the performance of an indicator, i.e. its ability to measure driver sleepiness (or the absence thereof), will depend on the values of its parameters. Thus, in the work presented here, stochastic optimization methods have been used for finding the best parameter settings for each indicator, and for setting the weights of the ANNs used for combining several indicators to form a detection system.

Optimization algorithms

Two kinds of population-based stochastic optimization methods have been used, namely a genetic algorithm (GA) [9] and particle swarm optimization (PSO) [11]. Unlike several classical optimization methods, these algorithms do not rely on gradients and therefore are less likely to get stuck in local optima. The limited space available here does not permit a complete description of population-based stochastic optimization methods. Suffice it to say that such methods maintain a population of candidate solutions (referred to as *individuals*) to the problem in question. After evaluating all individuals, and thus assigning fitness values according to the objective function (described in detail in the next section), new individuals are formed by modifying the existing ones. The procedure for modifying individuals distinguishes GAs from PSO. In a GA, individuals are selected stochastically in proportion to the fitness values. The properties of the selected individuals are then combined in a procedure referred to as *crossover* that results in new individuals, which are then slightly modified (randomly), in a process referred to as *mutation*, before being inserted into the population. By contrast, in (standard) PSO, there is no selection step. Instead, the notion of velocity is introduced in the parameter space. After evaluation, new velocities are calculated (according to an equation that will not be given here) for each individual, and the new individuals are formed by moving the existing ones in parameter space, according to the computed velocities. For a more thorough description of stochastic optimization algorithms, see e.g. [22].

Regardless of the optimization algorithm used, one must be careful to avoid the problem of *overfitting*, resulting from excessive training on a data set of limited size. Real-world data sets are generally of limited size and invariably contain noise. Thus, if the optimization algorithm is allowed to run for too long, it will eventually start fitting the parameters to the noise, i.e. adapting the parameter settings to obtain excellent performance on the limited amount of data used during training, but also, gradually, much *worse* performance on previously unseen data. Of course, it is the performance on the latter kind of data that matters. There exists several methods for avoiding overfitting. In *holdout validation*, some of the data is placed in a separate validation set and, during optimization, the performance of the optimization algorithm is measured both over the training set and the validation set, but the algorithm is provided feedback only regarding its performance on the training set. The validation set, by contrast, is used for determining when to stop the training; the best parameter set is taken as the one giving best performance over the validation set. It is also common to maintain a third data set, the test set, which is not used until after the optimization has been completed, and therefore measures the performance over previously unseen data.

Optimization procedure

As described in the previous section, the data were divided into three separate data sets, namely a training set (used to guide the parameter search carried out by the optimization algorithm), a validation set (used for determining when to terminate the optimization), and a test set (used to obtain the performance on previously unseen data).

The data sets were generated as follows: as mentioned above, each TS estimated the level of sleepiness every five minutes. Thus, for each KSS estimate, five minutes of driving behavior data were available. The data corresponding to each KSS estimate were divided into five 60-second intervals, such that the interval $(t-60,t]$ was used in the test set, the interval $(t-$

$120, t-60]$ in the validation set, and the three intervals $(t-300, t-240]$, $(t-240, t-180]$ and $(t-180, t-120]$ in the training set.

Intervals during which the speed dropped below 60 km/h were discarded in order to remove episodes where a small town or a roundabout was passed. Intervals for which more than 20% of the samples had a confidence level below 20 were also removed. Furthermore, individual samples with confidence level below 20 were discarded as well[5]. Note that no samples were removed from the time series sampled from the steering wheel angle signal as this signal was always available and reliable. After this data reduction, the total number of retained samples equalled 1244, distributed in such a way that the training, validation and test sets contained 753, 247, and 244 elements, respectively. Class C_1 (alert) contained a total of 837 elements, and C_2, therefore, 407 elements.

As the indicators and detection systems considered here are real-valued, a threshold function was used to assign a 60-second driving interval to either of the two classes C_1 or C_2. Based on the value V obtained from an indicator or a detection system, the class C, for the corresponding interval, was assigned as

$$C = \begin{cases} C_1 & \text{if } V < T \\ C_2 & \text{otherwise} \end{cases} \tag{5}$$

where T is the threshold value separating the two classes.

The objective function F was computed as follows: first, the average f_i of sensitivity (i.e. the number of correctly classified samples in class C_2, divided by the number of samples in C_2) and specificity (i.e. the number of correctly classified samples in C_1 divided by the number of samples in C_1), was calculated for each test subject i, according to

$$f_i = \frac{1}{2}\left(\frac{n_{i,2}}{N_{i,2}} + \frac{n_{i,1}}{N_{i,1}}\right), \tag{6}$$

where $n_{i,j}$ denotes the number of correctly classified samples in class C_j (for test subject i) and $N_{i,j}$ denotes the total number of samples in class C_j (also for test subject i). Finally F was formed as the average of the f_i measures:

$$F = \frac{1}{k}\sum_{i=1}^{k} f_i, \tag{7}$$

where k denotes the number of test subjects.

The performance of an indicator (or a detection system) is dependent not only on its parameters; it also depends on the threshold value T used for classifying all the 60-second driving intervals. Whereas the parameters were optimized using stochastic optimization, the threshold value (for any given set of parameters) was set by exhaustive search such that the chosen value provided the best possible performance on the training set.

A more detailed description of the data reduction procedure can be found in [19] (in which the same objective function was used).

RESULTS

Using the data presented above, several indicators from the literature were evaluated. Next, following the procedure described above, a large number of optimization runs were carried out.

---

[5] Discarding individual samples was possible, since none of the indicators considered here depends on the availability of consecutive samples.

Indicator	Training	Validation	Test
1. Standard deviation of lateral position	0.53	0.52	0.54
2. Standard deviation of heading angle	0.56	0.57	0.53
3. Standard deviation of steering wheel angle	0.56	0.58	0.60
4. Large steering wheel reversals	0.53	0.58	0.61
5. Medium steering wheel reversals	0.58	0.60	0.60
6. Small steering wheel reversals	0.52	0.51	0.51
7. Optimized steering wheel reversals	0.64	0.64	0.64
8. GVI of lateral position	0.70	0.69	0.70
9. GVI of steering wheel angle	0.58	0.62	0.60
10. GVI of heading angle	0.58	0.59	0.55
11. GVI2 of lateral position	0.81	0.81	0.80
12. Perclos	0.50	0.50	0.50
13. SWP	0.85	0.85	0.84

Table 1: *The performance F of the individual indicators. Rows 1-6: indicators from the literature; Rows 7-11: optimized indicators (GVI = Generic variability indicator. GVI2 = GVI with different parameter settings for daytime and nighttime driving, respectively); Row 12: the Perclos blink indicator; Row 13: the Sleep/Wake Predictor.*

The results are summarized in Table 1. The table shows the performance F of each indicator over the training, validation, and test sets.

As can be seen, the performance of the indicators from the literature (rows 1 through 6) was not very impressive. The performance of the optimized indicators on rows 7 through 10 was a bit better in some cases, particularly the GVI applied to the lateral position.

Note that a random prediction would, on average, give a performance score of 0.50. In fact, the poor performance of the first indicator (the standard deviation of lateral position) was especially puzzling, since several studies [23, 21, 4], including one of our own [18, 19], have shown a better performance score than the measly 0.54 achieved here, for previously unseen test data. In fact, in our earlier study, for which the same performance measure was used, this indicator obtained a score of 0.71 on test data[6].

After careful analysis of the data, it was discovered that this simple measure is not suitable for data obtained from trucks. Presumably, this is so, since a (large) truck responds less rapidly to a movement of the steering wheel than a passenger car. As can be seen in the upper left panel of Fig. 2, there is no discernible trend in the standard deviation of lateral position as a function of the KSS value.

On the other hand, the standard deviation of the *steering wheel angle* reaches a score of 0.60, comparable to the values 0.55 and 0.63 obtained using the steering wheel angle and steering wheel velocity, respectively, in the study involving passenger cars [18, 19].

The data analysis provided yet another interesting piece of information; perhaps not so surprisingly, it turned out that the average lateral position, summarized in Table 2, varied significantly *both* with the driver's state (sleepy or alert) *and* with time-of-day, in such a way that a sleepy driver, at night, places the vehicle closer to the center of the road than an alert driver during day time. The difference in average lateral position between these two cases is signi-

[6] Note that, in our earlier study, samples corresponding to KSS = 7 were excluded. However, repeating the analysis from that study, with those samples included, a similar value (0.69) was obtained for the standard deviation of lateral position, i.e. still much higher than the value obtained in the present study.

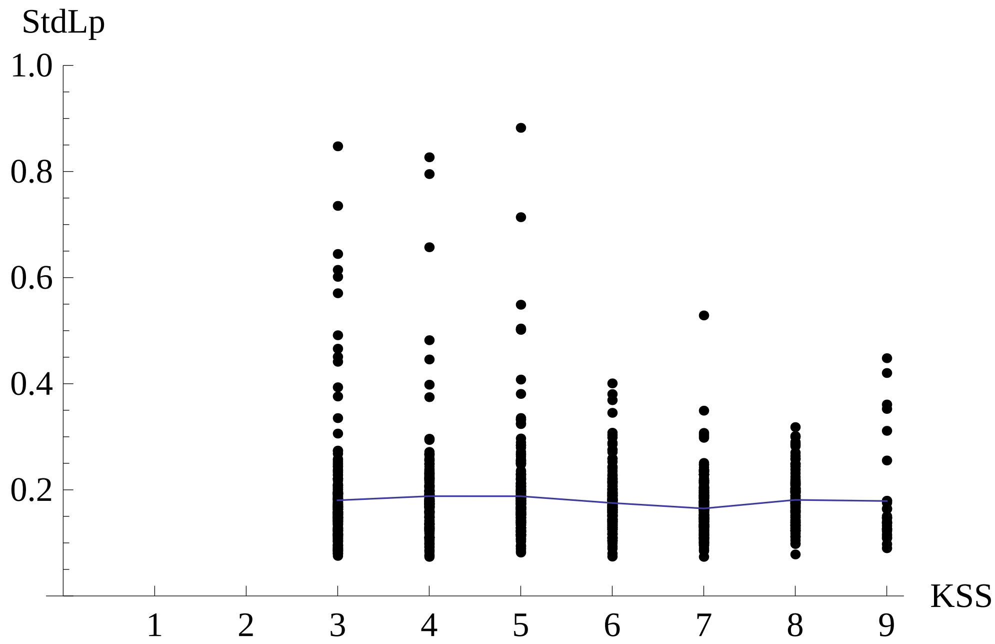

Figure 2: *The indicator values for the standard deviation of lateral position (StdLp), for all 1244 data samples. The solid line shows the mean of the indicator values for the various KSS values.*

	Day	Night
Alert	1.79	1.90
Sleepy	1.86	2.02

Table 2: *The drivers' average lateral position, shown in four different circumstances. The average lateral position is measured as the distance from the side of the road to the front center of the vehicle, such that the values become larger as the vehicle moves towards the center of the road. Note also that the entry for sleepy drivers during daytime (1.86) is based on only 41 samples, compared to more than 150 for the other three entries.*

ficant with a p-value of 0.05. For the other two combinations, namely alert driving at night and sleepy driving during day time, the drivers tend to place the vehicle at intermediate lateral positions.

This observation has important implications for some indicators, for example Lanex which depends on excursions from the lane and is therefore sensitive to the value of the average lateral position. Thus, since the GVI indicator *can* represent (generalized) Lanex (if the parameters are set appropriately) the GVI indicator may also be affected by differences in average lateral position.

As indicated in Table 2, using only the average lateral position, it is not possible to disentangle the effects of driver state (sleepy or alert) from the effects of the time-of-day. In other words, in order for an indicator to distinguish between a sleepy driver and an alert driver, the parameter settings may have to be different depending on whether the driving session takes place during the day or at night. Since the time-of-day easily can be measured, it is not difficult to apply two different parameter settings. As a complement to Table 2, Fig. 3 shows the values of the mean lateral position obtained for all of the 1244 elements in the three data sets (training, validation, and test), each element corresponding to a 60-second driving interval. As can be seen, there is a rising trend (albeit a weak one) in the mean lateral position for increasing KSS

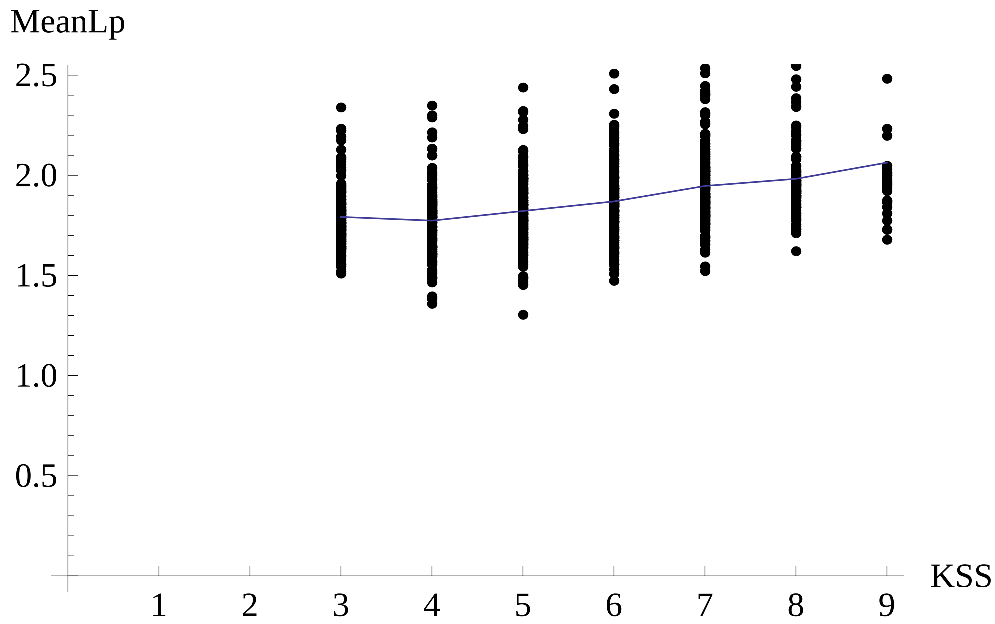

Figure 3: *The mean values of the lateral position (MeanLp) for all 1244 data samples, each corresponding to a 60-second driving interval. The solid line, in turn, shows the mean of the mean values obtained for each KSS value.*

Detection system	Training	Validation	Test
1. {GVI2Lp, ConfLp, SWARev, Perc, ConfPerc}	0.82	0.81	0.82
2. {GVI2Lp, ConfLp, SWARev, SWP}	0.86	0.86	0.85
3. {GVI2Lp, ConfLp, SWARev, Perc, ConfPerc, SWP}	0.88	0.87	0.88

Table 3: *The performance of the various optimized detection systems. GVI2Lp = the GVI2 indicator (with different parameter settings for day and night, respectively) applied to the lateral position signal, ConfLp = the confidence signal for the lateral position; SWARev = steering wheel angle reversals (optimized, corresponding to row 7 in Table 1); Perc = Perclos; ConfPerc = the confidence signal for Perclos; SWP = The Sleep/Wake predictor.*

values.

Returning to Table 1, row 11 shows the performance of the GVI2 indicator, which uses two different parameter settings. As can be seen, including information regarding the time-of-day increases the performance on unseen test data to 0.80.

By contrast, the overall performance of the Perclos indicator (row 12), is very poor indeed, reaching the same level as would a random prediction. However, as will be shown below, when combined with its corresponding confidence signal, Perclos can give a valuable (but small) positive contribution to a sleepiness detection system. Finally, the Sleep/Wake Predictor obtains a score of 0.84 on previously unseen test data. This is the best result obtained by any single indicator, outperforming the GVI2 indicator by 0.04 performance units.

The results obtained for detection systems (combining several indicators, using an ANN) are shown in Table 3. Note that some of the indicators used (e.g. GVI2Lp) have been optimized themselves, before being included in the detection system. The first system (row 1) combines indicators based on driving performance with the Perclos indicator, whereas the second system (row 2) combines indicators based on driving performance with the Sleep/Wake predictor. The final system (row 3) includes all four indicators. In all systems, the confidence signals are included, when available. As can be seen, the third system achieves the best result, outperforming the best single indicator (SWP) by 0.04 performance units on previously unseen test data.

DISCUSSION AND CONCLUSION

The main conclusion from this study is that, when applied to the problem of driver sleepiness classification, our optimization framework is capable both of (i) improving, through stochastic optimization, individual sleepiness indicators and (ii) generating and optimizing sleepiness detection systems, that combining several indicators non-linearly. For the data considered here, the best detection systems found outperform all individual indicators, reaching a top performance score of 0.88 on previously unseen test data.

However, one should keep in mind that the data used here contain certain fundamental limitations. For instance, the data are derived from driving sessions involving only eight test subjects, who all drove on the same road. Furthermore, due to legal requirements, the data do not include any sleep-deprived test subjects driving in daytime conditions. Nevertheless, there are, in fact, a few samples with high KSS values for daytime driving sessions, namely from drivers reporting a high level of sleepiness without being sleep-deprived.

When analyzing an indicator of driver sleepiness, it is important to make sure that the indicator considered actually measures the quantity that it is supposed to measure. For example, as mentioned above, it was found that the standard deviation of lateral position does not function very well as an indicator of sleepiness in the case of truck drivers, due to the slow response of a truck (relative to that of a passenger car) to steering wheel movements. However, when applying different thresholds for daytime and nighttime driving, respectively, the performance of this indicator shoots up to 0.70. Without taking into account the time-of-day effects, one may (incorrectly) conclude that the standard deviation of lateral position indicates sleepiness rather well, provided that two thresholds are used. However, a more careful analysis shows that the procedure for setting the threshold (between the sleepy and alert classes) simply generates an absurdly high threshold for daytime driving, placing *all* the corresponding data samples in the alert class, and an absurdly low threshold for nighttime driving, placing all those samples in the sleepy class. Thus, in this case, it turns out that the time-of-day, rather than the actual standard deviation, forms the basis for the sleepiness detection.

The best single indicator, i.e. the SWP, derives most of its success simply from distinguishing day from night (giving a score of 0.70), the remaning 0.14 performance units coming from the early stages of nighttime driving sessions, where the SWP correctly predicts that the driver is likely to be alert. The SWP output ranged from 3.76 to 3.79 for the daytime driving sessions, and from 6.52 to 7.50 for the nighttime sessions. The optimal threshold for SWP was found to be 6.8, for which all daytime intervals were assigned to class C_1, as were the first few intervals of the nighttime driving sessions. The remaining intervals for the nighttime sessions were all assigned to C_2.

In addition to the driving behavior data, physiological signals (namely EOG and EEG) were also measured during driving. However, as the analysis of the physiological data has yet to be completed, the KSS estimates, given by the test subjects, were used as the ground truth instead. However, this choice is well motivated, as KSS values have been shown to correlate well with physiological measures of sleepiness [2]. Furthermore, the use of KSS has allowed us to compare the results obtained in the simulator study [18, 19], for which KSS was used as well.

In the simulator study [18, 19], the class C_2 was based on driving periods for which the test subject estimated KSS to 8 or larger. By contrast, in the present study, driving periods with KSS = 7 were included as well. The reason for excluding KSS = 7 in the simulator study was to provide a clearer separation between the two classes. However, as many KSS estimates (especially during the night sessions) resulted in values of 7 in the present study, excluding the corresponding data would result in data sets with much fewer samples. Furthermore, as indicated in [2], physiological signs of sleepiness begin to show at KSS = 7, further motivating the inclusion of the corresponding data.

As a final point, one should note that the optimization and classification framework (and the corresponding software) developed in connection with this study (and earlier studies; see [18, 19]) is not limited to the study of driver sleepiness. In fact, the same framework can be applied to any classification task involving stochastic optimization. For example, the framework would be useful in a study of driver inattention based on, say, gaze direction data. This topic is a promising avenue for future research.

ACKNOWLEDGEMENT

The authors wish to thank the engineers and staff at Volvo Technology, Volvo Cars and the Swedish National Road and Transportation Research Institute (VTI) who carried out the study that provided the data used here.

References

[1] Åkerstedt, T., Connor, J., Gray, A. and Kecklund, G. (2008). Predicting road crashes from a mathematical model of alertness regulationThe Sleep/Wake Predictor, *Accident Analysis and Prevention*, vol. 40 , pp. 1480–1485.

[2] Åkerstedt, T. and Gillberg M. (1990). Subjective and objective sleepiness in the active individual, *International Journal of Neuroscience*, vol. 52, pp. 29–37.

[3] NTSB (1999). *Evaluation of us department of transportation efforts in the 1990s to address operator fatigue*, National Transportation Safety Board, Safety Report NTSB/SR-99/01.

[4] Brookhuis, K., De Waard, D. and Fairclough, S. (2003). Criteria for driver impairment, *Ergonomics*, vol. 46, pp. 433–445.

[5] Cybenko, G. V. (1989). Approximation by Superpositions of a Sigmoidal function, *Mathematics of Control, Signals and Systems*, vol. 2, pp. 303–314.

[6] Dinges, D. and Grace, R. (1998). *PERCLOS: A valid psychophysiological measure of alertness as assessed by psychomotor vigilance*, US Deptartment Transportation, Federal Highway Admin., Washington, DC, Tech. Rep. Publication No. FHWA-MCRT-98-006.

[7] Grace, R., Byrne, V., Bierman, D., Legrand, J., Gricourt, D., Davis, B., Staszewski, J. and Carnahan, B. (1998). A drowsy driver detection system for heavy vehicles. In Proceedings of the *17th Digital Avionics Systems Conference*.

[8] Hanowski, R., Bowman, D., Alden, A., Wierwille, W. and Carroll, R. (2008) *PERCLOS+: Moving Beyond Single-Metric Drowsiness Monitors*, SAE International.

[9] Holland, J. H. (1975). *Adaptation in natural and artificial systems*, University of Michigan Press.

[10] Horne, J. and Reyner L. (1995). Sleep related vehicle accidents, *British Medical Journal*, vol. 310 , pp. 565–567.

[11] Kennedy, J. and Eberhart, R. C. (1995). Particle swarm optimization, in Proceedings of the *IEEE International Conference on Neural Networks*, pp. 1942–1948.

[12] Klauer, S. G., Dingus, T. A., Neale, V. L. and Sudweeks, J. D. (2006). *The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data*, National Highway Traffic Safety Administration, DC, DOT HS, 810.

[13] Maycock, G. (1996). Sleepiness and driving: The experience of UK car drivers, *Journal of Sleep Research*, vol. 5, pp. 229–231.

[14] McLean, J. R. and Hoffman, E. R. (1971). Analysis of drivers control movements, *Human Factors*, vol. 17, pp. 407–418.

[15] Neale, V., Dingus, T., Klauer, S., Sudweeks, J. and Goodman, M. (2005). An overview of the 100-car naturalistic study and findings, in 19th *International Technical Conference on Enhanced Safety of Vehicles*, Washington, DC.

[16] Otmani, S., Pebayle, T., Roge, J. and Muzet, A. (2005). Effect of driving duration and partial sleep deprivation on subsequent alertness and performance of car drivers, *Physiological Behavior*, vol. 84, pp. 715–24.

[17] Reyner, L. and Horne, J. (1998). Falling asleep whilst driving: Are drivers aware of prior sleepiness?, *International Journal of Legal Medicine*, vol. 111, pp. 120–123.

[18] Sandberg, D. and Wahde, M. (2008). Particle swarm optimization of feedforward neural networks for the detection of drowsy driving, in *International Joint Conference on Neural Networks*, pp. 789–794.

[19] Sandberg, D. and Wahde, M (2009), Detecting driver sleepiness using optimized nonlinear combinations of sleepiness indicators, Submitted to *IEEE Transactions on Intelligent Transportation Systems*.

[20] Stutts, J., Wilkins, J., Osberg, J. and Vaughn, B. (2003). Driver risk factors for sleep-related crashes, *Accident Analysis & Prevention*, vol. 35, pp. 321–331.

[21] Tijerina, L., Gleckler, M., Stoltzfus, D., Johnston, S., Goodman, M. and Wierwille, W. W. (1999) *A Preliminary Assessment of Algorithms for Drowsy and Inattentive Driver Detection on the Road*, US Dept. of Transportation, National Highway Safety Administration.

[22] Wahde, M. (2008). *Biologically inspired optimization methods: An introduction*, WIT Press.

[23] Wierwille, W. W., Ellsworth, L. A., Wreggit, S. S., Fairbanks, R. J. and Kim, C. L. (1994). *Research on vehicle-based driver status/performance monitoring: Development, validation, and refinement of algorithms for detection of driver drowsiness*, National Highway Traffic Safety Administration Final Report: DOT HS, 808.