ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

Multi-User Blood Alcohol Content Estimation in a Realistic Simulator using Artificial Neural Networks and Support Vector Machines

ROBINEL Audrey & PUZENAT Didier
{arobinel, dpuzenat}@univ-ag.fr *

Laboratoire LAMIA, Université Antilles Guyane
Campus de Fouillole - Guadeloupe (France)

Abstract. We instrumented a realistic car simulator to extract low level data related to the driver's use of the vehicle controls. After proceeding these data, we generated features that were fed to a Multi-Layer Perceptron (MLP) and Support Vector Machines (SVM). Our goal was determine if the driver's Blood Alcohol Content (BAC) was over $0.4g.l^{-1}$ or not, and even estimate the BAC value. Our device process the vehicle's controls data and then outputs the user BAC. We discuss the results of the prototype using the MLP and SVM algorithms in both single-user and multi-user context for detection of drunk drivers and estimation of the BAC value. The prototype performed better with single user base than with multi-user, and provided comparable results with MLP and SVM. This paper corrects a small error in our previous publication in ESANN'12 [3].

1 Introduction

1.1 Problematic

Driving under influence affects the subject's behaviour, by impairing the required skills. It should thus be possible to detect and classify the users according to their behaviour. In [1], a generic method for behaviour analysis has been proposed, using Artificial Neural Networks (ANN). We applied this method to the Blood Alcohol Content (BAC) estimation problematic in [2], using a video game. Later, we used a realistic car simulator for the same task in ESANN'12 [3]. Those papers describe the methodology used in depth. In the current paper and when compared to ESANN'12, we increased the accuracy of the results with more examples (twice as many examples), and added new machine learning algorithms with Support Vector Machines (SVM) and Support Vector Regression (SVR). We also improved the MLP results accuracy with a more extensive topology search, and compared these new results of the ANN with those of the SVM and SVR. Furthermore, we studied the system's behaviour when having multiple users example base compared to single user base (more on MLP single user results in [4]). Section 1.2 presents a summary of the experimental methodology used to collect the examples and to create the learning base, and then details the machine learning algorithms used. In section 2.2 we will present some results obtained with a single-user base using ANN and SVR. In sections 2.3 and 2.4 we discuss the performances of the ANN in both classification and regression on two multi-user bases, and compare it with our

*This work has been funded by *ApportMédia* (www.apportmedia.fr), *la Région Guadeloupe* (www.cr-guadeloupe.fr), and *the European Social Fund* (ec.europa.eu/esf).

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

Fig. 1: The“realistic car simulator software”(left) have been used with a Logitech G27 and its “force feedback” steering wheel with pedals (centre), and was used to collect low level data related to the subjects behaviour in the simulation (right).

results with SVM and SVR on the same bases. We will then conclude on the multi-user aspect, and on the compared accuracies of the ANN and SVM/SVR.

1.2 Methodology summary

Multiple subjects drove in a realistic car simulator (“Stars AF 2011”, presented in [3], and provided by “ApportMédia”-fig 1-) with various Blood Alcohol Content (BAC). For each subject driving on the simulator software (we call that a “run”), we collected low level data on the use of the controls (steering wheel, pedals, etc). The BAC was measured with a consumer class breathalyzer. After some processing on the raw data, we generated “features” (e.g. using the position of the steering wheel, we generated the feature “average amplitude of steering wheel corrections”) that could be fed to the ANN, SVM, or SVR to estimate the BAC of the driver. We used the data collected from our runs to create a base with only one subject, then with multiple subjects. We selected some features with a methodology described in [4] (a complete description of the instrumentation and features creation methodology is presented in this reference), and generalized in classification and regression.

1.3 Machine Learning algorithms used

For our ANN, we used a classical Multi-Layer Perceptron (MLP) with back-propagation learning based on the FANN library [6]. It was used either in regression (real output) or binary classification (binary output). We developed a program to test large amounts of networks topologies to obtain the best result for each set of features: from 1 to 9 hidden layers, with 1 to 32 neurones each, with some hyper-parameters search. For Support Vector Machines, we used the libSVM library [9], which enabled us to do both classification (using a C-SVC SVM) or regression (using epsilon-SVR).We used the default kernel function (Radial Basis Function), and used the scripts provided (only slightly modified) with libSVM to perform a grid search of the optimal hyper-parameters for both C-SVC and epsilon-SVR: a grid of value for each hyper-parameter is tested against a grid of values for each other one.

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

2 Results

2.1 Estimation of the device performances

For the single user experiments, we used leave-one-out cross-validation in order to test the system due to the lower number of examples. For the multi-user bases, we used 4-Fold cross validation. For regression purposes, we had to introduce a maximal tolerated error, ε. For each value returned by the ANN or SVR, we compute the distance between this value and the expected value (absolute error). If it is below a fixed epsilon, we count a success. Otherwise, we count a failure. We then compute the success rate of the system in generalization. For classification, we count a success when the output matches the class of the example. Our two classes are: class 0 for sober subjects ($BAC < 0.4g.l^{-1}$) and class 1 for drunk subjects ($BAC \geq 0.4g.l^{-1}$). We used $0.4g.l^{-1}$ as a threshold in order to have a similar amount of examples in both classes. We counted True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). We used those to compute the "sensibility", which is in our case the ability to detect that a subject is drunk, and "specificity" which is the ability to detect that a subject is not drunk. We used the following the following formula :
$sensibility = \frac{TP}{TP+FN}$ and $specificity = \frac{TN}{TN+FP}$

2.2 Single user base

Our first subjects performed 28 "runs" (approximately 90 minutes of driving) in the simulator software. We used this subject to perform single-user experiments. More detailed results of the ANN in this setup are presented in [4]. Due to the fact that most of the examples had BAC values below $0.5g.l^{-1}$, we could not create an unbiased classification base, and then only present regression results, using the ANN and the SVR. Using 4 features (based on the steering wheel use, the position of wheels on the road, the wheels sleep angle, and on the lateral acceleration of the vehicle), the success rate of the MLP reaches 89% with $\varepsilon = 0.1$. We obtain the same results with and with 6 features and 8 features. When using epsilon-SVR, we obtain comparable but slightly lower results: a 82.14% success rate for 4 features, and the same with 6 or more features. However , the average absolute error is slightly lower with SVR ($\pm 0.061g.l^{-1}$) than with the MLP ($\pm 0.07g.l^{-1}$).

2.3 Balanced base

With promising results with a single user, we wanted to study the performances of our prototype with multiple users. However, it was not possible to collect 90 minutes of driving for each subject (as we did with the first one), but only 20 to 30 minutes, corresponding to 8-10 runs. In order to create an unbiased base, we kept only 10 runs from the first subject, and then added 8 to 10 for each new subject. When using the same features as in the previous base, in regression with the MLP we obtain 65% of estimations with an error lower than $\pm 0.2g^{l}-1$. The average error is much higher than in single user mode, but we must also consider the fact that this base contains many high BAC value (up to $1.06g.l^{-1}$) hence the higher error. However, even when considering

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

that the range of values doubled, and when doubling ε as well, the results are still lower. We could not improve those results with epsilon-SVR, the success rates and average absolute error being worse than with the MLP.

In classification, with the MLP, we obtain a 78% success rate, with a sensibility of 89% and a specificity of 64%. The system detects most of the drunk subjects, with very few undetected. However, the proportion of sober subjects that are incorrectly detected as drunk is superior. When using the SVM, we obtain similar results considering the success rate (75%). However, the sensibility reaches 100%, and the specificity 73%.

2.4 Unbalanced base with all available examples

In the end, we tested the network with all of the available examples.

By keeping the same inputs as in ESANN2012 [3], we reach a 66% success rate in regression with $\varepsilon = 0.2$ and an average absolute error of $\pm 0.19g/l-1$. Those results can not be compared with [3], as there was an error in the success rate computing algorithm. Increasing the count of used features did not improve the results, with a 65% and an average absolute error of $\pm 0.193g.l^{-1}$ With the SVM, we could improve the results, reaching 68% an average error of 0.187112, but with 8 features.

In classification with the MLP we obtain 70% for the success rate, 82% for sensibility and 58% for the specificity. The SVM reaches a 70% success rate, with 75% sensibility and 66% specificity. Increasing the amount of features used did not significantly improve the performances neither for the MLP nor for the SVM. In this configuration we obtain lower results in classification than with the homogenous base, which was expected. Furthermore, the bias in the base causes a significant drop in sensibility.

3 Conclusions

3.1 Single user regression

With a single user, the prototype performs quite well in regression. Considering that only 90 minutes of driving were required for this result, it does not seem unfeasible to embed such a system in a real car. The subject would have to drive for less than two hours to train the system. However, it would be complicated to drive under the influence of alcohol (it would have to be done on a circuit, with a sober driver and duplicated controls).

3.2 Multi-User regression

The ideal case would be to obtain a device already trained for generic users. In that case, with a balanced base, we obtain lower results in both regression and classification. However, when using all the examples available, we obtain higher results, despite the biased base. We probably need more examples to construct a base big enough to provide higher results in regression. The used breathalyzer was of consumer class, and proved to provide noisy measures degrading the overall accuracy of the system. In order to circumvent those problems we would require law enforcement class breathalyzer (or

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

even blood sample analysis) and more subjects than we could afford with our limited funds.

3.3 Multi-User binary classification

Considering binary classification, we obtain much higher sensibility than specificity. It would be interesting to improve the specificity, but the high sensibility is important, as very few drunk subjects remain undetected. If a sober user is detected as drunk, there are no consequences. On the contrary, a drunk person that is detected as drunk is more likely to cause an accident. However reducing false alarms is important to keep the users confidence in the system: someone that is often detected drunk when he is sober may think that it is a false alarm when it really is the case.

3.4 On the use of Support Vector Machines

In this experiment, we also used SVM and SVR to perform the tasks devoted to the MLP. In most cases, we obtained similar results, often slightly lower, some times slightly better. In the end, for this problematic, using SVM did not bring significant improvement over ANN. However, the SVM have proven to be able to provide satisfying results, and could be used as an alternative to ANN, or even in combination. In both cases, feature selection remains crucial. Hyper parameters optimization is also a necessity in both cases. Overall, those machine learning algorithms had similar constraints and have perform at a comparable level in our context.

3.5 Global conclusions

We have demonstrated the ability of our device to detect "drunk" drivers and to perform blood alcohol content estimation, and thus reached our goal. In classification, we reached a much higher "drunk" detection than "sober" detection. It is now possible to improve the results using the same methodology but with more means.

The system was developed to be generic, and we should be able to use it for other problematic easily. When the determination of the state of the subject requires complex and/or invasive measurement, our method can be useful: the costs of the system would have to be spent once, for creating the learning base. A production device would be nearly invisible for the end user, and have a really marginal cost. Of course, the tradeoff is that more time must be spent in the development phase to ensure the quality of the base, the quality of the measures, and the selection of the most efficient combination of features.

4 Perspectives

Our next goal will be to proceed with experimentation on other problematic using the same software, but in the hardware simulator (featuring a realistic car cockpit and controls, and triple screens for panoramic vision). Instrumenting will be done on the same basis, but with more available data (such as gear ratio and pedal, etc.). The use of the simulator hardware should provide a driving experience closer to real cars, and enable

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

us to collect more accurate data. We already have begun working on this hardware version of the simulator that has been provided by "ApportMédia".

We are also considering the use of data related to events rather than the average behaviour of the subject, like variation of parameters when specific events occur (e.g. an accident, a dangerous situation, a change of the driving conditions,etc.). In the long term, we are also looking forward to conduct similar experiments into real cars or trucks, during an upcoming partnership with "ApportMédia" and "Ediser".

We will later be experimenting on combination of multiple classifiers, in order improve the weak points of our system.

References

[1] D. Puzenat and I. Verlut, Behavior analysis through games using artificial neural networks. In *proceedings of the Third International Conferences on Advances in Computer-Human Interactions* (ACHI 2010), pages 134-138, Sint Maarten (Netherlands).

[2] A. Robinel and D. Puzenat, Real time drunkenness analysis through games using artificial neural networks. In *proceedings of the Fourth International Conferences on Advances in Computer-Human Interactions* (ACHI 2011), pages 206-211, Gosier (France).

[3] A. Robinel and D. Puzenat, Real Time Drunkenness Analysis in a Realistic Car Simulation. In *ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0 pages 85-90.

[4] A. Robinel and D. Puzenat, . In *proceedings of the Sixth International Conferences on Advances in Computer-Human Interactions* (ACHI 2013), to be published, Nice (France).

[5] ApportMédia, "StarsAF2011" realistic car simulator product online documentation `www.simucar.com/logiciels/STARS_AF.html`.

[6] S. Nissen, Implementation of a Fast Artificial Neural Network library (FANN). Technical report, Department of Computer Science University of Copenhagen, October 2003.

[7] C. M. Bishop, Neural Networks for Pattern Recognition. Clarendon Press, Oxford, 1995.

[8] E. Bogen, Drunkenness, a quantitative study of acute alcoholic intoxication. California and Western Medecine, XXVI(6) : 778-783, 1927.

[9] Chang, Chih-Chung and Lin, Chih-Jen,LIBSVM: A library for support vector machines, in ACM Transactions on Intelligent Systems and Technology,volume 2, issue 3, pages 27:1–27:27,2011.