Real-time detection of driver distraction: random projections for pseudo-inversion-based neural training

DOI:10.1007/s10115-019-01339-0

(Article begins on next page)

Knowledge and Information Systems

Real-time Detection of Driver Distraction: Random Projections for Pseudoinversion-based Neural Training

--Manuscript Draft--

Manuscript Number:	KAIS-D-17-00423
Full Title:	Real-time Detection of Driver Distraction: Random Projections for Pseudoinversion-based Neural Training
Article Type:	Regular Paper
Keywords:	Random projections; Pseudoinverse matrix; Genetic algorithms; Driver's Distraction Recognition.
Corresponding Author:	Rossella Cancelliere Universita degli Studi di Torino Torino, ITALY
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Universita degli Studi di Torino
Corresponding Author's Secondary Institution:	
First Author:	Marco Botta
First Author Secondary Information:	
Order of Authors:	Marco Botta
	Rossella Cancelliere
	Leo Ghignone
	Fabio Tango
	Patrick Gallinari
	Clara Luison
Order of Authors Secondary Information:	
Funding Information:	Artemis Joint Undertaking (332933) — Prof. Marco Botta
Abstract:	There is an accumulating evidence that distracted driving is a leading cause of vehicle crashes and accidents. In order to support safe driving, numerous methods of detecting distraction have been proposed, which are empirically focused on certain driving contexts and gaze behavior. This paper aims at illustrating a method for the non-intrusive and real-time detection of visual distraction based on vehicle dynamics data and environmental data, without using eye-tracker information. Experiments are carried out in the context of the automotive domain of the European project Holides, which addresses development and qualification of Adaptive Cooperative Human-Machine Systems, and is co-funded by ARTEMIS Joint Undertaking and Italian University, Educational and Research Department. The collected data are analised by a single layer feedforward neural network trained through pseudo-inversion methods, characterized by direct determination of output weights given randomly set input weights and biases. One main feature of our work is the convenient setting of input weights by the so called random projections: the presence of a great number of null elements in the involved matrices makes especially parsimonious the use at run time of the trained network. Moreover, we use a genetic approach to better explore the input weights network space. The obtained results show both better performance with respect to classical methods and effective and parsimonious use of memory resources.

Knowledge and Information Systems manuscript No.
(will be inserted by the editor)

Real-time Detection of Driver Distraction: Random Projections for Pseudoinversion-based Neural Training.

Marco Botta[1] · Rossella Cancelliere[1] · Leo Ghignone[1] · Fabio Tango[2] · Patrick Gallinari[3] · Clara Luison[1]

Received: date / Accepted: date

Abstract There is an accumulating evidence that distracted driving is a leading cause of vehicle crashes and accidents. In order to support safe driving, numerous methods of detecting distraction have been proposed, which are empirically focused on certain driving contexts and gaze behavior. This paper aims at illustrating a method for the non-intrusive and real-time detection of visual distraction based on vehicle dynamics data and environmental data, without using eye-tracker information. Experiments are carried out in the context of the automotive domain of the European project Holides, which addresses development and qualification of Adaptive Cooperative Human-Machine Systems, and is co-funded by ARTEMIS Joint Undertaking and Italian University, Educational and Research Department. The collected data are analised by a single layer feedforward neural network trained through pseudo-inversion methods, characterized by direct determination of output weights given randomly set input weights and biases. One main feature of our work is the convenient setting of input weights by the so called *sparse* random projections: the presence of a great number of null elements in the involved matrices makes especially parsimonious the use at run time of the trained network. Moreover, we use a genetic approach to better explore the input weights network space. The obtained results show both better performance with respect to classical methods and effective and parsimonious use of memory resources.

Keywords Random projections · Pseudoinverse matrix · Genetic algorithms · Drivers Distraction Recognition

This work was supported by the EU Artemis Joint Undertaking research project HoliDes, grant no.332933. HoliDes addresses development and qualification of adaptive cooperative human-machine systems (AdCoS) where many humans and many machines act together, cooperatively, in a highly adaptive way to guarantee fluent and cooperative task achievement. http://www.holides.eu

Rossella Cancelliere
Tel.: +39-011-6706737
E-mail: cancelli@di.unito.it
[1]Department of Computer Sciences, University of Turin, Turin, Italy
[2]Department E/E Systems, Centro Ricerche Fiat (CRF), Turin, Italy
[3]Laboratory of Computer Sciences LIP6, University Pierre et Marie Curie, Paris, France

1 Introduction

The "distracted driving" is a big issue that draws attention from the public, policymakers and researchers. Driver distraction is, according to [1], the diversion of attention away from activities critical for safe driving towards a competing activity. A vast variety of activities performed inside the vehicle can become potential distraction, including operating in navigation and entertainment systems [2], whose rapid development can contribute more and more to distracted driving, as a threat to road safety.

In this paper we address these topics and in particular we started by an initial research question: "Is it possible to reliably detect and recognize driver state, so that a supporting system would give just as much assistance as the driver needs"? For instance, the intervention of a forward collision warning system or of an emergency braking system can be triggered, based on the driver state: if distraction is detected the function strategies can be adjusted accordingly (e.g. braking is modulated differently or warning signals are anticipated). Such a smart assistance, which recognizes driver state, would allow for a greater safety margin, without irritating the driver with false alarms or inappropriate interventions in normal driving conditions and so enhancing the user acceptability. Therefore, in recent years, several methods have been published, which aims at estimating distracted driving (for example [3–6]) or also which concentrate on the detection and modeling of fatigue or stress as fundamental causes for driver inattention (like [7,8]).

Given this context, one goal of this paper is to present a non-intrusive approach for a real-time system to detect and classify distracted driving, using only vehicle dynamic data and environmental data (e.g. the other traffic agents) as inputs to the model. This is done in the context of the automotive domain of the European project Holides, which addresses development and qualification of Adaptive Cooperative Human-Machine Systems, and is co-funded by ARTEMIS Joint Undertaking and Italian University, Educational and Research Department.

In particular, here we mainly address the driver visual distraction that has been considered an important aspect in the investigated maneuvers. In this context, looking away for a short while (at least 1.8 seconds) can be considered as a driver visual distraction from her/his main activity. Machine Learning (ML) and in particular Neural Networks (NN) technologies may be able to provide the right algorithms to cope with such a challenge. This technology can be applied to build a discrimination model that captures the differences in behavior when people drive normally and when they are distracted.

In the past two decades, single hidden layer feedforward neural networks (SLFNs) have been one of the most important subject of study and discussion among neural researchers [9–11]. Methods based on gradient descent are mainly used for training, and among them the large family of techniques based on backpropagation (BP), widely studied in its variations. The start-up of these techniques assigns random values to the weights connecting input, hidden and output nodes, and these weights are then iteratively adjusted, so originating learning methods typically slow.

Some non iterative procedures have been proposed in litarature as learning algorithms for SLFNs based on random assignation of input weights and output weights evaluation through the use of generalized inverse (or pseudo-inverse) matrices: in the framework of radial basis function neural networks (RBF) the idea that random selection of hidden neurons from the domain of a data space is sufficient to allow universal approximation was proposed by Broomhead and Lowe in their classic 1988 paper [12], and later extended by Lowe [13] (also see textbooks by Haykin [14] and Ham and Kostanic [15]). Pao et al. (see e.g., [16–18]) proposed the random vector functional-link (RVFL) network as a special case of their more general functional-link neural networks (FLN), described in Pao's popular textbook [19]: in this model too, hidden neuron parameters are randomly selected and only the weights of the output layer need to be trained (e.g. with pseudoinverse or gradient descent procedures). Many theorethical and application-oriented studies in the last years were devoted to the use of these single-pass techniques, easy to implement and computationally fast; among the most known we recall the ELM method [21] while others important works are e.g. [20, 22–26].

The second goal of this paper is to use, in this context, random projections for input weight setting. The theoretical rationale for this approach can be found in many studies, showing random projections as a powerful method for dimensionality treatment [27–29]; in particular we analyse the so called *sparse* random projections, characterised by a majority of null elements, because they appear to be an useful tool for making more parsimonious and effective the run-time data processing. Moreover, we also explore the possibility to replace the completely random sampling of input weights space, typical of pseudoinversion-based methods, by a search procedure driven by genetic algorithm techniques. This approach has already been exploited in several papers ([30–33]), but the concomitant use of random projections, in our opinion makes our proposed method original and more effective.

The paper is organized as follow: we first look into the main ideas on distracted driving and randomized algorithms for training neural networks in sections 2 and 3.

In section 4 the basics of random projections approach, as well as some requirements on their application, are summarised; in section 5 we explain how genetic algorithms are used for input weights setting.

Section 6 presents the experiments carried out to collect data and finally in section 7 the performance of the proposed training method is assessed through computational experiences on the collected data.

2 The problem of distracted driving

Advanced Driver Assistance Systems (ADAS) and Autonomous Driving (AD) cars have attracted significant attention: the research on driver distraction started back in early 1990s, when distractions caused by cell phones were found to significantly affect driver capability of responding to critical situations [34]. It is likely that the problem will increase as more wireless or mobile technologies find their way into vehicles [35], [36], [37]. Although in the last few years many European countries have prohibited the use of - for example - mobile phones when driving, nonetheless it should not be expected that the amount of driving distraction will necessarily decrease.

In addition, the 100-Car Naturalistic Driving Study found that almost 80% of all crashes and 65% of all near-crashes involved driver distraction [35]. In fact, a large body of both empirical studies [38] and naturalistic field research [39], [40] have further established the relationship between driver distraction and degraded driving performance. With the rapid development of many on-board systems (more or less integrated in the vehicle, but able to interact with drivers) distracted driving is likely to continuously grow as a threat to road safety. Many efforts have been made to tackle the problem of distracted driving in different ways, such as raising the public awareness and enhancing the government legislation: another active and promising approach is to develop real-time driver distraction countermeasures, developing specific classifiers able to identify distraction when occurs and so mitigate its effect by modifying accordingly the strategies of the system or even by providing a direct feedback to the driver. In this context, early and accurate detection of driver distraction is essential for successful real-time countermeasures.

However, in literature, there is not a unique and commonly agreed definition of distraction, but several ones very often overlapped and mixed with inattention or with other drivers states, such as drowsiness and workload. For what concerns the definition of distraction adopted in our research, we have considered the taxonomy proposed by Regan et al. [41] and by Lee et al. [1]. In particular, we start from the following definition: driver distraction is the diversion of attention away from activities critical for safe driving toward a competing activity and it is a significant risk factor that can cause accidents [42]

This has been extended by Regan et al. [41], adding the concept of driver inattention, which means insufficient or no attention to critical activities for safe driving toward a competing activity. As mentioned, note that distraction differs from fatigue [43], which is defined as a state of exhaustion that makes a person unable to continue an activity [44].

Although existing data is inadequate and not representative of the driving population, it is estimated that drivers engage in potentially distracting secondary tasks approximately 30% of the time their vehicles are in motion (conversation with passengers is the most frequent secondary task followed by eating, smoking, manipulating controls, reaching inside the vehicle, and using cell phone).

Numerous researchers have developed driver distraction monitoring systems that aim at promoting safe driving by considering different types and levels of distraction [43]. These methods can be divided into the following five categories [45] according to the factors being measured:

- subjective report measures
- driver biological measures
- driving performance measures
- driver physical measures
- hybrid measures

In this paper we focus on the third point: steering, braking, and other related driving behaviors are suitable for detecting visual distraction [46]. In addition, we employed physical measures of drivers; in controlled settings, such as a driving simulator, the detection response tasks such as DRT or SURT(SUrrugate visual-Research Task) are promising methods of measuring visual (and cognitive) distraction [47], [48], [49], [50]. In the SURT approach, used in this research, subject is required to respond, via a tactile device, to visual stimuli. Since we have considered the same category of visual distraction also used by Lee et al.[1], this secondary task seems to be very appropriate, because, to accomplish it, the subject needs to diverge the visual attention away from the road.

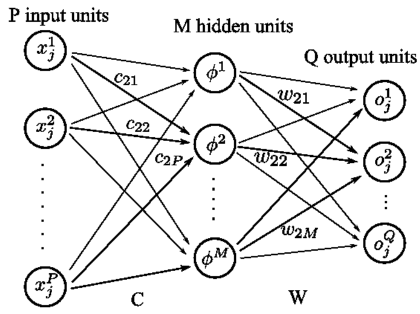

Fig. 1 A Single Layer Feedforward Neural Network.

3 Neural model and pseudo-inversion based training

In this section we introduce notation and we recall basic idea concerning the use of pseudo-inversion for neural training.

Fig. 1 shows a standard SLFN with P input neurons, M hidden neurons and Q output neurons, non-linear activation functions ϕ in the hidden layer and linear activation functions in the output layer.

Considering a dataset of N distinct training samples of (input, output) pairs $(\mathbf{x}_j, \mathbf{t}_j)$, where $\mathbf{x}_j \in \mathbb{R}^P$ and $\mathbf{t}_j \in \mathbb{R}^Q$, the learning process for a SLFN aims at producing the matrix of desired outputs $T \in \mathbb{R}^{N \times Q}$ when the matrix of all input instances $X \in \mathbb{R}^{N \times P}$ is presented as input.

As stated in the introduction, in the state of the art pseudoinverse approach input weights c_{ij} (and hidden neurons biases) are randomly sampled from a uniform distribution in a fixed interval and no longer modified.

After having determined the input weights matrix C, the use of linear output units allows to determine output weights w_{ij} as the solution of the linear system $H\,W = T$, where $H \in \mathbb{R}^{N \times M}$ is the hidden layer output matrix of the neural network, $H = \Phi(X\,C)$.

Regularisation methods have to be used [51,52] to turn the original problem into a well-posed one, i.e. roughly speaking into a problem insensitive to small changes in initial conditions. Among them, Tikhonov regularisation is one of the most common [53,54]: it minimises the error functional

$$E \equiv E_D + E_R = ||HW - T||_2^2 + \lambda ||W||_2^2 \tag{1}$$

where λ is known as the Tikhonov regularisation parameter.

The regularised solution $\hat{W}$ that minimises the error functional (1) has the form (see e.g. [55]):

$$\hat{W} = (H^T H + \lambda I)^{-1} H^T T. \tag{2}$$

If $\lambda = 0$ we have $\hat{W} = (H^T H)^{-1} H^T T \equiv H^+ T$, where H^+ is the Moore-Penrose generalised inverse (or pseudoinverse) of the matrix H.

With regularisation we introduce a penalty term that not only improves on stability, but also contains model complexity avoiding overfitting, as largely discussed in [56].

A suitable value for the Tikhonov parameter λ has therefore to derive from a compromise between having it sufficiently large to control ill-posedness while avoiding an excess of the penalty term in eq.(1): its value is therefore crucial.

The method OCReP ([57]) provides an optimal value for λ by defining a convenient regularized matricial formulation, in which the regularization parameter is derived under the constraint of condition number minimization. This value is proven to be very effective in terms of stabilisation and generalisation, as evidenced by comparison with implementations of other approaches from the literature, including cross-validation, so that we use this technique in the following.

4 Main ideas concerning random projections and input weight setting

If $X_{N\times P}$ is the original set of N P-dimensional observations,

$$X^{RP}_{N\times K} = X_{N\times P} C_{P\times K} \tag{3}$$

is the projection of the data onto a new K-dimensional space.

Strictly speaking, a linear mapping such as (3) is not a projection because C is generally not orthogonal and it can cause significant distortions in the data set. However, and unfortunately, orthogonalizing C is computationally expensive. Instead, we can rely on a result presented by Hecht-Nielsen [58]: in a high-dimensional space, there exists a much larger number of *almost orthogonal* than strictly orthogonal directions. Besides, Bingham and Mannila [59] performed an extensive experimentation which allows them to claim that vectors having random directions might be sufficiently close to orthogonality and equivalently that C^TC would approximate an identity matrix. They estimate the mean squared difference between C^TC and the identity matrix is about $1/K$ per element.

This key idea is confirmed also by the Johnson-Lindenstrauss lemma [60]: if a set of points in a vector space is randomly projected onto a selected space of suitable dimension, then the original distances between the points are approximately preserved in the new space, with only minimal distortions. For a simple proof of this result, see [61]. This property appears to be really appealing because suggests the possibility to preserve the topological structure of the initial input space while allowing the creation of a new optimal data representation in the hidden layer space, able to easy the classification/diagnosis task and to increase performance.

Therefore we can use random projections to project the original P-dimensional data into a K-dimensional space, using a random entries matrix $C_{K\times P}$ whose columns have unit norm.

Besides, random projection is very simple from a computational standpoint: the process of forming the random matrix C and projecting the data matrix X into K dimensions has complexity of order $\mathrm{O}(PKN)$; moreover, if the data matrix X is sparse with about G nonzero entries per column, the complexity is of order $\mathrm{O}(GKN)$.

Actually, a large variety of zero mean, unit variance distributions of elements c_{ij} result in a mapping that still satisfies the Johnson-Lindenstrauss lemma: among

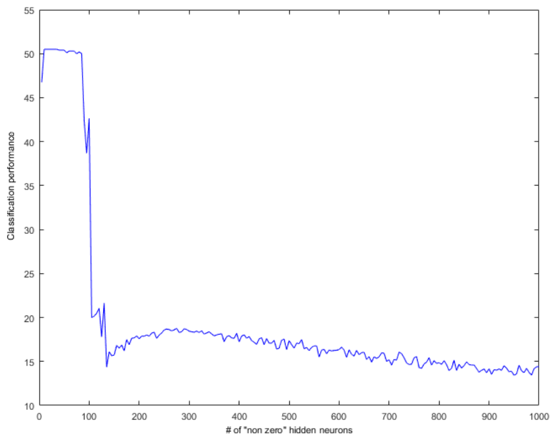

Fig. 2 Performance vs. sparsity degree

them, entries of C can be randomly sampled from a gaussian distribution. Another appealing possibility is using sparse random projections which have only a small fraction of nonzero elements. For example, Achlioptas [62] shows that generating random entries c_{ij} by

$$c_{ij} = \sqrt{3} \cdot \begin{cases} +1 & \text{with probability } 1/6 \\ 0 & \text{with probability } 2/3 \\ -1 & \text{with probability } 1/6 \end{cases} \tag{4}$$

one obtains a valid random projection with (expected) density 33%.

Because of these features, we decided to verify if sparse random projections matrices C can be used as suitable input weights values.

A difficulty arises because random projections are mainly used for linearly separable tasks although many real world problems are not linearly separable. Neural networks feature among the tools available to deal with the latter class of problems, so we propose to join these techniques using random projections matrices for the setting of input weights while the subsequent processing by hidden nodes nonlinear activation function will account for the non-linearity of the problem.

Some preliminary tests showed that the same classification performance obtained initializing a full random projection matrix could be reached by networks in which this matrix has a high level of sparsity (see Fig.2). We selected therefore input weight matrices having 90% of their values set to zero, and the others chosen from a normal distribution with variance $\sigma = \sqrt{(10)}$. So doing, we obtain 0-mean and 1-variance matrix columns, which satisfy the conditions required to have a random projection matrix.

5 Genetic pseudoinversion-based approach

As already discussed, in our approach input weights are random values: aiming at improving our model we decided to experiment a genetic algorithm strategy. In this framework, a starting population of random input weights matrices is generated and the network is then trained according to the method described in section 3. So doing an iterative process can start: in each iteration a new population (generation) of input weights is produced by applying selection, crossover and mutation on the previous generation, while the corresponding output weights are always computed by regularised pseudoinversion.

5.1 Selection

In the selection phase, a number of networks are chosen from the population in order to build the new generation: the chance each network has of being chosen depends on the error it had on the training set. A *fitness* score is calculated for each network by a linear scaling of the errors, they are then normalized to turn them into probabilities (their sum is made equal to 1) and a process of extraction with replacement is used to generate a new set of networks with the same magnitude as the starting population, and probably some repeated elements.

5.2 Crossover

Crossover is a procedure that generates new networks by combining previous ones. The networks selected in the previous phase are randomly paired with each other, then each couple generates a new couple of offsprings which will be part of the next generation. The crossover algorithm that gave the best results was the uniform crossover, done at neuron level: in this case each neuron of the hidden layer of the first offspring has a 50% chance of being inherited from each of its parents, independently from other neurons. The second offspring receives all the neurons wich haven't been chosen for the first one, producing in this way a recombination of the two parents without losing any "genetic material".

5.3 Mutation

Since selection and crossover work by recombining the original parameters, mutation is needed in order to explore other regions of the parameter space. After producing the new generation, each neuron of each network will have a chance to undergo a process of mutation: if the neuron is mutated, all its weights will be recomputed as if it was generated for a new random network. The probability of mutation is set to a low value in order to have on average less than one neuron mutated per network. A particular case happens when the two parents chosen for crossover are identical: in this case each of the two offsprings are forced to mutate at least one neuron, in order to avoid identical individuals in the new population.

6 Experimental Setup

The data related to distraction, the environment and the vehicle dynamic have been collected by means of dedicated experiments using a *real-prototype vehicle*.

In this section we describe the experiments carried out to collect these data as well as the database and the signals we used.

6.1 Subjects

Data from the system dynamics during driving have been collected from dedicated session performed on an equipped vehicle provided by CRF. Twenty-nine (29) test subjects drove for about 1 hour (i.e. 55km) on extra-urban and motorway roads. A minimum amount of driver experience was required, in particular i) at least 2 years of driving license and ii) at least 6000 km driven per year. Gender has been controlled (8 female and 22 male).

6.2 Procedure

Participants were asked to drive on the dedicated test-site in real-traffic situations, while completing a secondary task session. Studying driver distraction and designing assistance systems to counteract requires research tools that can be utilized to reliably induce distraction in simulator and real traffic studies. Distraction (visual and manual) has been induced by means of a secondary visual research task called SuRT, reproduced on an in-vehicle touch screen (7 TFT touch screen installed on the right-hand side of the car cabin). The SuRT is specified in the ISO/TS 14198 [63]. SURT was chosen with the aim at evaluating the interferences caused by a generic visual search task rather than a specific IVIS (In Vehicle Information System), which can be simulated in such a way. Like most commercial In-Vehicle Information Systems, it requires visual perception and manual response: such activities, according to Wickens multiple resources model [64], requires the same mental resources of the driving task and is therefore more likely to interfere, possibly causing a degradation of driving task performances.

In more details, SuRT is a distracting secondary task consisting of visually and manually demanding parts; participants are presented with a set of stimuli on a touch screen (e.g. a tablet or a smart phone) which can be mounted on the right side of the steering wheel in reach of the drivers right arm. The interested reader can find more details about SuRT in [63].

The time interval between two consecutive screens was pseudo-randomized between 3 and 9 seconds. The output data are the reaction times and the error rates.

6.3 Data Collection and preprocessing

In our work we have combined the eyes-position of subjects as derived by an internal camera with the presence of distractor tool, when the user was engaged in the secondary task (i.e. the SuRT). We adopted a supervised learning approach so that targets for training the classifiers has been defined in the following way:

– When subjects eyes were out of the road for 2s and secondary task was active, the distraction label was set to 1 (**driver is distracted**)

– When subjects eyes were on the road and secondary task was not active, the distraction label was set to 0 (**driver is not distracted**)
– In all the other situations, data have not been considered, since it was not possible to define the cognitive state of the driver in a sufficiently precise way for training purposes

For what concerns the vehicle dynamic data and the environment, the following variables have been collected:

– Speed [m/s]
– Steering Angle [deg]
– Lateral Position [m]
– Yaw Rate [deg/s]
– Lane Width [m]
– Road Curvature [1/m]
– Heading Angle [deg]
– Position of the accelerator pedal [%]
– Use of the brake pedal[1]
– X,Y coordinates of car in front [m] (if any)
– Speed of car in front [m/s] (if any)

These values are directly available on the prototype vehicle CAN bus or can be derived from those (e.g., time to collision is computed from vehicle speed and car in front data). The frequency of data collection was 20 Hz (1 data-point each 0.05s), which is the output rate of the vehicle bus.

It is worth to note here that only these variables constitute the inputs of the classifiers: the eye-movements data (from the internal camera) do not appear, since they have been used only to construct the target set.

Finally, for each numerical parameter in the list, we computed the mean and standard deviation, the gradient of variation between first and last value in a window of size 2s, as a method to group (summarize) the data and reduce noise.

Following the ordinary procedure for supervised learning, we use 10-fold cross-validation on 90% of the instances, while 10% of each dataset has been kept as a separate test set, on which the learned models have been tested.

7 Experimental results

Given the current state of the art and with reference to our previous works (see [5] and [65]]) we decided to compare a sparse standard pseudoinversion approach with our pseudoinversion-based sparse genetic method. We remark here that the former differs from the well known ELM method because no random projections (or sparse settings of weights) are used in ELM original technique.

All experiments are performed using GNU Octave, version 4.2.1 on a Unix workstation.

In our model the input layer is composed by as many nodes as there are variables describing a state (62), while the hidden layer size is determined through a procedure of 10-fold cross validation. In particular we chose 750 hidden neurons

[1] 0 = Not pressed; 1 = Pressed.

for all subjects but 27 and 28, for which we use 500 neurons, due to observed overfitting.

All neurons use the standard logistic function $f(x) = \frac{1}{1+e^{-x}}$ as activation function. The output layer is composed by two nodes, one for each of the two categories to be recognized.

For the standard pseudoinversion method we ran the learning procedure for 1000 initializations of the input weights and averaged the test error rates. In order to perform a fair comparison, for the genetic pseudoinversion we evaluated 1000 individuals (i.e. weight assignments) by running genetic algorithm for 50 generation on a population of 20 individuals, and averaged the test errors.

Moreover, we computed for both approaches the test error values generated by the weights configurations which gave rise to minimum training errors values; these quantities are reported as minimun test errors.

In Table 1 we compare performances computed on the separate test sets: the best performing method is reported in bold for each subject.

As can be noted the average test error of the genetic pseudoinversion is better in all cases but two than the standard pseudoinversion approach. For what concerns the minimum test error, the standard pseudoinversion approach is better in four cases, probably because sometimes the genetic approach needs more generations to find an equivalent performing weight configuration.

Table 1 Comparison of average and minimum test errors for all subjects (sparse approach is used in both cases)

	Average Test Error		Minimum Test Error	
	Standard Pseudoinv.	Genetic Pseudoinv.	Standard Pseudoinv.	Genetic Pseudoinv.
Sub.1	7.24	**4.64**	7.14	**4.65**
Sub.10	2.08	**1.76**	**1.56**	1.71
Sub.11	5.03	**3.39**	3.20	**2.48**
Sub.12	4.24	**2.20**	2.70	**1.85**
Sub.13	3.74	**2.94**	**1.65**	3.76
Sub.14	7.60	**5.12**	6.46	**6.06**
Sub.15	7.24	**4.55**	5.49	**3.39**
Sub.16	3.96	**2.48**	3.44	**2.31**
Sub.17	5.09	**3.06**	6.00	**3.59**
Sub.18	6.55	**5.41**	**4.19**	5.19
Sub.19	**6.01**	6.93	**3.92**	6.63
Sub.21	6.34	**5.61**	7.86	**5.38**
Sub.22	6.42	**5.12**	8.41	**4.5**
Sub.24	8.69	**5.84**	8.07	**5.32**
Sub.25	4.78	**2.86**	4.53	**3.22**
Sub.26	5.03	**4.21**	6.64	**4.43**
Sub.27	14.08	**6.91**	11.51	**6.52**
Sub.28	10.17	**5.97**	10.90	**5.96**
Sub.29	**5.90**	6.52	5.99	**5.13**

We also compared the time needed to perform the matrix product between input data and input weight matrices when the latter are implemented as sparse or with usual matrix representation. Fig. 3 reports the time (s.) taken vs the sparsity degree of the matrix for three hidden layer size dimensions (500, 750,

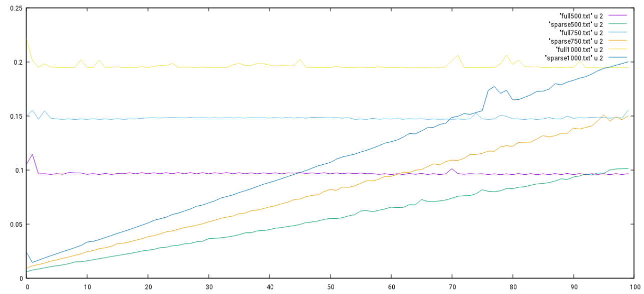

Fig. 3 Comparison of computational cost: sparse vs usual representation

1000 neurons). In all cases the sparse representation performs better, allowing to save a large amount of computational resources.

8 Conclusions

In this paper we present a method for non-intrusive and real-time detection of visual distraction, based on vehicle dynamics data and environmental data.

A single layer feedforward neural network trained through pseudo-inversion is used to process data: we propose an original approach which benefits from matrix sparsity, showing lower computational times with respect to standard implementations. Moreover, our genetic based technique outperforms the usual one for the majority of the considered subjects.

Acknowledgements The activity has been partially carried on in the context of the Visiting Professor Program of the Gruppo Nazionale per il Calcolo Scientifico (GNCS) of the Italian Istituto Nazionale di Alta Matematica (INdAM).

Conflict of Interest: The authors declare that they have no conflict of interest.

References

1. J.D. Lee, K.L. Young, M.A. Regan, 2008. Defining driver distraction. In: Regan, M.A., Lee, J.D., Young, K.L. (Eds.), Driver Distraction: Theory, Effects, and Mitigation. CRC Press Taylor & Francis Group, Boca Raton, FL, USA, pp. 31-40.
2. M. Bayly, K. L. Young, and M. A. Regan, Sources of distraction inside the vehicle and their effects on driving performance, in Driver Distraction: Theory, Effects, and Mitigation. Boca Raton, FL, USA: CRC Press, 2008, pp. 119-213
3. A. Pentland and A. Liu, Modeling and prediction of human behavior, Neural Comput., vol. 11, no. 1, pp. 229-242, Jan. 1999.
4. L. Qiao, M. Sato, and H. Takeda, Learning algorithm of environmental recognition in driving vehicle, IEEE Trans. Syst., Man, Cybern., vol. 25, no. 6, pp. 917-925, Jun. 1995.
5. Tango F., Botta M. Evaluation of Distraction in a Driver-Vehicle-Environment Framework: an application of different Data-mining techniques. In Proc. 9th Industrial Conference on Data Mining (ICDM09). Leipzig, Germany, 2009. Springer-Verlag Eds.
6. C. Blaschke, F. Breyer, B. Frber, J. Freyer, and R. Limbacher, Driver distraction based lane-keeping assistance, Transp. Res. Part F: Traffic Psychol. Behav., vol. 12, no. 4, pp. 288-299, Jul. 20

7. H. Zhang, M. R. H. Smith, and G. J. Witt, Identification of real-time diagnostic measures of visual distraction with an automatic eye-tracking system, Hum. Factors, vol. 48, no. 4, pp. 805-821, 2006
8. J. Hoel, M. Jaffard, P. Van Elslande, 2010. Attentional competition between tasks and its implications. Paper presented at the European Conference on Human Centred Design for Intelligent Transport Systems (29-30 April). Retrieved from http://www.conference2010.humanist-vce.eu/.
9. D. E. Rumellhart, G. E. Hinton, and R. J.Williams, Learning internal representations by error propagation, Parallel Distrib. Process.: Explanations Microstructure of Cognition, vol. 1, pp. 318-362, 1986.
10. C. Xiang, S. Q. Ding, and T. H. Lee, Geometrical interpretation and architecture selection of MLP, IEEE Trans. Neural Netw., vol. 16, n. 1, pp. 84-96, 2005.
11. J. Park and I. W. Sandberg, Universal approximation using radialbasis- function networks, Neural Comput., vol. 3, pp. 246-257, 1991.
12. D. S. Broomhead and D. Lowe, Multivariable functional interpolation and adaptive networks, Complex Syst., no. 2, pp. 321-355, 1988.
13. D. Lowe, Adaptive radial basis function nonlinearities, and the problem of generalisation, Proc. 1st Inst. Electr. Eng. Int. Conf. Artif. Neural Netw., pp. 171-175, 1989
14. Haykin, S.: Neural Networks, a comprehensive foundation, Prentice Hall, U.S.A (1999).
15. F. M. Ham and I. Kostanic, Principles of Neurocomputing for Science & Engineering, Boston, MA: McGraw-Hill, 2001.
16. Y. H. Pao and Y. Takefuji, Functional-link net computing, theory, system architecture, and functionalities, IEEE Comput., vol. 25, no.5, pp. 76-79, 1992.
17. Y. H. Pao, G. H. Park, and D. J. Sobajic, Learning and generalization characteristics of random vector functional-link net, Neurocomputing, vol. 6, pp. 163-180, 1994.
18. B. Igelnik, Y.-H. Pao, S. R. LeClair, and C. Y. Shen, The ensemble approach to neural-network learning and generalization, IEEE Trans. Neural Netw., vol. 10, no. 1, pp. 19-30, 1999.
19. Y. H. Pao, Adaptive Pattern Recognition and Neural Networks, Reading, MA: Addison-Wesley, 1989.
20. Halawa, K.: A method to improve the performance of multilayer perceptron by utilizing various activation functions in the last hidden layer and the least squares method. Neural Processing Letters 34, pp. 293-303 (2011)
21. G. Huang, G.-B. Huang, S. Song, K. You: Trends in extreme learning machines: A review, Neural Networks, vol. 61, pp. 32-48, 2015.
22. Nguyen, T. D., Pham, H. T. B., Dang, V. H.: An efficient Pseudo Inverse matrix-based solution for secure auditing. In: IEEE International Conference on Computing and Communication Technologies, Research, Innovation, and Vision for the Future (2010)
23. Kohno, K., Kawamoto, M., Inouye, Y.: A Matrix Pseudoinversion Lemma and Its Application to Block-Based Adaptive Blind Deconvolution for MIMO Systems. Circuits and Systems I: Regular Papers, IEEE Transactions on, 57, no.7, pp.1449-1462 (2010)
24. Ajorloo, H., Manzuri-Shalmani, M. T., Lakdashti, A. : Restoration of damaged slices in images using matrix pseudo inversion. In: 22nd International Symposium on Computer and Information Sciences (2007)
25. R. Cancelliere, A. Gosso, A.Grosso: Neural networks for wind power generation forecasting: A case study. In: 10th IEEE International Conference on Networking, Sensing and Control (ICNSC), pp. 666-671, 2013
26. L. Ghignone, R. Cancelliere: Neural Learning of Heuristic Functions for General Game Playing. In 2nd International Workshop on Machine Learning, Optimization and Big Data, Lecture Notes in Computer Science, vol. 10122, Springer-Verlag, (2016).
27. R. I. Arriaga and S. Vempala. An algorithmic theory of learning: robust concepts and random projection. In Proc. 40th Annual Symp. on Foundations of Computer Science, pp. 616-623. IEEE Computer Society Press, 1999.
28. S. Vempala. Random projection: a new approach to VLSI layout. In Proc. 39th Annual Symp. on Foundations of Computer Science. IEEE Computer Society Press, 1998.
29. P. Indyk and R. Motwani. Approximate nearest neighbors: towards removing the curse of dimensionality. In Proc. 30th Symp. on Theory of Computing, pp. 604-613. ACM, 1998.
30. Tiago M., Rui A., Carlos Henggeler A., Dulce G., Genetically optimized extreme learning machine, IEEE 18th Conference on Emerging Technologies & Factory Automation (ETFA), 2013.

31. Alexander V., Annamalai P., An Elitist Genetic Algorithm Based Extreme Learning Machine. In: Senthilkumar M., Ramasamy V., Sheen S., Veeramani C., Bonato A., Batten L. (eds) Computational Intelligence, Cyber Security and Computational Models. Advances in Intelligent Systems and Computing, vol 412. Springer, Singapore, 2016.
32. R. Sharma; A. S. Bist, Genetic algorithm based weighted extreme learning machine for binary imbalance learning, International Conference on Cognitive Computing and Information Processing(CCIP), 2015.
33. H. Liu, B. Jiao, L. Peng,T. Zhang, Extreme learning machine based on improved genetic algorithm, 5th International Conference on Information Engineering for Mechanics and Materials (ICIMM) 2015.
34. A. J. McKnight and A. S. McKnight, The effect of cellular phone use upon driver attention, Accid. Anal. Prev., vol. 25, no. 3, pp. 259-265, Jun. 1993.
35. T. A. Dingus, S. G. Klauer, V. L. Neale, A. Petersen, S. E. Lee, J. Sudweeks, M. A. Perez, J. Hankey, D. Ramsey, S. Gupta, C. Bucher, Z. R. Doerzaph, J. Jermeland, and R. R. Knipling, The 100-car naturalistic driving study, phase II-results of the 100-car field experiment, Nat. Highway Traffic Safety Admin., Washington, DC, Dept. Transp., HS 810 593, 2006.
36. M. Woeller, C. Blaschke, T. Schhindl, B. Schuller, B. Faerber, S. Mayer and B. Trefflich. Online Driver Distraction Detection using long short-term Memory in IEEE Transaction on Intelligent Transportation System , vol. 12, No. 2, June 2011, IEEE Society Editor.
37. D. Haigney and S. J. Westerman. Mobile (cellular) phone use and driving: A critical review of research methodology in Ergonom., vol. 44, no. 2, pp. 132-143, Feb. 2001.
38. K. Young and M. Regan, Driver distraction: A review of the literature, in Distracted driving, Australian College Road Safety, Sydney, NSW, Australia, 2007, pp. 379-405.
39. S. G. Klauer, T. A. Dingus, V. L. Neale, J. D. Sudweeks, and D. J. Ramsey, The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data, Nat. Highway Traffic Safety Admin. (NHTSA), Washington, DC, USA, Tech. Rep. DOT HS 810 594, 2006.
40. S. G. Klauer et al., Distracted driving and risk of road crashes among novice and experienced drivers, New England J. Med., vol. 370, no. 1, pp. 54-59, Jan. 2014.
41. M.A. Regan, C. Hallet and C.P. Gordon. Driver Distraction and Driver Inattention: definition, relationship and taxonomy. In Accident Analysis and Prevention journal, Elsevier, vol. 43, 2011.
42. T. A. Ranney, W. R. Garrott, and M. J. Goodman, NHTSA driver distraction research: Past, present, and future, National Highway Traffic Safety Administration, pp. 1-8, 2001.
43. Y. Dong, Z. Hu, K. Uchimura, and N. Murayama, Driver inattention monitoring system for intelligent vehicles: a review, IEEE Transactions on Intelligent Transportation Systems, IEEE Transactions on Intelligent Transportation Systems, vol. 12, no. 2, pp. 596-614, 2011.
44. H. Croo, M. Bandmann, G. Mackay, K. Rumar, and P. Vollenhoven, The role of driver fatigue in commercial road transport crashes, European Transport Safety Council, 2001.
45. T. Hirayama, K. Mase, C. Miyajima, and K. Takeda. Classification of Drivers Neutral and Cognitive Distraction States Based on Peripheral Vehicle Behavior in Drivers.
46. Y. Liang and J. D. Lee, Combining cognitive and visual distraction: less than the sum of its parts, Accident Analysis & Prevention, vol. 42, no. 3, pp. 881-890, 2010.
47. N. Merat and A. H. Jamson, Multisensory signal detection: a tool for assessing driver workload during IVIS management, Proceedings of the 4th International Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, 2007.
48. N. Merat, E. Johansson, J. Engstrom, E. Chin, F. Nathan, and T. Victor, Specification of a secondary task to be used in safety assessment of IVIS, Adaptive Integrated Driver-Vehicle Interface, 2007.
49. G. L. Rupp, ed., Performance metrics for assessing driver distraction: the quest for improved road safety, SAE International, 2010.
50. L. Hsieh, R. Young, and S. Seaman, Development of the enhanced peripheral detection task: a surrogate test for driver distraction, SAE International Journal of Passenger Cars-Electronic and Electrical Systems, vol. 5, no. 1, pp. 317-325, 2012.
51. Badeva, V., Morosov, V.: Problemes incorrectements posès, thèorie et applications, (in French). Masson, Paris (1991)
52. Cancelliere, R., De Luca, R., Gai, M., Gallinari, P., Rubini L.: An analysis of numerical issues in neural training by pseudoinversion. Computational and Applied Mathematics, 36(1), pp.599-609 (2015)

53. Tikhonov, A.N., Arsenin, V.Y.: Solutions of Ill-Posed Problems. Winston, Washington, DC (1977)
54. Tikhonov, A.N.: Solution of incorrectly formulated problems and the regularization method. Soviet Mathematics 4, pp. 1035-1038 (1963)
55. G. Golub and C. van Loan: Matrix computations (3rd ed.), Johns Hopkins University Press, Baltimore, MD, USA, 1996.
56. Gallinari, P. , Cibas, T.: Practical complexity control in multilayer perceptrons. Signal Processing 74, pp. 29-46 (1999)
57. R. Cancelliere, M. Gai, P. Gallinari, L. Rubini. OCReP: An Optimally Conditioned Regularization for pseudoinversion based neural training, Neural Networks, vol. 71, pp. 76-87, (2015)
58. R. Hecht-Nielsen, Context vectors: general purpose approximate meaning representations self-organized from raw data. In J.M. Zurada, R.J. Marks II, and C.J. Robinson, editors, Computational Intelligence: Imitating Life, pp. 43-56. IEEE Press, 1994.
59. E. Bingham and H. Mannila, Random projection in dimensionality reduction: Applications to image and text data. In Proc. of the conference Knowledge Discovery and Data Mining KDD 2001, San Francisco CA, USA, 2001.
60. W.B. Johnson and J. Lindenstrauss, Extensions of Lipshitz mapping into Hilbert space. In Conference in modern analysis and probability, vol. 26 of Contemporary Mathematics, pp. 189-206. Amer. Math. Soc., 1984.
61. S. Dasgupta and A. Gupta, An elementary proof of the Johnson-Lindenstrauss lemma. Technical Report TR-99-006, International Computer Science Institute, Berkeley, California, USA, 1999.
62. D. Achlioptas, Database-friendly random projections. In Proc. ACM Symp. on the Principles of Database Systems, pp. 274-281, 2001.
63. https://www.iso.org/standard/54496.html
64. Wickens C.D. Multiple Resources and performance prediction. Theoretical Issues In Ergonomics Science, 3(2), 159-17 (2002).
65. Tango, F., Minin L., Montanari, R., and Botta, M. (2010, August, 16-19). Non-intrusive Detection of Driver Distraction using Machine Learning Algorithms. In the proceeding of the XIX European Conference on Artificial Intelligence (ECAI). Lisbon, Portugal (ISBN: 978-1-60750-605-8; Editor: IOS Press Amsterdam, The Netherlands, The Netherlands).

Click here to access/download
Supplementary material
Cancelliere.pdf