# Drowsy Driver Detection using Representation Learning

Kartik Dwivedi, Kumar Biswaranjan and Amit Sethi
Department of Electronics and Electrical Engineering
Indian Institute of Technology
Guwahati, India

***Abstract*—The advancement of computing technology over the years has provided assistance to drivers mainly in the form of intelligent vehicle systems. Driver fatigue is a significant factor in a large number of vehicle accidents. Thus, driver drowsiness detection has been considered a major potential area so as to prevent a huge number of sleep induced road accidents. This paper proposes a vision based intelligent algorithm to detect driver drowsiness. Previous approaches are generally based on blink rate, eye closure, yawning, eye brow shape and other hand engineered facial features. The proposed algorithm makes use of features learnt using convolutional neural network so as to explicitly capture various latent facial features and the complex non-linear feature interactions. A softmax layer is used to classify the driver as drowsy or non-drowsy. This system is hence used for warning the driver of drowsiness or in attention to prevent traffic accidents. We present both qualitative and quantitative results to substantiate the claims made in the paper.**

***Keywords— Driver Drowsiness, Artificial Intelligence, Feature learning, Deep learning, Convolutional Neural Networks***

## I. INTRODUCTION

Driver fatigue is a significant factor in a large number of vehicle accidents. Fatalities have occurred as a result of car accidents related to driver inattention, such as distraction, fatigue, and lack of sleep. Studies and experiments have substantiated the fact that driving performance deteriorates with increased drowsiness [1]. The US National Highway Traffic Safety Administration has estimated approximately 100,000 crashes each year caused mainly due to driver fatigue or lack of sleep [2].

Autonomous systems designed to analyze driver exhaustion and detect driver drowsiness can be an integral part of the future intelligent vehicle so as to prevent accidents caused by sleep. A variety of techniques have been employed for vehicle driver fatigue and exhaustion detection. Driver operation and vehicle behavior can be implemented by monitoring the steering wheel movement, accelerator or brake patterns, vehicle speed, lateral acceleration, and lateral displacement. These are non-intrusive ways of driver drowsiness detection, but are limited to the type of vehicle and driver conditions [3]. Another set of techniques focuses on monitoring of physiological characteristics of the driver such as heart rate, pulse rate, and Electroencephalography (EEG) [4]. Research in these lines have suggested that as the alertness level decreases EEG power of the alpha and theta bands increase [5], hence providing indicators of drowsiness. Although the use of these physiological signals yields better detection accuracy, these are not accepted widely because of less practicality. A third set of techniques is based on computer vision systems which can recognize the facial appearance changes occurring during drowsiness [6, 7, 8]. Physiological feature–based approaches are intrusive because the measuring equipment must be attached to the driver. Thus, visual feature–based approaches have recently become preferred because of their non-intrusive nature. In this paper, we propose a new scheme based on extraction of visual features from the data without human intervention. These visual features have been learnt using a model of deep learning known as convolutional neural networks. The feature maps produced by convolving the learnt weights with input image act as the features for driver drowsiness detection. Using these set of features a soft-max layer classifier is used to finally classify the frames extracted as drowsy or non-drowsy. Further, a set of extra methodologies are suggested that could be combined with the scheme in the future to make the technique more robust.

## II. RELATED WORK

There are some significant previous studies about drowsiness detection and fatigue monitoring. Many computer vision based schemes have been developed for non-intrusive, real-time detection of driver sleep states with the help of various visual cues and observed facial features. An observed pattern of movement of eyes, head and changes in facial expressions are known to reflect the person's fatigue and vigilance levels. Eye closure, head movement, jaw drop, eyebrow shape and eyelid movement are examples of some features typical of high fatigue and drowsy state of a person. To make use of these visual cues, a remote camera is usually mounted on the dashboard of the vehicle which, with the help of various extracted facial features, analyses driver's physical conditions and classifies the current state as drowsy/non-drowsy. It has been concluded that computer vision techniques are non-intrusive, practically acceptable and hence are most promising for determining the driver's physical conditions and monitoring driver fatigue [9].

Most of the published researches based on computer vision techniques are image based real-time schemes for fatigue monitoring using typical facial features. Singh et al. [10] developed a vision based scheme based on eye blink duration using the proposed mean sift algorithm. Saito et al. [11] uses driver's line of sight to detect the mental and physical

conditions. Horng et al. [12] uses edge information for localizing eyes and dynamical template matching for eye tracking for driver fatigue detection. Smith et al. [13] describes an algorithm which relies on optical flow and color predicates to robustly track a person's head and facial features. Their study showed that the performance of their system is comparable with those of techniques using physiological signals.

New techniques are based on machine learning algorithms to detect driver drowsiness levels. Vural et al. [14] creates Automatic classifiers for 30 facial actions from the Facial Action Coding system using machine learning on a separate database of spontaneous expressions to finally categorize driver drowsiness. Vural et al. [15] proposes a system that applies automated measurement of the face during actual drowsiness to discover new signals of drowsiness in facial expression and head motion. Ji et al. [16] demonstrates that the simultaneous use of multiple visual cues and their systematic combination yields a much more robust and accurate fatigue characterization than using a single visual cue by using a Bayesian network.

Modern day algorithms exploiting multiple visual cues and using novel machine learning strategies for drowsiness detection have certainly resulted in significant improvement of such intelligent systems. However, all the work done in the field of visual cues based driver drowsiness detection uses only hand-picked features. Hand engineered features constitute eye blink, eye closure, expression detection features – mixture of face wrinkles, eye brow, lip and cheek shapes etc. Although novel machine learning based algorithms use multiple cues, they are unable to exploit the complex relationship between various features. In the proposed work, we demonstrate the effect of using facial features derived from a convolutional neural network based representation feature learning scheme. Rather than using human expertise and ingenuity to design features, representation learning believes models learning features from the data can exploit the feature space more intelligently and represent the perplex relationship of raw data with output by combining features of features. Apart from exploiting the perplex relationship between various features learnt using successive hidden layers, it is also able to extract some useful latent features that are difficult to acknowledge using hand engineered methods..

## III. REPRESETATION FEATURE LEARNING

### *A. Introduction to CNN*

Recent years has seen many significant improvements in the area of representation feature learning by introduction of many models such as Deep Boltzman Machines(DBM) [17], Deep Belief Networks(DBN)[18], Convolutional neural networks (CNN)[19], Restricted Boltzman Machine(RBM) [20, 21], Recurrent Neural Networks(RNN) [22, 23], Autoencoders [24] and others. The underlying driving force behind the success of these models is the learning of feature representation which is capable of capturing more intelligent features from the unlabeled input data. Most of the models use multiple hidden layers to learn complex, non-linear, high dimensional representation which are fed to a classifier for high level of classification task.

Convolutional neural nets are a variation of feed forward neural nets which incorporates three unique features: local receptive fields, sharing of weights and sometimes spatial or temporal pooling [19]. All the filters of convolutional net share the weights with all the pixels of input image. By restricting the weights to take the same value for different local regions ensures the detection of a shifted feature at different locations of an image and also reduces the number of parameters to be learned by a huge amount and acts as a regularizer. The convolution operation at each layer distinguishes it from other neural net models. In the context of image the input is presented as 2D vector on which filters are convoluted capturing the local features more efficiently and resulting a set of feature maps which becomes input to the next layer. A pooling operation can be performed at the output of each layer to extract shift invariant features up to certain extent. The pooling can be either a subsampling or a max-pooling operation. In a max-pooling operation it registers the highest response of a region. The weights of the convolutional neural net is shared across all the pixels which reduces the number of parameters to be learned making training faster.

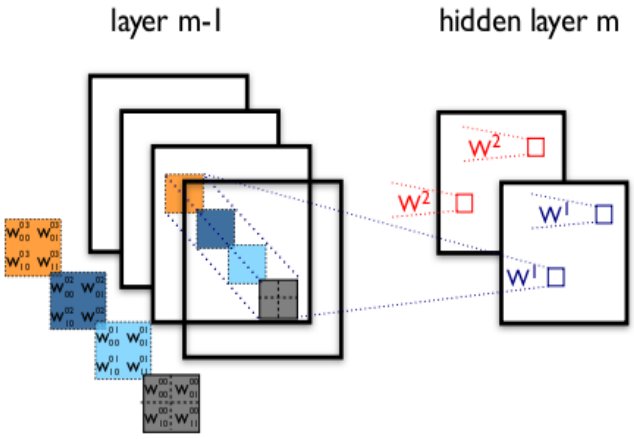



Fig. 1 Example of a convolutional layer.
Source: http://deeplearning.net/tutorial/_images/cnn_explained.png

The 1-d convolution of a input sequence x[n] with a filter f[n] is given by,

$$o[n] = f[n] * x[n] = \sum_{u=-\infty}^{\infty} f[u]x[n-u] = \sum_{u=-\infty}^{\infty} f[n-u]x[u] \quad (1)$$

The convolution can be extended to 2-D by the following equation

$$o[m,n] = f[m,n] * x[m,n] = \sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} f[u,v]x[u-m, v-n] \quad (2)$$

Here f[m,n] is the 2-D filter map convolved with input x[m,n] produces a feature map o[m,n]. Similarly this operation can be extended to set of filters to produce set of feature maps. Another feature of a convolutional layer is the Max Pooling operation. From a specified set of non-overlapped rectangular regions, the maximum response is given as output. It is a form of non-linear subsampling which allowed our feature to be locally translation invariant and reduced the dimension of our features.

### B. Layers of CNN

We used a model consisting of two convolutional layers along with max-pooling operation followed by a hidden layer of sigmoid which is fully connected to a logistic regression layer for classification. Sigmoid layer applies a non-linear transformation to the features from convolutional layers. The logistic regression layer has two nodes each for predicting the probability of drowsiness given the input and weights and other one similarly for non-drowsiness case. The two convolutional layers perform identical operation. They convolve a set of filters with the input data followed by a nonlinearity operation and a subsampling resulting into a set of feature maps which serves as input to the next layer.

Let f(x) be the features extracted by convolutional layer for input image x, $W_{ch}$ and $b_{ch}$ be the weights and biases connected from convolutional layer to the sigmoid hidden layer. Then activation of sigmoid layer is given by,

$$h = \frac{1}{1+\exp(-{W_{ch}}^T f(x)+b_{ch})} \quad (3)$$

Let $W_{hl}$ and $b_{hl}$ are the weights and biases from hidden layer to logistic regression layer, $W_{hl}^{(i)}$ and $b_{hl}^{(i)}$ are the weights and biases from sigmoid hidden layer to the logistic unit corresponding ith output, then the probability of ith output being true is given by,

$$P(y^{(i)}) = \frac{\exp\left({W_{hl}^{(i)}}^T h+b_{hl}^{(i)}\right)}{\exp\left({W_{hl}^{(1)}}^T h+b_{hl}^{(1)}\right)+\exp\left({W_{hl}^{(2)}}^T h+b_{hl}^{(2)}\right)} \quad (4)$$

Where i=1, 2 for drowsy and non-drowsy case respectively. The output of the model given the probabilities of both class is calculated by taking argmax over both class

$$y = \underset{i}{argmax} P(y^{(i)}) \quad (5)$$

Our objective function is consisted of minimizing the negative log likelihood cost function averaged over a mini-batch of images. Let D be the set of images for a single mini-batch, n be the number of training samples in a mini-batch, $t_{(n)}$ be the true output of nth training sample, $h_{(n)}$ be the activation hidden layer for nth training image, $W_{(hl)}$, $b_{(hl)}$ are the weights and bias from hidden layer to logistic regression layer and $logP(Y = t_{(n)}|h, W_{(hl)}, b_{(hl)})$ be the probability of $y^{t_{(n)}}$th is being the true output given all the parameters for nth sample image, then the objective function is given by,

$$O = min \frac{-1}{|D|} \sum_{n=1}^{|D|} log\, P(Y = t_{(n)}|h_{(n)}, W_{(hl)}, b_{(hl)}) \quad (6)$$

### C. Model Parameters

We trained our model using cross validation by dividing the whole dataset into five folds out of which one fold was used for validation and remaining four for training. A batch of 50 images of size (48*48) were fed to the first layer which convolved 20 filters of size (5*5) producing a set of 20 feature maps of size (44*44) for each image in the batch. Each feature map was down-sampled using (2*2) max pooling operation which resulted in 20 feature maps of size (22*22) for each image. All the down-sampled feature maps were fed to the second convolutional layer consisting 50 filters of size (5*5). After convolution 50 feature maps of size (18*18) were produced which down-sampled to size (9*9). All the features produced were flattened to a single 1-D vector for each image and fed to hidden layer of 1000 sigmoid units. The output 1000 features per image were given to logistic regression layer for classification.

## IV. METHOD

### A. Driving task and data collection

Due to lack of easy availability of standard datasets for driver drowsiness detection, a dataset was created so as to train the classifier and evaluate the performance of the scheme. Subjects were made to play an open source driving and obstruction avoidance game (Figure 3) after midnight at different fatigue levels. A diverse dataset has been created involving 30 subjects (Figure 2) with different physical attributes including variety in skin tone, eye size, fatigue level, facial structure, hair fringes and facial hair. Different illumination conditions were adopted to make dataset even more universal, keeping in mind the varying brightness conditions in real life scenarios. Thus, the classifier would become more robust and efficient in all circumstances. Subjects also wear eye glasses in few video sequences to further add to the diverse nature and difficulty of the dataset.

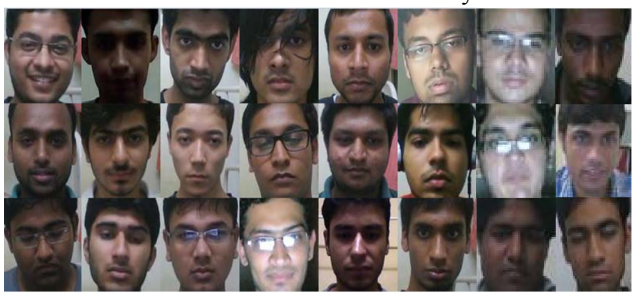


Fig. 2. Diverse nature of the dataset including 30 subjects with different skin tone, eye shape and size, face width and height, hair fringes, spectacles in different illumination conditions.

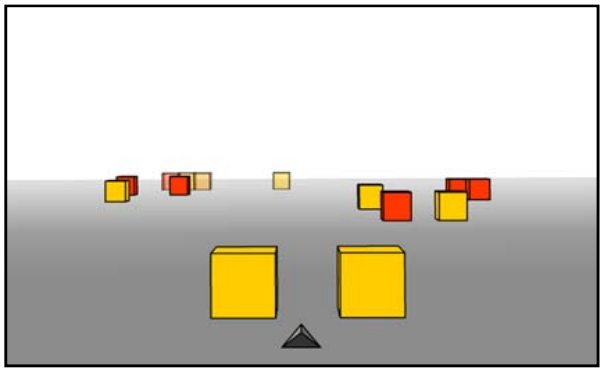


Fig. 3. A Typical scene from a famous open source online game – Cube field used as obstruction avoidance video game for driver vigilance/drowsiness detection. Source: http://www.yoarcade.net/ability/cubefield_content.html

### *B. Proposed Scheme*

The proposed method aims to classify frames in videos based on special facial features learnt via convolutional neural network. Figure 4. gives an overview of the training and testing procedure adopted in the scheme.

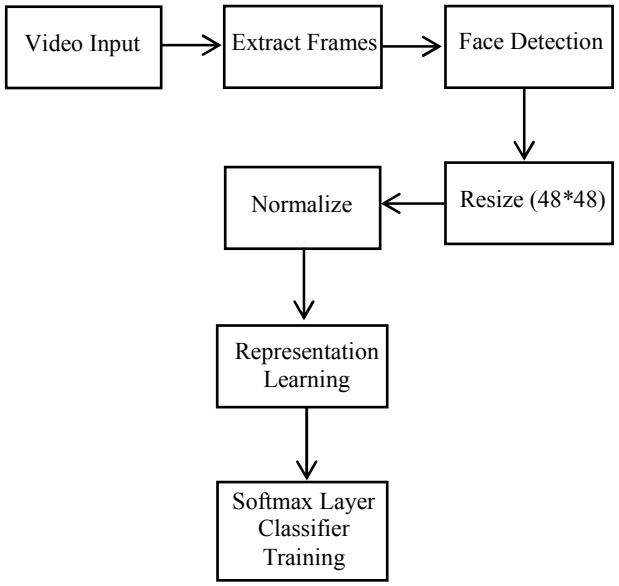


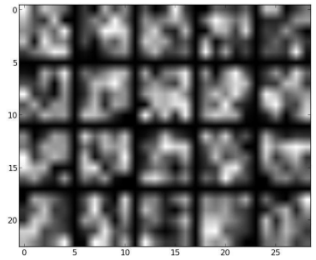

Fig 4. An outline of the proposed algorithm based on representation facial feature learning.

Firstly, frames are extracted from the video. These frames are fed to a Viola and Jones Haar-like features based face detector. The detected faces are cropped and resized to 48* 48 square images. These cropped images are normalized by subtracting each pixel by the mean followed by division with its standard deviation. Normalized images of 80 percent subjects are further fed to a multi-layer convolutional neural network. The outputs of the hidden layer are considered as the extracted features. On the basis of these features, the softmax layer classifier was trained. Once the classifier has been trained, the rest twenty percent of the images extracted earlier are tested on the trained classifier.

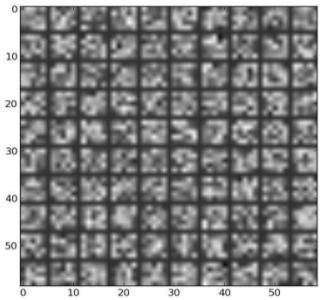
The above scheme describes drowsy driver detection at the frame level. A binary signal for each frames in the form of drowsy or non-drowsy face is been obtained. For an alert signal to be delivered to a driver, at least 40 out of 60 frames should be detected as drowsy. A buffer of 60 recent frame outputs is maintained and a warning is sent to the driver in the form of an alerting sound. Thus, the driver is being successfully alerted and assisted by the intelligent system based on non- intrusive vision scheme.

## V. RESULTS

Deep learning based feature learning methods are known to provide excellently designed features especially in cases of image or visual data. The convolutional neural network model is used to learn the features. The feature learning process can be described as a weight learning procedure. Some of the weights learnt at some layers are shown in Figure 5.



(a)



(b)

Fig 5. Weights learnt at the end of (a) layer 1. (b) layer 2.

The input being provided to the first layer and the output (drowsy/non drowsy label) provided to the output of the last layer, all weights are learnt, all the learned weights acts a learned feature detectors for driver drowsiness and these feature detectors are convolved with input images to produce the final features used for classification. The dataset collected from the 30 different subjects in diverse conditions was divided into training and validation data randomly. All the extracted faces from the frames were labeled manually as drowsy and non-drowsy. The trained classifier worked efficiently as it gave 92.33% validation accuracy. Considering the fact that a car is driven by the same single person most of the time, an experiment is carried out as the driver is made to drive his vehicle for hours together in artificially simulated conditions and a training video is recorded and manually

labeled later. Later, the test is done on the same driver. The average accuracy within subjects was 88%. Furthermore, another experiment was carried out in which we train the classifier on a set of subjects and the testing is done on absolutely different variety of people having different physical and facial characteristics. A satisfactory average result of 78% accuracy across subjects was found in such a case. Thus, the proposed deep learning based classifier detects the driver drowsiness based on only visual facial features efficiently on a diverse dataset.

## VI. FUTURE WORK

Although, the proposed deep learning based driver drowsiness detection is able to successfully give reasonable results on a diverse dataset, still there is a scope for improvement in its performance. Drowsiness induces involuntary rolling or falling of the driver's head which could act as a valuable cue for successful detection of drowsiness. Also, most of such accidents occur during nocturnal hours. An I.R. LED based tracking approach could be employed to help in detection of sleepiness in such situation thus making the scheme usable in all illumination conditions. Moreover, the proposed scheme makes decision on frame level by application of 2D convolutional neural networks on each frame for feature extraction. A 3D convolutional network could be applied for robust sleep state detection of driver by making use of spatio-temporal relationship.

## VII. CONCLUSION

This paper proposes an algorithm for driver drowsiness detection using representation learning. A new perspective towards driver sleep detection is presented as features responsible for decision making are produced by leveraging multi-layer convolutional neural networks. Previous approaches could only make decisions based on features such as eye blinks, eye closure, forehead strain marks or even eye brow shapes. Other modern approaches were based on carefully hand engineered features detecting driver drowsiness based on human facial expressions. Convolutional neural networks based representation feature learning approach provides an automated and efficient set of features which help us to classify the driver as drowsy or non-drowsy very accurately. Deep learning based methods are very well known to exploit the latent relationships in an image data and learn excellent features for better representation of raw data. The scheme was tested on a diverse dataset. Both quantitative and qualitative result were provided and found to be in support of the proposed scheme.

## REFERENCES


[1] P. S. Rau, "Drowsy drivers detection and warning system for commercial vehicle drivers: Field proportional test design, analysis, and progress", Proc. - *19th International Technical Conference on the Enhanced Safety of Vehicles, Washington, D.C.,* 2005

[2] United States Department of Transportation., "Saving lives through advanced vehicle safety technology" http://www.its.dot.gov/ivi/docs/AR2001.pdf.

[3]Y. Takei, Y. Furukawa, "Estimate of driver's fatigue through steering motion," in *Man and Cybernetics, IEEE International Conference*, Volume: 2, pp. 1765- 1770 Vol. 2. 2005

[4] W.A. Cobb., "Recommendations for the practice of clinical neurophysiology.," *Elsevier*, 1983.

[5] K. Hong, Chung, "Electroencephalographic study of drowsiness in simulated driving with sleep deprivation.," *International Journal of Industrial Ergonomics.*, Volume 35, Issue 4, April 2005, pp. 307- 320.

[6] M. Eriksson and N.P. Papanikolopoulos, "Eye-tracking for detection of driver fatigue", *IEEE proc. Intelligent Transport System, Boston, MA*, pp. 314-319, 1997

[7] Perez, Claudio A. et al., "Face and eye tracking algorithm based on digital image processing", *IEEE System, Man and Cybernetics 2001 Conference,* vol. 2, pp1178-1188. 2001

[8] S. Singh. and N. P. Fapanikolopaulas, "Monitoring driver fatigue using facial analysis technologies", *IEEE International conference on the Intelligent Transportation Systems.* pp.316-318, 1999

[9] Conf. Ocular Measures of Driver Alertness, Washington, DC, Apr.26–27, 1999.

[10] M.Singh, G.Kaur, "Drowsy detection on eye blink duration using algorithm" *International Journal of Emerging Technology and Advanced Engineering* ISSN 2250-2459, Volume 2, Issue 4, April 2012

[11] H. Saito, T. Ishiwaka, M. Sakata and S. Okabayashi, "Applications of driver's line of sight to automobiles – what can driver's eye tell". *Proceedings of 1994 Vehicle Navigation and Information Systems Conference*, Yokohama, Japan, August 1994, pp. 21-26

[12] W. Horng, C. Chen, Y. Chang, "Driver fatigue detection based on eye tracking and dynamic template matching". *Proceedings of the IEEE International Conference on Networking, Sensing & Control* 2004

[13] P. Smith, M. Shah, and N.V. Lobo, "Monitoring head/eye motion for driver alertness with one camera" *Proceedings. 15th International Conference on Pattern Recognition* (Volume:4 ) September 2000

[14] E. Vural, M. Cetin, A. Ercil, G. Littlewort, M. Barlett, "Drowsy Driver detection using facial movement analysis" Proc. of the *IEEE international conference on Human-computer interaction* pp. 6-18 Springer-Verlag Berlin, Heidelberg 2007

[15] E. Vural, M.S. Bartlett, G. Littlewort, M. Cetin, E. Ercil, and J. Movellan, "Discrimination of moderate and acute drowsiness based on spontaneous facial expressions" *IEEE International Conference on Pattern Recognition* 2010

[16] Q. Ji, Z. Zhu, P. "Lan real-time nonintrusive monitoring and prediction of driver fatigue" *IEEE Transactions on vehicular technology*, vol. 53, No. 4, July 2004

[17] R. Salakhutdinov and G.E. Hinton, "An efficient learning procedure for deep Boltzmann machines", *Neural Computation* August 2012, Vol. 24, No. 8: 1967 — 2006.

[18] G. E. Hinton, S. Osindero, and Y. Teh. "A fast learning algorithm for deep belief nets", Neural Computation 18:1527-1554, 2006

[19] Y. LeCun, Y. Bengio. "Convolutional networks for images, speech, and time series". The handbook of brain theory and neural networks, 3361. 1995

[20] H. Chen, A. F. Murray, "Continuous restricted Boltzmann machine with an implementable training algorithm". *Vision, Image and Signal Processing, IEEE Proceedings*- Vol. 150, No. 3, pp. 153-158. IET, June 2003

[21] R. Salakhutdinov, A.Mnih, G. Hinton, "Restricted Boltzmann machines for collaborative filtering". In Proc. of the *24th international conference on Machine learning* pp. 791-798. ACM, June 2007

[22] P. J. Angeline, G. M. Saunders, J. B. Pollack, "An evolutionary algorithm that constructs recurrent neural networks". *IEEE Transactions on Neural Networks*, 5(1), 54-65, 1994

[23] D. P. Mandic, J. Chambers, "Recurrent neural networks for prediction: learning algorithms, architectures and stability". *John Wiley & Sons, Inc,* 2001

[24] G. E. Hinton, R.S. Zemel, "Autoencoders, minimum description length, and Helmholtz free energy". *Advances in neural information processing systems*, 3-3, 1994