Echo State Network for Classification of Human Eye Movements During Decision Making

Petia Koprinkova-Hristova, Miroslava Stefanova, Bilyana Genova, Nadejda Bocheva

Echo State Network for Classification of Human Eye Movements during Decision Making

Petia Koprinkova-Hristova[1][0000-0002-0447-9667], Miroslava Stefanova[2], Bilyana Genova[2], and Nadejda Bocheva[2][0000-0001-7792-7742]

[1] Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, Sofia, Bulgaria
[2] Institute of Neurobiology, Bulgarian Academy of Sciences, Sofia, Bulgaria
pkoprinkova@bas.bg, mirad_st@abv.bg, b.henova@abv.bg, nadya@percept.bas.bg

Abstract. The paper develops further a recently proposed author's approach for classification of dynamic data series using a class of Recurrent Neural Network (RNN) called Echo state network (ESN). It exploits the Intrinsic Plasticity (IP) tuning of ESN reservoir of neurons to fit their dynamics to the data fed into the reservoir input. A novel approach for ranking of a data base of dynamic data series into groups using the length of the multidimensional vector of reservoir state achieved after consecutive feeding of each time series into the ESN is proposed here. It is tested on eye tracker recordings of human eye movements during visual stimulation and decision making process. The preliminary results demonstrated the ability of the proposed technique to discriminate dynamic data series.

Keywords: Echo State Network, Time Series, Classification, Eye Tracking.

1   Introduction

In previous author's work [1] a novel approach for unsupervised clustering of static multidimensional data sets using a class of RNNs called Echo state networks [2, 3] was proposed. Next it was successfully tested on numerous practical examples and the results were summarized in [4]. In [5] the approach was upgraded and applied for classification of dynamic data series too.

The core of the approach, proposed first in [1], was to use the ESN to extract more informative features from multidimensional data sets. For this aim equilibrium states of the ESN reservoir neurons corresponding to every multidimensional data item presented to the ESN input were used. As it was shown in [6], the fitting of the ESN reservoir dynamics to reflect the input data structure can be achieved by an approach for ESN reservoir tuning called Intrinsic Plasticity (IP) [7, 8] that is aimed at achieving the desired distribution of the ESN reservoir output.

Since the number of the new extracted features depends on the size of the ESN, question how to choose the most proper among them is still under investigation. Initially [1, 4] it was proposed to choose only two of all possible neurons steady states

based on their distribution in one- or two-dimensional space [4]. Next in [9] we tried to extend the number of representative neurons until the accuracy of data clustering increases. However, this approach increased computational burden too much. That is why here another approach was proposed: to use the geometric size of the vector of all reservoir states and then to rank the time series data based on that single feature.

The approach was tested on dynamic series of eye movements data collected during psycho-physiological experiments with humans observing specific visual stimuli and making decisions. The preliminary results demonstrated the ability of the proposed approach to rank the time series of human eye movements in dependence on their characteristics.

The paper is organized as follows: next chapter describes briefly the ESN structure and its IP tuning and the newly proposed feature extraction approach; next the experimental set-up and the collected time series data are described; the results from classification by the proposed algorithm are presented and discussed in section four; the paper finished with concluding remarks and directions for future work.

2   Clustering Algorithm

2.1   Echo State Network and IP Tuning

ESN, shown on Fig. 1, is a type of recurrent neural network that belongs to the novel and fast developing family of reservoir computing approaches [2, 3]. The ESN output for the current time instance k is the vector $out(k)$ with size n_{out}. It is a linear function f^{out} (usually identity) of the vectors of the current states of the input $in(k)$ (with size n_{in}) and the reservoir neurons $X(k)$ (with size n_X):

$$y(k) = f^{out}\left(W^{out}\left[in(k), \quad X(k) \right] \right) \tag{1}$$

Here W^{out} is a trainable $n_{out} \times (n_{in} + n_X)$ matrix. The neurons in the reservoir have a simple sigmoid output function f^{res} (usually hyperbolic tangent) that depends on both the ESN input $in(k)$ and the previous reservoir state $X(k-1)$:

$$X(k) = f^{res}\left(W^{in} in(k) + W^{res} X(k-1) \right) \tag{2}$$

Here W^{in} and W^{res} are $n_{in} \times n_X$ and $n_X \times n_X$ randomly generated weight matrices that are not trainable.

The main reason for development of such a type of RNN is to simplify their training algorithm. However, it appeared that although non-trainable weights can be random, there is need to tune them initially. For this aim different approaches were proposed [2, 3]. In [7, 8] an algorithm called intrinsic plasticity (IP) was proposed. Its aim was to increase the entropy of the reservoir neurons outputs by minimization of the Kullback-Leibler divergence:

$$D_{KL}\Big(p(X),p_d(X)\Big)=\int p(r)\log\left(\frac{p(X)}{p_d(X)}\right) \tag{2}$$

that is a measure for the difference between the actual $p(X)$ and the desired $p_d(X)$ probability distribution of reservoir neurons states X.
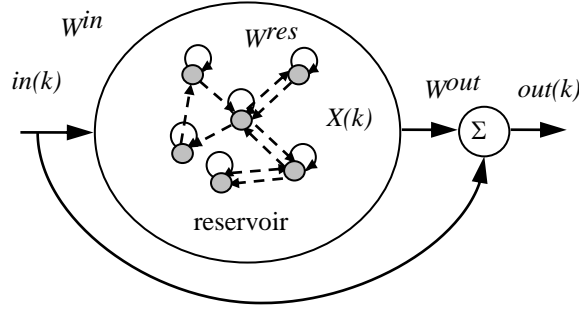

Fig. 1. Echo state network structure.

It was proven that for the commonly used hyperbolic tangent at the reservoir neurons output the proper target distribution has to be the Gaussian one. For this aim two additional reservoir parameters, gain a and bias b (both vectors with size n_X), were introduced in [8] as follows:

$$X(k)=f^{res}\Big(diag(a)W^{in}in(k)+diag(a)W^{res}X(k-1)+b\Big) \tag{3}$$

The IP training is gradient descent algorithm [8] minimizing the Kullback-Leibler divergence by adjustment of the vectors a and b.

2.2 Classification Approach for Dynamic Data Series

In [6] was demonstrated that besides its initial aim, the IP tuning also fits the reservoir connections matrix to the structure of the input data presented to the ESN. Moreover, the equilibrium states of reservoir neurons corresponding to each one of the input data items used during IP tuning reflect the overall data structure [1, 4]. Thus collected in this way features can be used for further classification or clustering.

In [5] it was demonstrated that the reservoir state $X(N)$ reached after feeding of non-constant (time varying) sequence of inputs from $in(0)$ to $in(N)$ to the IP tuned ESN:

$$X(N) = f^{res}\left(diag(a)W^{in}in(N) + diag(a)W^{res}X(N-1) + b \right)$$

$$X(N-1) = f^{res}\left(diag(a)W^{in}in(N-1) + diag(a)W^{res}X(N-2) + b \right) \qquad (4)$$

$$\dots$$

$$X(1) = f^{res}\left(diag(a)W^{in}in(0) + diag(a)W^{res}X(0) + b \right)$$

depends on dynamic characteristics of the time series *in* and can be exploited as a set of classification features.

Since the choice of neurons whose states are the best feature set for each particular data subject to classification or clustering is non-trivial [4, 9], here we propose another approach: to calculate the size of the vector containing all collected reservoir neurons states:

$$R = \sqrt{\sum_{i=1}^{n_X} x_i(N)^2}, \quad X(N) = \begin{bmatrix} x_1(N) & x_2(N) & \cdots & x_{n_X}(N) \end{bmatrix} \qquad (5)$$

and to use it as single discriminating data feature.

3   Experimental Set-up

The time series data used to test the idea described above were collected by eye tracing device that recorded the human eye movements during a behavioral experiment performed with the participation of volunteer human subjects observing series of visual stimuli.

Each stimulus is composed by a sequence of consecutive frames. A frame contains of 50 dots presented in a circular aperture with a radius of 7.5cm in the middle of the computer screen. The dots were grouped in 25 pairs placed at 2cm distance from each other. Each pair of dots had a limited lifetime of 3 frames. On every frame one-third of the pairs changed position. Each frame lasted 33ms. The orientation of the virtual lines connecting the dots in 18 pairs intersected in a common point considered as the center of each frame, while the rest 7 pairs had random orientations. The mean position of the centers of all frames in a stimulus sequence determines its "imaginary" center. We generated 14 different types of stimuli having centers at 7 positions shifted left and 7 positions shifted right from the screen midpoint. All shifts were in horizontal direction and varied between 0.67cm and 4.67cm with step of 0.67cm. Ten different patterns for each center position were generated.

The stimuli were presented on a gray screen with mean luminance 50cd/m² using 20.1″ NEC MultiSync LCD monitor with NvidiaQuadro 900XGL graphic board at a refresh rate of 60 Hz and screen resolution 1280×1024 pixels. The experiments were controlled by a custom program developed under Visual C++ and OpenGl.

The subject sat at 57cm from the monitor screen. Each stimulus presentation was preceded by a warning sound signal. A red fixation point with size of 0.8cm appeared in the center of the screen for 500ms. The stimuli were presented immediately after the disappearance of the fixation point. The Subject's task was to continue looking at the position where the fixation point was presented until he/she made a decision where the center of the pattern was and to indicate this position by a saccade (fast eye movement). The subjects also had to press the left or the right mouse button depending on the perceived position of the center - to the left or to the right from the middle of the screen. If the subject could not make a decision during the stimulus presentation (3.3s for 100 consecutive frames), the stimulus disappeared and the screen remained gray until the subject made a response.

The eye movements of the participants in the experiment were recorded by a specialized hardware – Jazz novo eye tracking system (Ober Consuting Sp. Z o.o.). All recordings from all the sensors of the device for one session per person were collected with 1 KHz frequency and the information is stored in files. These include: the calibration information; records of horizontal and vertical eye positions in degrees of visual angle eye_x and eye_y; screen sensor signal for presence/absence of a stimulus on the monitor; microphone signal recording sounds during the experiment; information about tested subjects (code) and type of the experimental trail for each particular record.

The raw data were processed to extract only the records during presence of a stimulus on the screen. The data between the stimuli was excluded since it is not relevant to the eye movements during task performance.

Three age groups took part in the experiment: young (from 20 to 35 years), elderly (from 57 to 84 years) and middle age group (from 25 to 55 years). From all collected experimental data we observed big variety of eye movement behaviors varying not only between three age groups but also within each group. So it was very hard to classify test subjects only on the basis of this information. Hence we decided to try whether the proposed above dynamic data discrimination approach can yield some reasonable results.

The input to the ESN was two dimensional vector composed by the visual angles data series recorded during presence of a stimulus on the screen, i.e. $in(k) = [eye_x(k) \quad eye_y(k)]$. We tuned three ESNs with reservoir sizes 10, 50 and 100 neurons using the IP algorithm described above. The extracted in this way feature of each dynamic data series was R calculated according to the equation (5). Thus the tested subjects were ordered based on the obtained value of R from their recorded eye movement data series.

4 Classification Results and Discussion

First we selected a representative group of four experienced test subjects from different age groups. These subjects took part in experimental set-up preparation so they were able to perform the behavioral tasks strictly and their eye movement recordings were clear from outliers due to improper behavior like looking to the mouse before

clicking when decision was taken or keeping fixation. Such noisy behavior was observed with other volunteer subjects especially during their first trails.

Fig. 2 represents the eye movements' data series collected from these four "experienced" test subjects who performed first the described above experiment. Subject 1 is the youngest, subject 3 is middle aged and other two subjects (2 and 4) belong to the elderly group.
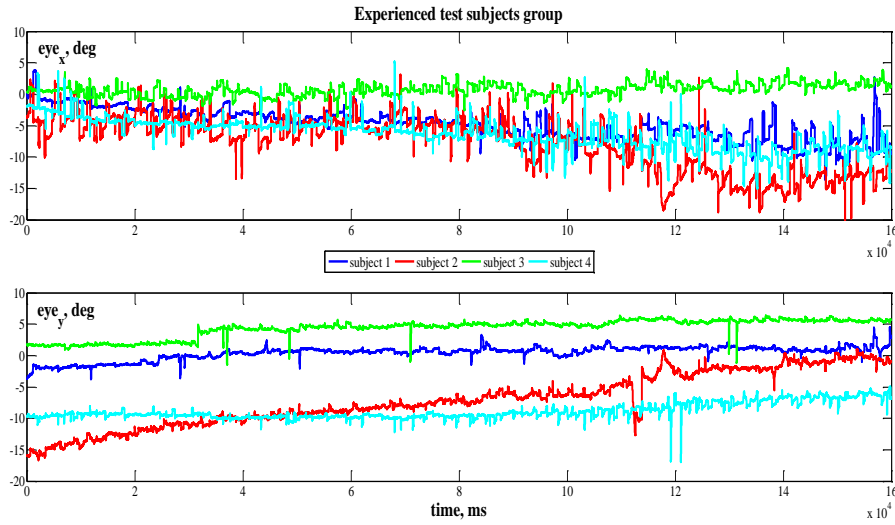

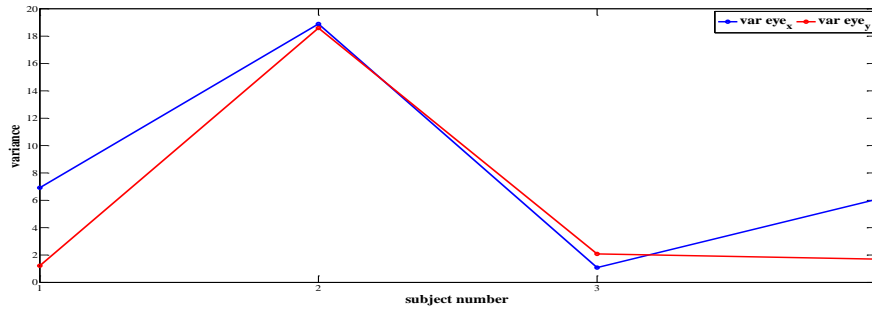**Fig. 2.** Experienced 4 subjects' recordings from eye tracker.

Fig. 3. Variances of eye movement coordinates from Fig. 2.

We can easily distinguish the middle aged subject while similarities between subjects 2 and 4 are not so obvious. From Fig. 3, representing the variances of the data series, we can conclude that age differentiation by this characteristic is also a hard task even for such small group.

Then we applied described above classification approach to these data series. In order to prove the expected effect of IP tuning of the ESN reservoir, we compare the obtained feature value before (Fig. 4) and after (Fig. 5) its application to the three randomly generated ESN reservoirs containing 10, 50 and 100 neurons.

From Fig. 5 we can conclude that IP tuning definitely helps to classify our four subjects by their age no matter of the reservoir size.
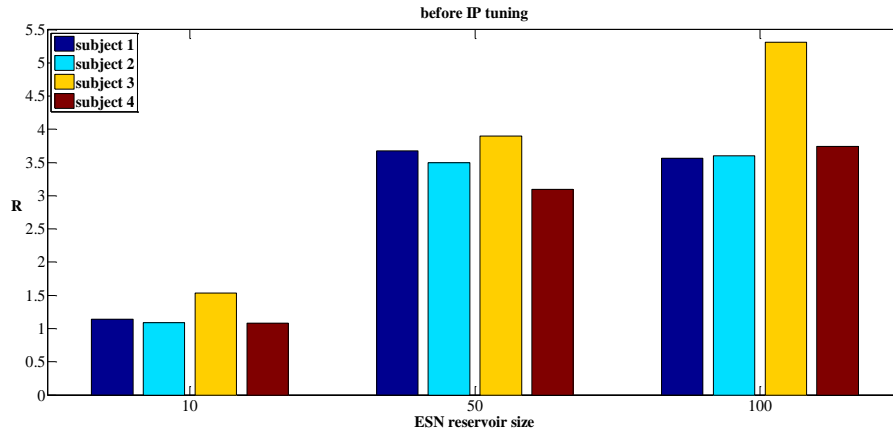

Fig. 4. Size of the reservoir state vector R achieved after presentation of input time series for four experienced test subjects before IP tuning.
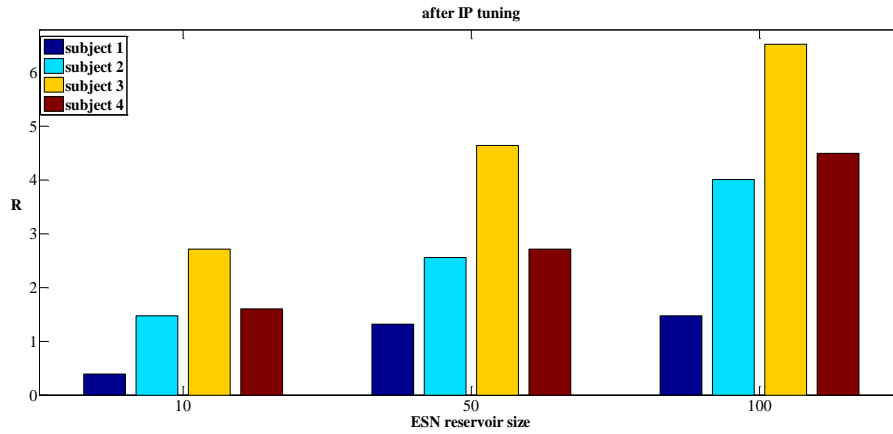

Fig. 5. Size of the reservoir state vector R achieved after presentation of input time series for four experienced test subjects after IP tuning.

As it was observed, the middle age subject was the best during experiments and his eye movements on Fig. 2 were significantly different from the other persons in the

group. This was confirmed by our algorithm according to which the middle aged subject 3 is clearly differentiated from the other three experienced subjects of different ages. Moreover, the two elderly subjects 2 and 4 were classified as close to each other while the younger person 1 was clearly differentiated too.
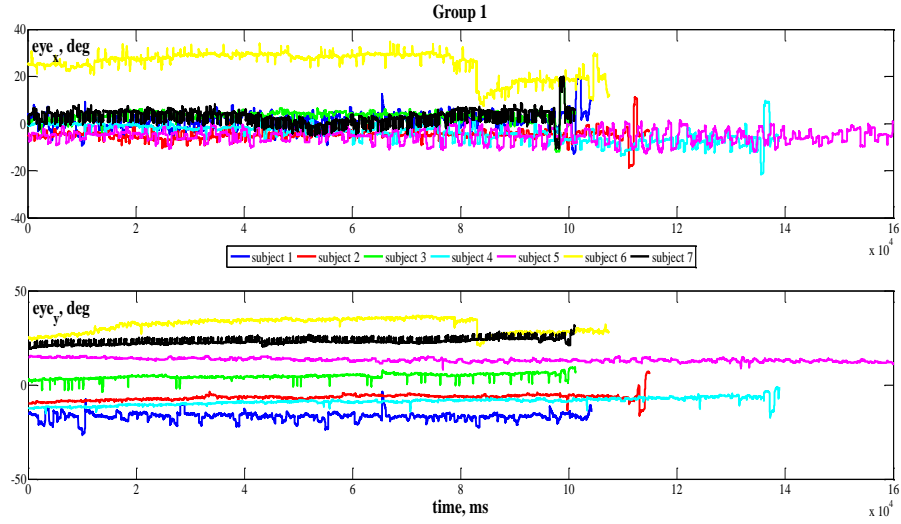

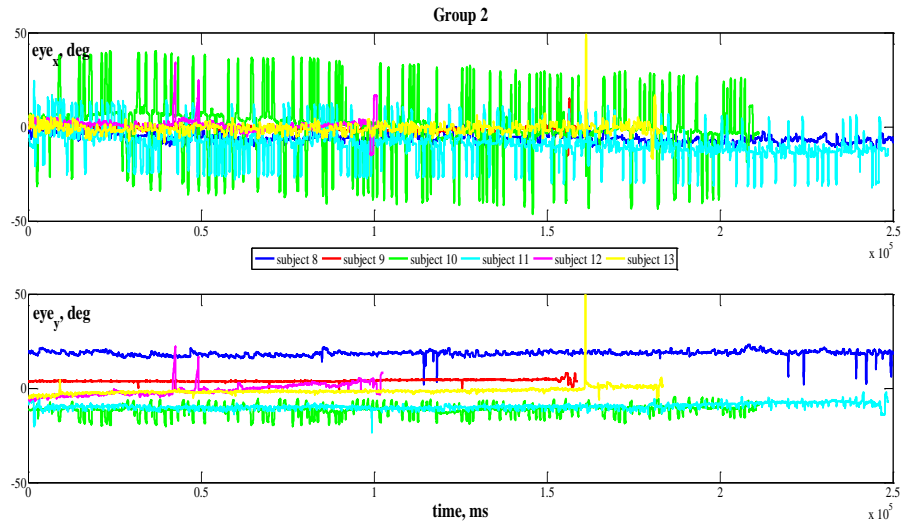**Fig. 6.** Eye movements recorded from young test subjects.

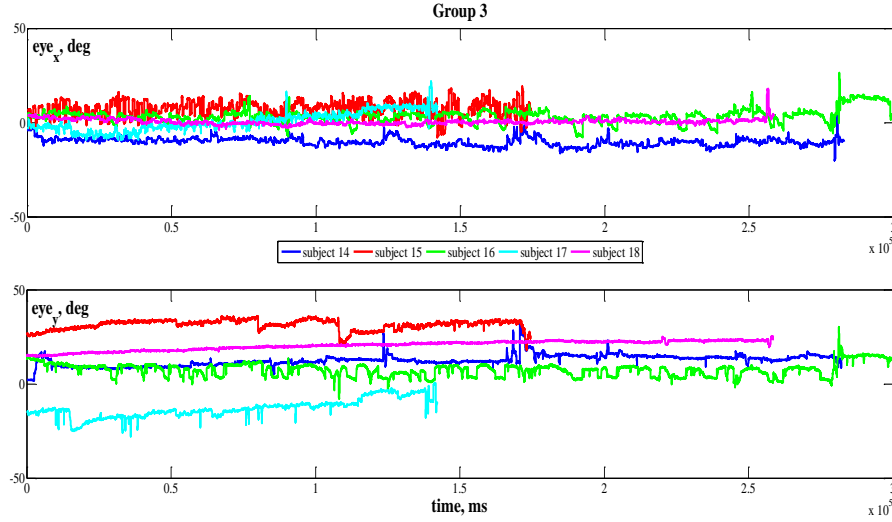**Fig. 7.** Eye movements recorded from the middle aged test subjects.

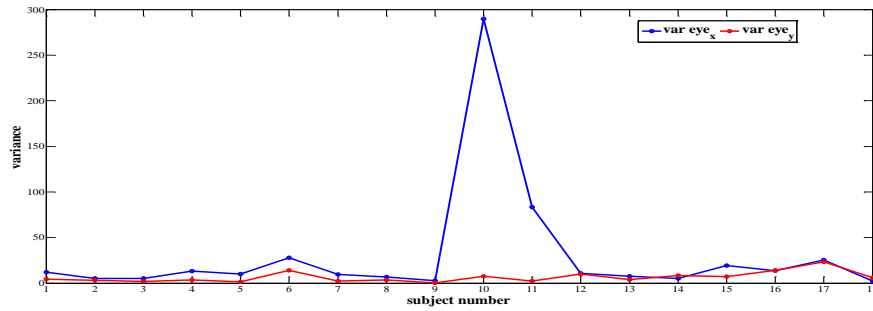**Fig. 8.** Eye movements recorded from the elderly test subjects.

Fig. 9. Variances of eye movement coordinates for all test subjects.

Then we proceed with data recorded from 18 volunteers from three age groups as shown on figures 6, 7 and 8 respectively. Fig. 9 represents the variances of all data series from figures 6-8.

Although we observe some similarities and differences between the three groups, there are also significant dissimilarities between subjects from the same group. For example, we can conclude that young subject 6 behaved significantly different from other members of Group 1; in Group 2 we can distinguish subjects 10 and 11 from others while in Group 3 subject 15 seems to have different eye movement behavior.

Here we decided to IP tune two ESN reservoirs (denoted here as ESN_1 and ESN_2) from each size since initial connections of ESN were randomly generated, and to compare the obtained results.
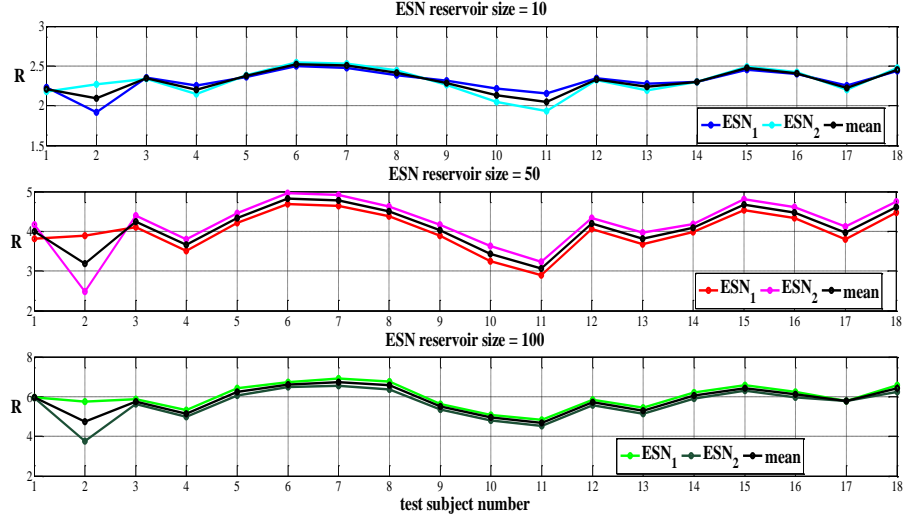
Fig. 10. Size of the reservoir state vector *R* for the three ESN reservoirs and all test subjects. ESN₁ and ESN₂ denote the results from the first and second ESN respectively; mean is the mean value between ESN₁ and ESN₂.
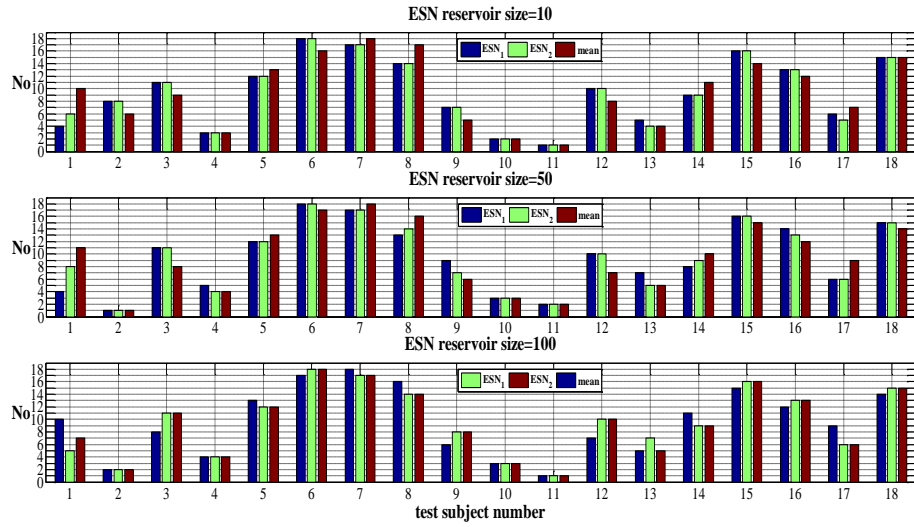

Fig. 11. Rank of the subjects according to the size of the reservoir vector *R* for the three ESN reservoirs and all test subjects. ESN₁ and ESN₂ denote the results from the first and second ESN respectively; mean is their mean value.

Fig. 10 represents the achieved values of R from ESN_1 and ESN_2 as well as its mean value. It is clear that for both initial ESN reservoirs we obtained similar results after IP tuning, especially for the middle group.

Fig. 11 represents the corresponding rank number of the subjects in dependence on the value of R from Fig. 10. The obtained subject order is similar for both ESN initial reservoirs. For most of the subjects two generations of ESN yielded the same rank. The same are the results from the ranking according to the mean value of R. The bigger differences are observed for the first and second subjects only.

However the obtained results showed that differentiation of tested subjects by age using only recorded eye movement behaviors is not possible since there are subjects from different age groups that received close ranks. Nevertheless, the approach seems promising for classification of types of eye movements during decision making that could be related to other psycho-physiological peculiarities of the tested subjects.

5 Conclusions

In conclusion, the proposed approach for classification of time series using the geometric size of the ESN reservoir vector state appears promising since it decreases significantly computational burden of the algorithm.

Concerning the classification of human subjects based on their eye movements, it is obvious that obtained ranking is not enough to separate the people into the groups related to their age. Since the achieved ranking of subjects remain stable using different initializations and sizes of ESN reservoir, we can search other similarities between tested subjects that this classification might reveal.

Another explanation of the reported results might be presence of outliers in the collected data since in these preliminary investigations the raw data was used. Hence further refinement of the collected experimental data base could help to reveal some age-related similarities of the recorded eye movements.

Another direction for future work can be inclusion of the additional information collected during the experiments like amplitude, velocity and acceleration of saccades performed during decision making as well as accuracy of the persons' response and the corresponding reaction time for each individual stimulus. All these characteristics can serve as features to support classification of test subjects in our future work.

Acknowledgment

12

References

1. Koprinkova-Hristova, P., Tontchev, N.: Echo state networks for multi-dimensional dataclustering. In: Villa, A.E.P., Duch, W., Érdi, P., Masulli, F., Palm, G. (eds.) Int. Conf. on Artificial Neural Networks 2012, LNCS vol. 7552, pp. 571–578. Springer, Heidelberg (2012)
2. Jaeger, H.: Tutorial on training recurrent neural networks, covering BPPT, RTRL, EKF and the "echo state network" approach. GMD Report 159, German National Research Center for Information Technology (2002)
3. Lukosevicius, M., Jaeger, H.: Reservoir computing approaches to recurrent neural network training. Computer Science Review 3, 127–149 (2009)
4. Koprinkova-Hristova, P.: Multi-dimensional data clustering and visualization via Echo state networks. In: Kountchev, R., Nakamatsu, K. (eds.) New Approaches in Intelligent Image Analysis, Intelligent Systems Reference Library vol. 108, pp. 93–122. Springer, Cham (2016)
5. Koprinkova-Hristova, P., Alexiev, K.: Echo state networks in dynamic data clustering. In: V. Mladenov et al. (eds.) Int. Conf. on Artificial Neural Networks 2013, LNCS vol. 8131, pp. 343–350. Springer, Heidelberg (2013)
6. Koprinkova-Hristova, P.: On effects of IP improvement of ESN reservoirs for reflecting of data structure. In: Proc. of the International Joint Conference on Neural Networks (IJCNN) 2015, IEEE, Killarney, Ireland, DOI: 10.1109/IJCNN.2015.7280703 (2015)
7. Steil, J.J.: Online reservoir adaptation by intrinsic plasticity for back-propagation-decoleration and echo state learning. Neural Networks 20, 353–364 (2007)
8. Schrauwen, B., Wandermann, M., Verstraeten, D., Steil, J.J., Stroobandt, D.: Improving reservoirs using intrinsic plasticity. Neurocomputing 71, 1159–1171 (2008)
9. Bozhkov, L., Koprinkova-Hristova, P., Georgieva, P.: Reservoir computing for emotion valence discrimination from EEG signals. Neurocomputing 231, 28–40 (2017)