IMPROVING LEARNING ABILITY OF RECURRENT NEURAL NETWORKS
Experiments on Speech Signals of Patients with Laryngopathies

Jarosław Szkoła, Krzysztof Pancerz
Institute of Biomedical Informatics, University of Information Technology and Management, Rzeszów, Poland

Jan Warchoł
Medical University of Lublin, Lublin, Poland

Keywords: Recurrent neural networks, Learning of neural networks, Laryngopathies, Temporal patterns.

Abstract: Recurrent neural networks can be used for pattern recognition in time series data due to their ability of memorizing some information from the past. The Elman networks are a classical representative of this kind of neural networks. In the paper, we show how to improve learning ability of the Elman network by modifying and combining it with another kind of a recurrent neural network, namely, with the Jordan network. The modified Elman-Jordan network manifests a faster and more exact achievement of the target pattern. Validation experiments were carried out on speech signals of patients with laryngopathies.

1 INTRODUCTION

Our research concerns designing effective methods for computer support of a non-invasive diagnosis of selected larynx diseases. Computer-based clinical decision support (CDS) systems play an important role in modern medicine (Greenes, 2007). The non-invasive diagnosis is based on an intelligent analysis of distinguished parameters of a patient's speech signal (phonation). Some approaches were considered in (Warchoł, 2006), (Szkoła et al., 2010), and (Warchoł et al., 2010). There exist various approaches to analysis of bio-medical signals (cf. (Semmlow, 2009)). In general, we can distinguish three groups of methods according to a domain of the signal analysis: analysis in a time domain, analysis in a frequency domain (spectrum analysis), analysis in a time-frequency domain (e.g., wavelet analysis). Therefore, in our research, we are going to build a specialized computer tool for supporting diagnosis of laryngopathies based on a hybrid approach. One part of this tool, playing an important role in a preliminary stage, will be based on the patients' speech signal analysis in the time domain. Hybridization means that a decision support system will have a hierarchical structure based on multiple classifiers working on signals in time and frequency domains. Results presented in this paper will be helpful for selection of suitable methods for the planned tool.

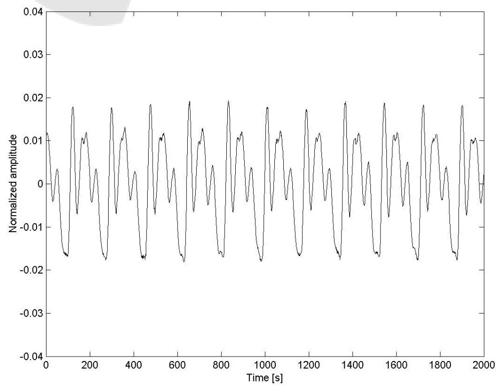

Figure 1: An exemplary speech signal (fragment) for a patient from a control group.

Preliminary observations of signal samples for patients from a control group and patients with a confirmed pathology clearly indicate deformations of standard articulation in precise time intervals (see Figures 1 and 2).

Designing the way of recognition of temporal patterns and their replications becomes the key element. It enables detecting all non-natural disturbances in articulation of selected phonemes. For the time domain

Szkoła J., Pancerz K. and Warchoł J..
IMPROVING LEARNING ABILITY OF RECURRENT NEURAL NETWORKS - Experiments on Speech Signals of Patients with Laryngopathies.
DOI: 10.5220/0003292603600364
In *Proceedings of the International Conference on Bio-inspired Systems and Signal Processing* (BIOSIGNALS-2011), pages 360-364
ISBN: 978-989-8425-35-5

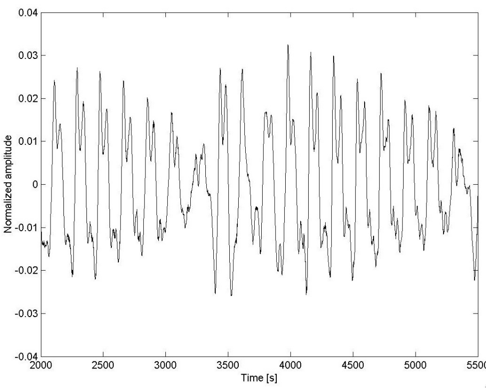

Figure 2: An exemplary speech signal (fragment) for a patient with polyp.

analysis, we propose using neural networks with the capability of extracting the phoneme articulation pattern for a given patient (articulation is an individual patient feature) and the capability of assessment of its replication in the whole examined signal. Preliminary observations show that significant replication disturbances in time, appear for patients with the clinical diagnosis of disease.

The capabilities mentioned are possessed by recurrent neural networks. One class of them are the Elman neural networks (Elman, 1990). In real time-decision making, an important thing is to speed up a learning process for neural networks. Moreover, accuracy of learning of signal patterns plays an important role. Therefore, in this paper, we propose some improvement of learning ability of the Elman networks by combining them with another kind of recurrent neural networks, namely, the Jordan networks (Jordan, 1986) and by providing some additional modification.

Our paper is organized as follows. After introduction, we shortly describe a structure and features of the modified Elman-Jordan neural network used for supporting diagnosis of laryngopathies (Section 2). In Section 3, we present results obtained by experiments done on real-life data. Some conclusions and final remarks are given in Section 4.

2 RECURRENT NEURAL NETWORKS

In most cases, neural network topologies can be divided into two broad categories: feedforward (with no loops and connections within the same layer) and recurrent (with possible feedback loops). The Hopfield network, the Elman network and the Jordan network are the best known recurrent networks. In the paper we are interested in the two last ones.

In the Elman network (Figure 3) (Elman, 1990), the input layer has a recurrent connection with the hidden layer. Therefore, at each time step the output values of the hidden units are copied to the input units, which store them and use them for the next time step. This process allows the network to memorize some information from the past, in such a way to detect periodicity of the patterns in a better manner. Such capability can be exploited in our problem to recognize temporal patterns in the examined speech signals. The Jordan networks (Jordan, 1986) are similar to the Elman networks. The context layer is, however, fed from the output layer instead of the hidden layer. To accelerate a learning (training) process of the Elman neural network we propose a modified structure of the network. We combine the Elman network with the Jordan network and add another feedback for an output neuron as it is shown in Figure 4.

The pure Elman network consists of four layers:

- an input layer (in our model: the neuron I_1),
- a hidden layer (in our model: the neurons H_1, H_2, ..., H_{40}),
- a context layer (in our model: the neurons C_1, C_2, ..., C_{40}),
- an output layer (in our model: the neuron O_1).

z^{-1} is a unit delay here.

To improve some learning ability of the pure Elman networks, we propose to add additional feedbacks in network structures. Experiments described in Section 3 validate this endeavor. We create (see Figure 4):

- feedback between an output layer and a hidden layer through the context neuron (in our model: the neuron C_{41}), such feedback is used in the Jordan networks,
- feedback for an output layer.

A new network structure will be called the modified Elman-Jordan network.

The Elman network, according to its structure, can store an internal state of a network. There can be values of signals of a hidden layer in time unit $t-1$. Data are stored in the memory context. Because of storing values of a hidden layer for $t-1$ we can make prediction for the next time unit for a given input value. In the case of learning neural networks with different architectures, we can distinguish three ways for making prediction for $x(t+s)$, where $s > 1$:

1. Training a network on values $x(t)$, $x(t-1)$, $x(t-1)$, ...

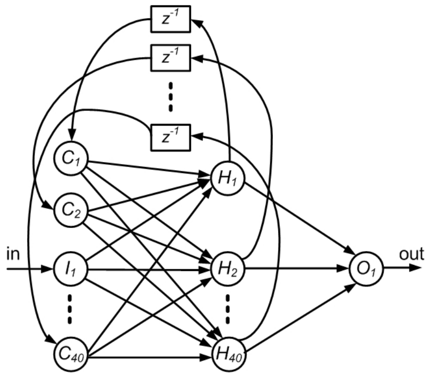

Figure 3: A structure of the trained Elman neural network.

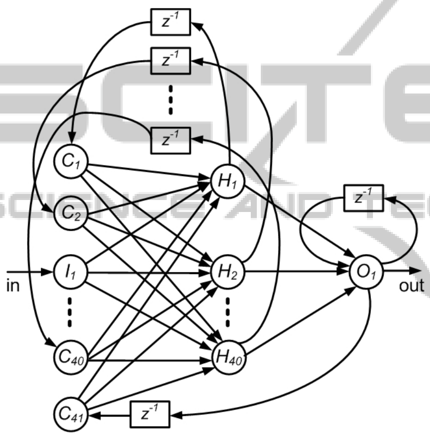

Figure 4: A structure of the trained Elman-Jordan neural network.

2. Training a network on each value $x(t+i)$, where $1 \geq i \geq s$. This way manifests good results for small s.
3. Training a network only on a value $x(t+1)$, going iteratively to $x(t+s)$ for any s.

In our case, we have used method 2.

The Jordan network can be classified as one of variants of the NARMA (Nonlinear Autoregressive Moving Average) model (Mandic and Chambers, 2001), where a context layer stores an output value for $t-1$. It is assumed that a network with this structure does not have a memory. It processes only a value taken previously from the output. In the NARMA model, a context layer operates as a subtractor for an input value.

If we pass a single value to the network input in a given time unit t, then the Elman network stores the copies of values from a hidden layer for $t-1$ in a context layer. The size of a hidden layer does not depend on the size of an output layer. In the case of the Jordan network, an output value for $t-1$ is passed to a context layer. Therefore, the size of this layer depends on the size of an output layer. If a network has only one input and one output, then we have only one neuron in the context layer. In comparison with the Elman network, the Jordan network learns slower and requires a bigger structure. Therefore, the pure Jordan network cannot be used in solving our problem. In the modified Elman-Jordan network proposed by us, the network has feedbacks between selected layers. We provide additional information for the hidden layer. The hidden layer has an access to an input value, previous values of the hidden layer as well as an output value. Additional information has a big impact on modifying weights of the hidden layer. It leads to shortening a learning process and decreasing a network structure compared to the classical Elman network.

3 EXPERIMENTS

The experiments were carried out on two groups of patients. The first group included patients without disturbances of phonation. It was confirmed by phoniatrist opinion. The second group included patients with clinically confirmed dysphonia as a result of Reinke's edema or laryngeal polyp. Experiments were carried out by a course of breathing exercises with instruction about a way of articulation. The task of all examined patients was to utter separately some vowels with extended articulation as long as possible, without intonation, and each on separate expiration.

The obtained speech signals were analyzed using the Elman network and the modified Elman-Jordan network. Articulation is an individual patient feature. Therefore, we cannot train a neural network on the independent patterns of phonation of individual vowels. For each patient, a recorded speech signal was used for both training and testing of a neural network. We can distinguish the following steps of our computer procedure:

- Division of the speech signal of an examined patient into time windows corresponding to phonemes.
- Random selection of a number of time windows.
- Taking one time window for training the neural network and testing of the neural network on the remaining ones.

The network learns a selected time window. If the remaining windows are similar to the selected one in terms of the time patterns, then for such windows an error generated by the network in a testing stage is small. If significant replication disturbances in time appear for patients with the larynx disease, then an

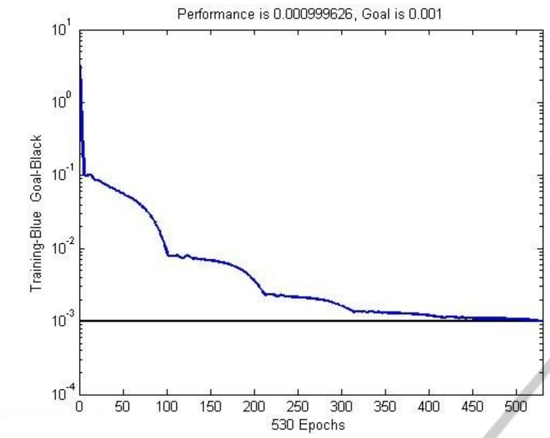

Figure 5: An exemplary learning process for the Elman network.

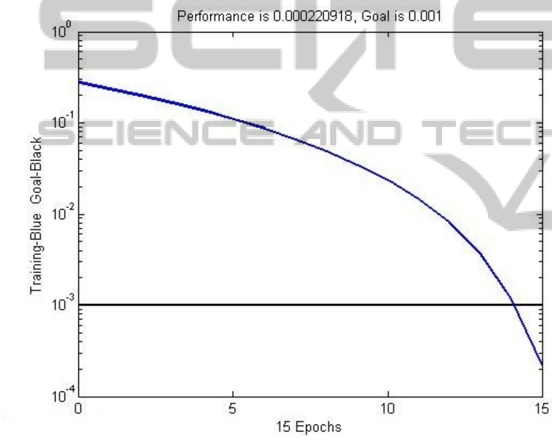

Figure 6: An exemplary learning process for the modified Elman-Jordan network.

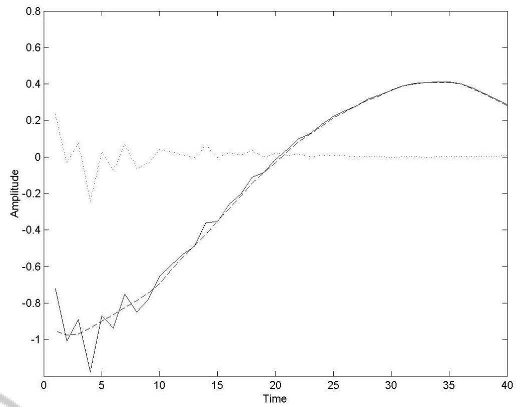

Figure 7: Exemplary errors for the Elman network (desired output - dashed line, obtained output - solid line, error - dotted line).

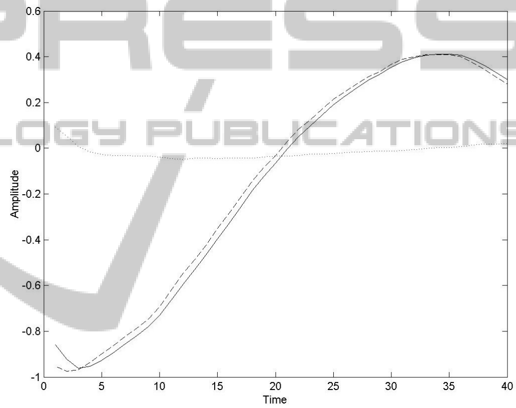

Figure 8: Exemplary errors for the modified Elman-Jordan network (desired output - dashed line, obtained output - solid line, error - dotted line).

error generated by the network is greater. In this case, the time pattern is not preserved in the whole signal. Therefore, the error generated by the network reflects non-natural disturbances in the patient phonation.

In our experiments. we used network structures shown in Section 2. In each case, we required the learning error less than or equal to 0.001. Figures presented further show important features of neural network models used in experiments. In Figures 5 and 6, exemplary learning processes for both neural networks are shown. For the Elman network, we can observe fast correction of weights (at the beginning of a learning process) and next (in the second phase) slow decreasing of an error. Sometimes an error increases. For the modified Elman-Jordan network, at the beginning, an error slowly but systematically decreases, and next (in the second phase) it significantly decreases.

In Figures 7 and 8, the quality of mapping of patterns learnt by networks is shown for both neural networks. For the modified Elman-Jordan network, the learnt pattern is smoother than for the pure Elman network. It leads to better and more exact mapping of a shape of the pattern. This feature is very important in pattern recognition and pattern searching in the whole group of examined signals.

It is worth noting that presented graphs are representative of most cases examined, but there exist exceptions. An important thing is a number of epochs needed for learning a neural network. Tables 1 and 2 include selected results of experiments for women from the control group and women with laryngeal polyp, obtained using the Elman network and the modified Elman-Jordan network. We give consecutively an average mean squared error $\overline{e}_{EN}$ ($\overline{e}_{EJN}$) and an average number $\overline{n}_{EN}$ ($\overline{n}_{EJN}$) of epochs needed to learn the network.

It is easy to see that neural network models have

Table 1: Selected results of experiments for women from the control group obtained using the Elman network and the modified Elman-Jordan network.

ID	$\overline{e}_{EN}$	$\overline{n}_{EN}$	$\overline{e}_{EJN}$	$\overline{n}_{EJN}$
w_{CG1}	0.0068	389	0.0061	88
w_{CG2}	2.4523	2335	0.0111	92
w_{CG3}	0.0170	501	0.0178	107
w_{CG4}	0.0109	597	0.0115	96
w_{CG5}	0.0332	662	0.0301	146
w_{CG6}	0.0178	609	0.0166	104
w_{CG7}	0.0096	428	0.0086	78
w_{CG8}	0.0068	318	0.0068	108
w_{CG9}	0.0080	490	0.0080	162
w_{CG10}	0.0084	553	0.0087	119

Table 2: Selected results of experiments for women with laryngeal polyp obtained using the Elman network and the modified Elman-Jordan network.

ID	$\overline{e}_{EN}$	$\overline{n}_{EN}$	$\overline{e}_{EJN}$	$\overline{n}_{EJN}$
w_{P1}	0.1720	331	0.1677	92
w_{P2}	0.2764	536	0.3107	191
w_{P3}	0.0518	566	0.0542	96
w_{P4}	0.0268	504	0.0258	142
w_{P5}	0.0418	646	0.0423	239
w_{P6}	0.2107	444	0.2134	71
w_{P7}	0.0921	1040	0.0877	40
w_{P8}	0.0364	993	0.0351	72
w_{P9}	0.0380	541	0.0370	180
w_{P10}	0.1461	363	0.1411	160

similar ability to distinct between normal and disease states (compare columns $\overline{e}_{EN}$ and $\overline{e}_{EJN}$), but the modified Elman-Jordan network needs a smaller number of epochs, sometimes by 50 per cent or more (compare columns $\overline{n}_{EN}$ and $\overline{n}_{EJN}$). Such observations are very important for further research, especially in the context of a created computer tool for diagnosis of larynx diseases. The experiments also showed that sometimes the Elman network does not have the capability of learning a given pattern. Ten thousand epochs set in experiments were not enough for learning with an error goal of 0.001. Such situation was observed for w_{CG2} (see Table 1). The modified Elman-Jordan network reached a goal in a set number of epochs for each sample.

4 CONCLUSIONS

In the paper, we have shown improving learning ability of recurrent neural networks used in speech signal analysis of patients with laryngopathies. A combined structure of two neural networks (the Elman network with the Jordan network) speeds up a learning process and that is important if a diagnostic decision should be made in real time. The presented results will be helpful for selection of suitable techniques for a created computer tool supporting diagnosis of larynx diseases.

ACKNOWLEDGEMENTS

This research has been supported by the grant No. N N516 423938 from the Polish Ministry of Science and Higher Education.

REFERENCES

Elman, J. (1990). Finding structure in time. *Cognitive Science*, 14:179–211.

Greenes, R. (2007). *Clinical Decision Support. The Road Ahead*. Elsevier Inc.

Jordan, M. (1986). Serial order: A parallel distributed processing approach. Technical Report 8604, University of California, San Diego, Institute for Cognitive Science.

Mandic, D. and Chambers, J. (2001). *Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability*. John Wiley & Sons, Ltd.

Semmlow, J. (2009). *Biosignal and Medical Image Processing*. CRC Press.

Szkoła, J., Pancerz, K., and Warchoł, J. (2010). Computer-based clinical decision support for laryngopathies using recurrent neural networks. In *Proc. of the 10th International Conference on Intelligent Systems Design and Applications (ISDA'2010)*, Cairo, Egypt.

Warchoł, J. (2006). *Speech Examination with Correct and Pathological Phonation Using the SVAN 912AE Analyser (in Polish)*. PhD thesis, Medical University of Lublin.

Warchoł, J., Szkoła, J., and Pancerz, K. (2010). Towards computer diagnosis of laryngopathies based on speech spectrum analysis: A preliminary approach. In Fred, A., Filipe, J., and Gamboa, H., editors, *Proc. of the Third International Conference on Bio-inspired Systems and Signal Processing (BIOSIGNALS'2010)*, pages 464–467, Valencia, Spain.