

A Reduction Algorithm for Analysis of Linear Neuron

Mohammad Dweib and Yousef Abuzir
Department of Computer Information Systems,
Faculty of Technology and Applied Sciences,
Al-Quds Open University, Bethelehem, Palestine

Abstract: Problem statement: The aim of feature selection is to select a feature set that is relevant for a given application. Feature selection is complex and remains an important issue in many domains. In the field of neural networks, feature selection has been used in many applications and their methods have been employed. In this study we present neural network approaches to feature selection. **Approach:** In this study a reduction algorithm of the features vector dimension was described by eliminating its selected components on the basis of analyzing the results of teaching a neuron, which has a linear activation function of the type. In the presented algorithm, the value of the mean square error, which appears after the reduction, is the criterion, on the basis of which the components of the eliminated vector were selected. The algorithm is based on the analysis of the classifier of balances vector. **Results:** The results of calculations obtained when analyzing the data describing an example-task of medical diagnosis were presented as an illustration. Results from the experiment indicate that the elimination of components of features vector using the Reduction algorithm did not cause the increase of the value of mean square. **Conclusion:** Our study showed that, results provide experimental evidences on the effectiveness of the proposed approach for feature selection in the bioinformatics applications.

Key words: Neural network, reduction algorithm, regression analysis, factor analysis, feature selection

INTRODUCTION

Selecting features in the tasks of objects recognition is a well known operation and the use of which can be differently justified. It is obvious that each object is described by a wide set of parameters, which do not all have to be used in order to achieve the classification with a satisfactory result. Parts of them is described at the beginning and are not taken into consideration. There could be many reasons, because they differentiate the objects in an insignificant way, or because their measurement is too complicated or expensive, or because they obtained measuring-results are burdened with too large noise, or finally because the time, during which data could be obtained exceeds the time, which is assigned for taking the decision (Kekre *et al.*, 2010). The question about the criterion, which should be used when selecting the set of features used to analyse the given task, remains an essential question.

The base for the initial selection of features is of course the intuitive selecting of features, performed by the expert who knows the problem. In the selected collection it is still possible to indicate more or less essential parameters (Khor *et al.*, 2009). However, to do

that, an algorithm which would make it possible to perform such analysis, is needed.

Many algorithms exist now, which can be used to solve this problem. We can at least mention the factor analysis and regression analysis. Having a selected set of features, we can start projecting a suitable recognition algorithm, which will take decision on this basis. Usually, the acquaintance of all elements of the features vector is indispensable for the proper work of the classifier and introducing incomplete data will cause considerable distortion of the response, usually disqualifying such result. An attempt of taking the decision, having incomplete data, forces to repeat the construction of the classifier. Therefore, it seems interesting to work out an algorithm, which would indicate features that could be neglected when taking decision, without deteriorating the recognition quality (or with a slight, acceptable deterioration) without the necessity to repeat the teaching process.

Background: The main idea of feature selection is the process of choosing a subset of input variables or features relevant to the bioinformatics applications. In the feature selection process, a decision criterion is used

Corresponding Author: Yousef Abuzir, Department of Computer Information Systems, Faculty of Technology and Applied Sciences, Al-Quds Open University, P.O. Box 51800, Postal Code 97917, Jerusalem, Palestine

to eliminating irrelevant or redundant features or no predictive information (Romero and Sopena, 2008). Saeys *et al.* (2007) give a review of feature selection techniques in bioinformatics. There are many contributions of feature selection research in a set of well-known bioinformatics applications (Yang *et al.*, 2010; Ahmed *et al.*, 2011).

Recently, there has been a great deal of interest in the use of Multilayer Perceptron, Radial Basis Function and Support Vector Machine as classifiers in pattern recognition problems and in the feature selection methodology (Saeys *et al.*, 2007).

Multilayer Perceptrons (MLPs) (Abdullah *et al.*, 2010; Amaroek *et al.*, 2010) are supervised networks so they require a desired response to be trained. They learn how to transform input data into a desired response, so they are widely used for pattern classification. They are feed forward neural networks trained with the standard backpropagation algorithm they have been shown to approximate the performance of optimal statistical classifiers in difficult problems.

Radial Basis Function (RBF) (Ajeel, 2010) networks have a static Gaussian function as the nonlinearity for the hidden layer processing elements. The Gaussian function responds only to a small region of the input space where the Gaussian is centred. The advantage of the radial basis function network is that it finds the input to output map using local approximations.

The final learning linear algorithms proposed in recent years is the Support Vector Machine (Gomathi and Thangaraj, 2010; Bharathi and Natarajan, 2011). The main advantage of the Support Vector Machine (SVM) is that its training is performed through the solution of a linearly constrained convex quadratic programming problem. Therefore, an efficient algorithm can find an approximate solution in a finite number of steps.

In the following part a features selection algorithm is presented and is based on the analysis of the balances vector of linear neuron. The basic assumption is not to deteriorate the quality of the classifier work based on such a neuron, measured by the value of mean square. The examined problem does not refer to the construction method of the recognition algorithm, so the aspects of teaching the linear neuron will be neglected while presenting the algorithm. The action of such classifier in the conditions of incomplete collection of features is substantial.

MATERIALS AND METHODS

In order to explain the used symbols, let us introduce the following assumption. The features selection algorithm is based on analysis of the balances

vector of linear neuron, which is taught on the basis of a teaching series that can be presented as follows Eq. 1:

$$S_N = \{(x_1, z_1), (x_2, z_2), \dots, (x_N, z_N)\} \quad (1)$$

where, $x_i = [x^{(1)}, x^{(2)}, \dots, x^{(r)}]^T$ is the i -vector of measured parameters (features vector of r dimension) from the collection of standards consisting of N elements, while z_i is the corresponding classification result given by the expert.

To illustrate this, we consider a case of recognition with two classes (the dichotomy task), in which the decision algorithm would be as follows Eq. 2:

$$\Psi(x) = \begin{cases} 1 & \text{where } y(x) > 0,5 \\ 2 & \text{where } y(x) \leq 0,5 \end{cases} \quad (2)$$

where, $y(x)$ is the response of linear neuron described as Eq. 3:

$$y(x) = \sum_{i=1}^r w^{(i)} x^{(i)} \quad (3)$$

In the above equation $w^{(i)}$ is the t -component of balances vector. The balances vector is equal to the features vector. The values of each of its components are selected during the teaching process, which will be ignored here.

If at the entrance of the classifier we give a features vector belonging to the teaching series (or the resting one), knowing the correct response we will be able to find the value of the error done during the presentation (of the i -standard) Eq. 4:

$$\delta_i = z_i - y_i = z_i - \sum_{i=1}^r w_i^{(i)} x_i^{(i)} \quad (4)$$

Repeating this process for all images in the teaching series, we will obtain the value of mean square error, which is widely used as a criterion to estimate the quality of the classifier action Eq. 5:

$$Q = \sum_{i=1}^N Q_i = \sum_{i=1}^N \frac{1}{2} \delta_i^2 = \sum_{i=1}^N \frac{1}{2} (z_i - y_i)^2 \quad (5)$$

While constructing a selection algorithm which would indicate features that can be eliminated from the features vector without deteriorating the recognition quality, it is interesting how the value of this error will change. While calculating this value let us assume that the elimination of k -component of the features vector means replacing it with zero ($x(k)=0$). Having such assumption, the previous equation can be presented as follows Eq. 6-8:

$$y^{(-k)}(x) = \sum_{\substack{t=1 \\ t \neq k}}^r w^{(t)} x^{(t)} \quad (6)$$

$$\delta_i^{(-k)} = z_i - y_i^{(-k)} = z_i - \sum_{\substack{t=1 \\ t \neq k}}^r w_i^{(t)} x_i^{(t)} \quad (7)$$

$$Q^{(-k)} = \sum_{i=1}^N Q_i^{(-k)} = \sum_{i=1}^N \frac{1}{2} \delta_i^{(-k)^2} = \sum_{i=1}^N \frac{1}{2} (z_i - y_i^{(-k)})^2 \quad (8)$$

In the above equations, the variable on top of which (-k) appeared, stands for the value of this variable at $x^{(k)} = 0$. As it was indicated in the introduction, to achieve elimination of a selected feature from the features vector, it is important that the value of the error (5) doesn't arise by more than a selected quantity Eq. 9:

$$Q^{(-k)} - Q \leq \alpha \quad (9)$$

where, α is the acceptable increase of the mean square error.

While solving the above inequality let us assume that $\alpha = 0$ and substitute Eq. 10:

$$\begin{aligned} y^{(-k)}(x) &= \sum_{\substack{t=1 \\ t \neq k}}^r w^{(t)} x^{(t)} = \sum_{t=1}^r w^{(t)} x^{(t)} - w^{(k)} x^{(k)} \\ &= y(x) - w^{(k)} x^{(k)} \end{aligned} \quad (10)$$

After putting and simple transformations we obtain Eq. 11:

$$w^{(k)} \left(w^{(k)} \sum_{i=1}^N (x_i^{(k)})^2 + 2 \sum_{i=1}^N x_i^{(k)} \delta_i \right) \leq 0 \quad (11)$$

The following should be fulfilled so that the inequality would be genuine Eq. 12:

$$w^{(k)} \in \left[\underline{w}^{(k)}, \overline{w}^{(k)} \right] = \begin{cases} \left[0, \frac{2 \sum_{i=1}^N x_i^{(k)} \delta_i}{\sum_{i=1}^N (x_i^{(k)})^2} \right] \\ \text{where } \sum_{i=1}^N x_i^{(k)} \delta_i \geq 0 \\ \left[\frac{2 \sum_{i=1}^N x_i^{(k)} \delta_i}{\sum_{i=1}^N (x_i^{(k)})^2}, 0 \right] \\ \text{where } \sum_{i=1}^N x_i^{(k)} \delta_i < 0 \end{cases} \quad (12)$$

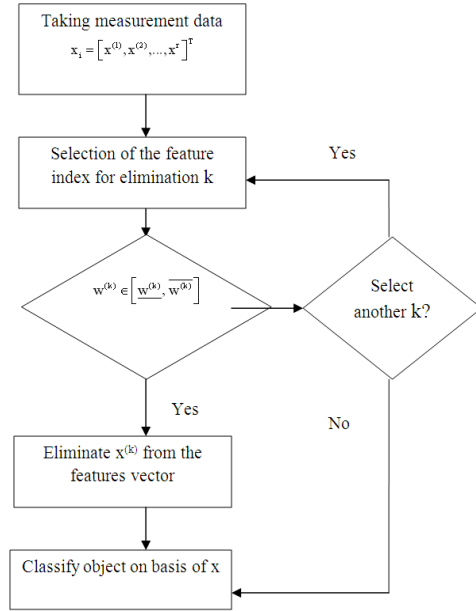


Fig. 1: The proposed algorithm-reduction algorithm

We will treat the genuineness of (12) as a criterion for the possibility of elimination of k-feature from the features vector during classification. Figure 1 presents the proposed procedure algorithm.

Taking measurement data:

$$x_i = [x^{(1)}, x^{(2)}, \dots, x^r]^T$$

Selection of the feature index for elimination k eliminate $x^{(k)}$ from the features vector:

$$w^{(k)} \in \left[\underline{w}^{(k)}, \overline{w}^{(k)} \right]$$

Classify object on basis of x select another k?

RESULTS

In the network architectures (Fig. 2) we have used multi-layered neuron, delta rule, momentum factor $\alpha = 0.9$, sigmoid function, learning rate $\eta_1 = 0.5$ for hidden layer, $\eta_2 = 0.37$ for input layer. The back-propagation network used three layers: the input layer with 28 neuron, the output layer with 1 neuron (1 means patient, 0 means healthy) and the hidden layer with 12 neurons.

The classification was achieved based on cardiological tests such as: pulse frequency, blood pressure, the height of ST- deflection in ECG. Summing up, the features vector had 28 components in total.

Table 1: Mean square error for reduction algorithm (Elimination of 5 features) Number of parameters in the

Features vector	The value of mean square error
28	15.63
23	10.90

Table 2: A Comparison between the three methods: Regression analysis, factor analysis and our algorithm (Reduction algorithm)

Methods	Number of eliminated parameters
Regression analysis	14
Factor analysis	7
Reduction algorithm	5

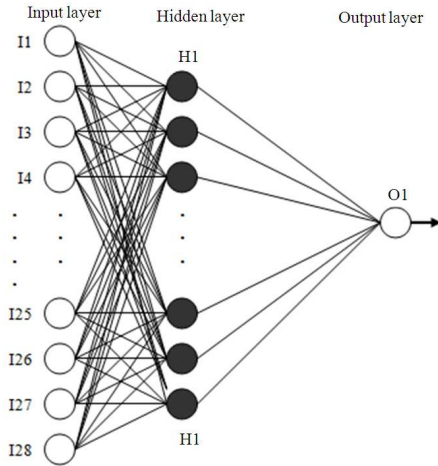


Fig. 2: Multi-layered neuron network architectures

Tests results of 95 patients were collected, on the basis of which a teaching series was created, by which teaching of the linear neuron was preceded. For a so-prepared classifier, recognition was made using objects from the teaching series and the value of mean square was calculated.

Afterwards, the features selection for elimination was conducted using the proposed algorithm. As a result of this operation, 5 components of features vector were discarded (during recognition, their value was set on zero) and the value of mean square was calculated again. The obtained results are presented in Table 1. Data from the table indicates that the elimination of 5 components of features vector, which were indicated using the presented algorithm, did not cause increasing of the value of mean square, which was the basic criterion accepted when constructing the features selection algorithm.

DISCUSSION

In this study we have used three methods to compare the best elimination. These methods are regression analysis, factor analysis and the presented algorithm.

We have eliminated the selected components by using the three methods. We can see the results of applying these three methods in Table 2. The table shows the number of eliminated parameters for each method.

The result was that each method used its own way in the reduction on the basis of analyzing the results of linear neuron.

After learning the linear neuron on the three methods, the best results were achieved through the presented algorithm.

CONCLUSION

In this study, we summarized our recent work in comparing different methods and our reduction algorithm for feature elimination. A multi-layer feed forward ANN is employed to perform this task, using the same training and testing sets for all the different methods, the regression analysis, factor analysis and reduction algorithm.

The reduction algorithm easily achieves high accuracy, regardless of whether all features or parameters are used, only the important features for each case are used. Based on the experiment the reduction algorithm can be seen as the best method. As we can see in Table 2, eliminating the parameters using the reduction algorithm didn't have any effects but the parameters that were eliminated in factor analysis and regression analysis had an effect on detaining the disease.

The experiment results show that training time and running time of the reduction run an order of magnitude faster than the other two methods. In addition to that, the Reduction Algorithm can train with larger number of patterns.

It should be marked that the presented algorithm elaborated for linear neuron, i.e., for a classifier which does not have too big possible applications in practical tasks.

The next step should be elaborating an adequate features selection rule for non-linear neuron, which would be the base to educe a method for multilayer, non-linear neural networks, which are frequently used tool object recognition.

REFERENCES

Abdullah, W.F.H., M. Othman, M.A.M. Ali and M.S. Islam, 2010. Knowledge representation of ion-sensitive field-effect transistor voltage response for potassium ion concentration detection in mixed potassium/ammonium ion solutions. *Am. J. Applied Sci.*, 7: 81-88. DOI: 10.3844/ajassp.2010.81.88

- Ahmed, A., A. Abdo and N. Salim, 2011. An enhancement of bayesian inference network for ligand-based virtual screening using features selection. *Am. J. Applied Sci.*, 8: 368-373. DOI: 10.3844/ajassp.2011.368.373
- Ajeel, S.A., 2010. A novel carbon steel pipe protection based on radial basis function neural network. *Am. J. Applied Sci.*, 7: 248-251. DOI: 10.3844/ajassp.2010.248.251
- Amaroek, S., N. Theera-Umpon, K. Wantanajittikul and S. Auephanwiriyaikul, 2010. Moisture content prediction of dried longan aril from dielectric constant using multilayer perceptrons and support vector regression. *Am. J. Applied Sci.*, 7: 1387-1392. DOI: 10.3844/ajassp.2010.1387.1392
- Bharathi, A. and A.M. Natarajan, 2011. Efficient classification of cancer using support vector machines and modified extreme learning machine based on analysis of variance features. *Am. J. Applied Sci.*, 8: 1295-1301. DOI: 10.3844/ajassp.2011.1295.1301
- Gomathi, M. and P. Thangaraj, 2010. A computer aided diagnosis system for lung cancer detection\using support vector machine. *Am. J. Applied Sci.*, 7: 1532-1538. DOI: 10.3844/ajassp.2010.1532.1538
- Khor, K.C., C.Y. Ting and S.P. Amnuaisuk, 2009. From feature selection to building of Bayesian classifiers: A network intrusion detection perspective. *Am. J. Applied Sci.*, 6: 1948-1959. DOI: 10.3844/ajassp.2009.1948.1959
- Kekre, H.B., S.D. Thepade and A. Maloo, 2010. CBIR feature vector dimension reduction with eigenvectors of covariance matrix using row, column and diagonal mean sequences. *Int. J. Comput. Appl.*, 3: 39-46. DOI: 10.5120/786-1114
- Romero, E. and J.M. Sopena, 2008. Performing feature selection with multilayer perceptrons. *IEEE Trans. Neural Netw.*, 19: 431-441. DOI: 10.1109/TNN.2007.909535
- Saeyns, Y., I. Inza and P. Larranaga, 2007. A review of feature selection techniques in bioinformatics. *Bioinformatics*, 23: 2507-2517. DOI: 10.1093/bioinformatics/btm344
- Yang, P., B.B. Zhou, Z. Zhang and A.Y. Zomaya, 2010. A multi-filter enhanced genetic ensemble system for gene selection and sample classification of microarray data. *BMC Bioinform.*, 11: S5-S5. DOI: 10.1186/1471-2105-11-S1-S5