UDC 004.032.26

Ye. Bodyanskiy, P. Mulesa, O. Vynokurova

KOHONEN NEURAL NETWORK LEARNING IN THE CLUSTERING-CLASSIFICATION TASKS

Abstract. In the paper, combined self-learning and learning method of self-organizing map (SOM-LVQ) is proposed. Such method allows to increase quality of information processing under condition of overlapping classes due to rational choice of learning rate parameter and introducing special procedure of fuzzy choice in the clustering-classification process, which occurs both with external learning signal ("supervised"), and without one ("unsupervised"). As similarity measure of neighborhood function or membership one, cosine structures are used, which allow to provide a high flexibility due to self-learning-learning process and to provide some new useful properties.

Key words: self-organizing map, learning vector quantization, fuzzy clustering, large data set.

Introduction

In Data Mining problems, associated with clustering, classification, fault detection, compression of information and etc., self-organizing maps (SOM) and neural networks of learning vector quantization (LVQ) are widespread. Such neural networks are proposed T. Kohonen [1, 2] and represented in fact by single-layer feedforward architecture, which provides mapping of input space $X \subset R^n$ using some operator F into the output space $Y \subset R^m$.

During operation, each neuron of SOM or LVQ gets information about analyzed input signal. After that, competition mode occurs in single network's layer (also known as the Kohonen layer), and single winner neuron with the maximum output signal is determined. The vector of neuron-winner synaptic weights is most similar in terms of the accepted metric (typically a Euclidean metric) to the input signal. This signal can provide the excitation of nearest "neighbors" of the winner and the reaction suppression of the far dispersed neurons by lateral relations. However, the decision about input vector-pattern membership to one or another class (cluster) is adopted uniquely by the rule of "Winner-Take-All» (WTA). Exactly this uniqueness can lead to the fact that in the case of overlapping classes accuracy of the task solution may be not high due to the fact that the same observation with different levels of membership (sometimes identical) can belong to some clusters at one time. So, it is reasonable to provide for SOM and LVQ properties and capabilities of fuzzy classification, in addition saved to sequential information processing, i.e. the possibility of operation in on-line mode.

1. Formulation of the problem

Let us consider single-layer neural network with lateral connections containing n receptors and m neurons in the Kohonen layer. Each of the neuron is characterized by it's own synaptic weights $w_j, j = 1,2,...,m$, at the same time during learning process input vector-pattern $x(k)$ is fed on the inputs of all neurons (usually adaptive linear associators) (here $k = 1,2,...$ - either the number of observation in a table "object-properties", or current discrete time for on-line processing mode) and neurons produce the scalar signals on their outputs

$$y_j(k) = w_j^T(k)x(k),\ \ j = 1,2,\dots,m\,, \qquad (1)$$

which depend on current values of synaptic weights vectors $w_j(k)$, that tuned using accepted algorithm for determined input space domain X_j. Similar in the sense of accepted metric input vectors can activate either one and the same neuron w_j, or and its neurons-neighbors, for example, w_j and w_p.

Self-organization procedure is based on the competitive learning approaches (self-learning), and the procedure begins from the initialization of network synaptic weights, selecting usually sufficiently randomly, at that preferably, for each of neurons the normalization condition is performed

$$\|x(k)\|^2 = x^T(k)x(k) = \|w_j(k)\|^2 = w_j^T(k)w_j(k) = 1. \qquad (2)$$

The goal of this paper is introducing to the self-organizing process the fourth stage – fuzzy inference, which allowed in on-line mode to classify data (in the context of conventional SOM – LVQ architecture), belonging to several clusters at the same time.

2. SOM learning algorithm

The competition process is started to analysis of current pattern $x(k)$, which is fed to all neurons of Kohonen's layer from receptive (zero) layer. For the each neurons the distance is computed in form

$$D(w_j(k), x(k)) = \left\| x(k) - w_j(k) \right\|, \tag{3}$$

at that, if inputs and synaptic weights are normalized accordingly (2) and Euclidian metric is used as distance, than the proximity measure between the vectors $w_j(k)$ and $x(k)$ can be inner product in form

$$w_j^T(k)x(k) = y_j(k) = \cos(w_j(k), x(k)) = \cos\theta_j(k). \tag{4}$$

Using relation (4), we can replace metric (3) to more simple construction, referred to as a measure of similarity [3]

$$\psi(w_j(k), x(k)) = \max\left\{0, \cos(w_j(k), x(k)\right\} = \max\left\{0, \cos\theta_j(k)\right\}. \tag{5}$$

Further we define neuron-winner that is the most similar to the input vector

$$\psi(w_j^*(k), x(k)) = \max_l \psi(w_l(k), x(k)), \tag{6}$$

after that, temporarily omitting the stage of cooperation, tuning of synaptic weights using WTA self-learning rule is realized in form

$$w_j(k+1) = \begin{cases} w_j^*(k) + \eta(k)(x(k) - w_j^*(k)), \text{ if } j-\text{th neuron} \\ \text{won in the competition}, \\ w_j(k) \quad \text{otherwise}. \end{cases} \tag{7}$$

Requirement of monotonic decreasing of the parameter $\eta(k)$ leads to expression in form

$$\eta(k) = r^{-1}(k), \ \eta(k) = \alpha r(k-1) + \left\| x(k) \right\|^2, \ 0 \le \alpha \le 1. \tag{8}$$

One of the SOM features is the presence of cooperation stage in the self-learning process, when neuron-winner defines local domain of topological neighbourhood, in which is trained not only himself, but and his nearest environment, at that neurons, that are more close to winner, adjusts their weights more than a remote. This topological domain is defined by neighborhood function $\varphi(j,p)$, which depends from distance $D(w_j^*(k), w_p(k))$ between winner $w_j^*(k)$ and any of Kohonen's layer neurons $w_p(k)$, $p = 1, 2, \ldots, m$ and some parameter σ, which sets its width.

Using of neighborhood functions leads to the self-learning rule in form

$$w_p(k+1) = w_p(k) + \eta(k)\varphi(j,p)(x(k) - w_p(k)), \ p = 1, 2, \ldots, m, \tag{9}$$

which minimizes criterion

$$E_p^k = \sum_{\tau=1}^{k} \varphi(j,p) \left\| x(\tau) - w_p \right\|^2 \tag{10}$$

of produced criterion "Winner Takes More" (WTM) type.

The necessity to maintain of condition (2) leads to the expression in form

$$\begin{cases} w_p(k+1) = \dfrac{w_p(k) + \eta(k)\varphi(j,p)(x(p) - w_p(k))}{\left\| w_p(k) + \eta(k)\varphi(j,p)(x(p) - w_p(k)) \right\|}, \ p = 1,2,\ldots,m, \\ \eta(k) = r^{-1}(k), \ \ r(k) = \alpha r(k-1) + 1, \ 0 \le \alpha \le 1. \end{cases} \tag{11}$$

In many real tasks clusters can overlap. In this case vector $x(k)$ with proportional membership level $\cos\theta_j(k)$ belongs *j-th* cluster, and with proportional level $\cos\theta_p(k)$ - to *p*-th one. Synaptic weights, which situated in crosshatched region, don't have relationship to the pattern $x(k)$ according to (5).

Using similarity measure, we can introduce the membership estimate of pattern $x(k)$ to *j*-th class in form:

$$0 \le \mu_{w_j(k)}(x(k)) = \frac{\psi(w_j(k), x(k))}{\sum_{l=1}^{m} \psi(w_l(k), x(k))} \le 1. \tag{12}$$

Learning vector quantization neural networks in contrast to self-learning SOM adjust their parameters based on external learning signal (reference signal), which fixed the membership of each pattern $x(k)$ to one or another class. The mean idea of LVQ neural network is the possibility of compact representation of large data sets in the form of restricted prototypes set, or by centroids $w_j(k)$, $j = 1,2,\ldots,m$, which good enough approximate the initial space X.

For each normalized according to (2) input vector $x(k)$ the neuron-winner is defined, in which the synaptic weights $w_j^*(k)$ correspond to the prototype of the certain class. In other words, winner is neuron with minimal distance to the input vector (9) or, which is the same, with maximal similarity measure (6).

Since the learning is controllable (with supervisor), than membership of the vector $x(k)$ to specific domain X_j of the space X is known,

 ISSN 1562-9945

that allow to consider two typical situation occurring in the trained linear vector quantization:

- the input vector $x(k)$ and neuron-winner $w_j^*(k)$ belong to the same cell of Voronoy [2];

- the input vector $x(k)$ and neuron-winner $w_j^*(k)$ belong to the different cells of Voronoy.

Than corresponding learning LVQ-rule can be written in form:

$$w_j(k+1) = \begin{cases} \dfrac{w_j^*(k) + \eta(k)(x(k) - w_j^*(k))}{\left\| w_j^*(k) + \eta(k)(x(k) - w_j^*(k)) \right\|}, \text{if } x(k) \text{ and } w_j^*(k) \\ \text{belong to the same cell,} \\ \dfrac{w_j^*(k) - \eta(k)(x(k) - w_j^*(k))}{\left\| w_j^*(k) - \eta(k)(x(k) - w_j^*(k)) \right\|}, \text{if } x(k) \text{ and } w_j^*(k) \\ \text{belong to the different cells,} \\ \\ w_j(k) \text{ for neurons, which are not won in instant } k. \end{cases} \quad (13)$$

If the first and third expression of formula (13) are completely identical to WTA – self-learning algorithm, than it should be stayed more details on “push” procedure (second expression of formula (13)).

Let’s considered to situation, when neuron $w_j^*(k)$ won in competition, although presented vector-pattern $x(k)$ belongs to the class with centroid $w_p(k)$, and neuron $w_j(k)$ didn’t win in competition. Naturally it is necessary to «push» vector $w_j^*(k)$ so, that $x(k)$ was equidistant from $w_j^*(k)$ and from $w_p(k)$.

For that, without comments, we make to the elementary transformation over the expression in vector form:

$$w_j(k+1) = w_j^*(k) - \eta(k)(x(k) - w_j^*(k)), \quad (14)$$

$$x^T(k)w_j(k+1) = x^T(k)w_j^*(k) - \eta(k)\left\| x(k) \right\|^2 - \eta(k)x^T(k)w_j^*(k), \quad (15)$$

$$\cos(w_j(k+1), x(k)) = \cos(w_j^*(k), x(k)) - \eta(k)(1 + \cos(w_j^*(k), x(k))) \quad (16)$$

at that, for providing condition

$$\cos(w_j(k+1), x(k)) = \cos(w_p(k), x(k)) \quad (17)$$

it is necessary to set $\eta(k)$ in form

$$\eta(k) = \frac{\cos(w_j^*(k), x(k)) - \cos(w_p(k), x(k))}{\cos(w_j^*(k), x(k)) + 1} =$$

$$= \frac{\cos(w_j^*(k), x(k)) - \cos(w_p(k), x(k))}{\cos(w_j^*(k), x(k)) + \cos(x(k), x(k))} = \quad (18)$$

$$= \frac{\cos\theta_j(k) - \cos\theta_p(k)}{\cos\theta_j(k) + 1} = \frac{x^T(k)w_j^*(k) - x^T(k)w_p(k)}{x^T(k)w_j^*(k) - x^T(k)x(k)}.$$

3. Compatible learning of SOM and LVQ

In the enough wide class of Data Mining tasks, for example, which are connected to the medical diagnostics, in table «object-properties» for part of the features vectors $x(k)$ the diagnosis is known, and for some one is or ambiguous, or nonunique, or not defined at all. In this case it is possible to tune the synaptic weights with unified architecture with lateral connections using different learning methods. Each of such learning methods initializes according to level of apriori information about membership or not membership $x(k)$ to one or another class.

As a result general learning and self-learning algorithm can be written in form:

$$w_j(k+1) = \begin{cases} \dfrac{w_j^*(k) + \eta(k)(x(k) - w_j^*(k))}{\left\| w_j^*(k) + \eta(k)(x(k) - w_j^*(k)) \right\|}, \\ \delta_L \dfrac{w_j^*(k) - \eta(k)(x(k) - w_j^*(k))}{\left\| w_j^*(k) - \eta(k)(x(k) - w_j^*(k)) \right\|}, \text{if } x(k) \text{ and } w_j^*(k) \\ \text{belong to the different cells}, \\ \delta_L = \begin{cases} 1, & \text{if the network works in supervised model}, \\ 0, & \text{otherwise}, \end{cases} \\ w_j(k) \text{ for neurons, which are not won in instant } k. \end{cases} \quad (19)$$

Conclusion

The combined self-learning procedure for Kohonen neural network is proposed. Such method allows data processing under the overlapping classes condition, when memberships of training data to specific classes can be unknown at all, and have both crisp and fuzzy nature. This method is based on using similarity measure, recurrent optimization and fuzzy inference and differs with high speed, possibility of operating in on-line mode and simplicity of computational realization.

LITERATURE

1. Kohonen, T. Self-Organizing Maps. – Berlin: Springer-Verlag, 1995 – 362 p.
2. Haykin, S. Neural Networks: A Comprehensive Foundation. - Upper Saddle River, N.Y.: Prentice Hall, 1999. – 842 p.
3. Sepkovski, J.J. Quantified coefficients of association and measurement of simularity // J. Int. Ass. Math. - 1974. – 6 (2). – P. 135-152.