Real time clustering of high dimensional data by Projective Adaptive Resonance Theory

Roman Krakovský[1], Radoslav Forgáč[2], Igor Mokriš[2]

[1] Catholic University, Department of Informatics, Ružomberok, Slovakia,
[2] Institute of Informatics Slovak Academy of Sciences, Bratislava, Slovakia,
Roman.Krakovsky@ku.sk, Radoslav.Forgac@savba.sk, Igor.Mokris@savba.sk

Abstract. The paper presents one of approaches to the real time clustering of the high dimensional data by modification of the Projective Adaptive Resonance Theory (PART) named as PART with buffers (BPART). PART is designed for the high dimensional data clustering, but the clustering accurancy depends on optimal parameters setting and on the appropriate order of input patterns. The modified model of PART with buffers eliminates these dependecies, so this neural network is suitable for real time clustering in the high dimensional data space.

Keywords: Projective ART, neural network, clustering algorithm, clustering accuracy, buffer management.

1 Introduction to PART

The neural networks belong to the versatile tools suitable for solving tasks of data classification and clustering or for process prediction [3, 4, 5, 10]. They found the application in areas where it is necessary to process of inaccurate data, such as various forms of the word, variety of occurrences of keywords, changed shapes and sizes of images and the like.

Recently, in the tasks of high dimensional data clustering was applied a relatively new neural network with learning without the teacher PART (Projective Adaptive Resonance Theory) [2, 3]. PART is based on an alghoritm PROCLUS (Projective Clustering) [1] and ART (Adaptive Resonance Theory) [4].

PART solves the main problems of PROCLUS and others clustering algorithms, for example, at the beginning of clustering must be known overall number of clusters [1, 5, 10, 12], the clustering accurancy depends on optimal parameters setting or on the order of input patterns as well as the computational complexity that increases with the volume of processed data. However, in applications with real data clustering is very often not known in advance the number of clusters or the multiplicity of patterns in these clusters.

The essential characteristic of the PART is the ability to switch between plastic and stable mode without information loss. Another important feature of this neural network is projective clustering. The projective clustering allows to find such clusters which are sufficiently closely linked to the corresponding subset of dimensions.

The PART neural network consists of an input, so-called comparative layer F_1 and output, so-called clustering layer F_2 (Fig. 1). Neurons in the comparative layer F_1, are denoted as U_i. Neurons in the clustering layer F_2, are denoted as V_j. The activation values of neurons in the comparative layer F_1 are denoted as x_i, where $i = 1 \dots m$ is an index of a neuron in the input layer F_1. The activation values of neurons in the clustering layer F_2 are denoted as y_j, where $j = 1 \dots n$ is an index of a neuron in the output layer F_2. Neural connections between layers F_1 and F_2 are defined by the bottom-up w_{ij} and top-down z_{ji} weights. The neuron in the clustering layer F_2 is referred as committed, if there was learned some input pattern according to committed neuron in the recognition phase of processing. Otherwise, the neuron is noncommitted. Only the committed neurons accept signals from the F_1 layer in the comparative phase.

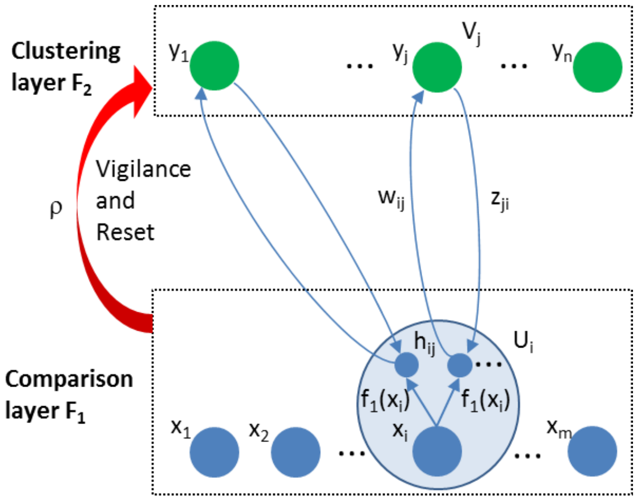

Fig. 1. The structure of PART

In the PART neural network layer F_1 selectively sends signals to neurons of F_2 layer. In other words, the neuron in the layer F_1 may be active to a certain neuron in the layer F_2, but inactive to the other neurons from layer F_2.

The main difference between the PART and ART is in introducing of the selective output signal h, which is a function of the output signaling function $f_1(x_i)$ and top-down weights z_{ji}. Neurons in layer F_1 must be linked with neurons in layer F_2 with using a similarity test between top-down weights and the value of the function $f_1(x_i)$ generated in a neuron U_i [2, 3, 7, 8, 11]:

$$h_{ij} = h(x_i, w_{ij}, z_{ji}) = h_\sigma(f_1(x_i),\ z_{ji}).l(w_{ij}) \tag{1}$$

where

$$h_\sigma(f_1(x_i),\ z_{ji}) = \begin{cases} 1, & if\ d(f_1(x_i),\ z_{ji}) \le \sigma \\ 0, & if\ d(f_1(x_i),\ z_{ji}) > \sigma \end{cases} \tag{2}$$

where d is distance function

$$d(f_1(x_i),\ z_{ji}) = |f_1(x_i) - z_{ji}| \tag{3}$$

and $l(w_{ij})$ is the step function

$$l(w_{ij}) = \begin{cases} 1, & \text{if } w_{ij} > \theta \\ 0, & \text{if } w_{ij} \le \theta \end{cases} \tag{4}$$

where θ is a threshold value of bottom-up weights and σ is a distance parameter. Just this similarity test plays the key role in the PART clustering of the high dimensional data.

2 PART with buffers for real time clustering

Traditional clustering algorithms fail by processing of high dimensional data in the large databases [5, 10, 12]. PART neural network is designed precisely to create clusters in high dimensional data space. However, the practice has been verified that the PART clustering accuracy depends on optimal parameters setting and on the appropriate order of input patterns. Mentioned deficiencies inhibit PART networks in their application, e.g. data processing in real-time systems such as banking, reservation systems, biological networks, multimedia and so on. In these systems, it is required that the data processing was carried out in the shortest possible time without loss of time caused by parameters setting or by structure setting of the neural network.

The PART with buffers (BPART) enables data clustering in real time [9]. This model includes a mechanism that prevents the data clustering to a single cluster through the vigilance parameter ρ and function avg_j (Fig. 2). The function avg_j reflects to the average degree of similarity between the winning neuron and corresponding cluster. Thanks avg_j and ρ, the BPART can successfully cluster the similar but noisy data using sub-optimal parameters of the neural network. The buffered control process can additionally manage large amounts of data in the input data sets through a small buffer space. The BPART contains the following additions in comparison to the original model of PART:

1. The storage manager controls the selection of the winner neuron. If the similarity of the patterns is less than the vigilance parameter ρ and at the same time greater than the average similarity avg_j, the pattern is assigned to the cluster C_j.
2. If the input pattern is assigned to the cluster C_j, the degree of avg_j similarity is recalculated. Otherwise it remains integrated into the buffer and the resonance process is continuing to seek a new winner neuron.
3. The buffer control mechanism dynamically controls the degree of similarity between patterns stored in the buffer and the average degree of similarity avg_j in cluster C_j. If the degree of similarity is greater than or equal to avg_j, the pattern is moved from the buffer to the corresponding cluster. In the case of buffer overflow the patterns with the minimum value of avg_j are automatically moved to the outlier cluster.

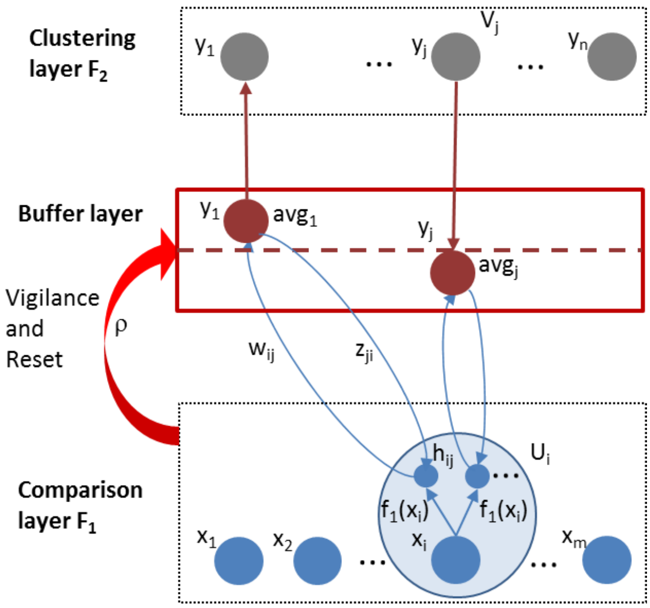

Fig. 2 PART with buffers

When a suitable candidate is looking for winning neuron then the degree of similarity between pattern and committed neuron must be greater than or equal to avg_j. It is dynamically calculated by equation

$$avg_j(t) = avg_j(t-1)\frac{n_j - 2}{n_j} + q\frac{2}{n_j} \quad (5)$$

where n_j is the current number of patterns in the j-th cluster, t is time, q is the degree of similarity between each input vector and the expected winner neuron v_j. This means that a stable value avg_j reflects to the maximum number of projective dimensions associated with a particular cluster even with suboptimal vigilance parameter ρ.

According to [9], a series of experiments was carried out with BPART, which showed that in any regrouping of the input patterns it was achieved the same result. In the case of the original PART were the results different. This proves that BPART is invariant to the change of the order of input patterns. The experimental results also confirmed that there is no need to look for the optimal value of the vigilance parameter ρ.

3 Conclusion

The experiments aimed to real time clustering of high dimensional data by PART confirmed that the original model PART published by Cao and Wu needs to be modified. Modification can be realized by adding a new layer, by changing the learning top-down and bottom-up weights or by applying one of the known statistical methods or metrics. The choice of the modification method depends on the type and characteristics of the high dimensional data.

The basic idea of real time data clustering by BPART described in this paper consists of adding a buffer with its own management to the original PART structure. Thanks for the average similarity avg_j and vigilance parameter ρ, the BPART can cluster with the required accuracy the similar but noisy data using sub-optimal parameters of the neural network.

In conclusion, the above advantages of BPART predetermine this neural network for clustering and classification of high dimensional data in various types of applications.

Acknowledgment
This work is supported by Slovak Science Agency VEGA No. 2/0054/12.

References

1. Aggarwal, C. C., Procoius, C., Wolf, J.L., Yu, P.S., Park, J.S.: Fast Algorithm for Projected Clustering. Proceeding of Conference SIGMOD '99, Philadelphia, USA, 1999, pp. 61--72.
2. Cao, Y., Wu, J.: Dynamics of Projective Adaptive Resonance Theory Model: the Foundation of PART Algorithm. Neural Networks, Vol. 15, pp. 245—260, (2004).
3. Cao, Y., Wu, J.: Projective ART for Clustering Data sets in High Dimensional Spaces. Neural networks, Vol. 15 (1), pp. 105—120, (2002).
4. Grossberg, S., Carpenter, G.A.: Adaptive Resonance Theory. The Handbook of Brain Theory and Neural Networks. MIT Press, 2002.
5. Han, J., Kamber, M.: Data Mining - Concepts and Techniques. Morgan Kaufman, (2000).
6. Huang, A.: Similarity Measures for Text Documents Clustering. In New Zealand Computer Science Research Student Conference, pp. 49—56, (2008).
7. Hunter, J.D., Wu, J., Milton, J. G.: Clustering Neural Spike Trains with Transient Responses. Proceedings of IEEE Conference on Decision and Control, Cancun, Mexico, pp. 2000-2005, (2008).
8. Krakovský, R., Forgáč, R., Mokriš, I.: Modified Clustering Algorithm for Projective ART Neural Network. Proceedings of INES 2014 - 18th IEEE International Conference on Intelligent Engineering Systems, Budapest, pp. 245—250, (2014).
9. Liu, L., Huang, L.: Projective ART with Buffers for the High Dimensional Space Clustering and an Application to Discover Stock Associations. Neurocomputing, Elsevier, pp. 1283—1295, (2009).
10. Parson, L., Haque, E. - Liu, H.: Subspace Clustering for High Dimensional Data. SIKDD Explorations, pp. 90—105, (2004).
11. Takahashi, H., Kobayashi, T., Honda, H.: Gene Screening Method for Prognostic Prediction Using Projective ART model. Genome Informatics 14, pp. 320-321, (2003).
12. Xu, S, Zhang, J.: A hybrid paralel Web document clustering algorithm and its performance study. The Journal of Supercomputing No 30, pp 117—131, (2004).