A Simple "Possibilistic" Clustering Neural Network

O. Yadid-Pecht and M. Gur

Department of Biomedical Engineering, Technion-Israel Institute of Technology, Haifa 32000, Israel

Abstract

A simple "possibilistic" clustering method, i.e. clustering where each datum has a degree of possibility of belonging to the cluster, using a neural net, is suggested. The implementation consists of simple "neurons", requiring only a small number of local connections, collectively performing a diffusion-like process. In spite of its simplicity, this implementation has several advantages over commonly used fuzzy clustering methods. Specifically, it provides the "typicality" notion that is lacking in the well known Fuzzy C Means (FCM) and its derivatives, and is less sensitive to noise.

1. General Description

We propose a neural network implementation for clustering where each data point has a degree of possibility of belonging to the cluster [1]. The implementation requires basic elements that work in parallel which can be viewed as simple "neurons", with only a small number of local connections (see Fig. 1). The neurons are located on a grid of points, dimensioned as the number of features in the data. In the two dimensional case, each point has two features, its x and y coordinates. The system input is a set of data points which are described as positively initialized values on the two-dimensional grid of neurons. Then, a "diffusion" like process is implemented by the network. This process uses lateral excitation to spread activity locally, generating long-range interactions over time. Depending on the extent of the diffusion process, i.e. the time elapsed, it generates peaks over areas of growing size (see Fig. 2). A maxima selection net uses lateral inhibition to localize the apex of those local peaks as cluster center points [2]. These processes are implemented in neural networks and yield the proposed clustering system.

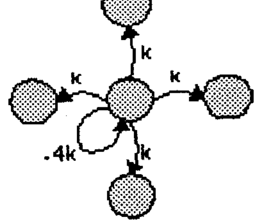

Fig. 1: A basic element in the proposed neural network receives diffusion values from its four neighbors and inhibits itself accordingly.

The diffusion process description and the analytic convergence results are found in [3]. The experimental results and discussion are followed.

2. Experimental Results

When using the FCM algorithm, presence of noise in the data (see Fig. 3a) can drastically influence the estimates of the class prototypes.

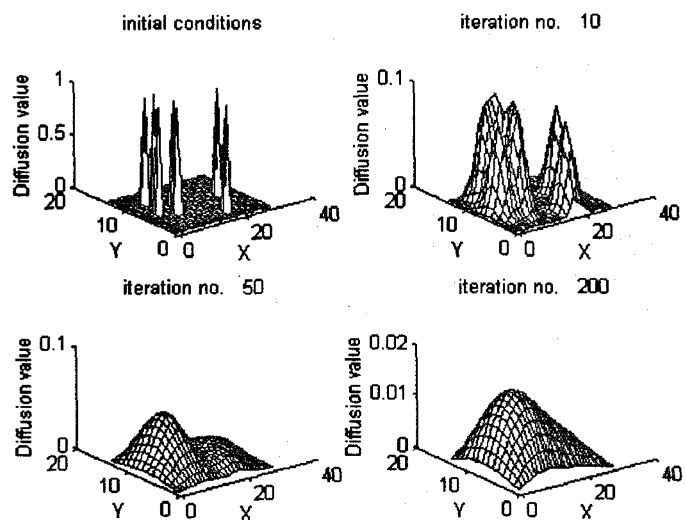

Fig. 2: Evolving maxima during the diffusion process. X and Y axes describe feature vaues. Z axis describes the diffusion value.

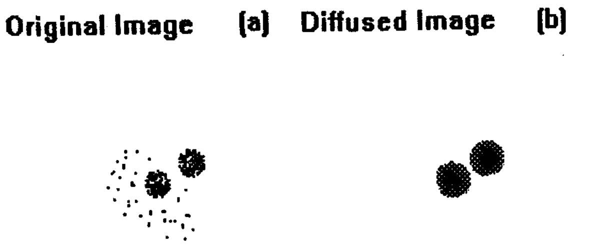

Fig. 3: (a) A data set consisting of two classes with noise added. (b) Results of applying the diffusion-like process to the data set of Fig. 3a.

Results of implementing the diffusion process on a noisy input set (Fig. 3a) are illustrated in Fig. 3b. If only two clusters are desired, the proposed clustering can identify and eliminate noise outliers from the data set in a natural way, by choosing the two largest maxima and determining the clusters accordingly. In Fig. 3b, the darker the point- the higher its diffusion value. The points that are not parts of the clusters are treated as "noise". Subsequently, by neglecting noise points, the risk of biasing the results is reduced.

If cluster validity is not performed, then another problem encountered in using the FCM algorithm is that membership values which are unrepresentative of the degree of belonging can be assigned [1]. Consequently, membership values cannot distinguish between a moderately atypical member and an extremely atypical one (see Fig. 4a).

Implementing the diffusion process for clustering the data set results in two clusters which form quite quickly and are very stable, i.e. their centers remain static for quite a long period (see Fig. 4b,4c). After many iterations (see Fig. 4d), point X can be included in one of the two clusters, although its diffusion value is very small (lightest shade in Fig. 4d). Point Y still remains an outlier (its diffusion value is below threshold to be seen). In Fig. 4e both points may not be considered outliers, yet point Y, naturally, has a lower diffusion value than point X.

3. Discussion

We have presented a neural network that accomplishes clustering by implementing a diffusion-like process.

Our approach is similar to that of Krishnapuram and Keller [1] in which all points have membership values that are interpreted as degrees of "typicality". The diffusion process is also similar to the "melting" procedure suggested by Wong [4]. However, these algorithms require massive computations which are avoided in our approach.

There are several benefits using this method. First, it requires only simple local calculations. Second, it offers

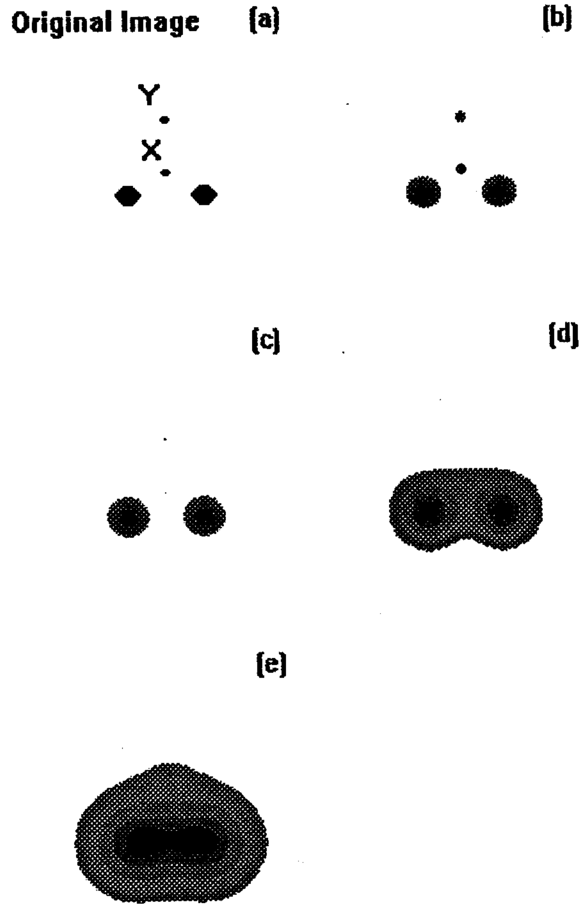

Fig. 4: (a) Example of a data set with two clusters and two outlying points X, Y. Diffusion results at 50 (b), 100 (c), 500 (d) and 900 (e) iterations.

the notion of "typicality". This feature was already shown to be useful [1]. Third, the process is less sensitive to noise than commonly used fuzzy C means algorithms. Finally, the framework can be enlarged to get information concerning the history of the points stabilization to decide on the number of clusters in the data.

5. References

1. R. Krishnapuram and J.M. Keller, "A possibilistic approach to clustering", IEEE Trans. on Fuzzy Systems, Vol. 1, No. 2, pp. 98–110, May 1993.
2. O. Yadid–Pecht and M. Gur, "A biologically inspired modified Maxnet", IEEE Trans. on NN, in press.
3. O. Yadid–Pecht, Ph.D. thesis, in preparation.
4. Y. Wong, "Clustering data by melting", Neural Computation, Vol. 5, pp. 89–104, 1993.