ADAPTIVE CLUSTERING OF INCOMPLETE DATA USING NEURO-FUZZY KOHONEN NETWORK

Yevgeniy Bodyanskiy, Alina Shafronenko, Valentyna Volkova

***Abstract**: The clustering problem for multivariate observations often encountered in many applications connected with Data Mining and Exploratory Data Analysis. Conventional approach to solving these problems requires that each observation may belong to only one cluster, although a more natural situation is when the vector of features with different levels of probabilities or possibilities can belong to several classes. This situation is subject of consideration of fuzzy cluster analysis, intensively developing today.*

In many practical tasks of Data Mining, including clustering, data sets may contain gaps, information in which, for whatever reasons, is missing. More effective in this situation are approaches based on the mathematical apparatus of Computational Intelligence and first of all artificial neural networks and different modifications of classical fuzzy c-means (FCM) method.

But these methods are effective only in cases when the original data set is given beforehand and does not change during data processing. At the same time there is a wide class of problems when the data are fed to processing sequentially in on-line mode as it occurs in self-organizing Kohonen networks training. At the same time apriori it is not known which of the vectors-images contain gaps.

*In this paper the problem of probabilistic and possibilistic on-line clustering of data with gaps using Partial Distance Strategy is discussed and solved, self-organizing neuro-fuzzy Kohonen network and new self-learning algorithm that is the hybrid of "Winner-takes-more" rule and recurrent fuzzy clustering procedures are proposed and investigated***.**

***Keywords**: Fuzzy clustering, Kohonen self-organizing network, learning rule, incomplete data with gaps.*

***ACM Classification Keywords**: 1.2.6 [Artificial Intelligence]: Learning – Connectionism and neural nets; 1.2.8 [Artificial Intelligence]: Problem Solving, Control Methods, and Search – Control theory; 1.5.1 [Pattern Recognition]: Clustering – Algorithms.*

Data with gaps clustering on the basis of neuro-fuzzy Kohonen network

The clustering of multivariate observations problem often occurs in many applications associated with Data Mining. The traditional approach to solving these tasks requires that each observation may relate to only one cluster, although the situation more real is when the processed feature vector with different levels of probabilities or possibilities may belong more than one class. This situation is subject of fuzzy cluster analysis fast-growing at the present time [Bezdek, 1981; Hoeppner 1999; Xu, 2009].

However, in numerous problems of Data Mining, including, of course, clustering, the original data sets may contain gaps, information in which for some reasons, is missing. In this situation more effective are approaches based on the mathematical apparatus of computational intelligence, and especially, artificial neural networks [Marwala, 2009] and modifications of the classical c-means method (FCM) [Hathaway, 2001].

Notable approaches and solutions are efficient only in cases when the original array data set has batch form and does not change during the analysis. However there is enough wide class of problems when the data are fed to the processing sequentially in on-line mode as this occurs when training Kohonen self-organizing maps [Kohonen, 1995]. In this case, however, it is not known beforehand which of the processed vector-images contains gaps (missing values). This paper is devoted to solving the problem on-line clustering of data based on the Kohonen neural network, adapted for operation in presence of overlapping classes.

Adaptive algorithm for probabilistic fuzzy clustering

Baseline information for solving the tasks of clustering in a batch mode is the sample of observations, formed from N n-dimensional feature vectors $X=\{x_1,x_2,...,x_N\}\subset R^n, x_k\in X, k=1,2,...,N$. The result of clustering is the partition of original data set into m classes $(1\le m\le N)$ with some level of membership $U_q(k)$ of k-th feature vector to the q-th cluster $(1\le q\le m)$. Incoming data previously are centered and standardized by all features, so that all observations belong to the hypercube $[-1,1]^n$. Therefore, the data for clustering form array $\tilde{X}=\{\tilde{x}_1,...,\tilde{x}_k,...,\tilde{x}_N\}\subset R^n$, $\tilde{x}_k=(\tilde{x}_{k1},...,\tilde{x}_{ki},...,\tilde{x}_{kn})^T$, $-1\le\tilde{x}_{ki}\le 1$, $1<m<N$, $1\le q\le m$, $1\le i\le n$, $1\le k\le N$.

Introducing the objective function of clustering [Bezdek, 1981]

$$E(U_q(k),w_q)=\sum_{k=1}^{N}\sum_{q=1}^{m}U_q^{\beta}(k)D^2(\tilde{x}_k,w_q)$$

with constraints $\sum_{q=1}^{m}U_q(k)=1,\ 0<\sum_{k=1}^{N}U_q(k)<N$ and solving the nonlinear programming problem, we get the probabilistic fuzzy clustering algorithm [Hoeppner, 1999; Xu, 2009]

$$\begin{cases} U_q^{(\tau+1)}(k)=\dfrac{(D^2(\tilde{x}_k,w_q^{(\tau)}))^{\frac{1}{1-\beta}}}{\sum_{l=1}^{m}(D^2(\tilde{x}_k,w_l^{(\tau)}))^{\frac{1}{1-\beta}}}, \\ w_q^{(\tau+1)}=\dfrac{\sum_{k=1}^{N}(U_q^{(\tau+1)})^{\beta}\tilde{x}_k}{\sum_{k=1}^{N}(U_q^{(\tau+1)}(k))^{\beta}}, \end{cases} \tag{1}$$

where w_q - prototype (centroid) of q-th cluster, $\beta>1$ - parameter that is called fuzzyfier and defines "vagueness" the boundaries between classes, $D^2(\tilde{x}_k,w_q)$ - the distance between $\tilde{x}_k$ and w_q in adopted metric, $\tau=0,1,2,...$ - index of epoch information processing which is organized as a sequence of $w_q^{(0)}\to U_q^{(1)}\to w_q^{(1)}\to U_q^{(2)}\to...$. The calculation process continues until satisfy the condition

$$\left\|w_q^{(\tau+1)}-w_q^{(\tau)}\right\|\le\varepsilon\ \forall\ 1\le q\le m,$$

where ε - defines threshold of accuracy. Choosing $\beta=2$ and taking the Euclidean distance, we get a popular algorithm of Bezdek's fuzzy c-means (FCM)

$$\begin{cases} U_q^{(\tau+1)}(k)=\dfrac{\left\|\tilde{x}_k-w_q^{(\tau)}\right\|^{-2}}{\sum_{l=1}^{m}\left\|\tilde{x}_k-w_l^{(\tau)}\right\|^{-2}}, \\ w_q^{(\tau+1)}=\dfrac{\sum_{k=1}^{N}(U_q^{(\tau+1)}(k))^{2}\tilde{x}_k}{\sum_{k=1}^{N}(U_q^{(\tau+1)}(k))^{2}}. \end{cases}$$

The process of fuzzy clustering can be organized in on-line mode as sequentially processing. At this situation batch algorithm (1) can be rewritten in recurrent form [Bodyanskiy, 2005]

$$\begin{cases} U_q(k+1)=\dfrac{(D^2(\tilde{x}_{k+1},w_q^{(k)}))^{\frac{1}{1-\beta}}}{\sum\limits_{l=1}^{m}(D^2(\tilde{x}_{k+1},w_l(k)))^{\frac{1}{1-\beta}}}, \\ w_q(k+1)=w_q(k)+\eta(k+1)U_q^{\beta}(k+1)(\tilde{x}_{k+1}-w_q(k)), \end{cases} \qquad (2)$$

(here $\eta(k+1)$ - learning rate parameter), which is a generalization of the clustering gradient procedure of Park-Dagher [Park, 1984] and the learning algorithm of Chung-Lee [Chung, 1994]. If the data are fed to the processing with high-frequency, recalculation of epochs is not made, if this frequency is low, between the instants k and $k+1$ it is possible to organize several epochs in an accelerated time.

It should be noted that the first expression in (2) can be rewritten in the form

$$U_q(k+1)=\frac{(D^2(\tilde{x}_k,w_q(k)))^{\frac{1}{1-\beta}}}{\sum\limits_{l=1}^{m}(D^2(\tilde{x}_k,w_l(k)))^{\frac{1}{1-\beta}}}=$$

$$=\frac{(D^2(\tilde{x}_k,w_q(k)))^{\frac{1}{1-\beta}}}{(D^2(\tilde{x}_k,w_q(k)))^{\frac{1}{1-\beta}}+\sum\limits_{\substack{l=1\\ l\neq q}}^{m}(D^2(\tilde{x}_k,w_l(k)))^{\frac{1}{1-\beta}}}=$$

$$=\frac{1}{1+(D^2(\tilde{x}_k,w_q(k)))^{\frac{1}{\beta-1}}\sum\limits_{\substack{l=1\\ l\neq q}}^{m}(D^2(\tilde{x}_k,w_l(k)))^{\frac{1}{1-\beta}}},$$

for the Euclidean metric and $\beta=2$ taking the form of the Cauchy function with a parameter of width σ^2:

$$U_q(k+1)=\frac{1}{1+\dfrac{\left\|\tilde{x}_k-w_q(k)\right\|^2}{\sigma^2}},$$

$$\sigma^2=(\sum_{\substack{l=1\\ l\neq q}}^{m}\left\|\tilde{x}_k-w_l(k)\right\|^{-2})^{-1}.$$

This fact allows us to rewrite the second expression in (2) with $\beta = 2$ in the form

$$\begin{aligned} w_q(k+1) &= w_q(k) + \eta(k+1)U_q^2(k+1)(\tilde{x}_{k+1} - w_q(k)) = \\ &= w_q(k) + \eta(k+1)\varphi_q(k+1)(\tilde{x}_{k+1} - w_q(k)) \end{aligned}$$

where $U_q^2(k+1) = \varphi_q(k+1)$ - the bell-shaped neighborhood function of neuro-fuzzy Kohonen network [Gorshkov, 2009] designed for solving the fuzzy clustering task [Shafronenko, 2011] using the principle "winner-takes-more» (WTM).

Adaptive probabilistic fuzzy clustering algorithm for data with gaps

In the situation if the data in the array $\tilde{X}$ contain gaps, the approach discussed above should be modified accordingly. For example, in [Hathaway, 2001] it was proposed the modification of the FCM-procedure based on partial distance strategy (PDS FCM). Thus introducing, additional arrays:

$X_F = \{\tilde{x}_k \in \tilde{X} \mid \tilde{x}_k \text{- } vector\ \ containing\,all\ \ components\}$;

$X_P = \{\tilde{x}_{ki}, 1 \le i \le n, 1 \le k \le N \mid\ values\ \tilde{x}_k,\ available\ in\ \tilde{X}\}$;

$X_G = \{\tilde{x}_{ki} = ?, 1 \le i \le n, 1 \le k \le N \mid\ values\ \tilde{x}_k, absent\ in\ \tilde{X}\}$

and taking instead of the traditional Euclidean metric partial distance (PD):

$$D_P^2(\tilde{x}_k, w_q) = \frac{n}{\delta_{k\Sigma}} \sum_{i=1}^{n} (\tilde{x}_{ki} - w_{qi})^2 \delta_{ki},$$

the objective function of clustering

$$E(U_q(k), w_q) = \sum_{k=1}^{N} \sum_{q=1}^{m} U_q^{\beta}(k) \frac{n}{\delta_{k\Sigma}} \sum_{i=1}^{n} (\tilde{x}_{ki} - w_{qi})^2 \delta_{ki}$$

(here $\delta_{ki} = \begin{cases} 0 \mid \tilde{x}_{ki} \in X_G, \\ 1 \mid \tilde{x}_{ki} \in X_F, \end{cases}$

$\delta_{k\Sigma} = \sum_{i=1}^{n} \delta_{ki}$)

and solving nonlinear programming problem, we obtain the algorithm

$$\begin{cases} U_q^{(\tau+1)} = \dfrac{(D_P^2(\tilde{x}_k, w_q^{(\tau)}))^{\frac{1}{1-\beta}}}{\sum\limits_{l=1}^{m}(D_P^2(\tilde{x}_k, w_q^{(\tau)}))^{\frac{1}{1-\beta}}}, \\ w_{qi}^{(\tau+1)} = \dfrac{\sum\limits_{k=1}^{N}(U_q^{(\tau+1)}(k))^{\beta}\delta_{ki}\tilde{x}_{ki}}{\sum\limits_{k=1}^{N}(U_q^{(\tau+1)}(k))^{\beta}\delta_{ki}} \end{cases} \qquad (4)$$

which is a generalization of the standard FCM-procedure (1).

Algorithm (4) can be rewritten in recurrent form

$$\begin{cases} U_q(k+1) = \dfrac{(D_P^2(\tilde{x}_{k+1}, w_q(k)))^{\frac{1}{1-\beta}}}{\sum\limits_{l=1}^{m}(D_P^2(\tilde{x}_{k+1}, w_q(k)))^{\frac{1}{1-\beta}}}, \\ w_{qi}(k+1) = w_{qi}(k) + \eta(k+1)U_q^{\beta}(k+1)(\tilde{x}_{k+1,i} - w_{qi}(k))\delta_{ki}, \end{cases} \qquad (5)$$

with the second relation (5) that can be represented as learning algorithm for neuro-fuzzy Kohonen network:

$$w_q(k+1) = w_q(k) + \eta(k+1)\phi_q(k+1)(\tilde{x}_{k+1} - w_q(k)) \square\ \delta_k, \qquad (6)$$

where $\phi_q(k+1) = U_q^{\beta}(k+1)$ - bell-shaped neighborhood function, $\delta_k = (\delta_{k1}, \ldots, \delta_{kn})^T$, $\square$ -symbol of direct product.

Thus, using a standard Kohonen network architecture and algorithm of its tuning (6) in on-line mode it is possible to solve the problem of fuzzy clustering data with gaps.

Adaptive algorithm for possibilistic fuzzy clustering

The main disadvantage of probabilistic algorithms is connected with the constraints on membership levels which sum has to be equal unity. This reason has led to the creation of possibilistic fuzzy clustering algorithms [Krishnapuram, 1993].

In possibilistic clustering algorithms the objective function has the form

$$E(U_q(k), w_q, \mu_q) = \sum_{k=1}^{N}\sum_{q=1}^{m} U_q^{\beta}(k) D^2(\tilde{x}_k, w_q) + \sum_{q=1}^{m}\mu_q \sum_{k=1}^{N}(1 - U_q(k))^{\beta} \qquad (7)$$

where the scalar parameter $\mu \geq 0$ determines the distance at which level of membership equals to 0.5, i.e. if $D^2(\tilde{x}_k, w_q) = \mu_q$, then $w_q(k) = 0.5$.

Minimizing (7) relatively $U_q(k)$, w_q and μ_q we get the solution

$$\begin{cases} U_q^{(\tau+1)}(k) = \dfrac{1}{1+\left(\dfrac{D^2(\tilde{x}_k, w_q^{(\tau)})}{\mu_q^{(\tau)}}\right)^{\frac{1}{\beta-1}}}, \\ w_q^{(\tau+1)} = \dfrac{\sum_{k=1}^{N}(U_q^{(\tau+1)}(k))^{\beta}\tilde{x}_k}{\sum_{k=1}^{N}(U_q^{(\tau+1)}(k))^{\beta}}, \\ \mu_q^{(\tau+1)} = \dfrac{\sum_{k=1}^{N}(U_q^{(\tau+1)}(k))^{\beta}D^2(\tilde{x}_k, w_q^{(\tau+1)})}{\sum_{k=1}^{N}(U_q^{(\tau+1)}(k))^{\beta}}, \end{cases} \tag{8}$$

which with $\beta = 2$ and Euclidean metric has the form

$$\begin{cases} U_q^{(\tau+1)}(k) = \dfrac{1}{1+\dfrac{\left\|\tilde{x}_k - w_q^{(\tau)}\right\|^2}{\mu_q^{(\tau)}}}, \\ w_q^{(\tau+1)} = \dfrac{\sum_{k=1}^{N}(U_q^{(\tau)}(k))^{2}\tilde{x}_k}{\sum_{k=1}^{N}(U_q^{(\tau)}(k))^{2}}, \\ \mu_q^{(\tau+1)} = \dfrac{\sum_{k=1}^{N}(U_q^{(\tau)}(k))^{2}\left\|\tilde{x}_k - w_q^{(\tau+1)}\right\|^2}{\sum_{k=1}^{N}(U_q^{(\tau)}(k))^{2}}. \end{cases} \tag{9}$$

Information processing in the on-line mode (8), (9) can be written as [Bodyanskiy, 2005; Gorshkov, 2009]

$$\begin{cases} U_q(k+1)=\dfrac{1}{1+(\dfrac{D^2(\tilde{x}_{k+1},w_q(k))}{\mu_q(k)})^{\frac{1}{\beta-1}}}, \\ w_q(k+1)=w_q(k)+\eta(k+1)U_q^{\beta}(k+1)(\tilde{x}_{k+1}-w_q(k)), \\ \mu_q(k+1)=\dfrac{\sum\limits_{p=1}^{k+1}U_q^{\beta}(p)D^2(\tilde{x}_p,w_q(k+1))}{\sum\limits_{p=1}^{k+1}U_q^{\beta}(p)} \end{cases} \qquad (10)$$

and

$$\begin{cases} U_q(k+1)=\dfrac{1}{1+\dfrac{\left\|\tilde{x}_k-w_q(k)\right\|^2}{\mu_q(k)}}, \\ w_q(k+1)=w_q(k)+\eta(k+1)U_q^{2}(k+1)(\tilde{x}_{k+1}-w_q(k)), \\ \mu_q(k+1)=\dfrac{\sum\limits_{p=1}^{k+1}U_q^{2}(p)\left\|\tilde{x}_p-w_q(k+1)\right\|^2}{\sum\limits_{p=1}^{k}U_q^{2}(p)}. \end{cases} \qquad (11)$$

It's easily to see that relations (10), (11) are the Kohonen's self-learning WTM-rule with Cauchy functions as a neighborhood ones.

Adaptive algorithm for possibilistic fuzzy clustering of data with gaps

Adopting instead of Euclidean metric partial distance (PD), we can write the objective function of the type (7) as

$$E(U_q(k),w_q,\mu_q)=\sum_{k=1}^{N}\sum_{q=1}^{m}U_q^{\beta}(k)\frac{n}{\delta_{k\Sigma}}\sum_{i=1}^{n}(\tilde{x}_{ki}-w_{qi})^2\delta_{ki}+\sum_{q=1}^{m}\mu_q\sum_{k=1}^{N}(1-U_q(k))^{\beta}$$

and then solving the equations system

$$\begin{cases}\dfrac{\partial E(U_q(k),w_q,\mu_q)}{\partial U_q(k)}=0,\\ \dfrac{\partial E(U_q(k),w_q,\mu_q)}{\partial \mu_q}=0,\\ \nabla_{w_q}E(U_q(k),w_q,\mu_q)=\vec{0},\end{cases}$$

get the procedure of type (8), which can be rewritten in the recurrent form

$$\begin{cases}U_q(k)=\dfrac{1}{1+(\dfrac{D_P^2(\tilde{x}_{k+1},w_q(k))}{\mu_q(k)})^{\frac{1}{\beta-1}}},\\ w_{qi}(k+1)=w_{qi}(k)+\eta(k+1)U_q^{\beta}(k+1)(\tilde{x}_{k+1,i}-w_{qi}(k))\delta_{ki},\\ \mu_q(k+1)=\dfrac{\sum\limits_{p=1}^{k+1}U_q^{\beta}(p)D_P^2(\tilde{x}_p,w_q(k+1))}{\sum\limits_{p=1}^{k+1}U_q^{\beta}(p)}.\end{cases}$$

The second relation can be rewritten as

$$w_q(k+1)=w_q(k)+\eta(k+1)U_q^{\beta}(k+1)(\tilde{x}_{k+1}-w_q(k))\square\ \delta_k$$

coinciding with the learning procedure (6).

Thus, the process of fuzzy possibilistic clustering data with gaps can also be realized by using neuro-fuzzy Kohonen network.

Conclusion

The problem of probabilistic and possibilistic on-line fuzzy clustering of data with gaps based on the strategy of partial distances is considered. It is shown that it can be solved on the basis of self-organizing neuro-fuzzy Kohonen network. Proposed learning algorithm is a hybrid of rule "winner-takes-more" and recurrent fuzzy clustering algorithms.

Bibliography

[Bezdek, 1981] J.C. Bezdek. Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York, 1981.

[Hoeppner, 1999] F Hoeppner, F. Klawonn, R. Kruse, T. Runker. Fuzzy Clustering Analysis: Methods for Classification, Data Analysis and Image Recognition. Chichester, John Wiley &Sons, 1999.

[Xu, 2009] R. Xu, D.C. Wunsch. Clustering. Hoboken, N.J. John Wiley & Sons, Inc., 2009.

[Marwala, 2009] T Marwala. Computational Intelligence for Missing Data Imputation, Estimation, and Management: Knowledge Optimization Techniques. Hershey-New York: Information Science Reference, 2009.

[Hathaway, 2001] R.J. Hathaway, J.C Bezdek. Fuzzy c-means clustering of incomplete data. IEEE Trans on Systems, Man, and Cybernetics, №5, 31, 2001, P. 735-744.

[Kohonen, 1995] T. Kohonen. Self-Organizing Maps. Berlin: Springer-Verlag, 1995.

[Bodyanskiy, 2005] Ye. Bodyanskiy. Computational intelligence techniques for data analysis. Lecture Notes in Informatics. Bonn: GI, 2005, V. P-72, P. 15-36.

[Park, 1984] D.C. Park, I. Dagher. Gradient based fuzzy c-means (GBFCM) algorithm. Proc. IEEE Int. Conf. on Neural Networks, 1984, P.1626-1631.

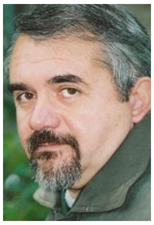
[Chung, 1994] F.L. Chung, T. Lee. Fuzzy competitive learning. Neural Networks, 1994, 7, №3, P.539-552.

[Gorshkov, 2009] Ye. Gorshkov, V. Kolodyazhniy, Ye. Bodyanskiy. New recursive learning algorithms for fuzzy Kohonen clustering network. Proc. 17th Int. Workshop on Nonlinear Dynamics of Electronc Systems. (Rapperswil, Switzerland, June 21-24, 2009) Rapperswil, Switzerland, 2009, P. 58-61.

[Shafronenko, 2011] A. Y. Shafronenko, V.V. Volkova, Ye. Bodyanskiy. Adaptive clustering data with gaps. Radioelectronics, informatics, control. – 2011. - №2. – P. 115-119 (in Russian)

[Krishnapuram, 1993] R. Krishnapuram, J.M. Keller. A possibilistic approach to clustering. Fuzzy Systems, 1993, 1, №2, P.98-110.

Authors' Information



Yevgeniy Bodyanskiy – Professor, Dr. – Ing. habil., Scientific Head of Control Systems Research Laboratory, Kharkiv National University of Radio Electronics, 14 Lenin Ave., Office 511, 61166 Kharkiv, Ukraine; e-mail: bodya@kture.kharkov.ua

Major Fields of Scientific Research: Artificial neural networks, Fuzzy systems, Hybrid systems of computational intelligence

Alina Shafronenko – intern-researcher of Control Systems Research Laboratory, Kharkiv National University of Radio Electronics, 14 Lenin Ave., Office 517, 61166 Kharkiv, Ukraine; e-mail: alinashafronenko@gmail.com

Major Fields of Scientific Research: neural networks, neural network processing of data with gaps, fuzzy clustering, clustering of data

Valentyna Volkova - Candidate of Technical Science (Ph.D.), Senior lecturer in Artificial Intelligence dept., Kharkiv National University of Radioelectronics Lenin Ave., 14, Kharkiv, 61166, Ukraine; e-mail: volkova@kture.kharkov.ua

Major Fields of Scientific Research: neural networks, fuzzy clustering, clustering of data