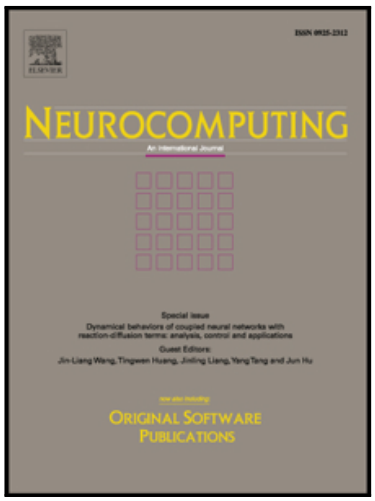

An Evolving Connectionist System for Data Stream Fuzzy Clustering and Its Online Learning

Yevgeniy V. Bodyanskiy , Oleksii K. Tyshchenko , Daria S. Kopaliani

PII: S0925-2312(17)30978-5
DOI: 10.1016/j.neucom.2017.03.081
Reference: NEUCOM 18508

To appear in: *Neurocomputing*

Received date: 12 July 2016
Revised date: 19 January 2017
Accepted date: 6 March 2017

Please cite this article as: Yevgeniy V. Bodyanskiy , Oleksii K. Tyshchenko , Daria S. Kopaliani , An Evolving Connectionist System for Data Stream Fuzzy Clustering and Its Online Learning, *Neurocomputing* (2017), doi: 10.1016/j.neucom.2017.03.081

An Evolving Connectionist System for Data Stream Fuzzy Clustering and Its Online Learning

Yevgeniy V. Bodyanskiy, Oleksii K. Tyshchenko, Daria S. Kopaliani

Control Systems Research Laboratory, Kharkiv National University of Radio Electronics, 14 Nauky Ave, Kharkiv, 61166, Ukraine, tel.: +380577021890

Abstract

An evolving cascade neuro-fuzzy system and its online learning procedure are considered in this paper. The system is based on conventional Kohonen neurons. The proposed system solves a clustering task of non-stationary data streams under uncertainty conditions when data come in the form of a sequential stream in an online mode. A quality estimation process is defined by finding an optimal value of the used cluster validity index.

Keywords – Evolving connectionist system, neuro-fuzzy network, data stream, fuzzy clustering

Introduction

The clustering task (classification without a reference signal) is quite common in many Data Mining applications, i.e. when some original unlabeled data array should be partitioned into homogeneous groups (in the accepted sense) in a self-learning mode. A traditional approach to a clustering task assumes that each observation belongs to a single cluster, while it's a more natural situation when each vector of observations belongs to several classes at the same time with different membership levels. This situation is subject to fuzzy cluster analysis [1-8], and the computational intelligence techniques as well as the neuro-fuzzy approach [9] are widely used to solve the task [10-13]. Most of the fuzzy clustering algorithms process data in a batch mode, i.e. when a whole data array to be processed is given a priori. Initial information for this task is a sample of observations which consists of N $m-$dimensional feature vectors $x(1), x(2), \ldots, x(k), \ldots x(N)$. To ease numerical implementation, initial data are previously transformed in such a way that all the observations should belong to the hypercube $[-1, 1]^n$ or the unit hypersphere $\|x(k)\|^2 = 1$.

A result of this clustering is a partition of the initial data array to M classes with some membership level $u_J(k)$ of the $k-$th feature vector $x(k)$ to the $J-$th cluster, $J = 1, 2, \ldots, M$. It is assumed that N, m as well as clustering parameters (first of all, a fuzzifier) are given a priori, and they don't change while the data are processed.

At the same time, there's a wide class of Dynamic Data Mining and Data Streams Mining tasks, when data come in the form of a sequential stream in an online mode. It's clear that a sample volume N in this case is not limited, and it acquires the meaning of the current discrete time.

Kohonen artificial neural networks [14] are well adapted to solve the clustering task when data come sequentially. These neural networks have one layer with lateral connections, and they are learnt with the «winner takes all» or «winner takes more» principles. These neural networks have demonstrated their effectiveness solving different tasks with non-overlapping (separable) classes. The necessity of solving the fuzzy clustering tasks in real time (as data streams come to a system) has led to self-learning neuro-fuzzy systems, which are hybrid systems consisting of the Kohonen self-organizing map (SOM) and the fuzzy C-means algorithm by Bezdek (FCM)