Available online at www.sciencedirect.com
SCIENCE DIRECT®

ELSEVIER

Neurocomputing 67 (2005) 313–320

NEUROCOMPUTING

www.elsevier.com/locate/neucom

Letter

GreyART network for data clustering

Ming-Feng Yeh*, Shao-Shan Chiang

Department of Electrical Engineering, Lunghwa University of Science and Technology, Taoyuan, Taiwan

Received 20 December 2004; received in revised form 13 January 2005; accepted 13 January 2005
Available online 8 March 2005
Communicated by R.W. Newcomb

Abstract

This paper attempts to incorporate grey relational analysis into the ART 2 (adaptive resonant theory) network to construct a GreyART network. GreyART networks not only possess the structure and learning ability of an ART-based network, but also use grey relational analysis to process the grey information among the data for clustering. The problems of determining the optimal number of clusters and the optimal locations of cluster centres are also considered. Additionally, the network is used to solve two data clustering problems for illustration. Simulation results demonstrate the effectiveness and feasibility of the GreyART network in solving the data clustering problems.

Keywords: Grey relational analysis; Adaptive resonant theory; Data clustering

1. Introduction

According to the grey system theory proposed by Deng in 1982 [3], most of the practical systems are grey systems, in which the information cannot be completely known. Here, the word "grey" means poor, incomplete, or uncertain. Grey relational analysis [4] is an essential topic in grey system theory. For a given reference sequence and a given set of comparative sequences, grey relational analysis can be used to determine the relational grade between the reference and each comparative

*Corresponding author. Tel.: +886 2 82093211; fax: +886 2 82099728.
E-mail address: mfyeh@mail.lhu.edu.tw (M.-F. Yeh).

0925-2312/$ - see front matter
doi:10.1016/j.neucom.2005.01.003

sequence in the given set. Then the best comparative one can be found by further analysing the resultant grey relational grades. Therefore, grey relational analysis can be performed as a similarity measure for finite sequences. Since the relationship between data may be viewed as grey information, the clustering problem can be solved by a grey system. Studies [2,5,8] have successfully shown that grey relational analysis can be applied to cluster analysis.

Adaptive resonant theory (ART) was introduced by Grossberg early in 1976, and thereafter different ART-type networks were subsequently developed by Carpenter et al. [1]. ART architectures are neural networks that carry out stable self-organization of recognition codes for arbitrary sequences of input patterns. Without preliminary training, ART networks not only allow category templates to adapt to current circumstances, but also allow on-line creation of categories during classification sessions. This twofold flexibility is very useful for solving the unsupervised data clustering problem.

The proposed GreyART network, a neural-network architecture combining grey relational analysis and ART 2 network, is designed to find the underlying structure of a given data set. During each clustering session, each category template is viewed as a comparative sequence while the input data is a reference sequence. The similarity measurement obtained by grey relational analysis is further examined by the vigilance test to determine whether the input data belongs to the existing categories or becomes the template of a new category. The final category templates are the estimated clustering centres. Besides, the problem of determining optimal threshold vigilance is also introduced. This value can furthermore determine the optimal number of clusters and the optimal locations of cluster centres for the considered data set.

The remainder of this paper is organized as follows. Section 2 presents some background material. The proposed GreyART network and simulation results are shown in Sections 3 and 4, respectively. Finally, Section 5 contains some conclusions of GreyART network.

2. Grey relational analysis and ART 2 network

Grey relational analysis is a similarity measure for finite sequences with incomplete information [4]. Assume that the reference sequences are defined as $x_i = \langle x_i(1), x_i(2), \ldots, x_i(N)\rangle$, $i \in I = \{1, 2, \ldots, m\}$ and the comparative sequences are given by $y_j = \langle y_j(1), y_j(2), \ldots, y_j(N)\rangle$, $j \in J = \{1, 2, \ldots, n\}$. For a specified reference sequence x_i and all comparative sequences, the grey relational coefficient between x_i and y_j at the kth datum, $k \in K = \{1, 2, \ldots, N\}$, is defined as follows:

$$r(x_i(k), y_j(k)) = \frac{\Delta_{i,\min} + \xi \cdot \Delta_{i,\max}}{\Delta_{ij}(k) + \xi \cdot \Delta_{i,\max}}, \tag{1}$$

where $\Delta_{ij}(k) = |x_i(k) - y_j(k)|$, $\Delta_{i,\max} = \max_{j\in J} \max_{k\in K} \Delta_{ij}(k)$, $\Delta_{i,\min} = \min_{j\in J} \min_{k\in K} \Delta_{ij}(k)$, and $\xi \in (0, 1]$, which is a distinguishing coefficient to control the resolution between $\Delta_{i,\max}$ and $\Delta_{i,\min}$.