

Chapter 2

Fuzzy Cluster Validation Based on Fuzzy PCA– Guided Procedure

K. Honda

Osaka Prefecture University, Japan

T. Matsui

Osaka Prefecture University, Japan

A. Notsu

Osaka Prefecture University, Japan

H. Ichihashi

Osaka Prefecture University, Japan

ABSTRACT

Cluster validation is an important issue in fuzzy clustering research and many validity measures, most of which are motivated by intuitive justification considering geometrical features, have been developed. This paper proposes a new validation approach, which evaluates the validity degree of cluster partitions from the view point of the optimality of objective functions in FCM-type clustering. This approach makes it possible to evaluate the validity degree of robust cluster partitions, in which geometrical features are not available because of their possibilistic natures.

INTRODUCTION

Fuzzy clustering (or fuzzy cluster analysis) is a most active research area in the fuzzy systems field. The fuzzy c-means (FCM) algorithm (Bezdek, 1981) and its variants (Höppner et al., 1999; Miyamoto et al., 2008) have been proved to be useful for data summarization. In the FCM-type clustering models, the objective function is given as the fuzzy membership-weighted inner-cluster errors between data points and cluster prototypes. Although the algorithms are designed

well for finding local optimal solutions based on the iterative optimization scheme, they often derive several different local solutions in the multi-starting strategy. Additionally, the optimal cluster number is not known a priori. Therefore, we need the validation measure for selecting the optimal cluster partition from multiple solutions.

Many validation measures have been proposed, some of which were designed for finding compact and separate clusters from the view point of intuitive geometrical features. Xie-Beni index (Xie & Beni, 1987) and other indices based

on similar concepts (Dunn, 1974, Fukuyama & Sugeno, 1989) measures the cluster separateness using distances among cluster centers. Another approach considers cluster overlapping without using prototypes (Bezdek, 1981; Kim et al., 2003). Although these measures have been intuitively justified, it is not necessarily guaranteed that the validation measures really suit the evaluation of ‘local optima’ of objective functions. So, Rankler (2007) considered the pareto optimal solutions in multi-objective problems of objective functions and validation measures. In this paper, the cluster validity is discussed considering only the optimality of objective functions.

Another topic considered in this paper is noise rejection mechanism in the FCM-type clustering. Noise fuzzy clustering (Davé, 1991) uses an additional “noise cluster” so that noise samples are dumped into it. Masulli and Rovetta (2006) proposed a technique for soft transition from the conventional FCM constraint to a robust situation where noise samples are rejected (or ignored). Although noise rejection mechanisms are useful for obtaining cluster structures for contaminated data, they also create difficulties when we apply conventional validity measures designed for fuzzy clustering.

In this paper, a new cluster validation approach is considered based on principal component analysis (PCA)-guided procedures. The PCA-guided k -means (Ding & He, 2004a) is a deterministic method for finding the optimal k -means partition, in which a relaxed cluster indicator is estimated from a rotated principal component score matrix, though the rotation matrix cannot be explicitly estimated. Honda et al. (2010) introduced the noise rejection mechanism and proposed fuzzy PCA-guided robust k -means (FPR k -means). Because these PCA-guided procedures can find the optimal cluster structure considering only the optimality of objective functions, we can validate candidate cluster partitions if we estimate the rotation matrix for reconstructing the cluster indicators. In the proposed validation approach,

a candidate for the pseudo-optimal indicator is first reconstructed by Procrustean transformation (Procrustean rotation) and a fair deviation from the pseudo-optimal solution is calculated for the current partition. Then, the cluster partition having least deviation is selected.

The remainder of the article is organized as follows. The next section describes some related works. The following section outlines the proposed validation approach. Experimental results are then shown. The final section outlines a summary and the conclusions.

RELATED WORK

FCM-Type Fuzzy Clustering and Cluster Validation

Let $\mathbf{x}_i, i=1, \dots, n$ be n samples characterized by m dimensional observation. The FCM objective function for partitioning samples into C fuzzy clusters is defined as the generalized sum of within-cluster errors.

$$L_{fcm} = \sum_{c=1}^C \sum_{i=1}^n u_{ci}^\theta \|\mathbf{x}_i - \mathbf{b}_c\|^2 \quad (1)$$

where $u_{ci} \in [0, 1]$ represents the membership degree of sample i to cluster c . Usually, the sum of u_{ci} with respect to c is constrained to be 1 and the constraint is called the “probabilistic constraint”. θ is the fuzzification index called “fuzzifier” that tunes the degree of fuzziness of memberships. \mathbf{b}_c is the prototypical centroid vector of cluster c . When $\theta=1$, the FCM model is reduced to k -means (or hard c -means) (MacQueen, 1967), in which $u_{ci} \in \{0, 1\}$ is given based on the nearest prototype assignment.

In order to determine the optimal fuzzy partition in the iterative optimization scheme, we must select the optimal cluster number and the suitable

10 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/fuzzy-cluster-validation-based-fuzzy/67479

Related Content

Vehicular Networks Security: Attacks, Requirements, Challenges and Current Contributions

Christian Tchepnda, Hassnaa Moustafa, Houda Labiodand Gilles Bourdon (2009). *International Journal of Ambient Computing and Intelligence* (pp. 39-52).

www.irma-international.org/article/vehicular-networks-security/1371

Using Semantic Technologies for the Support of Engineering Design Processes

Sebastian C. Brandt, Marcus Schlüterand Matthias Jarke (2008). *Intelligent Information Technologies and Applications* (pp. 222-248).

www.irma-international.org/chapter/using-semantic-technologies-support-engineering/24267

Relative Superiority of Key Centrality Measures for Identifying Influencers on Social Media

Yifeng Zhangand Xiaoqing Li (2014). *International Journal of Intelligent Information Technologies* (pp. 1-23).

www.irma-international.org/article/relative-superiority-of-key-centrality-measures-for-identifying-influencers-on-social-media/123941

AI-Driven Crop Yield Prediction and Disease Detection in Agroecosystems

Muhammad Waseem, Aamir Razaand Abdul Malik (2024). *Maintaining a Sustainable World in the Nexus of Environmental Science and AI* (pp. 229-258).

www.irma-international.org/chapter/ai-driven-crop-yield-prediction-and-disease-detection-in-agroecosystems/355513

Transforming Birmingham City with Smart Applications

Luis Hernandez-Munoz, Gerald Feldman, Vahid Javidroozi, Annette C. Kingand Raj Mack (2016). *International Journal of Conceptual Structures and Smart Applications* (pp. 16-37).

www.irma-international.org/article/transforming-birmingham-city-with-smart-applications/176585