Available online at www.sciencedirect.com

ScienceDirect

Fuzzy Sets and Systems 281 (2015) 272–279

FUZZY sets and systems

www.elsevier.com/locate/fss

Fuzzy clustering: More than just fuzzification

Frank Klawonn [a,b,*], Rudolf Kruse [c], Roland Winkler [d]

[a] *Department of Computer Science, Ostfalia University of Applied Sciences, Salzdahlumer Str. 46/48, 38302 Wolfenbuettel, Germany*
[b] *Biostatistics, Helmholtz Centre for Infection Research, Inhoffenstr. 7, 38124 Braunschweig, Germany*
[c] *Department of Computer Science, Otto-von-Guericke-University, Universitaetsplatz 2, 39106 Magdeburg, Germany*
[d] *Leibniz Institute of Astrophysics Potsdam (AIP), An der Sternwarte 16, 14482 Potsdam, Germany*

Received 6 January 2015; received in revised form 15 May 2015; accepted 15 June 2015

Available online 16 July 2015

Abstract

The initial idea of extending the classical k-means clustering technique to an algorithm that uses membership degrees instead of crisp assignments of data objects to clusters led to the invention of a large variety of new fuzzy clustering algorithms. However, most of these algorithms are concerned with cluster shapes or outliers and could have been defined without any problems in the context of crisp assignments of data objects to clusters. In this paper, we demonstrate that the use of membership degrees for these algorithms – although it is not necessary from the theoretical point of view – is essential for these algorithms to function in practice. With crisp assignments of data objects to clusters these algorithms would get stuck most of the time in a local minimum of their underlying objective function, leading to undesired clustering results. In other contributions it was shown that the use of membership degrees can avoid this problem of local minima but it also introduces new problems, especially for clusters with varying density and for high-dimensional data, at least if fuzzy clustering is carried out with the simple standard fuzzifier.

Keywords: Fuzzy cluster analysis; Fuzzifier; Local optima; High-dimensional data

1. Introduction

Cluster analysis is an unsupervised data analysis task, trying to group data objects into clusters, such that similar data objects are assigned to the same cluster whereas dissimilar data objects should belong to different clusters. This informal definition of cluster analysis already implies that a notion of similarity is needed to cluster data. It is out of the scope of this paper, to provide an overview on the large variety of existing approaches to clustering data. We confine our discussion to clustering multivariate real-valued data and to extensions of the classical k-means clustering algorithm (see for instance [1]), since until today the large majority of fuzzy clustering approaches were based on

* Corresponding author at: Department of Computer Science, Ostfalia University of Applied Sciences, Salzdahlumer Str. 46/48, 38302 Wolfenbuettel, Germany. Tel.: +49 5331 939 31100; fax: +49 5331 939 31004.
E-mail address: f.klawonn@ostfalia.de (F. Klawonn).

http://dx.doi.org/10.1016/j.fss.2015.06.024
0165-0114/

the initial generalisation of k-means clustering to fuzzy c-means clustering as it was proposed by Dunn [2] and Bezdek [3,4].

An obvious question to be asked is why there is or was a need to extend an algorithm like k-means clustering to allow for membership degrees instead of crisp assignments of data objects to clusters. First of all, the idea of "soft" assignments of data objects to clusters in terms of probabilities has been around long before in the form of mixture models, especially the well-known Gaussian mixture models (for an overview see for instance [5]). Gaussian mixture models assume that the clusters originate from multivariate normal distributions, so that the whole data set can be described as a sample from a mixture of normal distributions. Such mixture models estimate the parameters of the underlying multivariate normal distributions as well as the a priori probabilities for each cluster or multivariate normal distribution to generate a data point. On this basis, mixture models compute (posterior) probabilities that a specific data point was generated by a specific multivariate normal distribution, i.e. the probability whether it belongs to the corresponding cluster. In a Gaussian mixture model, data points that can be assigned very well to a specific cluster will have a probability close to one for this cluster and a probability close to zero for the other clusters, whereas ambiguous data points between clusters will have moderate probabilities to two or more clusters. One could argue that the concept of fuzziness is different from that of probability and therefore a membership degree to a cluster is different from the probability to belong to a cluster. But at least the original fuzzy extension of the k-means clustering algorithm puts a probabilistic constraint on the membership degrees. So this cannot be the main reason for the use of fuzzy clustering.

Connecting fuzzy rule-based systems to data is another motivation for fuzzy clustering. Sugeno and Yasukawa [6] used fuzzy clustering to cluster the output or target attribute of a fuzzy system to generate fuzzy rules. Since in this case only one-dimensional data are considered for clustering, this is essentially a fuzzified discretisation technique. But could this not also be done by standard discretisation techniques like equi-width, equi-frequency or V-optimal discretisation (for an overview see for example [7]) and then using the resulting intervals to define corresponding fuzzy sets? In [8,9] multidimensional fuzzy clusters are used to construct fuzzy rules. But there are specifically tailored methods to derive fuzzy rules from data, for instance neuro-fuzzy methods [10].

A clear argument in favour of fuzzy clustering is the possibility of new cluster validity measures like Bezdek's partition coefficient or Bezdek's partition entropy [4] that only make sense in the context of fuzzy clustering. Such validity measures are used to judge whether a clustering result is meaningful and also to determine the number of clusters. In [11], visualisation techniques based on the membership degrees were used to judge the quality of clusters.

But there is one very important reason speaking in favour of fuzzy clustering that is discussed in more detail in this paper. Fuzzy clustering can help to avoid algorithmic problems from which methods like the k-means clustering algorithm suffer. The result of k-means clustering highly depends on the initialisation of the algorithm, leading to undesired clustering results. This is not true for fuzzy clustering. In this sense, fuzzy clustering is not only an improvement of k-means clustering. It also opens the possibility to introduce more flexible and sophisticated clustering models than the simple k-means algorithm, still avoiding the problem of undesired clustering results.

After a brief introduction of the fuzzy extension of the k-means algorithm in Section 2, Section 3 briefly reviews a number of extensions to the fuzzy c-means algorithm that would theoretically but not practically function if the same extensions were applied to hard k-means clustering. Section 4 demonstrates and explains why these algorithms – although essentially not being fuzzy – require the concept of fuzzy clustering. In Section 5 we discuss alternatives to the original fuzzification of the k-means clustering algorithm before we briefly conclude the paper in Section 6.

2. From crisp to fuzzy clustering

Although first ideas on fuzzy clustering were already proposed at the end of the sixties of the last century by Ruspini [12], the breakthrough for fuzzy clustering came with the fuzzification of the k-means algorithm. In the k-means algorithm, each cluster is represented by a so-called prototype which is supposed to be the centre of the corresponding cluster. Given a data set $\{x_1, \dots, x_n\} \subset \mathbb{R}^q$, the algorithm is based on the objective function

$$f = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij} d_{ij} \tag{1}$$

that should be minimised under the constraints