Chapter 1
Fuzzy C-Means in High Dimensional Spaces

Roland Winkler
German Aerospace Center, Germany

Frank Klawonn
Ostfalia University, Germany

Rudolf Kruse
Otto-von-Guericke University, Germany

ABSTRACT

High dimensions have a devastating effect on the FCM algorithm and similar algorithms. One effect is that the prototypes run into the centre of gravity of the entire data set. The objective function must have a local minimum in the centre of gravity that causes FCM's behaviour. In this paper, examine this problem. This paper answers the following questions: How many dimensions are necessary to cause an ill behaviour of FCM? How does the number of prototypes influence the behaviour? Why has the objective function a local minimum in the centre of gravity? How must FCM be initialised to avoid the local minima in the centre of gravity? To understand the behaviour of the FCM algorithm and answer the above questions, the authors examine the values of the objective function and develop three test environments that consist of artificially generated data sets to provide a controlled environment. The paper concludes that FCM can only be applied successfully in high dimensions if the prototypes are initialized very close to the cluster centres.

INTRODUCTION

Clustering high dimensional data has many interesting applications. For example clustering similar music files, semantic web applications, image recognition or biochemical problems. Many tools today are not designed to handle hundreds of dimensions, or in this case, it might be better to call it degrees of freedom. Many clustering approaches work quite well in low dimensions, especially the fuzzy c-means algorithm (FCM) (Dunn, 1973; Bezdek, 1981; Höppner et al., 1999; Kruse et al., 2007) seems to fail in high dimensions. Hence FCM is so useful in cases where data object arrangements that are not crisp, this paper is dedicated to give some insight into the behaviour of FCM in high dimensions.

DOI: 10.4018/978-1-4666-1870-1.ch001

Copyright © 2013, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

One of the problems of FCM is, that the prototypes tend to run into the centre of gravity of the complete data set, almost independently of the initialisation of the prototypes. Figure 1 shows on the right hand side a 50 dimensional artificial data set, projected on 2 dimensions, containing 100 clusters, 100 data objects each. The lines represent the way the prototypes took from their initial position. The left hand side shows FCM in 2 dimensions with 4 clusters where it works quite well. It appears that the structural problem of FCM is independent from the data set, this question is addressed later in the paper. The 4 questions presented in the abstract will be answered in this paper:

- How many dimensions are at least necessary to cause an ill behaviour of FCM?
- How does the number of prototypes influence ill behaviour?
- Why has the objective function a local minima in the centre of gravity?
- How must FCM be initialised to avoid the local minima in the centre of gravity?

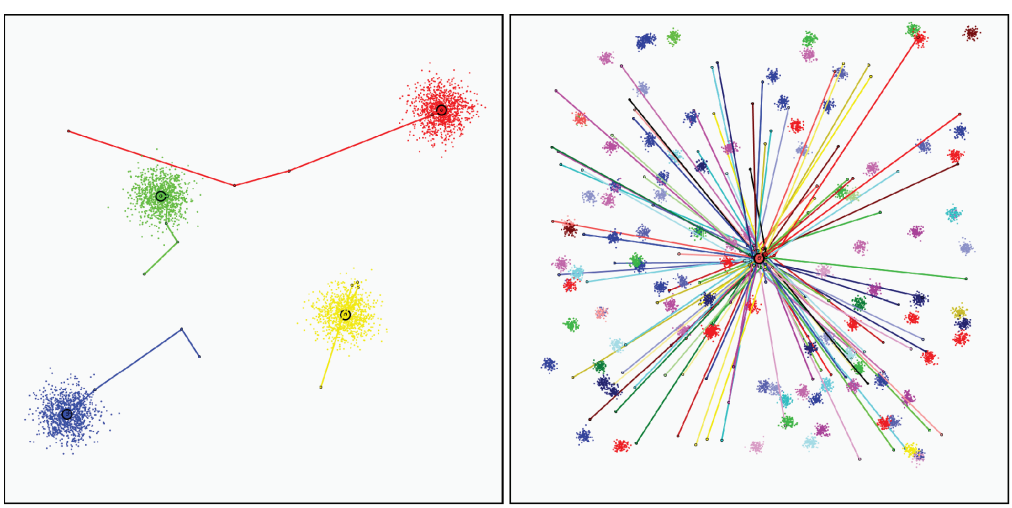
These questions are independent of the actual data set FCM is applied to. The idea is to use test environments that are optimal for FCM, so that it is likely that FCM fails on all data sets if it fails on the optimal ones. Four data sets D_1 through D_4 are used which have different properties and are introduced Q_1. is answered using D_1, Q_2 is answered using D_2, Q_3 by analysing the objective function and for Q_4 the data sets D_3 and D_4 are used. The answers to the questions are found by observing the objective function of FCM.

Related Work

The curse of dimensionality was often addressed in literature. Most of the time this term appears is, to invent a clustering algorithm that does not suffer from it by design. For example in Hinneburg and Keim (1999) a method is proposed to subdivide the input space by creating a non-axis parallel grid. This subdivides the input space in such a way, that the grid boarders do not cross areas of high density but rather follow empty space areas. Also Steinbach et al. (2004) gives a good overview about a collection of methods that address clustering in high dimensional spaces.

In this paper, outliers are not addressed. However, since there are many methods to detect them in lower dimensions, they tend to fail in higher

Figure 1. Fuzzy c Means in 2 dimensions (left) and 50 dimensions (right)



14 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-means-high-dimensional-spaces/67478

Related Content

Design Consideration of Sociomaterial Multi-Agent CSCW Systems
Tagelsir Mohamed Gasmelseid (2015). *Recent Advances in Intelligent Technologies and Information Systems (pp. 83-103).*
www.irma-international.org/chapter/design-consideration-of-sociomaterial-multi-agent-cscw-systems/125505

User Relevance Feedback in Semantic Information Retrieval
Antonio Picarielloand Antonio M. Rinaldi (2007). *International Journal of Intelligent Information Technologies (pp. 36-50).*
www.irma-international.org/article/user-relevance-feedback-semantic-information/2417

The Role of AI in Crypto Currency Market Expansion With Cyber Security to Analyze Cryptography Methods and Protocols to Make Systems Secure
J. Vijaya, Jajam Nagaraju, J. Manokaran, P. Akhiland Prachee Dewangan (2026). *Governance and Automated Security in AI-Driven Cybersecurity (pp. 307-350).*
www.irma-international.org/chapter/the-role-of-ai-in-crypto-currency-market-expansion-with-cyber-security-to-analyze-cryptography-methods-and-protocols-to-make-systems-secure/413087

Empowering Learner Autonomy for Developing Writing Skills Through AI-Enhanced Reflective Journals
Sena Könezand Selami Aydin (2026). *Harnessing AI's Potential to Support Student Success and Teaching Excellence (pp. 217-244).*
www.irma-international.org/chapter/empowering-learner-autonomy-for-developing-writing-skills-through-ai-enhanced-reflective-journals/385749

Low Dimensional Data Privacy Preservation Using Multi Layer Artificial Neural Network
R. VidyaBanuand N. Nagaveni (2012). *International Journal of Intelligent Information Technologies (pp. 17-31).*
www.irma-international.org/article/low-dimensional-data-privacy-preservation/69388