

## COLLABORATIVE AND KNOWLEDGE-BASED FUZZY CLUSTERING

WITOLD PEDRYCZ

Department of Electrical and Computer Engineering  
University of Alberta  
Edmonton, Canada

Systems Research Institute  
Polish Academy of Sciences  
Warsaw, Poland  
pedrycz@ee.ualberta.ca

Received July 2006; revised October 2006

*ABSTRACT.* Clustering is commonly regarded as a synonym of unsupervised learning aimed at the discovery of structure in highly dimensional data. With a plethora of existing algorithms, the area offers an evident diversity of possible approaches along with their underlying features and potential applications. When augmented by fuzzy sets, fuzzy clustering has become an integral component of Computational Intelligence (CI) and is now broadly exploited in fuzzy modeling, fuzzy control, pattern recognition, and exploratory data analysis. A lot of pursuits of CI are human-centric in the sense they are either initiated or driven by some domain knowledge or the results generated by the CI constructs are made easily interpretable. In this sense, to follow the tendency of human-centricity so profoundly visible in the CI domain, the very concept of fuzzy clustering needs to be carefully revisited. We propose a certain paradigm shift that brings us to the idea of knowledge-based clustering in which the development of information granules – fuzzy sets is governed by the use of data as well as domain knowledge supplied through an interaction with the developers, users and experts. In this study, we elaborate on the concepts and algorithms of knowledge-based clustering by considering the well known scheme of Fuzzy C-Means (FCM) and viewing it as an operational model using which a number of essential developments could be easily explained. The fundamental concepts discussed here involve clustering with domain knowledge articulated through partial supervision and proximity-based knowledge hints. Furthermore we exploit the concepts of collaborative as well as consensus driven clustering. Interesting and useful linkages between information granularity and privacy and security of data are also discussed.

**Keywords:** Knowledge – guidance in data analysis, Data mining, Collaborative clustering

**1. Introduction.** The human-centric facet of Computational Intelligence (CI) becomes profoundly visible in a significant number of developments. One could mention here system modeling, pattern recognition, and decision-making. In data analysis tasks completed in the setting of the CI, the phenomenon of human-centricity manifests quite vividly in several ways and the needs there are well articulated. First, the results are presented at some suitable level of abstraction secured by the use of information granules. Likewise the semantics of the information granules that are used to organize findings about data is conveyed in the language of fuzzy sets whose interpretation is quite intuitive. In this