Expert Systems With Applications 46 (2016) 274–285

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Design of rule-based models through information granulation

Jeremy Kerr-Wilson*, Witold Pedrycz

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

A R T I C L E   I N F O

Keywords:
Fuzzy model
Hierarchical clustering
Information granulation
Fuzzy rules
Fuzzy C-Means (FCM)

A B S T R A C T

In this study, we explore the combination of two well defined topics in fuzzy systems research: fuzzy rule based systems, and information granulation. Rule based systems are a powerful and well-studied form of knowledge representation, due to their approximation abilities and interpretability. In recent years, these types of systems have become increasingly powerful with regards to modeling accuracy; however, many of these improvements come at the cost of model interpretability. This recent direction of research has left an unexplored avenue towards the generation of increasingly interpretable fuzzy rule based models, which we intend to explore. Information granulation is a relatively new, yet very promising area of research in human centric systems. As a form of knowledge representation, information granulation is very well suited to fuzzy rule based systems, where rules represent linguistic quantities in a, intuitively understandable format. It is notable that the combination of these two concepts has been left largely unstudied. We aim to explore this union by defining a methodology for the construction of a partially granular fuzzy rule based model. The aim of this novel model format is to provide a first step in the improvement of fuzzy model interpretability, through the use of information granulation. We are additionally interested in studying new ways of generating fuzzy rules; hence, we will also look at the use of hierarchical clustering as a potential alternative to the tried and tested Fuzzy C Means clustering algorithm. The models created using hierarchical clustering are then compared with those generated using Fuzzy C Means to evaluate the effectiveness of this algorithm. As a result of these experiments, we demonstrate that partially granular fuzzy rules are capable of providing a significant improvement to fuzzy rule interpretability, and we believe that granular fuzzy models present an exciting avenue of future research in human centric systems.

1. Introductory notes

Rule based models have been a popular method of knowledge representation for many years in machine learning research. Since their introduction, rule based models have been extensively studied in a variety of forms, and through a variety of algorithms and methodologies. Fuzzy rule based systems are one such example which continue to be an active area of research, with many studies over recent years including those by Kung and Su (2007), Chen (2007), Fernández, Calderón, Barrenechea, Bustince, and Herrera,(2010), Štěpnička and Baets (2013), and Dvořák, Štěpnička, and Štěpničková (2014). Recently, the bulk of research attention has focused on small adjustments and improvements to established rule based model formats. Such studies include those examining hierarchical rule based structures (Wang, Kwong, Jin, Wei, & Man, 2005) (D'Andrea, & Lazzerini, 2013), those defining fuzzy rule networks (Gegov, Arabikhan, & Petrov, 2015), and those examining the challenges of handling of large amounts of data (López, del Río, Benítez, & Herrera, 2014). Each of the listed studies introduces a novel contribution to the field of fuzzy rule based systems; however, it is interesting to note that each study leaves the format of fuzzy rules unchanged, continuing to work with traditional Takagi–Sugeno style fuzzy rules.

One young but promising field of research is the study of rough sets and information granulation. These areas have develop relatively recently, but have still garnered significant attention in human centric computing. In the late 1990's, a study by Zadeh pioneered the concept of using information granulation in fuzzy modeling, employing information granules as linguistic quantities (Zadeh, 1997). Other recent research on this topic includes that of (Kang, Li, Wang, & Qu, 2012), (Qian, Wang, Cheng, Liang, & Dang, 2014), (Pedrycz & Bargiela, 2012), (Balamash, Pedrycz, Al-Hmouz, & Morfeq, 2015), and (Pedrycz, 2014). Information granulation is an intriguing tool in machine learning, as it provides a way to improve the interpretability and human readability of otherwise obfuscated mathematical models. As fuzzy modeling has a significant goal of providing a human centric knowledge representation, the use of information granules in this type of modeling research seems logical and intuitive.

In this study, we seek to extend the well-established, traditional, format of fuzzy rule based models, specifically Mamdani (1974) or

* Corresponding author. Tel.: +1 780 217 9734; fax: +1 780 492 1811.
E-mail addresses: kerrwils@ualberta.ca (J. Kerr-Wilson), wpedrycz@ualberta.ca (W. Pedrycz).

http://dx.doi.org/10.1016/j.eswa.2015.10.030
0957-4174/

Takagi and Sugeno (1985) style fuzzy rules, by transforming the standard numeric or functional conclusion of a fuzzy rule into granular format. The motivation for this change is rooted in a desire to create a more human centric knowledge representation, wherein rules provide readers with more information about the system than is traditionally discernable at a glance. As information granules are a well-established human centric data structure, their use in this manner seems logical and intuitive. Standard evaluation metrics for fuzzy rule based systems are largely incompatible with the evaluation of granular results; hence, we will also benefit from the definition of some simple evaluation metrics designed specifically for the evaluation of our granular rule format.

While both fuzzy rule based models and fuzzy granulation are well studied on their own, their combination has been left largely unexplored. This becomes a curious gap in current literature as both rule base models, and fuzzy granulation have received substantial attention, and their combination would seem a logical progression in human centric modeling. We recognize this as a key area of novelty, which we intend to explore in this paper.

While many forms of clustering have been used in the generation of rule based models, hierarchical clustering has been left unexplored as a tool for the generating of fuzzy models. This exclusion may be due to a number of factors including the crisp nature of the algorithm, or the increased computational complexity of the currently available implementations. In this study, we explore the use of hierarchical clustering in the generation of fuzzy rule based models, with the intention of identifying its candidacy as a vehicle for future study.

We therefore propose two novel aspects of fuzzy modeling which we examine in this study. First, we propose the extension of traditional Mamdani style fuzzy rules through the use of information granulation. Our goal with this modification is to progress fuzzy rule based systems in a more human centric direction. Second, we examine the effectiveness of hierarchical clustering as an unexplored tool for the generation of fuzzy rules. Both of these aspects provide significant novelty to our study, by the establishment of a new granular rule format, and through the exploration of new potential rule generation methods. We must note that this study limits itself with regards to the complexity of the methods presented, notably only applying granularization to fuzzy rule conclusions, and keeping the algorithms involved in granule calculation very simple. Hence, we leave the optimization of interval creation and the study more complex rule granularizations to future studies.

The following paper will be organized in the following manner. First, we will review generic fuzzy rule based models. Second we will introduce the use of hierarchical clustering in fuzzy rule generation, and will provide some numerical analysis of hierarchical clustering performance for this task, notably in contrast to FCM. Next we introduce a method for the granulation of a rule based model conclusions, and introduce two evaluation metrics for this new type of model. Finally we will present some example results using the methods and metrics described, and evaluate them accordingly.

2. The design of generic fuzzy rule-based models

In this section, we briefly review the commonly used and well-established form of fuzzy rule-based models and highlight the main design phases. Rules realizing a mapping from $\boldsymbol{R}^n$ to $\boldsymbol{R}$ and coming in their generic format read as follows

$$-\text{if } \boldsymbol{x} \textit{ is } A_i \text{ then } y \textit{ is } b_i \tag{1}$$

$I = 1, 2, \ldots c$. The condition parts of the rules are information granules and in case of fuzzy rule-based models, they are fuzzy sets. More specifically, A_i standing in (1) denotes a fuzzy set expressed in the n-dimensional input space, that is A_i: $R^n \to [0,1]$, and b_i are constants (constant functions) denoting the conclusion of the corresponding rules. It is worth noting that the rules of this format are a realization of the Mamdani type of rules (Mamdani, 1974) “if $\boldsymbol{x}$ is A_i then y is D_i” with the decoding being implemented by taking the numeric representative of the fuzzy set of conclusion D_i (such as its mean or modal value). At the same time, these rules can be regarded as the simplest version of the Takagi–Sugeno rules (Takagi & Sugeno, 1985) “if $\boldsymbol{x}$ is A_i then y is $f_i(\boldsymbol{x})$” with the functional format of the conclusion (local functions) being the constant ones ($f_i(\boldsymbol{x}) = b_i$). For any $\boldsymbol{x}$ in the input space, the inference (mapping) mechanism (Di Nola, Pedrycz, Sessa, & Sanchez, 1991) realized in fuzzy rule-based models takes into account a collection of activation levels of the rules associated with this input, namely $A_i(\boldsymbol{x})$, and aggregates those with the constant conclusions. In this way we obtain

$$y = \sum_{i=1}^{c} A_i(\mathbf{x}) b_i \tag{2}$$

The design of rule-based models entails two fundamental development phases:

(i) *Granulation of input space*. The input space is granulated by forming a collection of fuzzy sets (A_i) and in this way revealing a structure in the space of input data. Commonly fuzzy clustering is used here. In fuzzy rules, Fuzzy C-Means (FCM) (Gustafson & Kessel, 1979) is the dominant algorithmic vehicle to build fuzzy sets. The role of this method is profoundly visible in the literature [5–8]. From the design perspective, the main parameters of the FCM such as the number of clusters c (the number the rules) and the fuzzification coefficient m (assuming values greater than 1) can be adjusted to optimize the resulting fuzzy model. The distance function ||.|| used in the FCM is predominantly the Euclidean one or it comes as its weighted (normalized) counterpart involving standard deviations of the individual input variables regarded as the weight coefficients. Obviously, the choice of the distance function impacts the geometry of the clusters and subsequently entails modeling capabilities supported by the ensuing rule-based models. There are also a number of more specialized versions of the FCM investigated in fuzzy models (Setnes, 1999) (Kung & Su, 2007). The common feature of the design tools used in this phase is that those are techniques of unsupervised learning.

(ii) Determination of the conclusions of the rules. This design phase follows the formation of information granules in $\boldsymbol{R}^n$ and is concerned with a specification of a class of local models (type of functions f_i) and estimation their parameters. Given that the type of the function has been already fixed (m_i), the values of these constants are optimized in a supervised mode by minimizing a certain performance index Q in such a way that a distance between the data and the corresponding results produced by the model are made as close as possible. The supervised mode of learning (estimation of the parameters) is based on input-output data ($\boldsymbol{x}_k$, $target_k$), k=1, 2, …, N. Quite commonly Q is expressed as a sum of squared errors

$$Q = \sum_{k=1}^{N} (target_k - y_k)^2 \tag{3}$$

where y_k is the output of the rule-based model produced for $\boldsymbol{x}_k$; y_k = FM($\boldsymbol{x}_k$) where FM is governed by (2). Likewise one can consider a RMSE version of (3) taking on the form $\sqrt{\frac{1}{N}\sum_{k=1}^{N}(target_k - y_k)^2}$.

For the design purpose, the data are split into training and testing parts or eventually training-validation-testing parts (in case of optimization of some structure-implied architecture of the model) and the corresponding values of RMSE index serve as indicators of the performance of the model.