Chapter 3

Supervised and Unsupervised Information Granulation: A Study in Hyperbox Design

Andrzej Bargiela
The University of Nottingham, UK

Witold Pedrycz
Polish Academy of Sciences, Poland

ABSTRACT

In this study, we are concerned with information granulation realized both in supervised and unsupervised mode. Our focus is on the exploitation of the technology of hyperboxes and fuzzy sets as a fundamental conceptual vehicle of information granulation. In case of supervised learning (classification), each class is described by one or more fuzzy hyperboxes defined by their corresponding minimum and maximum vertices and the corresponding hyperbox membership function. Two types of hyperboxes are formed, namely inclusion hyperboxes that contain input patterns belonging to the same class, and exclusion hyperboxes that contain patterns belonging to two or more classes, thus representing contentious areas of the pattern space. With these two types of hyperboxes each class fuzzy set is represented as a union of inclusion hyperboxes of the same class minus a union of exclusion hyperboxes. The subtraction of sets provides for efficient representation of complex topologies of pattern classes without resorting to a large number of small hyperboxes to describe each class. The proposed fuzzy hyperbox classification is compared to the original Min-Max Neural Network and the General Fuzzy Min-Max Neural Network and the origins of the improved performance of the proposed classification are identified. When it comes to the unsupervised mode of learning, we revisit a well-known method of Fuzzy C-Means (FCM) by incorporating Tchebyschev distance using which we naturally form hyperbox-like prototypes. The design of hyperbox information granules is presented and the constructs formed in this manner are evaluated with respect to their abilities to capture the structure of data.

DOI: 10.4018/978-1-60566-324-1.ch003
1. INTRODUCTION

Fuzzy hyperbox classification derives from the original idea of Zadeh (1965) of using fuzzy sets for representation of real-life data. Such data frequently is not crisp (has a binary inclusion relationship) but rather has a property of a degree of membership. In this case the use of traditional set theory introduces unrealistic constraint of forcing binary decisions where the graded response is more appropriate. An early application of fuzzy sets to the pattern classification problem (Bellman et al., 1966) proves the point that fuzzy sets represent an excellent tool simplifying the representation of complex boundaries between the pattern classes while retaining the full expressive power for the representation of the core area for each class. By having classes represented by fuzzy set membership functions it is possible to describe the degree to which a pattern belongs to one class or another.

Bearing in mind that the purpose of classification is the enhancement of interpretability of data or, in other words, derivation of a good abstraction of such data the use of hyperbox fuzzy sets as a description of pattern classes provides clear advantages. Each hyperbox can be interpreted as a fuzzy rule. However, the use of a single hyperbox fuzzy set for each pattern class is too limiting in that the topology of the original data is frequently quite complex (Zadeh, 1965) (and incompatible with the convex topology of the hyperbox). This limitation can be overcome by using a collection (union) of hyperboxes to cover each pattern class set (Simpson, 1992, 1993), (Gabrys, Bargiela, 2000). Clearly, the smaller the hyperboxes the more accurate cover of the class set can be obtained. Unfortunately, this comes at the expense of increasing the number of hyperboxes thus eroding the original objective of interpretability of the classification result. We have therefore a task of balancing the requirements of accuracy of coverage of the original data (which translates on the minimization of misclassifications) with the interpretability of class sets composed of many hyperboxes. These concerns are not unique to hyperboxes and they demonstrate themselves also in the context of other topologies of information granules, (Bargiela, 2001).

The tradeoff originally proposed by Simpson (1992) was the optimization of a single parameter defining the maximum hyperbox size as a function of misclassification rate. However, the use of a single maximum hyperbox size is somewhat restrictive. For class sets that are well separated from each other the use of large hyperboxes is quite adequate while for the closely spaced class sets, with a complex partition boundary, there is a need for small hyperboxes, so as to avoid high misclassification rates. One solution to this problem, proposed in (Gabrys, Bargiela, 2000), is the adaptation of the size of hyperboxes so that it is possible to generate larger hyperboxes in some areas of the pattern space while in the other areas the hyperboxes are constrained to be small to maintain low misclassification rates. The adaptation procedure requires however several presentations of data to arrive at the optimum sizes of hyperbox sizes for the individual classes.

In this paper we consider an alternative approach to achieving low misclassification rate while maintaining good interpretability of the classification results. Rather than trying to express the class sets as a union of fuzzy hyperbox sets (Gabrys, Bargiela, 2000), (Simpson, 1992), we represent them as a difference of two fuzzy sets. The first set is a union of hyperboxes produced in the standard way and the second set is a union of intersections of all hyperboxes that belong to different classes. We will refer to the first type of hyperboxes as inclusion hyperboxes and the second type as exclusion hyperboxes. By subtracting the exclusion hyperboxes from the inclusion ones it is possible to express complex topologies of the class set using fewer hyperboxes. Also, the three steps of the Min-Max clustering (Gabrys, Bargiela, 2000), (Simpson, 1992), namely expansion, overlap test and contraction can be reduced to...