Chapter V

A New Approach to Classification of Imbalanced Classes via Atanassov’s Intuitionistic Fuzzy Sets

Eulalia Szmidt
Polish Academy of Sciences, Poland

Marta Kukier
Polish Academy of Sciences, Poland

ABSTRACT

We present a new method of classification of imbalanced classes. The crucial point of the method lies in applying Atanassov’s intuitionistic fuzzy sets (which are a generalization of fuzzy sets) while representing the classes during the first training phase. The Atanassov’s intuitionistic fuzzy sets are generated according to an automatic and mathematically justified procedure from the relative frequency distributions representing the data. Next, we use the information about so-called hesitation margins (which besides membership and non-membership values characterize Atanassov’s intuitionistic fuzzy sets) making it possible to improve the results of data classification. The results obtained in the testing phase were examined not only in the sense of general error/accuracy but also by using confusion matrices, that is, exploring a detailed behavior of the intuitionistic fuzzy classifiers. Detailed analysis of the errors for the examined examples has shown that applying Atanassov’s intuitionistic fuzzy sets gives better results than the counterpart approach via fuzzy sets. Better performance of the intuitionistic fuzzy classifier concerns mainly the recognition power of a smaller class. The method was tested using a benchmark problem from UCI machine learning repository.
INTRODUCTION

Imbalanced and overlapping classes are a real challenge for the standard classifiers. The problem is not only theoretical but it concerns many different types of real tasks. Examples are given by Kubat, Holte, and Matwin (1998); Fawcett and Provost (1997); Japkowicz (2003); Lewis and Catlett (1994); and Mladenic and Grobelnik (1999).

The problem of imbalanced classes occurs when the training set for a classifier contains far more examples from one class (majority illegal class) than the other (minority legal class).

To deal with the imbalance problems, up-sampling and down-sampling usually are used. Alas, both methods interfere in the structure of the data, and in a case of overlapping classes even the artificially obtained balance does not solve the problem (some data points may appear as valid examples in both classes). As the problem is still open, the new methods are investigated and trying to be improved (Chawla, Hall, & Kegelmeyer, 2002; Maloof, 2003; Visa & Ralescu, 2004; Zhang & Mani, 2003).

In this chapter we propose a new approach to the problem of classification of imbalanced and overlapping classes. The method proposed uses Atanassov’s intuitionistic fuzzy sets (A-IFSs for short) (Atanassov, 1983, 1986, 1999). We consider a two-class classification problem (legal and illegal class).

Atanassov’s theory of intuitionistic fuzzy sets (Atanassov, 1983, 1986, 1999) is one among many extensions of fuzzy sets (Zadeh, 1965) which has gained popularity in recent. Basically, it introduces, for each element of a universe of discourse, a degree of membership and a degree of non-membership, both from interval [0,1], but which do not sum up to 1 as in the conventional fuzzy sets. Such an extended definition can help more adequately represent situations when, for instance, decision makers abstain from expressing their testimonies, some assessments can not be classified but also can not be discarded, and so on.

Therefore, A-IFSs provide a richer apparatus to grasp the inherent imprecision of information than the conventional fuzzy sets by assigning to each element of the universe besides membership and non-membership functions also the corresponding lack of knowledge called hesitation margin, or intuitionistic fuzzy index (Atanassov, 1999).

The classification method which will be presented here (using A-IFSs) has its roots in the fuzzy set approach given in (Baldwin, Lawry, & Martin 1998). In that approach the classes are represented by fuzzy sets. The fuzzy sets are generated from the relative frequency distributions representing the data points used as examples of the classes (Baldwin et al., 1998). In the process of generating fuzzy sets a mass assignment based approach is adopted (Baldwin, Martin, & Pilsworth, 1995), (Baldwin et al., 1998). For the obtained model (fuzzy sets describing the classes), using a chosen classification rule, a testing phase is performed to assess the performance of the proposed method.

The approach proposed in this paper is similar to the above one in the sense of the same steps are performed. The main difference lies in using A-IFSs for the representation of classes, and next—in exploiting the structure of A-IFSs to obtain a classifier better recognizing the smaller classes.

The crucial point of the method is in representing the classes by A-IFSs (first, training phase). The A-IFSs are generated from the relative frequency distributions representing the data points used as examples of the classes. A-IFSs are obtained according to the automatic, and mathematically justified procedure given in (Szmidt, & Baldwin, 2005, 2006).

Having in mind recognition of the smaller class as good as possible we use the information about hesitation margins making it possible to improve the results of data classification in the (second) testing phase. The results obtained in the testing phase were examined not only in the sense of general error/accuracy but also with using confusion matrices making possible to explore a