



# Multi-label classification using a fuzzy rough neighborhood consensus

Sarah Vluymans<sup>a,b,c,\*</sup>, Chris Cornelis<sup>a</sup>, Francisco Herrera<sup>c,d</sup>, Yvan Saeys<sup>a,b</sup>

<sup>a</sup> Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Belgium

<sup>b</sup> Data Mining and Modelling for Biomedicine, VIB Center for Inflammation Research, Ghent, Belgium

<sup>c</sup> Department of Computer Science and Artificial Intelligence, University of Granada, Spain

<sup>d</sup> Faculty of Computing and Information Technology, King Abdulaziz University, Saudi Arabia

## ARTICLE INFO

### Article history:

Received 15 May 2017

Revised 16 November 2017

Accepted 23 December 2017

Available online 25 December 2017

### Keywords:

Fuzzy rough set theory

Multi-label classification

Nearest neighbor classification

## ABSTRACT

A multi-label dataset consists of observations associated with one or more outcomes. The traditional classification task generalizes to the prediction of several class labels simultaneously. In this paper, we propose a new nearest neighbor based multi-label method. The nearest neighbor approach remains an intuitive and effective way to solve classification problems and popular multi-label classifiers adhering to this paradigm include the MLKNN and IBLR methods. To classify an instance, our proposal derives a consensus among the labelsets of the nearest neighbors based on fuzzy rough set theory. This mathematical framework captures data uncertainty and offers a way to extract a labelset from the dataset that summarizes the information contained in the labelsets of the neighbors. In our experimental study, we compare the performance of our method with five other nearest neighbor based multi-label classifiers using five evaluation metrics commonly used in multi-label classification. Based on the results on both synthetic and real-world datasets, we are able to conclude that our method is a strong competitor to nearest neighbor based multi-label classifiers like MLKNN and IBLR.

© 2017 Elsevier Inc. All rights reserved.

## 1. Introduction

Research in machine learning concerns the ability to learn a confident prediction model based on a set of observations. For example, in a classification context, a learner trains a classification model on the given elements, of which the outcomes are known, and uses the derived model to predict the outcome of previously unseen instances. In the traditional single-label setting, each observation is associated with one outcome, its class label. Multi-label learning [16,18,60] represents a more general approach, where an observation can belong to several classes at the same time, that is, more than one class label can be associated with the same instance. The total number of classes is known, but the number of labels per instance can differ across the dataset. Multi-label learning can be more challenging than single-label learning or learning all possible classes independently, as correlations between some classes may be present. An example of such a situation is the existence of a label hierarchy, which needs to be taken into account in the prediction process [37]. Application domains of multi-label classification include image processing (e.g. [24,49]), text categorization (e.g. [29,31]) and bioinformatics (e.g. [5,44]).

\* Corresponding author.

E-mail address: [sarah.vluymans@ugent.be](mailto:sarah.vluymans@ugent.be) (S. Vluymans).

In a multi-label dataset, every instance  $x$  is described by a number of input features and associated with a labelset. This labelset is represented as a binary vector  $L_x = \langle l_1(x), l_2(x), \dots, l_m(x) \rangle$ , with  $m$  the total number of possible class labels in the dataset and  $(\forall i)(l_i \in \{0, 1\})$ . The value  $l_i(x)$  indicates whether or not  $x$  belongs to class  $l_i$ . The task of a multi-label classifier is to predict the complete labelset of a target instance. This is inherently different from single-label classification, where only one outcome label needs to be predicted.

Several approaches to multi-label classification have been proposed in the literature. The recent overview book [18] discerns between two main families, the data transformation methods and method adaptation algorithms. The former group of methods applies a transformation to the multi-label dataset, such that it degenerates to one or more easier-to-handle single-label problems, on which a single-label classifier can be applied. Two well-known representatives of this family are the binary relevance (BR, [17]) and label powerset (LP, [3]) transformations. BR creates  $m$  binary single-label datasets, one for each class. Each dataset contains the same instances as the original multi-label dataset, but their labelsets are transformed to a single label. For the dataset associated with class  $l_i$ , an instance  $x$  receives the label ‘positive’ when  $l_i(x) = 1$  and ‘negative’ otherwise. The LP transformation on the other hand creates only one single-label dataset. Each possible labelset receives an identifier, such that labelsets that entirely coincide are associated with the same identifier. This identifier is used as the single new class label. The second family of multi-label classification algorithms handle the multi-label dataset directly and are often based on modifications or generalizations of existing single-label classification schemes. An example is the MLKNN method proposed in [59].

In this paper, we focus on nearest neighbor methods for multi-instance classification, of which MLKNN is an example. The nearest neighbor approach [11] is an intuitive way to predict an outcome for a new observation based on a set of known instances. In its simplest form, it requires no training phase, since no classification model is built. Instead, all known instances are stored in memory as prototypes. In order to predict the outcome of a target instance, its nearest element (or set of nearest elements) is extracted from the prototype set and the prediction is derived from the outcomes of these neighbors. In particular, to classify an instance  $x$ , the  $k$  nearest neighbor classifier (kNN) locates the  $k$  nearest elements among the stored instances and aggregates their class labels to a prediction for  $x$ . In single-label classification, this is commonly achieved by a majority vote. The kNN classifier is a simple and understandable algorithm and remains popular in the machine learning community [46]. Several multi-label classifiers based on or extending kNN have been proposed in the literature. We propose a new member of this family in this contribution and use a novel way to aggregate the labelsets of the  $k$  nearest neighbors to a prediction based on fuzzy rough set theory.

Fuzzy rough set theory [12] is an alternative to traditional set theory and models uncertainty in data. It covers two complimentary aspects of uncertainty, namely vagueness (fuzziness) and indiscernibility (roughness). The former relates to unclear descriptions of concepts, to which elements can belong to a certain degree. As an example, the set of elements that are *similar* to a given element  $x$  is necessarily fuzzy, since some elements are intrinsically more similar to  $x$  than others and making a strict division between similar and non-similar is difficult. The membership degree of an element to a fuzzy set is represented by a real number between 0 and 1. Roughness in a dataset concerns the issue when observations that are indiscernible with respect to their descriptive features have distinct outcomes. In such a situation, it is challenging to sharply delineate the outcome concept based on the input features. Instead, a lower and an upper approximation are provided. Fuzzy rough set theory was developed as a hybridization of fuzzy set theory [58] and rough set theory [32] and has been used successfully in a variety of machine learning techniques [40]. It provides a framework to approximate a concept by two fuzzy sets, the fuzzy rough lower and upper approximation.

The fuzzy rough approximation operators are essentially based on a similarity relation that measures the degree to which elements are similar to each other. As such, they are related to nearest neighbor approaches. In this paper, we use these operators to derive a consensus prediction from the labelsets of the  $k$  nearest neighbors of a target instance. In particular, each of the neighbors of the target instance may have a different labelset and the challenge is to aggregate this information to one predicted labelset. Fuzzy rough set theory forms an ideal means to do so. Based on the similarity of the neighbors to the target, an appropriate consensus labelset is derived. We will experimentally show that our approach can outperform state-of-the-art nearest neighbor based multi-label classifiers. Following the recent study of Reyes et al. [34], we limit the comparison of our proposal to the state-of-the-art within the same classifier family, the nearest neighbor methods

The remainder of the paper is structured as follows. In Section 2, we review the existing nearest neighbor based multi-label classifiers. Section 3 recalls fuzzy rough set theory and describes our proposed classification method. Our method is carefully evaluated in an experimental study, of which the set-up is described in Section 4. Section 5 lists and analyzes the experimental results. Finally, Section 6 concludes the paper. We note that additional content, including the full experimental results, is made available at our web page <http://www.cwi.ugent.be/sarah.php>.

## 2. Related work: nearest neighbor based multi-label classifiers

To focus our discussion, we consider a subgroup of multi-label classifiers, namely those based on the nearest neighbor paradigm. Several nearest neighbor based multi-label classifiers have been proposed in the literature. We provide an overview in this section. In our experimental study, we select a number of these methods to compare among each other and against our proposed classifier. This selection is made based on the popularity and performance of these methods in other experimental studies. We also take into account how many parameters should be set by the user. The more parameters that need to be set manually, the less attractive a method is for practical use.

Download English Version:

<https://daneshyari.com/en/article/6856718>

Download Persian Version:

<https://daneshyari.com/article/6856718>

[Daneshyari.com](https://daneshyari.com)