

# Fuzzy Multi-Label Learning Under Veristic Variables

Zoulficar Younes, Fahed Abdallah and Thierry Denœux

**Abstract**—Multi-label learning is increasingly required by many applications where instances may belong to several classes at the same time. In this paper, we propose a fuzzy  $k$ -nearest neighbor method for multi-label classification using the veristic variable framework. Veristic variables are variables that can assume simultaneously multiple values with different degrees. In multi-label learning, class labels can be considered as veristic variables since each instance can belong simultaneously to more than one class. Several applications on benchmark datasets demonstrate the efficiency of our approach.

## I. INTRODUCTION

In multi-label classification problems, each object may belong simultaneously to several classes, contrary to standard single-label problems where objects belong to only one class. Multi-label classification methods have been increasingly required by many applications where the target classes are not exclusive and an object may belong to an unrestricted set of classes instead of exactly one. For instance, in text categorization [7], each document may belong to multiple topics, such as health, sports and nutrition.

In [10], an approximate reasoning framework has been proposed for the representation and manipulation of knowledge concerning veristic variables. Veristic variables are variables that can take more than one value at the same time, such as languages spoken by a person. Due to the fact that knowledge about a variable may be uncertain and imprecise, the developed theory is based on fuzzy sets [16] [17] rather than crisp sets in order to make it rich enough to handle all kinds of information.

In multi-label learning tasks, class label of each instance can be considered as veristic variables. In this paper, we propose a fuzzy multi-label  $k$ -NN method that uses the approximate reasoning framework for veristic variables [10] in order to represent and combine knowledge about an unseen object and predict the corresponding set of labels. Each neighbor gives some information about the object to classify according to the distance between the two patterns. In addition, due to the fact that veristic framework is based on the fuzzy set theory, and label sets of commonly multi-labelled data are crisp sets, we used a technique to fuzzify class membership of training data and transform the crisp label sets into fuzzy ones. Our method can also be applied to train multi-label classifiers from imprecise data. Such situation occurs when the learning dataset is labelled by one or several experts. Due to lack of confidence and absence of ground truth, an expert may be undecided and unable to assign unambiguously a crisp label set to each instance. This

situation can be handled by assigning different classes to each instance with various degrees of membership.

The remainder of this paper is structured as follows. Section II presents related work. Section III introduces the approximate reasoning framework for veristic variables. Section IV describes the proposed method for multi-label learning. Section V describes experiments on several datasets and discusses the results. Finally, Section VI concludes this paper.

## II. RELATED WORK

Several methods have been proposed in the literature for multi-label learning. These methods can be categorized into two groups. A first group contains the *indirect* methods that transform a multi-label classification problem into binary classification problems (a binary classifier for each class or pairwise classifiers) [14] [1] or into multi-class classification problem (each subset of classes is considered as a new class) [9]. A second group consists in extending common learning algorithms and making them able to manipulate multi-label data *directly* [15].

Hereafter, we briefly describe some multi-label classification methods. Using the  $k$ -NN rule, a multi-label classification method named ML- $k$ NN has been proposed in [18]. Under Bayesian reasoning, the label set of each unseen instance is determined using maximum a posteriori principle. In [19], an adaptation of the traditional radial basis function (RBF) neural network for multi-label learning has been presented. For ML-RBF, the first layer of the corresponding neural network is constructed by clustering instances of each possible class; the weights of the second layer are then optimized through minimizing an error function.

In [4], multi-label generalization of support vector machines (SVM) has been introduced. The author defined a cost function and a special multi-label margin and then proposed an algorithm named Rank-SVM based on a ranking system combined with a label set size predictor.

In [15], an evidence-theoretic  $k$ -NN rule for multi-label classification has been presented. This rule is based on an evidential formalism for representing uncertainties on the classification of multi-labelled data and handling imprecise labels, which has been detailed in [2]. The formalism extends all the notions of Dempster-Shafer theory [8] to the multi-label case with only a moderate increase in complexity as compared to the classical case. Under this formalism, each piece of evidence about an instance to classify is represented by a pair of sets: a set of classes that surely apply to the unseen instance, and a set of classes that surely do not apply to this instance.

Zoulficar Younes, Fahed Abdallah and Thierry Denœux are with the UMR CNRS Heudiasyc Laboratory, Université de Technologie de Compiègne, France, (fax: +33 3 44234477; emails: zoulficar.younes@hds.utc.fr, fahed.abdallah@hds.utc.fr, thierry.denœux@hds.utc.fr).

### III. VERISTIC VARIABLES

In [10], a theory has been proposed for the expression within the language of approximate reasoning of statements involving veristic variables. In [11] and [12], the author showed the role of veristic variables within databases. Methods have been considered for representing and evaluating queries involving veristic variables. In this paper, the veristic variable framework is used to build a multi-label classifier. Hereafter, a description of this framework is given.

Let  $Y$  be a veristic variable taking values in the universe of discourse  $\Omega$ , i.e.,  $Y$  has a single value in the set  $I^\Omega$  of fuzzy subsets of  $\Omega$ . Let  $Y_0 \in I^\Omega$  denote the unknown true value of  $Y$  and  $A \in I^\Omega$  a fuzzy subset of  $\Omega$ . Recall that each fuzzy set  $A$  defined over  $\Omega$  is characterized by a *membership function*  $\mu_A : \Omega \rightarrow [0, 1]$ , where for each  $\omega \in \Omega$ ,  $\mu_A(\omega)$  represents the degree of membership of  $\omega$  in  $A$ . Given  $A$ , the following statements have been presented in [10] to associate variable  $Y$  with  $A$ :

1.  $Y \text{ isv } A$ , meaning that  $A \subseteq Y_0$ ;
2.  $Y \text{ isv}(n) A$ , meaning that  $Y_0 \subseteq \bar{A}$ ;
3.  $Y \text{ isv}(c) A$ , meaning that  $Y_0 = A$ ;
4.  $Y \text{ isv}(c, n) A$ , meaning that  $Y_0 = \bar{A}$ .

In the above expressions, the relation *isv* has two parameters:  $c$  for closed (or exclusive) and  $n$  for negative. The following example gives an illustration of these notations.

**Example 1:** For a multi-label classification problem, assume that instances are songs and classes are emotions generated by these songs, as in the emotion dataset used in the experiments below. Upon hearing a song, more than one emotion can be generated at the same time. Let  $Y$  be a variable that corresponds to the emotions evoked by a given song. Let  $A$  be the set containing emotions *sad* and *quiet*.

- $Y \text{ isv } A$ , means that the song evokes sadness and quietness but it can also generates other emotions such as anger, calm, surprise, etc. This statement represents an open affirmative (or positive) information;
- $Y \text{ isv}(n) A$ , means that the song does not evoke sadness and quietness. We have no idea about the remaining emotions. This is an open negative information;
- $Y \text{ isv}(c) A$ , means that the song *only* evokes sadness and quietness, no more emotions are generated by this song. This is a closed positive information;
- $Y \text{ isv}(c, n) A$ , means that the song *only does not* evoke sadness and quietness. This is an exclusive negative information.

As shown in [10], each of these basic types of statements can be interpreted as specifying a set  $W$  of fuzzy subsets of  $\Omega$ , i.e., a crisp subset of  $I^\Omega$ .  $W$  contains the possible values of  $Y$ . It is defined as follows for the four types of statements:

$$\begin{aligned} Y \text{ isv } A &\rightarrow W = \{B \in I^\Omega | B \supseteq A\}; \\ Y \text{ isv}(n) A &\rightarrow W = \{B \in I^\Omega | B \subseteq \bar{A}\}; \\ Y \text{ isv}(c) A &\rightarrow W = \{A\}; \\ Y \text{ isv}(c, n) A &\rightarrow W = \{\bar{A}\}; \end{aligned}$$

where  $\bar{A}$  denotes the complement of  $A$ .

Two functions are associated to  $W$  from  $\Omega$  to  $[0, 1]$  which give information about each element  $\omega \in \Omega$ . These functions are called the *verity* and *rebuff* distributions and defined as follows:

$$\begin{aligned} \text{Ver}(\omega) &= \min_{B \in W} \mu_B(\omega), \\ \text{Rebuff}(\omega) &= 1 - \max_{B \in W} \mu_B(\omega) = \min_{B \in W} 1 - \mu_B(\omega) \\ &= \min_{B \in W} \mu_{\bar{B}}(\omega). \end{aligned}$$

$\text{Ver}(\omega)$  can be viewed as the minimal support for  $\omega$  being one of the values taken by  $Y$ . In contrast,  $\text{Rebuff}(\omega)$  can be interpreted as the minimal support for  $\omega$  *not* being one of the values taken by  $Y$ . These distributions have the following expressions for the different types of veristic statements:

$$x \text{ isv } A \Rightarrow \text{Ver}(\omega) = \mu_A(\omega), \text{Rebuff}(\omega) = 0;$$

$$x \text{ isv}(n) A \Rightarrow \text{Ver}(\omega) = 0, \text{Rebuff}(\omega) = \mu_A(\omega);$$

$$x \text{ isv}(c) A \Rightarrow \text{Ver}(\omega) = \mu_A(\omega), \text{Rebuff}(\omega) = 1 - \mu_A(\omega);$$

$$x \text{ isv}(c, n) A \Rightarrow \text{Ver}(\omega) = 1 - \mu_A(\omega), \text{Rebuff}(\omega) = \mu_A(\omega).$$

In [11], a possibility distribution denoted by  $\text{Poss}(\omega) = 1 - \text{Rebuff}(\omega)$  is also introduced. It represents *maximal* support for  $\omega$  being one of the values taken by  $Y$ .

In the following, we will focus on the first two veristic statements in order to develop our multi-label classifier. The first one is the *open affirmative* statement, like  $Y \text{ isv } A$ . The second one is the *open negative* statement, like  $Y \text{ isv}(n) A$ . The remaining ones are *strong and exclusive* statements.

Given different pieces of knowledge about a veristic variable  $Y$ , the *conjunctive* combination of the corresponding veristic statements is defined as follows:

$$Y \text{ isv } A \text{ and } Y \text{ isv } B \equiv Y \text{ isv } A \cup B,$$

$$Y \text{ isv}(n) A \text{ and } Y \text{ isv}(n) B \equiv Y \text{ isv}(n) A \cup B.$$

The *disjunctive* combination of veristic statements is defined by:

$$Y \text{ isv } A \text{ or } Y \text{ isv } B \equiv Y \text{ isv } A \cap B,$$

$$Y \text{ isv}(n) A \text{ or } Y \text{ isv}(n) B \equiv Y \text{ isv}(n) A \cap B.$$

We notice the *unexpected* association of union ( $\cup$ ) and intersection ( $\cap$ ) with the conjunctive and disjunctive combination, respectively.

### IV. FUZZY MULTI-LABEL $k$ -NN RULE

#### A. Multi-label learning

Let  $\mathcal{X} = R^P$  denote the domain of instances and let  $\Omega = \{\omega_1, \omega_2, \dots, \omega_Q\}$  be the finite set of labels. A Multi-label classification problem can be formulated as follows. Given a set  $S = \{(\mathbf{x}_1, A_1), \dots, (\mathbf{x}_M, A_M)\}$  of  $M$  training examples, where  $\mathbf{x}_i \in \mathcal{X}$ ,  $A_i$  denotes a crisp or fuzzy subset of  $\Omega$ , the goal of the learning system is to build a multi-label classifier  $\mathcal{H}$  that associates a label set to each unseen instance.

In [5], a  $k$ -NN rule for single-label classification based on the fuzzy set theory has been presented. The fuzzy  $k$ -NN was shown to dominate its crisp counterpart. In this paper, we propose a fuzzy  $k$ -NN rule for multi-label learning using the theory of veristic variables presented in Section III. The proposed algorithm is called FV- $k$ NN for Fuzzy Veristic  $k$ -Nearest Neighbor rule. The  $k$ -NN rule is widely used in pattern recognition problems due to its simplicity and its competitiveness with other sophisticated learning methods. Two major tasks are related to the  $k$ -NN classifier [20]. The first one concerns the influence of the nearest neighbors on the classification of an unseen object  $\mathbf{x}$ . Instead of giving equal importance to all neighbors, a *weight* can be assigned to each nearest neighbor according to the distance to  $\mathbf{x}$  [5]. The second task concern class memberships of training data. Usually, input data are *perfectly* labelled, i.e., the label sets of all instances are crisp sets. Such situations are not always possible, especially when the learning dataset is labelled by one or several experts in absence of ground truth. In this situation, the class label of each instance may be represented by a fuzzy set.

In the next section, we describe an algorithm that can be used in order to generate fuzzy label sets for instances that have been originally labelled by crisp ones.

### B. Data fuzzifying

We used the algorithm proposed in [5] in order to fuzzify the label memberships of instances. First, for each training object  $\mathbf{x}_i$ , we determine its  $k$  nearest neighbors in  $\mathcal{S}$  represented by  $\mathcal{N}_{\mathbf{x}_i}$ , based on a certain distance function  $d$ , usually the Euclidean one. A fuzzy label set  $A_i$  is associated to  $\mathbf{x}_i$  and replaces the crisp label set. The membership function of  $A_i$  is generated according to the crisp label sets of the  $k$  nearest neighbors. For each class  $\omega_q \in \Omega$ , the degree of membership is computed as follows:

$$\mu_{A_i}(\omega_q) = \begin{cases} 0.51 + 0.49(\frac{n_q}{k}) & \text{if } \omega_q \text{ applies to } \mathbf{x}_i, \\ 0.49(\frac{n_q}{k}) & \text{otherwise.} \end{cases}$$

$n_q$  is the number of instances in  $\mathcal{N}_{\mathbf{x}_i}$  which belong to class  $\omega_q$ . We notice that the degrees of membership of all originally assigned classes of an instance are greater than 0.5. Note that the number of neighbors taken into account for generating fuzzy label sets is not necessarily equal to the number of neighbors considered to classify an unseen object.

### C. Description of the learning algorithm

Let  $\mathbf{x}$  be an instance for which we seek to estimate the unknown set of labels  $Y$ . The proposed method consists of the following steps:

- 1) Determine the  $k$  nearest neighbors of  $\mathbf{x}$  in  $\mathcal{S}$ , represented by  $\mathcal{N}_{\mathbf{x}}$ , using the distance function  $d$ .
- 2) Associate a weight  $w_i$  to each neighbor according to the distance  $d_i$  between  $\mathbf{x}$  and  $\mathbf{x}_i$ . There exist different ways for transforming distances into weights. Here, we adopted the method proposed in [3], which will be

explained in the following. Two versions for the  $k$ -NN rule are proposed:

- unweighted version :  $w_i = 1$  for all  $\mathbf{x}_i \in \mathcal{N}_{\mathbf{x}}$ ;
- weighted version :  $w_i = \frac{d_{k+1} - d_i}{d_{k+1} - d_1}$  for each  $\mathbf{x}_i \in \mathcal{N}_{\mathbf{x}}$ , with  $d_1 \leq d_2 \leq \dots \leq d_{k+1}$ .  $d_1$  are the distances to the  $k + 1$  nearest neighbors of  $\mathbf{x}$  arranged in increasing order.

- 3) Update the membership functions of the fuzzy label sets of the instances in  $\mathcal{N}_{\mathbf{x}}$  according to their weights. For each element  $(\mathbf{x}_i, A'_i)$  with weight  $w_i$ , the corresponding updated fuzzy set  $A_i$  is defined as:

$$\mu_{A_i}(\omega_q) = \mu_{A'_i}(\omega_q) \cdot w_i, \quad \omega_q \in \Omega.$$

- 4) Represent each piece of knowledge  $(\mathbf{x}_i, A_i)$  about the instance to classify by a veristic statement. We adopted two versions:

- Open affirmative statement:  $Y \text{ isv } A_i$ ;
- Open negative statement:  $Y \text{ isv } \bar{A}_i$ .

If the instances of the available dataset are originally labelled by crisp sets that can be afterwards transformed into fuzzy ones, then we have the choice to represent each piece of knowledge about an unseen instance by one of the two veristic statement types. Suppose now that data labeling have been done by a pool of experts, thus the label sets are fuzzy ones. If the experts give confidence degrees about labels that *apply* to each instance, then we have to represent each piece of knowledge by affirmative veristic statement. In contrast, if the experts give confidence degrees about labels that *do not apply* to each instance, the pieces of knowledge must be represented in this case by negative statements.

- 5) Combine the veristic statements corresponding to all pieces of knowledge. For instance, the combinations of open affirmative statements are defined as:

- Conjunctive combination:

$$Y \text{ isv } A_1 \text{ and } \dots \text{ and } Y \text{ isv } A_k = Y \text{ isv } \left( \bigcup_i A_i \right),$$

- Disjunctive combination:

$$Y \text{ isv } A_1 \text{ or } \dots \text{ or } Y \text{ isv } A_k = Y \text{ isv } \left( \bigcap_i A_i \right).$$

- 6) Estimate the label set of  $\mathbf{x}$ , denoted by  $\hat{Y}$ . The decision rule can be chosen depending on the type of veristic statements.

- Let  $Y \text{ isv } A^*$  be the resulted statement obtained after combining (conjunctively or disjunctively) the  $k$  pieces of knowledge represented by open affirmative veristic statements.  $\hat{Y}$  is then computed as follows:

$$\hat{Y} = \{\omega_q \in \Omega | \text{Ver}(\omega_q) = \mu_{A^*}(\omega_q) > t\},$$

where  $t$  is a threshold to be fixed manually or by cross validation. In the experiments,  $t$  was fixed empirically to 0.5.

- On the other hand, let  $Y_{isv(n)} A^*$  denote the resulted veristic statement obtained after combining the open negative statements that correspond to the  $k$  pieces of knowledge for  $\mathbf{x}$ . In this case,  $\hat{Y}$  is computed as follows:

$$\hat{Y} = \{\omega_q \in \Omega | 1 - \text{Rebuff}(\omega_q) = 1 - \mu_{A^*}(\omega_q) > t\}.$$

## V. EXPERIMENTS

### A. Datasets

Three datasets were used for experiments: the emotion, scene and yeast datasets<sup>1</sup>.

a) *Emotion dataset*: This dataset contains 593 songs, each represented by a 72-dimensional feature vector (8 rhythmic features and 64 timbre features) [6]. The emotional labels are: *amazed-surprised*, *happy-pleased*, *relaxing-calm*, *quiet-still*, *sad-lonely* and *angry-fearful*.

b) *Scene dataset*: This dataset contains 2000 natural scene images. Each image is associated with some of the six different semantic scenes: *sea*, *sunset*, *trees*, *desert* and *mountains*. For each image, spatial color moments are used as features. Images are divided into 49 blocks using a  $7 \times 7$  grid. The mean and variance of each band are computed corresponding to a low-resolution image and to computationally inexpensive texture features, respectively [1]. Each image is then transformed into a  $49 \times 3 \times 2 = 294$ -dimensional feature vector.

c) *Yeast dataset*: The yeast dataset contains 2417 genes, each represented by a 103-dimensional feature vector [4]. Functional classes of many genes have been already determined and classified into a hierarchy of functions. Each gene may have several functions at the same time. Thus, the problem of Yeast classification is a multi-label problem with 14 labels.

Table I summarizes the characteristics of the datasets used in the experiments. The label cardinality of a dataset is the average number of labels of the instances, while the label density is the average number of labels of the instances divided by the total number of labels [9].

### B. Evaluation metrics

A result given by a multi-label classifier can be fully correct, partly correct, or fully wrong. We give hereafter some of the most widely used evaluation measures providing a degree of correctness of the label sets predicted by a multi-label learning system that have been proposed in the literature. Let  $\mathcal{D} = \{(\mathbf{x}_1, Y_1), \dots, (\mathbf{x}_N, Y_N)\}$  be a multi-label evaluation dataset containing  $N$  labelled examples. Let  $\hat{Y}_i = \mathcal{H}(\mathbf{x}_i)$  be the predicted label set for the pattern  $\mathbf{x}_i$ , while  $Y_i$  is the ground truth label set for  $\mathbf{x}_i$ .

A first metric called *Accuracy* gives an average degree of similarity between the predicted and the ground truth label sets of all test examples [9]:

$$\text{Accuracy}(\mathcal{H}, \mathcal{D}) = \frac{1}{N} \sum_{i=1}^N \frac{|Y_i \cap \hat{Y}_i|}{|Y_i \cup \hat{Y}_i|}.$$

<sup>1</sup><http://mlkd.csd.auth.gr/multilabel.html>

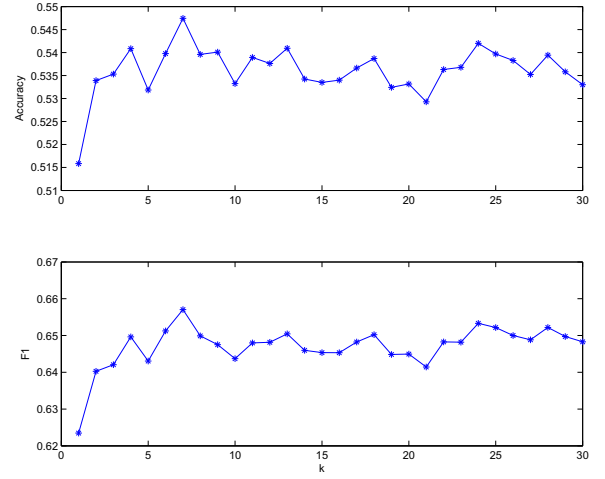


Fig. 1. Accuracy and F1 measures on the emotion dataset for FV- $k$ NN with conjunctively combined open affirmative veristic statements corresponding to weighted neighbors, for different values of the number of NN taken into account to generate the fuzzy label sets.

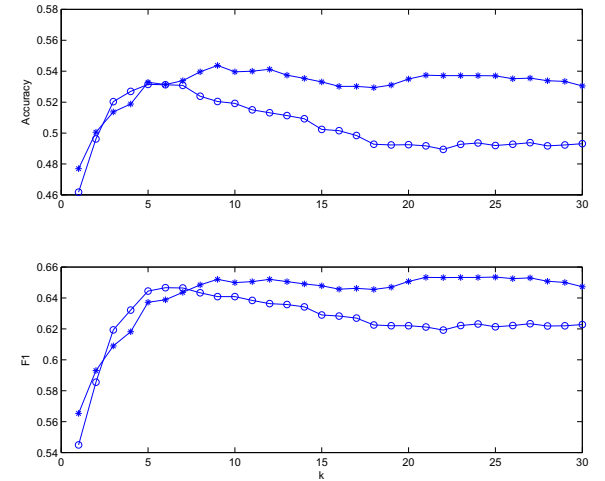


Fig. 2. Comparison between neighbor fuzzy (\*) and crisp (o) label sets using the Accuracy and F1 measures on the emotion dataset, for FV- $k$ NN with conjunctively combined open affirmative veristic statements.

A second evaluation criterion is the *F1* measure that is defined as the harmonic mean of the *Precision* and *Recall* metrics [13]. *Precision* metric computes the proportion of correct positive predictions, while *Recall* calculates the proportion of true labels that have been predicted as positives.

$$\text{Precision}(\mathcal{H}, \mathcal{D}) = \frac{1}{N} \sum_{i=1}^N \frac{|Y_i \cap \hat{Y}_i|}{|\hat{Y}_i|},$$

$$\text{Recall}(\mathcal{H}, \mathcal{D}) = \frac{1}{N} \sum_{i=1}^N \frac{|Y_i \cap \hat{Y}_i|}{|Y_i|},$$

TABLE I  
CHARACTERISTICS OF DATASETS

Dataset	Number of instances	Feature vector dimension	Number of labels	Training instances	Test instances	Label cardinality	Label density	maximum size of a label set
emotion	593	72	6	391	202	1.869	0.311	3
scene	2407	294	6	1211	1196	1.074	0.179	3
yeast	2417	103	14	1500	917	4.237	0.303	11

$$F1(\mathcal{H}, \mathcal{D}) = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}}.$$

The values of these evaluation criteria are in the interval  $[0, 1]$ . Larger values of these metrics correspond to higher classification quality.

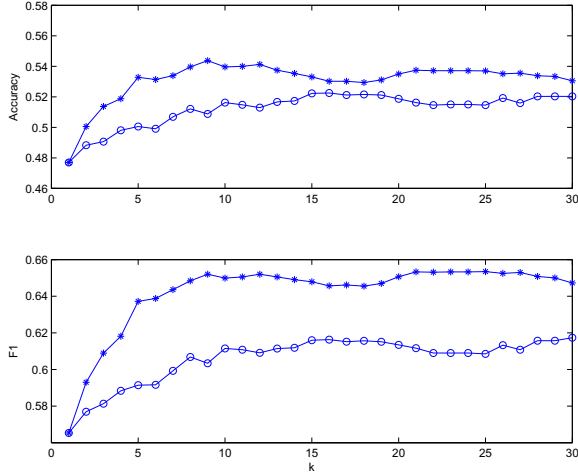


Fig. 3. Accuracy and F1 measures on the emotion dataset for FV- $k$ NN with conjunctive combination of open affirmative veristic statements (\*) and open negative ones (o), for different values of  $k$ .

### C. Results and discussions

1) *Configuration of our algorithm:* The proposed method was implemented in Matlab. We first carried out different experiments in order to configure our method, such fixing the type of veristic statement and the combination rule. Hereafter, train/test experiments were performed on the emotion dataset using different configurations of FV- $k$ NN in order to determine the best one. We focussed on the conjunctive combination (**and**) of veristic statements, which is more intuitive, especially when the different pieces of knowledge are considered as reliable. First, we tried to fix the number of nearest neighbors used to generate fuzzy label sets from the crisp ones as explained in Section IV-B. Figure 1 shows the values of the *Accuracy* and *F1* measures, for different values of the number of neighbors, obtained by using open affirmative veristic statements, and by fixing the number of neighbors considered for classifying an unseen instance to 10. In the following, the number of NN for data fuzzifying was fixed to 8.

In a second step, we tried to assess what we gain when using fuzzy label sets instead of crisp ones. Figure 2 shows

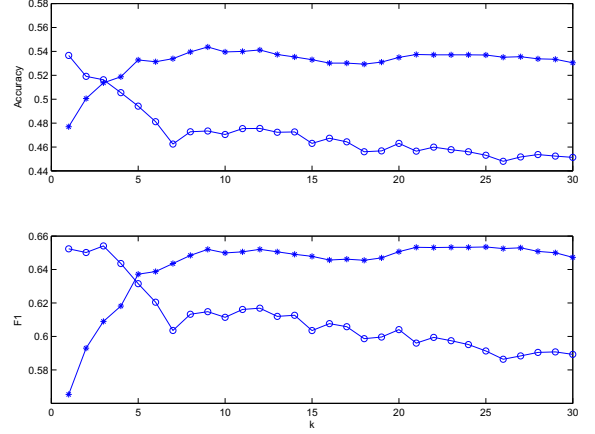


Fig. 4. Comparison between weighted (\*) and unweighted (o) nearest neighbors using Accuracy and F1 measures on the emotion dataset for FV- $k$ NN with conjunctively combined open affirmative veristic statements, for different values of  $k$ .

the results for different values of  $k$  for both types of label sets. We can deduce that it is more interesting to represent the class memberships of the training instances by fuzzy sets.

The question we have tried to answer here concerns the type of veristic statement to use for representing the information given by each neighbor on the classification of an unseen instance. Figure 3 shows that the representation of pieces of knowledge by open affirmative statements leads to better performance of FV- $k$ NN than the representation by negative affirmative ones.

Finally, we assigned equal weights to all nearest neighbors of an unseen instance  $\mathbf{x}$  instead of assigning a weight to each neighbor according to its distance to  $\mathbf{x}$ , as explained in Section IV-C. As shown in Figure 4, for values of  $k$  greater than 5, it is more beneficial to weight the neighbors for FV- $k$ NN.

2) *Comparison with other methods:* We compared our approach to three existing multi-label classification methods that were shown to have good performances: ML- $k$ NN [18], ML-RBF [19] and Rank-SVM [4]. For our method, fuzzy label sets were used to represent the class memberships of training instances. The pieces of knowledge about an object to classify that correspond to the different weighted nearest neighbors were represented by open affirmative veristic statements. These statements were afterwards conjunctively combined for making decision. The number of neighbors  $k$  was fixed to the moderate value 10. For ML- $k$ NN,  $k$  was

also fixed to 10 as in [18]. As shown in [19], the fraction parameter was set to 0.01 for ML-RBF, and the scaling factor to 1. For Rank-SVM, the best parameters reported in [4], i.e., polynomial kernels with degree 8, were used.

For each method, 10-fold cross validation was performed on each dataset. Figures 5, 6 and 7 show the box plots for the *Accuracy* and *F1* measures obtained by the four methods applied to the emotion, scene and yeast datasets, respectively. As we can see, the proposed method yields good performances on the three dataset according to the both evaluation criteria, and it is competitive with the three multi-label learning algorithms. The experimental results for FV- $k$ NN on the emotion dataset are close to those obtained by the ML- $k$ NN and ML-RBF methods. FV- $k$ NN performs better than the other algorithms on the scene dataset. Its performance is significantly better on the yeast dataset.

## VI. CONCLUSION

In this paper we have presented a fuzzy  $k$ -nearest neighbor rule for multi-label learning using the approximate reasoning framework of veristic variables. This framework allows us to represent pieces of knowledge about a veristic variable by different types of statements, and combine them conjunctively or disjunctively, in order to make decision about the values taken by this variable. For multi-label learning, class label of each instance can be considered as a veristic variable. Thus, we have used the veristic theory to build a multi-label classification method referred to as FV- $k$ NN. We have shown that our method is quite flexible as it can be configured in different ways, such as, weighting or not the neighbors of an instance to classify according to their distances to this one, fuzzifying or not the label sets of the neighbors, representing the information given by each piece of knowledge by open affirmative or negative veristic statement, etc. Experimental results with three datasets demonstrate the efficiency of our approach and prove its competitiveness with state-of-the-art methods.

## REFERENCES

- [1] M. R. Boutell, J. Luo, X. Shen and C. M. Brown, "Learning multi-label scene classification," *Pattern Recognition*, vol. 37(9), pp. 1757–1771, 2004.
- [2] T. Denœux, Z. Younes and F. Abdallah, "Representing uncertainty on set-valued variables using belief functions," *Artificial Intelligence*, vol. 174(7-8), pp. 479–499, 2010.
- [3] S. A. Dudani, "The distance-weighted  $k$ -nearest neighbor rule," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 6(4), pp. 325–327, 1976.
- [4] A. Elisseeff and J. Weston, "Kernel methods for multi-labelled classification and categorical regression problems," *Advances in Neural Information Processing Systems*, vol. 14, pp. 681–687, 2002.
- [5] J. M. Keller, M. R. Gray and J. A. Givens, "A fuzzy  $k$ -nearest neighbor algorithm," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 15(4), 1985.
- [6] T. Konstantinos, G. Tsoumakas, G. Kalliris and I. Vlahavas, "Multi-label classification of music into emotions," *Proc. 9th International Conference on Music Information Retrieval*, Philadelphia, PA, USA, September 2008.
- [7] R. E. Schapire and Y. Singer, "BoosTexter: A boosting-based system for text categorization," *Machine Learning*, vol. 39, pp. 135–168, 2000.
- [8] P. Smets, "The combination of evidence in the Transferable Belief Model," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 12(5), pp. 447–458, 1990.

- [9] G. Tsoumakas and I. Katakis, "Multi-Label Classification: An Overview," *International Journal of Data Warehousing and Mining*, vol. 3(3), pp. 1–13, 2007.
- [10] R. R. Yager, "Veristic variables," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 30(1), pp. 71–84, 2000.
- [11] R. R. Yager, "Querying databases containing multivalued attributes using veristic variables," *Fuzzy sets and systems*, vol. 129(2), pp. 163–185, 2002.
- [12] R. R. Yager, "Veristic Variables and Approximate Reasoning for Intelligent Semantic Web Systems," *Studies in Fuzziness and Soft Computing, Forging New Frontiers: Fuzzy Pioneers I*, Springer Berlin/Heidelberg, vol. 217, pp. 231–249, 2007.
- [13] Y. Yang, "An evaluation of statistical approaches to text categorization," *Journal of Information Retrieval*, vol. 1, pp. 78–88, 1999.
- [14] Z. Younes, F. Abdallah and T. Denœux, "Multi-label classification algorithm derived from  $k$ -nearest neighbor rule with label dependencies," *Proc. 16th European Signal Processing Conference*, Lausanne, Switzerland, August 2008, pp. 25–29.
- [15] Z. Younes, F. Abdallah and T. Denœux, "An Evidence-Theoretic  $k$ -Nearest Neighbor Rule for Multi-Label Classification," *Proc. 3rd International Conference on Scalable Uncertainty Management*, Washington, DC, USA, September 2009, LNAI-5785, Springer-Verlag, pp. 297–308.
- [16] L. A. Zadeh, "Fuzzy sets," *Information and control*, vol. 8, pp. 338–353, 1965.
- [17] L. A. Zadeh, "Fuzzy sets as a basis for a theory of possibility," *Fuzzy sets and systems*, vol. 1, pp. 3–28, 1978.
- [18] M. L. Zhang and Z. H. Zhou, "ML-KNN: A lazy learning approach to multi-label learning," *Pattern Recognition*, vol. 40(7), pp. 2038–2048, 2007.
- [19] M. L. Zhang, "ML-RBF: RBF neural networks for multi-label learning," *Neural Processing Letters*, vol. 29(2), pp. 61–74, 2009.
- [20] W. Zuo, D. Zhang and K. Wang, "On kernel difference-weighted  $k$ -nearest neighbor classification," *Pattern Analysis & Applications*, vol. 11(3-4), pp. 247–257, 2008.

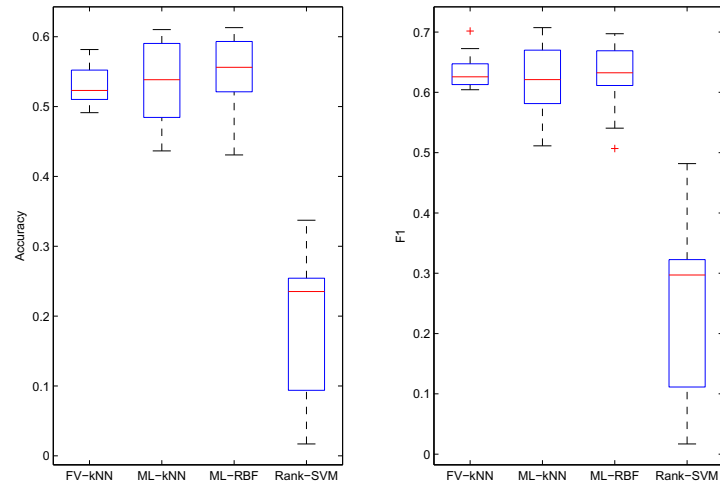


Fig. 5. Accuracy and  $F1$  box plots for the emotion dataset with the following methods: FV- $k$ NN, ML- $k$ NN, ML-RBF and Rank-SVM.

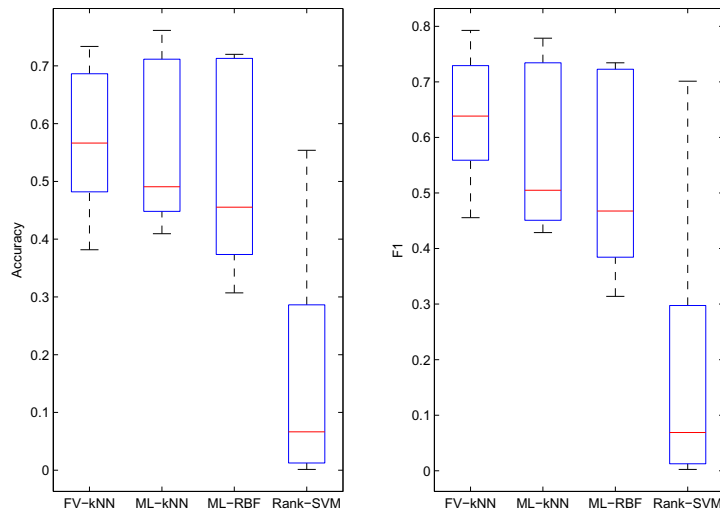


Fig. 6. Accuracy and  $F1$  box plots for the scene dataset with the following methods: FV- $k$ NN, ML- $k$ NN, ML-RBF and Rank-SVM.

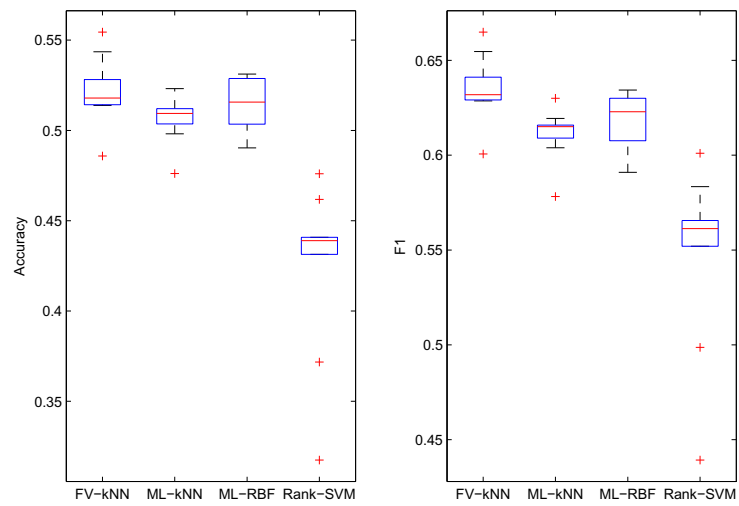


Fig. 7. *Accuracy* and *F1* box plots for the yeast dataset with the following methods: FV-*k*NN, ML-*k*NN, ML-RBF and Rank-SVM.