ORIGINAL PAPER

Fuzzy rule-based systems for recognition-intensive classification in granular computing context

Han Liu[1] · **Li Zhang[2]**

Abstract

In traditional machine learning, classification is typically undertaken in the way of discriminative learning using probabilistic approaches, i.e. learning a classifier that discriminates one class from other classes. The above learning strategy is mainly due to the assumption that different classes are mutually exclusive and each instance is clear-cut. However, the above assumption does not always hold in the context of real-life data classification, especially when the nature of a classification task is to recognize patterns of specific classes. For example, in the context of emotion detection, multiple emotions may be identified from the same person at the same time, which indicates in general that different emotions may involve specific relationships rather than mutual exclusion. In this paper, we focus on classification problems that involve pattern recognition. In particular, we position the study in the context of granular computing, and propose the use of fuzzy rule-based systems for recognition-intensive classification of real-life data instances. Furthermore, we report an experimental study conducted using 7 UCI data sets on life sciences, to compare the fuzzy approach with four popular probabilistic approaches in pattern recognition tasks. The experimental results show that the fuzzy approach can not only be used as an alternative one to the probabilistic approaches but also is capable to capture more patterns which probabilistic approaches cannot achieve.

Keywords Machine learning · Multi-task learning · Real-life data classification · Fuzzy classification · Fuzzy rule-based systems · Granular computing

1 Introduction

Classification is one of the most popular tasks of machine learning, which has been popularly involved in various application areas, such as sentiment analysis (Liu and Cocea 2017b; Pedrycz and Chen 2016; Jefferson et al. 2017), image processing (Liu et al. 2017a; Wang and Yu 2016), pattern recognition (Teng et al. 2007; Wu et al. 2011) and decision making (Liu and Gegov 2015; Xu and Wang 2016; Liu and You 2017).

In traditional machine learning, classification is typically conducted by training a classifier that discriminates one class from other classes towards uniquely classifying each instance, since the classification is based on the assumptions that different classes are mutually exclusive and that each instance is clear-cut and thus cannot belong to more than one class. However, the above assumptions do not always hold in real-life data classification. For example, it is very normal that the same movie can belong to different categories or the same book can be associated with different subjects. Also, while different classes are truly mutually exclusive, it is also possible that some instances are very complex and hard to distinguish, e.g. in handwritten digits recognition, the two digits '4' and '9' can be highly similar to each other, due to the diversity in handwriting styles from different people.

Furthermore, as introduced in Liu et al. (2017), classification is essentially a task of predicting the value of a discrete attribute. In the context of data science, discrete attributes can be specialized into several other types, such as nominal, ordinal and string (Tan et al. 2005). Due to the difference in types of discrete attributes, the nature of classification tasks

✉ Han Liu
  LiuH48@cardiff.ac.uk

  Li Zhang
  li.zhang@northumbria.ac.uk

[1] School of Computer Science and Informatics, Cardiff University, Queen's Buildings, 5 The Parade, Cardiff CF24 3AA, UK

[2] Department of Computer and Information Sciences, Faculty of Engineering and Environment, Northumbria University, Newcastle NE1 8ST, UK

can also be varied. In particular, classification tasks can be specialized into pattern recognition, rating and decision making (Liu et al. 2017), which indicates that a classification task is not necessarily aimed at discrimination between different classes, i.e. the purpose could be simply to identify patterns of a specific class from instances, without the need to distinguish the class from other classes.

In this paper, we focus on recognition-intensive classification in the setting of granular computing. In particular, we propose to adopt fuzzy rule-based systems in the context of multi-task classification, i.e. each class is viewed as an information granule, which involves a specific recognition task, in terms of the membership degree of an instance to the class. Also, the recognition task for each class is undertaken independently, i.e. the membership degree of an instance to each class is measured independently, in the context of generative classification.

The contributions of this paper include the following: (a) we point out the case that in recognition-intensive classification different classes generally involve some specific relationships rather than mutual exclusion, so it is not appropriate to undertake such a classification task in a discriminative way; (b) we show both theoretically and empirically that fuzzy approaches are more suitable than probabilistic ones for recognition-intensive classification, i.e. fuzzy approaches can not only be used as the alternative ones to probabilistic approaches in terms of classification performance, but also show the capability of capturing more patterns that cannot be discovered using probabilistic approaches.

The rest of this paper is organized as follows: Sect. 2 provides related work on recognition-intensive classification in the context of traditional machine learning and the concepts of granular computing. In Sect. 3, we illustrate the procedure of fuzzy rule-based classification in the context of multi-task learning. We also justify the significance and advantages of fuzzy classification of real-life data that involve recognition tasks. In Sect. 4, we report an experimental study conducted using 7 UCI data sets, and discuss the results critically and comparatively to show the advantages of fuzzy approaches for recognition-intensive classification, in comparison with probabilistic approaches. In Sect. 5, we summarize the contributions of this paper and suggest further directions towards advancing this research area in the future.

2 Related work

This section provides a review of recognition-intensive classification when traditional machine learning approaches are used. This section also presents an overview of granular computing concepts and techniques and shows how they can be used effectively and efficiently for dealing with real-life classification problems.

2.1 Review of recognition-intensive classification

As introduced in Liu et al. (2017), recognition can be either a binary or multi-class classification task. A popular example of binary classification for the purpose of recognition is gender identification (Guo 2014), which is aimed at judging that a person is male or female. In this context, both the male and female classes are of high interest, since it is required to distinguish clearly the two classes towards identifying accurately the gender of a person, which has motivated researchers to focus the research on discriminative approaches of classification according to Wu et al. (2011), Ali and Xavier (2014), Lin et al. (2016), Suykens and Vandewalle (1999). In other words, researchers aim to identify features that can discriminate effectively between male and female in the setting of discriminative learning.

However, there are also some examples of binary classification that only involve one class of interest, such as cyberbullying detection (Zhao et al. 2016; Reynolds et al. 2011). In the context of cyberbullying classification, the aim is essentially at recognizing effectively any such offensive languages from online text posted via social media, i.e. it is to judge if the text is sent for the purpose of bullying. In reality, vast majority of textual instances posted via social media would normally belong to the 'no' class (i.e. a collected data set usually contains less than 10% cyberbullying instances), as mentioned in Reynolds et al. (2011), which indicates the case of class imbalance. Since discriminative classification has been mainly involved in traditional machine learning, some popular probabilistic approaches, such as support vector machine (SVM), naive Bayes (NB) and decision trees (DT), have been popularly used for cyberbullying detection (Zhao et al. 2016).

From a perspective of granular computing-based machine learning, the 'yes' class, which represents the case of cyberbullying, can be viewed as the only target class, since it is the only class of interest and the prediction accuracy for the 'no' class would be usually very high. From this point of view, it is only needed to extract a set of features that are highly relevant to the target class, such that a classifier is learned for recognizing the case of cyberbullying. In other words, the classifier output is the 'no' class by default, and the 'yes' class is provided as the output only when some features of cyberbullying are found from text.

On the other hand, there are also many examples of multi-class classification for the purpose of recognition, such as emotion identification (Teng et al. 2007). Due to the popularity of probabilistic approaches in traditional machine learning, SVM and NB have been used for discriminating one emotion from the other ones (Altrabsheh

et al. 2015). However, as argued in Liu et al. (2017), different emotions are not really mutually exclusive, i.e. it is normal that different emotions can be identified from the same person at the same time, so it is not really appropriate to learn classifiers towards discriminating between different classes. Instead, it would be necessary to treat identification of each emotion as an independent task. In this context, it is necessary to extract only features that are highly relevant to this specific emotion, such that a classifier is learned from these features for identifying whether a person has this specific emotion at a particular point.

2.2 Overview of granular computing

Granular computing is a paradigm of information processing. It is aimed at structural thinking from a philosophical perspective and is aimed at structural problem solving from a practical perspective (Yao 2005b).

In general, granular computing involves two operations, namely granulation and organization (Yao 2005a). The former operation is aimed at decomposing a whole into several (overlapping or non-overlapping) parts, whereas the latter operation is aimed at integrating several (overlapping or non-overlapping) parts into a whole. In the context of computer science, granulation and organization are typically used as the top-down and bottom-up approaches, respectively (Liu and Cocea 2017a). In other words, granulation means to divide a complex problem into several simpler sub-problems, whereas organization indicates that several modular problems are linked together into a more systematic problem.

In practice, two main concepts of granular computing have been popularly involved in the two operations of granular computing (granulation and organization), namely granule and granularity. A granule generally represents a large particle, which can be divided into several smaller particles that can form a larger unit. There are many real-life examples as follows:

– In the context of classification, each class can be viewed as a granule, since a class is essentially a collection of objects/instances.
– In the context of rule-based systems, each rule can be viewed as a granule, since a rule consists of a collection of rule terms as its antecedent.
– In the context of fuzzification of continuous attributes, each linguistic term can be viewed as a granule, since a linguistic term is essentially a fuzzy set that represents a collection of elements with different membership degrees to the fuzzy set.

In general, there are some specific relationships between granules in the same level or different levels, which leads to the need to involve the concept of granularity (Pedrycz and Chen 2015). In particular, granules, which are located at the same level of granularity, involve horizontal relationships (Liu and Cocea 2018), e.g. mutual exclusion, correlation and mutual independence.

In contrast, granules, which are located at different levels of granularity, involve hierarchical relationships (Liu and Cocea 2018; Liu et al. 2017), e.g. generalization/specialization and aggregation/decomposition. For example, in the context of classification, a class at a higher level of granularity may be specialized/decomposed into sub-classes at a lower level of granularity. Also, classes at a lower level of granularity may be generalized/aggregated into a super class at a higher level of granularity (Liu and Cocea 2017a). On the other hand, different classes may also be mutually exclusive, correlated or mutually independent, when these classes are at the same level of granularity (Liu et al. 2017b).

In practice, granular computing concepts have been popularly used in various areas, such as artificial intelligence (Wilke and Portmann 2016; Pedrycz and Chen 2011; Skowron et al. 2016), computational intelligence (Dubois and Prade 2016; Yao 2005b; Kreinovich 2016; Livi and Sadeghian 2016), machine learning (Min and Xu 2016; Peters and Weber 2016; Liu and Cocea 2017c; Antonelli et al. 2016; Chen et al. 2001), decision making (Xu and Wang 2016; Liu and You 2017; Chatterjee and Kar 2017), data clustering (Chen et al. 2009; Horng et al. 2005; Chen et al. 2011) and natural language processing (Zhang et al. 2007).

Furthermore, granular computing concepts have also been popularly used in ensemble learning techniques (Liu and Cocea 2017c). In particular, ensemble learning approaches, such as Bagging, involve granulation of information through decomposing a training set into a number of overlapping samples (different versions of training data), and also involve organization through combining the individual outputs derived from different base classifiers towards finally assigning a class to an unseen instance; there has also been a very similar perspective stressed and discussed in Hu and Shi (2009).

3 Fuzzy multi-task classification approach

In this section, we illustrate the procedure of fuzzy rule-based systems in the context of generative multi-task classification. Also, we justify the theoretical significance and advantages of using fuzzy approaches for recognition-intensive classification.

3.1 Procedure of fuzzy rule-based systems

A fuzzy rule-based system is essentially based on fuzzy logic and fuzzy set theory. Fuzzy logic is an extension of deterministic logic, i.e. fuzzy truth values are continuous, which are ranged from 0 to 1, unlike binary truth values (0 or 1).

In the context of fuzzy sets, each element x_i has a certain degree of membership to the set S, i.e. it partially belongs to the set. The value of the membership degree is determined by the membership function $f_s(x_i)$ defined for the fuzzy set S. There are various shapes of membership functions used in practice, such as trapezoid, triangle and rectangle. In general, trapezoidal membership functions can be seen as a generalization of triangular and rectangular membership functions. The definition of a membership function is essentially achieved by estimating four parameters a, b, c, d, as illustrated below and in Fig. 1.

$$f_T(x) = \begin{cases} 0, & \text{when } x \leq a \text{ or } x \geq d; \\ (x-a)/(b-a), & \text{when } a < x < b; \\ 1, & \text{when } b \leq x \leq c; \\ (d-x)/(d-c), & \text{when } c < x < d; \end{cases}$$

According to Fig. 1, the shape of the membership function would be triangle, if $b = c$, or the shape would be rectangle, if $a = b$ and $c = d$. The rectangle area between b and c is referred to as 'core area', which represents the case that elements fully belong to the set, whereas the whole trapezoid area, between a and d, is referred to as 'support area', which represents the case that elements may partially belong to the set.

A membership function can be defined using expert knowledge (Mamdani and Assilian 1999) or by learning statistically from data (Bergadano and Cutello 1993). More details on fuzzy sets and logic can be found in Zadeh (1965, Chen and Chang (2001), Chen and Chen (2011), Chen (1996).
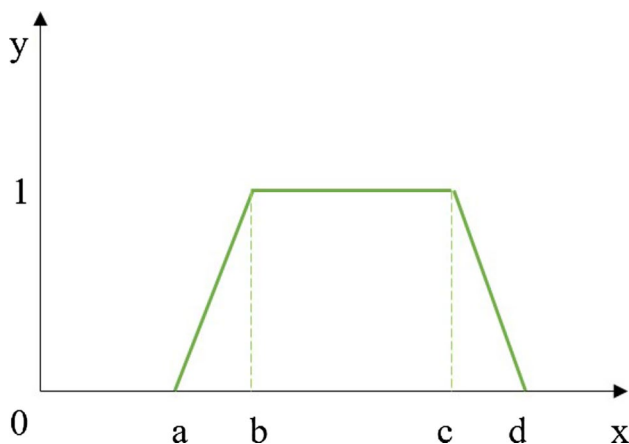

Fig. 1 Trapezoid membership function (Liu and Cocea 2017b)

Algorithm 1: Fuzzy Multi-task Classification

Input : An attribute vector I: numerical value v_{ij} of each attribute x_i, where i is the index of the attribute and j is the id of the attribute vector, A set of fuzzy rules: each rule is represented as R_m and is assigned a class c_t as its consequent, A set of fuzzy intervals for each attribute x_i: each interval is represented as T_{ik}, where k is the index of the interval

1 Fuzzification: **for** *each fuzzy rule R_m* **do**
2   **for** *each attribute x_i* **do**
3   covert the attribute value v_{ij} into the membership degree value f_{ijk}
4   **end**
5 **end**
6 Application: **for** *each fuzzy rule R_m* **do**
7   **for** *each attribute x_i* **do**
8   conjunct the membership degree value f_{ijk}infer the firing strength of the rule using the 'min' function
9   **end**
10 **end**
11 Implication: **for** *each fuzzy rule R_m* **do**
12   infer the membership degree $f_{R_m \to c_t}(I)$ of the attribute vector I to the class c_t based on this rule
13 **end**
14 Aggregation: **for** *each class c_t* **do**
15   disjunct the membership degree values derived from all fuzzy rules of the class c_t to infer the overall membership degree $f_{c_t}(I)$ of the attribute vector I to the class c_t by using the 'max' function
16 **end**

In the context of multi-task classification, a fuzzy rule-based system involves four main operations (Liu et al. 2017): fuzzification, application, implication, and aggregation, as shown in Algorithm 1. We illustrate the whole procedure using the following example of fuzzy rules:

- Rule 1: if x_1 is Young and x_2 is High then class = Impressive;
- Rule 2: if x_1 is Young and x_2 is Middle then class = Impressive;
- Rule 3: if x_1 is Young and x_2 is Low then class = Normal;
- Rule 4: if x_1 is Middle-aged and x_2 is High then class = Impressive;
- Rule 5: if x_1 is Middle-aged and x_2 is Middle then class = Normal;
- Rule 6: if x_1 is Middle-aged and x_2 is Low then class = Odd;
- Rule 7: if x_1 is Old and x_2 is High then class = Normal;
- Rule 8: if x_1 is Old and x_2 is Middle then class = Odd;
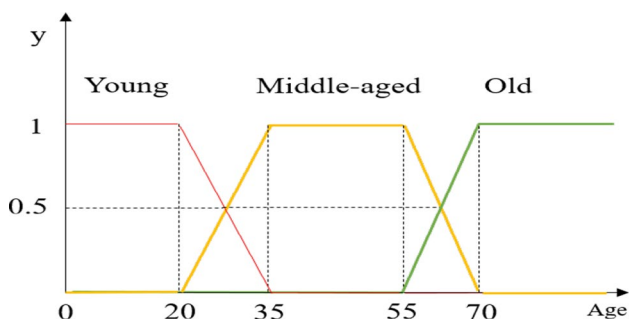- Rule 9: if x_1 is Old and x_2 is Low then class = Odd;

Fig. 2 Fuzzy membership functions for linguistic terms of attribute 'age' (Liu et al. 2017)
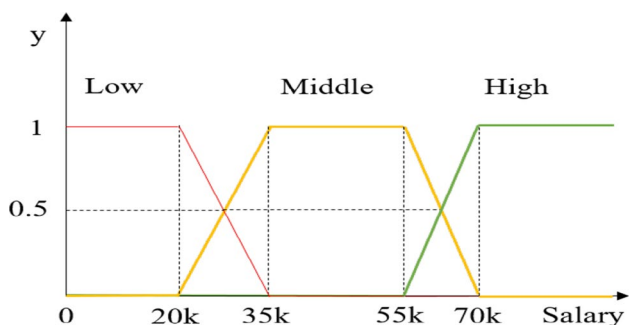

Fig. 3 Fuzzy membership functions for linguistic terms of attribute 'salary'

The fuzzy membership functions defined for the linguistic terms transformed from x_1 and x_2 are illustrated in Fig. 2 and Fig. 3, respectively.

According to Figs. 2 and 3, if $x_1 = 30$ and $x_2 = 60k$, then the following steps will be executed:

Fuzzification:

Rule 1: $f_{\text{Young}}(30) = 0.67$, $f_{\text{High}}(60k) = 0.33$;
Rule 2: $f_{\text{Young}}(30) = 0.67$, $f_{\text{Middle}}(60k) = 0.67$;
Rule 3: $f_{\text{Young}}(30) = 0.67$, $f_{\text{Low}}(60k) = 0$;
Rule 4: $f_{\text{Middle-aged}}(30) = 0.33$, $f_{\text{High}}(60k) = 0.33$;
Rule 5: $f_{\text{Middle-aged}}(30) = 0.33$, $f_{\text{Middle}}(60k) = 0.67$;
Rule 6: $f_{\text{Middle-aged}}(30) = 0.33$, $f_{\text{Low}}(60k) = 0$;
Rule 7: $f_{\text{Old}}(30) = 0$, $f_{\text{High}}(60k) = 0.33$;
Rule 8: $f_{\text{Old}}(30) = 0$, $f_{\text{Middle}}(60k) = 0.67$;
Rule 9: $f_{\text{Old}}(30) = 0$, $f_{\text{Low}}(60k) = 0$;

In the fuzzification step, the notation $f_{\text{High}}(60k) = 0.33$ represents that the membership degree of the numerical value '60k' to the fuzzy set defined with the linguistic term 'High' is 0.33. The fuzzification step is aimed at mapping the value of a continuous attribute to a value of membership degree to a fuzzy set (i.e. mapping to the value of a linguistic term transformed from the continuous attribute).

Application:

Rule 1: $f_{\text{Young}}(30) \wedge f_{\text{High}}(60k) = Min(0.67, 0.33) = 0.33$;
Rule 2: $f_{\text{Young}}(30) \wedge f_{\text{Middle}}(60k) = Min(0.67, 0.67) = 0.67$;
Rule 3: $f_{\text{Young}}(30) \wedge f_{\text{Low}}(60k) = Min(0.67, 0) = 0$;
Rule 4: $f_{\text{Middle-aged}}(30) \wedge f_{\text{High}}(60k) = Min(0.33, 0.33) = 0.33$;
Rule 5: $f_{\text{Middle-aged}}(30) \wedge f_{\text{Middle}}(60k) = Min(0.33, 0.67) = 0.33$;
Rule 6: $f_{\text{Middle-aged}}(30) \wedge f_{\text{Low}}(60k) = Min(0.33, 0) = 0$;
Rule 7: $f_{\text{Old}}(30) \wedge f_{\text{High}}(60k) = Min(0, 0.33) = 0$;
Rule 8: $f_{\text{Old}}(30) \wedge f_{\text{Middle}}(60k) = Min(0, 0.67) = 0$;
Rule 9: $f_{\text{Old}}(30) \wedge f_{\text{Low}}(60k) = Min(0, 0) = 0$;

In the application step, the conjunction of the two fuzzy membership degree values, respectively, for the two attributes 'x_1' and 'x_2' is aimed at deriving the firing strength of a fuzzy rule. For example, the antecedent of Rule 1 consists of x_1 is Young and x_2 is High, so the firing strength of Rule 1 is 0.33, while $f_{\text{Young}}(30) = 0.67$ and $f_{\text{High}}(60k) = 0.33$.

Implication:

Rule 1: $f_{\text{Rule1}\rightarrow\text{Impressive}}(30, 60k) = 0.33$;
Rule 2: $f_{\text{Rule2}\rightarrow\text{Impressive}}(30, 60k) = 0.67$;
Rule 3: $f_{\text{Rule3}\rightarrow\text{Normal}}(30, 60k) = 0$;
Rule 4: $f_{\text{Rule4}\rightarrow\text{Impressive}}(30, 60k) = 0.33$;
Rule 5: $f_{\text{Rule5}\rightarrow\text{Normal}}(30, 60k) = 0.33$;
Rule 6: $f_{\text{Rule6}\rightarrow\text{Odd}}(30, 60k) = 0$;
Rule 7: $f_{\text{Rule7}\rightarrow\text{Normal}}(30, 60k) = 0$;
Rule 8: $f_{\text{Rule8}\rightarrow\text{Odd}}(30, 60k) = 0$;
Rule 9: $f_{\text{Rule9}\rightarrow\text{Odd}}(30, 60k) = 0$;

In the implication step, the firing strength of a fuzzy rule derived in the application step can be used further to infer the value of membership degree of an input vector to one of the class labels 'Impressive', 'Normal' and 'Odd', depending on the actual consequent of the fuzzy rule. For example, $f_{\text{Rule1}\rightarrow\text{Impressive}}(30, 60k) = 0.33$ indicates that the consequent of Rule 1 is assigned the class label 'Impressive' and the input vector '(30, 60k)' has the membership degree value of 0.33 to the class label 'Impressive'. In other words, the input vector '(30, 60k)' gains the membership degree value of 0.33 to the class label 'Impressive', through the inference using Rule 1.

Aggregation:

$$f_{\text{Impressive}}(30, 60k) = f_{\text{Rule1}\rightarrow\text{Impressive}}(30, 60k)$$
$$\vee f_{\text{Rule2}\rightarrow\text{Impressive}}(30, 60k) \vee f_{\text{Rule4}\rightarrow\text{Impressive}}(30, 60k)$$
$$= Max(0.33, 0.67, 0.33) = 0.67$$

$$f_{\text{Normal}}(30, 60k) = f_{\text{Rule3}\rightarrow\text{Normal}}(30, 60k)$$
$$\vee f_{\text{Rule5}\rightarrow\text{Normal}}(30, 60k) \vee f_{\text{Rule7}\rightarrow\text{Normal}}(30, 60k)$$
$$= Max(0, 0.33, 0) = 0.33$$

$$f_{\text{Odd}}(30, 60k) = f_{\text{Rule6}\rightarrow\text{Odd}}(30, 60k)$$
$$\vee f_{\text{Rule8}\rightarrow\text{Odd}}(30, 60k) \vee f_{\text{Rule9}\rightarrow\text{Odd}}(30, 60k)$$
$$= Max(0, 0, 0) = 0$$

In the aggregation step, the membership degree value of the input vector to the class label ('Impressive', 'Normal' or 'Odd'), which is inferred through using a rule, is compared with the other membership degree values inferred through using the other rules, towards finding the maximum among all the membership degree values. For example, Rule 1, Rule 2 and Rule 4 are all assigned the class label 'Impressive' as their consequent and the input vector '(30, 60k)' gains the membership degree values of are 0.33, 0.67 and 0.33, respectively, to the class label 'Impressive', through the inference using the above three rules. As the maximum of the fuzzy membership degree values is 0.67, the input vector is judged finally to have the membership degree value of 0.67 to the class label 'Impressive'.

In traditional machine learning, it is usually needed to provide an crisp output as the classification outcome, so defuzzification is typically involved by choosing the class label with the maximum value of membership degree. When there is more than one class label with the maximum value of membership degree, defuzzification is achieved by randomly choosing one of these classes with the maximum membership degree. For the above illustrative example, the final classification outcome is to assign the class label 'Impressive' to the unseen instance '(30, 60k, ?)', since the value (0.67) of the membership degree to this class label is the maximum one. In contrast, generative multi-task classification is aimed at measuring independently the membership degree value of an instance to each class, so it is not necessary to involve the defuzzification step.

Besides, as mentioned above, definition of membership functions can be based on expert knowledge or real data. In the context of data-driven definition of membership functions, it is generally not applicable to assign each fuzzy set a linguistic term. Instead, each fuzzy set is provided with an ID, e.g. ID '0' represents the first fuzzy set. The representation of each fuzzy set is achieved through providing the actual parameters of the membership function defined for the fuzzy set, e.g. [20, 30, 55, 75] represents the four parameters (a, b, c, d) of a trapezoid membership function.

3.2 Justification

In the context of recognition-intensive classification, the purpose is essentially to discover the presence of a target class of instances. From this point of view, fuzzy multi-task classification is considered as a very suitable approach. For example, in the context of human activities recognition, there are several activities that need to be identified, and each of the activities is viewed as a target class, such that a set of fuzzy rules are learned for each target class and are used to identify the degree to which the activity (corresponding to the target class) is present, in the setting of fuzzy multi-task classification.

On the other hand, recognition-intensive classification can usually involve a large number of classes, e.g. human activities recognition can involve more than ten classes as indicated in Kalua et al. (2010). It is very likely that these classes are not mutually exclusive. For example, in activities recognition, the three classes 'sitting', 'sitting down' and 'sitting on the ground' are generally correlated to some extents. Also, the three classes 'standing up from lying', 'standing up from sitting' and 'standing up from sitting on the ground' would have some overlaps in terms of their features. From the above point of view, human activities recognition is not a black and white problem, so fuzzy approaches are capable to deal with this kind of problems in a grey manner, i.e. it is aimed at identifying independently the degree of presence of each activity.

Furthermore, as mentioned in Sect. 1, it is also possible in reality that an instance can fully belong to more than one class, since these classes are defined from different perspectives. For example, a student can be classified as an international student in terms of nationality, as a full-time student in terms of study mode, or as a undergraduate student in terms of degree levels. In this context, a student can fully belong to all the three classes above. Since fuzzy rule-based classification is generally done in a generative way, i.e. it treats each class equally and the membership degree of an instance to each class is measured independently, a fuzzy classifier is really capable to capture the case that an instance highly or even fully belongs to more than one class, i.e. an instance appears to have a very high membership degree (closer or even equal to 1) to more than one class.

On the basis of the above argumentation, it is necessary to propose the use of fuzzy approaches instead of probabilistic approaches, in terms of recognition-intensive classification. First, probabilistic approaches aim at learning classifiers that discriminate one class from other classes. However, when these classes are not mutually exclusive, probabilistic classifiers, such as DT (Quinlan 1993), would fail to identify that some instances actually belong to one class, due to the case that these instances have been recognized as instances of another class. Also, when the number of classes is grown and becomes massive, it would be more difficult for probabilistic approaches to train classifiers that discriminate effectively between classes, e.g. NB (Rish 2001) is generally not able to learn from a data set that involves a massive number of classes. In some cases, it is also possible to result in the case that some instances can not be classified. For example, as illustrated in Fig. 4, a linear SVM classifier (Cristianini 2000) is unable to classify any instances that lie in an area (area E) enclosed by three boundaries. In addition, due to a massive number of classes, when the K Nearest Neighbour (KNN) algorithm (Zhang 1992) is used, it would also be more likely to happen that multiple classes appear to be the
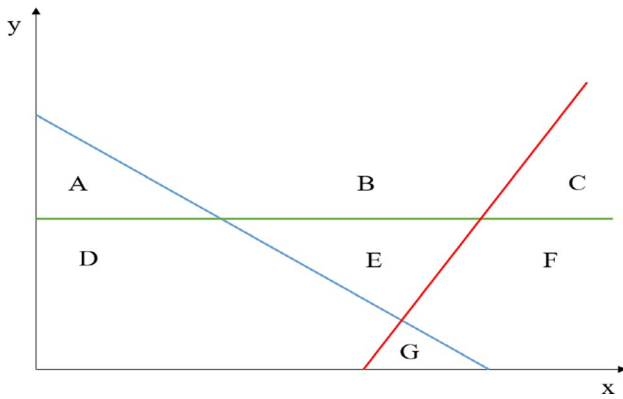
Fig. 4 Linear SVM for multi-class classification

most frequently occurring ones, leading to the uncertainty in classifying instances.

We will show experimental results to support the above argumentation in the context of fuzzy multi-task classification from granular computing perspectives.

4 Experiments, results and discussion

In this section, we report an experimental study on fuzzy multi-task learning for recognition-intensive classification. The experiments were conducted using 7 UCI data sets on

life sciences (Lichman 2013). The characteristics of the data sets are shown in Table 1.

In terms of classification accuracy, we compare the fuzzy approach with four popular probabilistic ones for pattern recognition, namely DT (Quinlan 1993), NB (Rish 2001), KNN (Zhang 1992) and SVM with the polynomial kernel (Cristianini 2000). Also, we show the membership degree values of some representative instances (selected from the test sets) to all the given classes, to indicate that the fuzzy approach is capable to capture more patterns than expected, i.e. an instance may also belong to other classes apart from the target class, or the set of given classes is not complete, so the instance cannot be classified and an extra class needs to be found.

The learning of fuzzy rule-based systems is based on the mixed fuzzy rule formation algorithm (Berthold 2003), which has been implemented on the KNIME platform (Berthold et al. 2013).

The results on classification accuracy are shown in Table 2 and indicate that the fuzzy approach outperforms all the probabilistic ones in two out of the seven cases (on the 'Forest-Type' and 'Glass' data sets). In the other five cases, the fuzzy approach performs marginally worse than the best performing one but still outperforms the majority of the probabilistic approaches. The results shown in Table 2 indicate that the fuzzy approach can fairly be used as the alternative one to these popular probabilistic approaches for recognition-intensive classification, without loss of classification accuracy.

Table 1 Characteristics of data sets

Dataset	Attribute types	#Attributes	#Instances	#Classes	References
Forest-type	Continuous	27	326	4	Johnson et al. (2012)
Anuran-calls-1	Continuous	22	7195	4	Colonna et al. (2015)
Anuran-calls-2	Continuous	22	7195	8	Colonna et al. (2015)
Anuran-calls-3	Continuous	22	7195	10	Colonna et al. (2015)
Cardiotocography-1	Continuous	23	2126	3	de Campos et al. (2000)
Cardiotocography-2	Continuous	23	2126	10	de Campos et al. (2000)
Glass	Continuous	10	214	6	Evett and Spiehler (1987)

Table 2 Classification accuracy

Dataset	DT (Quinlan 1993)	NB (Rish 2001)	KNN (Zhang 1992)	SVM (Cristianini 2000)	Fuzzy (Berthold 2003)
Forest-type	0.806	0.772	0.84	0.828	0.853
Anuran-calls-1	0.951	0.878	0.991	0.935	0.98
Anuran-calls-2	0.937	0.854	0.988	0.933	0.97
Anuran-calls-3	0.947	0.884	0.987	0.94	0.96
Cardiotocography-1	0.983	0.986	0.821	0.78	0.974
Cardiotocography-2	1	0.999	0.422	0.278	0.991
Glass	0.663	0.485	0.655	0.553	0.686

Table 3 Results sample on forest-type data

ID	Class	d	h	o	s	Prediction
1	d	0.25	0	0	0	d
2	s	0	0.63	0	1	s
3	s	0	0	0	1	s
4	d	1	0	0	0	d
5	h	0	1	0	0	h
6	o	0	0	1	0	o
7	d	0.91	0	0	0.28	d
8	d	0	0	0	0	?
9	s	0	0	0	0	?
10	s	0	1	0	1	s
11	s	1	0	0	1	s
12	s	0	0.77	0	1	s
13	d	1	0	0.23	0	d
14	d	1	0	0.44	0	d
15	o	0.2	0	1	0	o

Table 4 Results sample on glass data

Class	1	2	3	4	5	6
1	1	0	0.5	0	0	0
2	0.70	0.79	0	0	0	0
2	0	0.54	0	0	0	0
2	0.75	1	0.29	0	0	0
1	1	0	0	0	0	0
2	0	0	0	0	0	0
4	0	0	0	1	0	0
2	0.07	1	0	0.77	0.53	0
6	0	0.39	0	0.07	0	0.84
5	0	0	0	0	0	0

Table 5 Results sample on Anuran-calls-1 data

Class	B	D	H	L
D	0	1	0	0
L	0	0	0.63	1
L	0	0	1	1
H	0	0	1	1
L	0	0	0.45	1
D	0	0	0	0
D	0	1	0	0.72
L	0	0	0.20	1
L	0	0	0	0
L	0	0	0	0.73

However, the fuzzy approach is capable to capture more patterns which probabilistic approaches cannot achieve, as mentioned above. In particular, Tables 3, 4 and 5 are presented to show the membership degrees of each instance to different classes. For example, in Table 3, the first column represents the ID of an instance; the second column represents the class label that is assigned to each instance by experts, and the third to sixth columns represent the membership degrees of each instance to these corresponding classes (i.e. 'd', 'h', 'o' and 's'). In addition, the last column represents the prediction made by the fuzzy classifier for assigning a class to an instance in the setting of traditional machine learning, i.e. it is the output of the fuzzy classifier through defuzzification. However, as argued in Sect. 3, the defuzzification step is not needed in the setting of recognition-intensive classification from a granular computing perspective, and we include this column just for clarifying what outputs would be provided by the fuzzy classifier if the defuzzification step is involved.

The results shown in Table 3 indicate that an instance can match the features of more than one type of forest, i.e. an instance may have a very high membership degree (closer or even equal to 1) to more than one class, e.g. instances 10, 11 and 12.

Furthermore, the results show that an instance may not belong to any of the predefined classes, i.e. an instance has

the membership degree value of 0 to all the classes. In this case, the instance is unclassified, so it is labelled with "?" as shown in Table 3, but it is very different from the case of unclassification from a probabilistic classifier. In probabilistic classification, the above case is due to a normal distribution (e.g. 50/50 for a two-class classification problem) happening to an instance, i.e. maximum uncertainty is reached. In contrast, the phenomenon of the membership degree of 0 to all the classes indicates that the fuzzy classifier is confident that the instance does not belong to any classes, i.e. no evidence is found to assign the instance any non-zero values of membership degree to any one of the classes.

From a mathematical perspective, the above phenomenon can be explained by the case of incomplete mapping. In particular, a classifier is essentially a function that provides a discrete output after an input is provided. A function f is defined as a mapping from set A to set B, where A is the domain of f and the range of f is a subset of B. In this context, if a classifier does not represent a complete mapping, then there would be some truly existing classes (available in set B) but they are not in the range of this function f. In fact, real-life environments are generally imperfect, imprecise, incomplete and uncertain, so it is fairly possible that a set of predefined classes is not complete, and an extra class, which is not known yet, needs to be found to classify an instance.

For the above example on forest type identification, the four classes 's', 'h', 'd' and 'o' represent 'Sugi' forest, 'Hinoki' forest, 'Mixed deciduous' forest and 'Other' non-forest land, respectively. On the basis of the above argumentation, it is fairly possible that there may be other types of forest beyond experts' knowledge and cognition, i.e. these forest types are indeed existing but are not known yet. Also, the membership degrees shown in Table 3 for instances 13, 14 and 15 indicate that even identifying the case of forest or non-forest is not really a black and white problem, which again supports the argumentation made in Sect. 3.2 that some instances of mutually exclusive classes may still show some highly similar features, so the same instance may have non-zero values of membership degree to some or even all of these classes. The similar phenomenon (in terms of membership degrees of instances to different classes) can also be found in other recognition-intensive classification tasks as shown in Tables 4 and 5.

5 Conclusions

In this paper, we proposed the use of fuzzy rule-based systems for recognition-intensive classification in the setting of granular computing. In particular, we treated the recognition of each class of instances as an independent task of learning and classification, and the class is viewed as the target class. When there are several target classes of instances that

need to be recognized, fuzzy multi-task learning becomes very suitable to not only identify the presence of the patterns of each target class but also measure the degree to which the patterns of a target class are present. The features of fuzzy multi-task learning are highly required, especially when there is a large number of classes involved and the classification problem is not black and white.

The experimental results show that the classification performance of the fuzzy approach is fairly comparable to the ones of the probabilistic approaches (DT, NB, KNN and SVM), which indicate that the fuzzy approach can be used as the alternative ones to the probabilistic approaches. However, the probabilistic approaches would fall short in the aspects that are usually involved in recognition-intensive classification. In particular, the probabilistic approaches aim at learning classifiers that discriminate one class from other classes. As mentioned in Sect. 3.2, when the number of classes is very large or even massive, it would become very difficult to discriminate effectively between classes. Also, in the context of recognition-intensive classification, it is fairly possible that different classes are not mutually exclusive so there is no need to involve discrimination between classes.

In contrast, the fuzzy approach aims at training classifiers in the way of generative learning, i.e. each class is treated equally, and recognition of each class of instances is involved in an independent task of learning and classification, i.e. multi-task learning. Therefore, the fuzzy approach is capable to deal effectively with a massive number of classes and to discover that an instance does not only belong to the target class but also to other classes. Furthermore, the fuzzy approach can also discover the case that an instance does not belong to any of the given classes and an extra class thus needs to be discovered. In fact, the above case is fairly possible to appear in real-life environments that are imperfect, imprecise, incomplete and uncertain.

In future work, we will investigate further the use of fuzzy rule-based systems for identifying the relationships between classes in the setting of granular computing, i.e. it is to identify the relationships between information granules where each class is viewed as a granule. In particular, following the completion of fuzzy multi-task classification, all instances are assigned values of membership degree to the given classes. In this context, a secondary learning task for association (correlation) analysis can be undertaken, where each class is treated as an attribute (feature) and the membership degree value of each instance to this class is treated as a value of this feature. We will also look into the ensemble classification or data stream mining problems using fuzzy rule-based systems where the data instances are challenging, unpredictable and diverse embedded with newly arrived classes. In addition, it is worth to investigate the use of optimization techniques (Chen and Chien 2011; Chen and Kao 2013; Tsai et al. 2008, 2012; Chen and Chang 2011; Chen

et al. 2013; Chen and Chung 2006; Chen and Huang 2003) for tuning the shapes of membership functions towards obtaining better performance of prediction.

References

Ali N, Xavier L (2014) Person identification and gender classification using Gabor filters and fuzzy logic. Int J Electr Electron Data Commun 2(4):20–23

Altrabsheh N, Cocea M, Fallahkhair S (2015) Predicting students' emotions using machine learning techniques. Springer, Cham, pp 537–540

Antonelli M, Ducange P, Lazzerini B, Marcelloni F (2016) Multi-objective evolutionary design of granular rule-based classifiers. Granul Comput 1(1):37–58

Bergadano F, Cutello V (1993) Learning membership functions. In: European conference on symbolic and quantitative approaches to reasoning and uncertainty, Granada, Spain, pp 25–32

Berthold MR (2003) Mixed fuzzy rule formation. Int J Approx Reason 32:67–84

Berthold MR, Wiswedel B, Gabriel TR (2013) Fuzzy logic in knime: modules for approximate reasoning. Int J Comput Intell Syst 6(1):34–45

Chatterjee K, Kar S (2017) Unified granular-number-based ahp-vikor multi-criteria decision framework. Granul Comput 2(3):199–221

Chen S-M (1996) A fuzzy reasoning approach for rule-based systems based on fuzzy logics. IEEE Trans Syst Man Cybern Part B Cybern 26(5):769–778

Chen S-M, Chang T-H (2001) Finding multiple possible critical paths using fuzzy pert. IEEE Trans Syst Man Cybern Part B Cybern 31(6):930–937

Chen S-M, Chang Y-C (2011) Weighted fuzzy rule interpolation based on ga-based weight-learning techniques. IEEE Trans Fuzzy Syst 19(4):729–744

Chen S-M, Chen C-D (2011) Handling forecasting problems based on high-order fuzzy logical relationships. Expert Syst Appl 38(4):3857–3864

Chen S-M, Chien C-Y (2011) Parallelized genetic ant colony systems for solving the traveling salesman problem. Expert Syst Appl 38(4):3873–3883

Chen S-M, Chung N-Y (2006) Forecasting enrollments of students by using fuzzy time series and genetic algorithms. Int J Inform Manag Sci 17(3):1–17

Chen S-M, Huang C-M (2003) Generating weighted fuzzy rules from relational database systems for estimating null values using genetic algorithms. IEEE Trans Fuzzy Syst 11(4):495–506

Chen S-M, Kao P-Y (2013) Taiex forecasting based on fuzzy time series, particle swarm optimization techniques and support vector machines. Inform Sci 247:62–71

Chen S-M, Lee S-H, Lee C-H (2001) A new method for generating fuzzy rules from numerical data for handling classification problems. Appl Artif Intell 15(7):645–664

Chen S-M, Wang N-Y, Pan J-S (2009) Forecasting enrollments using automatic clustering techniques and fuzzy logical relationships. Expert Syst Appl 36(8):11070–11076

Chen S-M, Lee S-H, Lee C-H (2011) Fuzzy forecasting based on high-order fuzzy logical relationships and automatic clustering techniques. Expert Syst Appl 38(12):15425–15437

Chen S-M, Chang Y-C, Pan J-S (2013) Fuzzy rules interpolation for sparse fuzzy rule-based systems based on interval type-2 gaussian fuzzy sets and genetic algorithms. IEEE Trans Fuzzy Syst 21(3):412–425

Colonna JG, Cristo M, Jnior MS, Nakamura EF (2015) An incremental technique for real-time bioacoustic signal segmentation. Expert Syst Appl 42(21):7367–7374

Cristianini N (2000) An introduction to support vector machines and other Kernel-based learning methods. Cambridge University Press, Cambridge

de Campos DA, Bernardes J, Garrido A, de S JM, Pereira-Leite L (2000) Sisporto 2.0 a program for automated analysis of cardiotocograms. J Matern Fetal Med 9(5):311–318

Dubois D, Prade H (2016) Bridging gaps between several forms of granular computing. Granul Comput 1(2):115–126

Evett IW, Spiehler EJ (1987) Rule induction in forensic science. Technical report, central research establishment, home office forensic science service

Guo G (2014) Gender classification. In: Encyclopedia of biometrics. Springer, New York, pp 1–6

Horng Y-J, Chen S-M, Chang Y-C, Lee C-H (2005) A new method for fuzzy information retrieval based on fuzzy hierarchical clustering and fuzzy inference techniques. IEEE Trans Fuzzy Syst 13(2):216–228

Hu H, Shi Z (2009) Machine learning as granular computing. In: IEEE International conference on granular computing, Nanchang, Beijing, pp 229–234

Jefferson C, Liu H, Cocea M (2017) Fuzzy approach for sentiment analysis. In: IEEE International conference on fuzzy systems, Naples, Italy

Johnson B, Tateishi R, Xie Z (2012) Using geographically-weighted variables for image classification. Remote Sens Lett 3(6):491–499

Kalua B, Mirchevska V, Dovgan E, Lutrek M, Gams M (2010) An agent-based approach to care in independent living. In: International joint conference on ambient intelligence, pp 177–186

Kreinovich V (2016) Solving equations (and systems of equations) under uncertainty: how different practical problems lead to different mathematical and computational formulations. Granul Comput 1(3):171–179

Lichman M (2013) UCI machine learning repository. http://archive.ics.uci.edu/ml. Accessed 22 Oct 2017

Lin F, Wu Y, Zhuang Y, Long X, Xu W (2016) Human gender classification: a review. Int J Biom 8(3–4). https://doi.org/10.1504/IJBM.2016.082604

Liu H, Cocea M (2017a) Fuzzy information granulation towards interpretable sentiment analysis. Granul Comput 2(4):289–302

Liu H, Cocea M (2017b) Fuzzy rule based systems for interpretable sentiment analysis. In: International conference on advanced computational intelligence, Doha, Qatar, pp 129–136

Liu H, Cocea M (2017c) Granular computing based approach for classification towards reduction of bias in ensemble learning. Granul Comput 2(3):131–139

Liu H, Cocea M (2018) Granular computing based machine learning: a big data processing approach. Springer, Berlin

Liu H, Gegov A (2015) Collaborative decision making by ensemble rule based classification systems, Springer, Switzerland, pp 245–264

Liu H, Cocea M, Ding W (2017a) Decision tree learning based feature evaluation and selection for image classification. In: International conference on machine learning and cybernetics, Ningbo, China

Liu H, Cocea M, Mohasseb A, Bader M (2017b) Transformation of discriminative single-task classification into generative multi-task classification in machine learning context. In: International conference on advanced computational intelligence. Doha, Qatar, pp 66–73

Liu H, Cocea M, Ding W (2017) Multi-task learning for intelligent data processing in granular computing context. Granul Comput **(In press)**

Liu P, You X (2017) Probabilistic linguistic todim approach for multiple attribute decision-making. Granul Comput 2(4):332–342

Livi L, Sadeghian A (2016) Granular computing, computational intelligence, and the analysis of non-geometric input spaces. Granul Comput 1(1):13–20

Mamdani E, Assilian S (1999) An experiment in linguistic synthesis with a fuzzy logic controller. Int J Hum Comput Stud 51(2):135–147

Min F, Xu J (2016) Semi-greedy heuristics for feature selection with test cost constraints. Granul Comput 1(3):199–211

Pedrycz W, Chen S-M (2011) Granular computing and intelligent systems: design with information granules of higher order and higher type. Springer, Heidelberg

Pedrycz W, Chen S-M (2015) Information granularity, big data, and computational intelligence. Springer, Heidelberg

Pedrycz W, Chen S-M (2016) Sentiment analysis and ontology engineering: an environment of computational intelligence. Springer, Heidelberg

Peters G, Weber R (2016) Dcc: a framework for dynamic granular clustering. Granul Comput 1(1):1–11

Quinlan RJ (1993) C4.5: programs for machine learning. Morgan Kaufmann Publishers, San Francisco

Reynolds K, Kontostathis A, Edwards L (2011) Using machine learning to detect cyberbullying. In: Proceedings of the 10th international conference on machine learning and applications, pp 241–244

Rish I (2001) An empirical study of the naive bayes classifier. In: IJCAI 2001 workshop on empirical methods in artificial intelligence vol 3(22), pp 41–46

Skowron A, Jankowski A, Dutta S (2016) Interactive granular computing. Granul Comput 1(2):95–113

Suykens JA, Vandewalle J (1999) Least squares support vector machine classifiers. Neural Process Lett 9(3):293–300

Tan P-N, Steinbach M, Kumar V (2005) Introduction to data mining. Addison-Wesley Longman Publishing Co., Inc, Boston

Teng Z, Ren F, Kuroiwa S (2007) Emotion recognition from text based on the rough set theory and the support vector machines. In: International conference on natural language processing and knowledge engineering, Beijing, China, pp 36–41

Tsai P-W, Pan J-S, Chen S-M, Liao B-Y, Hao S-P (2008) Parallel cat swarm optimization. In: Proceedings of the 2008 international conference on machine learning and cybernetics, Kunming, China, vol 6, pp 3328–3333

Tsai P-W, Pan J-S, Chen S-M, Liao B-Y (2012) Enhanced parallel cat swarm optimization based on the taguchi method. Expert Syst Appl 39(7):6309–6319

Wang Y, Yu H, Dong J, Stevens B, Liu H (2016) Facial expression-aware face frontalization. In: LNCS Proceedings of Asian conference on computer vision, Taibei, Taiwan, pp 375–388

Wilke G, Portmann E (2016) Granular computing as a basis of human data interaction: a cognitive cities use case. Granul Comput 1(3):181–197

Wu J, Smith WA, Hancock ER (2011) Gender discriminating models from facial surface normals. Pattern Recognit 44(12):2871–2886

Xu Z, Wang H (2016) Managing multi-granularity linguistic information in qualitative group decision making: an overview. Granul Comput 1(1):21–35

Yao J (2005a) Information granulation and granular relationships. In: IEEE international conference on granular computing. Beijing, China, pp 326–329

Yao Y (2005b) Perspectives of granular computing. In: Proceedings of 2005 IEEE international conference on granular computing, Beijing, China, pp 85–90

Zadeh L (1965) Fuzzy sets. Inform Control 8(3):338–353

Zhang J (1992) Selecting typical instances in instance-based learning. In: Proceedings of the ninth international workshop on machine learning, Aberdeen, United Kingdom, pp 470–479

Zhang X, Yin Y, Yu H (2007) An application on text classification based on granular computing. Commun IIMA 7(2):1–8

Zhao R, Zhou A, Mao K (2016) Automatic detection of cyberbullying on social networks based on bullying features. In: Proceedings of the 17th international conference on distributed computing and networking