Multi-Level Fusion of Classifiers through Fuzzy Ensemble Learning

Han Liu
School of Computer Science and Informatics
Cardiff University
Cardiff, United Kingdom
e-mail: liuh48@cardiff.ac.uk

Shyi-Ming Chen
Department of Computer Science and Information Engineering
National Taiwan University of Science and Technology
Taipei, Taiwan
e-mail: smchen@mail.ntust.edu.tw

***Abstract*—Classification is a popular task of supervised machine learning, which can be achieved by training a single classifier or a group of classifiers. In general, the performance of each traditional learning algorithm which leads to the production of a single classifier is varied on different data sets, i.e., each learning algorithm may produce good classifiers on some data sets, but may produce poor classifiers on the other data sets. In order to achieve a more stable performance of machine learning, ensemble learning has been undertaken more popularly to produce a group of classifiers that can be complementary to each other. In this paper, we focus on advancing fuzzy classification through multi-level fusion of fuzzy classifiers in the setting of ensemble learning. In particular, we propose an ensemble learning framework that leads to creating a group of fuzzy classifiers that are complementary to each other. The experimental results show that the proposed ensemble learning framework leads to considerable advances in the performance of fuzzy classification, in comparison with using each single fuzzy classifier.**

***Keywords*-machine learning; fuzzy classification; ensemble learning; fuzzy rule; classifiers fusion;**

I. Introduction

Due to the vast and rapid increase in the amount of data, machine learning has been widely used in real-world applications, such as knowledge discovery and pattern recognition. A typical way of machine learning is referred to as supervised learning, which can be undertaken for classification and regression in practice. Since the vast majority of applications (e.g., pattern recognition and decision making) involve predicting the value of a discrete output, classification has become a popular task of supervised learning.

In the machine learning context, classification is achieved by training a single classifier or a group of classifiers. However, it is a commonly known issue that each traditional learning algorithm has its own advantages and disadvantages [1], leading to the case that the performance of each learning algorithm is varied on different data sets, i.e., a learning algorithm may produce high quality classifiers on some data sets but may produce low quality classifiers on the others. In order to address the above issue, some researchers are motivated to develop ensemble learning approaches, towards advancing the overall performance of classification.

On the other hand, since most real-world problems are not black-and-white but involves degrees of fuzziness [2], fuzzy approaches become more needed to achieve fuzzy ensemble classification. In this paper, we propose a new ensemble learning framework for creation of a group (ensemble) of fuzzy rule based classifiers (based on fuzzy sets [2]), towards advancing the overall performance of classification through multi-level fusion of fuzzy classifiers. The main contributions of this paper include:

- The proposed framework of ensemble learning leads to an ensemble of fuzzy classifiers that are diverse and complementary to each other.
- The adoption of the proposed ensemble learning framework achieves to advance the classification performance, in comparison with using each traditional learning algorithm for training a single fuzzy classifier.
- The adoption of the proposed framework can overcome the limitations of each single algorithm of fuzzy rule learning, e.g., the case that some instances are left unclassified by a single fuzzy classifier can be effectively avoided through the fusion of multiple fuzzy classifiers.

The rest of this paper is organized as followsIn Section II, we provide an overview of ensemble learning approaches. In Section III, we describe preliminaries of fuzzy logic and illustrate the proposed fuzzy ensemble learning framework. In Section IV, we show the experimental results on multi-level fusion of fuzzy classifiers in the setting of fuzzy ensemble learning and the results are presented and discussed. The conclusions of this paper are given in Section V.

II. Related Work

Ensemble learning aims at advancing the overall performance through fusing different classifiers for classification tasks. In general, successful setting of ensemble learning needs to meet two key points [1], i.e., (1) the performance of each single (base) classifier must not be too bad and (2) it is crucial that different base classifiers show high diversity to each other. The term 'diversity' means that the use of different classifiers leads to different sets of incorrectly classified instances [3]. In this case, different classifiers are likely to be complementary to each other, such that the fusion of these classifiers leads to an increased number of correct classifications. The ideal outcome of classifiers fusion is referred to as 'Oracle' [4], which indicates that each instance would be classified correctly by an ensemble of classifiers if at least one of the classifiers in the ensemble gives a correct classification for this instance.

In order to create an ensemble of diverse classifiers, the most popular approaches of ensemble learning include the Bagging approach [5] and the Boosting approach. [6]. The Bagging approach involves random sampling of training data with replacement, which indicates that some instances may be selected more than once and some others may never be selected, and each sample is expected to contain 63.2% of the instances in the original training set [5]. The above procedure results in n training samples and a base classifier is trained on each sample, i.e., n base classifiers are trained, respectively, on the n training samples in parallel to make up an ensemble. Since the n classifiers are learned from n different subsets of instances, it is very likely to encourage the diversity among the classifiers [7]. In contrast, the Boosting approach involves sequential training of n classifiers. In particular, a base classifier is trained at each iteration t, and the classifier h_t is then evaluated using a validation set to measure the weight w_t of h_t. At the next iteration $t+1$, another base classifier h_{t+1} is trained by focusing the learning task more on the instances classified incorrectly by the previous classifier h_t. Through n iterations, n base classifiers are trained to make up an ensemble. Since each base classifier is trained by focusing the learning task on a different set of incorrectly classified instances, it is very likely that the n classifiers in the ensemble are diverse [7].

Both the Bagging approach and the Boosting approach are designed to encourage the diversity among classifiers through manipulation of training data, but the same learning algorithm is used for training all the base classifiers [8]. On the other hand, the increase of the diversity among classifiers can be achieved without the data manipulation, but different learning algorithms are used to train the base classifiers on the same training set [1]. In this context, it is necessary that the employed learning algorithms involve different strategies of learning, such that the classifiers trained by using these algorithms are likely to be diverse. More detailed reviews of ensemble learning techniques and the diversity creation can be found in [3], [9]. A fuzzy ensemble learning method was proposed in [10], which aims at determining the final output of a fuzzy ensemble based on the weight assigned to each output from a single fuzzy classifier. However, this method does not involve the creation of diversity among fuzzy classifiers. In the next section, we will develop effective ways of diversity creation in fuzzy ensembles.

III. Fuzzy Ensemble Learning Framework

In this section, we propose a new ensemble learning framework based on fuzzy logic [2]. The proposed fuzzy ensemble learning framework is illustrated and the theoretical significance of the framework is also justified.

A. Preliminaries

In the context of machine learning, applications of fuzzy logic [2] typically involve (1) fuzzification of continuous attributes and (2) fuzzy classification.

Fuzzification of continuous attributes is essentially a process of fuzzy sets definition. Each fuzzy set S_{ik} is identified by assigning it a linguistic term T_{ik}, e.g., 'good' and 'bad'. In other words, following the fuzzification stage, each continuous attribute A_i can be described by a number of linguistic terms $T_{i1}, T_{i2}, ..., T_{im}$ and each value of the continuous attribute has a certain degree of membership to each of the linguistic terms (fuzzy sets). The membership degree of a numeric value to a fuzzy set is determined by a membership function constructed for the fuzzy set. The construction of a membership function can be done either by experts or through statistical learning from data.

Fuzzy classification has been typically undertaken by learning a set of fuzzy rules from data and following the fuzzification of continuous attributes. A set of fuzzy rules is typically represented in the following form:

- Rule 1: if A_1 is T_{11} and A_2 is T_{21} and ... and A_d is A_{d1} then class= c_1;
- Rule 2: if A_1 is T_{12} and A_2 is T_{22} and ... and A_d is A_{d2} then class = c_2;
- $\vdots$
- Rule r: if A_1 is T_{1m} and A_2 is T_{2m} and ... and A_d is T_{dm} then class = c_v;

where c_v represents a class label and v is the class index.

In order to undertake the fuzzy rule-based classification, we need to compute the firing strength $fs(R_j) \in [0,1]$ of each rule R_j based on the membership degree G_{ji} obtained for each rule term 'A_i is T_{ik}', as shown in Eq. (1).

$$fs(R_j) = T_{i=1}^{d}(G_{ji}) \tag{1}$$

where T is an operation (T-norm) of combining the membership degrees obtained for all the terms of Rule R_j.

$$G_{c_v} = S_{j=1}^{r}(fs(R_j)) \quad subject\ to \quad consequent(R_j) = c_v \tag{2}$$

where S is an operation (T-conorm) of combining the firing strengths of some of the fuzzy rules $R_1, R_2, ..., R_r$ for the class c_v and $consequent(R_j) = c_v$ denotes that the consequent of the rule R_j is c_v.

Following the above operation shown in Eq. (1), the firing strengths of the fuzzy rules $R_1, R_2, ..., R_r$ need to be concatenated further to compute the membership degree for each class c_v, as shown in Eq. (2). The above two operations (T-norm and T-conorm) shown in Eqs. (1) and (2) can be jointly defined as specific fuzzy norms, such as the Min/Max norm [11], the Product norm [11], the Lukasiewicz's norm [11] and the Yager[2.0]'s norm [11].

B. The Proposed Fuzzy Ensemble Learning Framework

The proposed fuzzy ensemble learning framework is illustrated in Fig. 1, which involves three parts of design, namely, (1) feature selection, (2) primary ensemble creation and (3) final ensemble creation, towards diversity creation in ensembles.

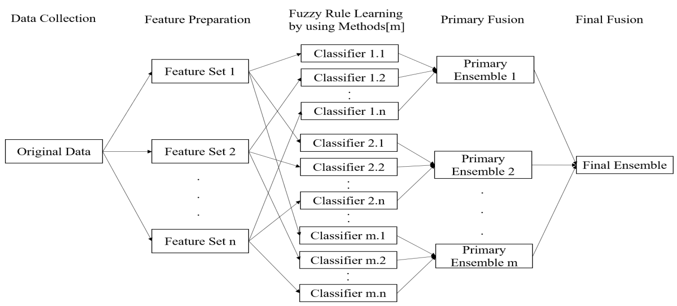

Figure 1. The proposed ensemble learning framework

In the feature selection part, the aim is to get a reduced feature set by selecting only highly relevant features. Since different feature selection methods usually involve different ways of feature relevance evaluation, it is thus likely to increase the diversity among different reduced feature sets that are obtained by using different feature selection methods.

In the primary ensemble creation part, since the base classifiers are trained on different feature sets by using the same learning algorithm, the classifiers in each primary ensemble are likely to be diverse leading to advances in classification performance. In Fig. 1, Methods[m] means that m learning algorithms are used to create m primary ensembles. In the final ensemble creation part, the aim is to encourage further diversity through fusion of the primary ensembles created through using different learning algorithms. In other words, different algorithms usually involve different learning strategies leading to diversity among the primary ensembles.

In Fig. 1, the creation of an ensemble is essentially the fusion of classifiers in the ensemble. The proposed framework involves multi-level fusion of classifiers since ensembles are created in different levels, e.g., primary ensembles and the final ensemble. There have been various rules used for fusion of probabilistic classifiers as introduced in [4], such as majority vote, mean, min and max. However, for fuzzy classifiers fusion, we propose to adopt the mean rule (averaging the membership degrees derived from different classifiers for each class), since the use of a single fuzzy classifier is likely to result in some instances being unclassified due to the sample representative issue [12]. The reason behind the above case is that an instance obtains a membership degree of 0 for each class. In order to address this issue, it is essential that the fusion of fuzzy classifiers can result in a non-zero membership degree for at least one class by using a suitable fusion rule.

IV. Experimental Results

In this section, we conduct the experiments on 5 UCI data sets [13] in order to evaluate the performance of the proposed fuzzy ensemble learning framework. In particular, the mixed fuzzy rule formation algorithm [14] is used to train different base classifiers by selecting different fuzzy norms and the area-based (border-based) shrink function [11] as the parameters of this algorithm. In our experimental setup, the Min/Max norm [11], the Product norm [11], the Lukasiewicz's norm [11] and the Yager[2.0]'s norm [11] are selected for obtaining diverse fuzzy classifiers that make up an ensemble.

On the other hand, we apply the correlation-based feature subset selection method [15] to each data set for getting diverse classifiers trained on different feature sets. In other words, we produce two feature sets for each data set – one contains all original features and the other one contains selected features only. In this way, two base classifiers are trained, respectively, on the two feature sets to make up a primary ensemble, by using each fuzzy norm as a parameter of the mixed fuzzy rule formation algorithm [14]. Therefore, the final (secondary) ensemble consists of four primary ensembles created by using the above four fuzzy norms.

All the experiments are conducted by using the 10-fold cross validation method [1], where the results are shown in Table I and Table II in terms of classification accuracy and number of unclassified instances, respectively. In Table I and Table II, 'Diabetes', 'Heart-Stalog', 'Liver-Disorders', 'Sonar' and 'Spambase' represent five UCI data sets [13] used in our experiments. Moreover, Min/Max 1 represents that the Min/Max norm is used for training a fuzzy classifier on all original features from a data set, whereas Min/Max 2 represents that a fuzzy classifier is trained on selected features only by using the Min/Max norm. Min/Max 3 represents the case of primary fusion of two classifiers resulting from using Min/Max 1 and Min/Max 2, respectively. The above description also applies to the cases when using the Product norm, Lukasiewicz's norm and Yager[2.0]'s norm.

Table I
Classification Accuracy

Methods	Diabetes	Heart-Stalog	Liver-Disorders	Sonar	Spambase
Min/Max 1	0.702	0.774	0.643	0.712	0.919
Min/Max 2	0.71	0.77	0.472	0.74	0.891
Min/Max 3	0.729	0.785	0.635	0.779	0.935
Product 1	0.716	0.759	0.594	0.716	0.923
Product 2	0.727	0.767	0.426	0.74	0.895
Product 3	0.742	0.781	0.568	0.716	0.937
Lukasiewicz 1	0.717	0.756	0.62	0.534	0.924
Lukasiewicz 2	0.706	0.796	0.464	0.654	0.894
Lukasiewicz 3	0.754	0.778	0.603	0.649	0.938
Yager[2.0] 1	0.725	0.763	0.623	0.649	0.929
Yager[2.0] 2	0.724	0.767	0.472	0.736	0.897
Yager[2.0] 3	0.742	0.781	0.652	0.716	0.938
The proposed method	**0.758**	**0.8**	**0.672**	**0.822**	**0.942**

From Table I, we can see that the proposed fuzzy ensemble learning framework leads to the best performance comparing with the use of each single fuzzy classifier or each primary ensemble of fuzzy classifiers, while the performance of each single classifier or each primary ensemble is varied on different data sets. The improvement of classification accuracy shows the effectiveness of our proposed framework on the creation of an ensemble of diverse classifiers.

Table II
NUMBER OF UNCLASSIFIED INSTANCES

Methods	Diabetes	Heart-Stalog	Liver-Disorders	Sonar	Spambase
Min/Max 1	59	19	36	10	284
Min/Max 2	37	10	0	11	174
Min/Max 3	0	0	0	0	0
Product 1	61	17	36	10	301
Product 2	43	17	0	10	160
Product 3	0	0	0	0	0
Lukasiewicz 1	97	20	54	206	412
Lukasiewicz 2	33	9	0	108	204
Lukasiewicz 3	0	0	0	0	0
Yager[2.0] 1	69	14	34	112	323
Yager[2.0] 2	30	8	0	27	171
Yager[2.0] 3	0	0	0	0	0
The proposed method	0	0	0	0	0

From Table II, we can see that the primary fusion of classifiers trained on different feature sets by using the same learning algorithm can effectively overcome the limitation that instances are left unclassified by using a single fuzzy classifier. Furthermore, the above statement means that training classifiers on different feature sets is an effective way to increase the diversity in ensembles. In other words, for each instance, at least one of the base classifiers does not leave it unclassified, such that the instance is not left unclassified following the fusion of the base classifiers.

V. CONCLUSION

In this paper, we have proposed a new fuzzy ensemble learning framework for advancing fuzzy classification through multi-level fusion of classifiers. In particular, we argued that each traditional fuzzy rule learning algorithm would have its own advantages and disadvantages, leading to the varied classification performance on different data sets. Also, we have identified that the fuzzy classifier trained by using a traditional algorithm is likely to leave some instances unclassified due to the sample representativeness issue.

The experimental results show that the adoption of the proposed fuzzy ensemble learning framework can effectively overcome the limitations of the traditional fuzzy rule learning algorithms, leading to considerable advances in the classification performance in comparison with using the traditional fuzzy rule learning algorithms separately.

In the future, we will investigate in depth how to better encourage the diversity among different fuzzy classifiers in the setting of extraction and selection of diverse features. It is also worth to investigate granular computing techniques [12] towards fusion of fuzzy classifiers in more depth.

ACKNOWLEDGMENT

This work is supported by the Ministry of Science and Technology, Republic of China, under Grant MOST 107-2221-E-011-122 -MY2.

REFERENCES

[1] Z. H. Zhou, *Ensemble Methods: Foundations and Algorithms*. London: Chapman and Hall/CRC, 2012.

[2] L. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.

[3] G. Brown and L. I. Kuncheva, "Good and bad diversity in majority vote ensembles," in *Proceedings of the 2010 International Workshop on Multiple Classifier Systems*, Cairo, Egypt, 7-9 April 2010, pp. 124–133.

[4] L. I. Kuncheva, "A theoretical study on six classifier fusion strategies," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 2, pp. 281–285, 2002.

[5] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.

[6] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," in *Proceedings of the Thirteenth International Conference on Machine Learning*, Bari, Italy, 1996, pp. 148–156.

[7] P. Melville and R. J. Mooney, "Creating diversity in ensembles using artificial data," *Information Fusion*, vol. 6, no. 1, pp. 99–111, 2005.

[8] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*. New Jersey: John Wiley and Sons, Inc., 2004.

[9] M. Wozniak, M. Grana, and E. Corchado, "A survey of multiple classifier systems as hybrid systems," *Information Fusion*, vol. 16, pp. 3–17, 2014.

[10] G. Nakai, T. Nakashima, and H. Ishibuchi, "A fuzzy ensemble learning method for pattern classification," *Journal of Japan Society for Fuzzy Theory and Intelligent Informatics*, vol. 15, no. 6, pp. 671–681, 2003.

[11] T. R. Gabriel and M. R. Berthold, "Influence of fuzzy norms and other heuristics on mixed fuzzy rule formation," *International Journal of Approximate Reasoning*, vol. 35, pp. 195–202, 2004.

[12] H. Liu and M. Cocea, "Fuzzy rule based systems for gender classification from blog data," in *Proceedings of the 2018 International Conference on Advanced Computational Intelligence*, Xiamen, China, 2018, pp. 79–84.

[13] M. Lichman, "UCI Machine Learning Repository, http://archive.ics.uci.edu/ml," 2013.

[14] M. R. Berthold, "Mixed fuzzy rule formation," *International Journal of Approximate Reasoning*, vol. 32, pp. 67–84, 2003.

[15] M. A. Hall and L. A. Smith, "Feature subset selection: a correlation based filter approach," in *Proceedings of the 1997 International Conference on Neural Information Processing and Intelligent Information Systems*. Berlin, Germany: Springer, 1997, pp. 855–858.