JOURNAL OF APPLIED INTELLIGENT SYSTEMS & INFORMATION SCIENCES
VOL. 1, ISSUE 1, PP. 24-34, FEBRUARY 2020.
Available at: www.journal.research.fanap.com
DOI: 10.22034/jaisis.2020.102462

MUTUAL INFORMATION BASED FUZZY INFERENCE SYSTEM FOR CLASSIFICATION PROBLEMS

Mahdi Illbeygi[1,*], Mohammad Reza Kangavari[2]
[1] *Department of Computer Engineering, Iran University of Science and Technology (IUST), Tehran, Iran;* mahdi_ilbeygi@iust.ac.ir
[2] *Department of Computer Engineering, Iran University of Science and Technology (IUST), Tehran, Iran.*

ABSTRACT

Fuzzy inference system (FIS) is one of the most powerful inference systems that is widely used in the field of classification. Indeed, in this approach, FIS is engaged to create a mapping from features (inputs) onto classes (outputs) using fuzzy set theory. So far many efforts have been made to improve classification accuracy performed by FIS. Generally, these efforts have been put in the following areas: efficient fuzzy rule generation, fuzzy membership function tuning, fuzzy rule weight tuning, feature selection for the antecedent part of fuzzy rules, and so on. In this paper, we consider this issue and propose a method based on mutual information for applying the impact factor of input parameters on the fuzzy inference process for improving the accuracy of fuzzy classification. Finally, we test our proposed method for boosting classification on six different problems using manual and auto-generated FIS. The method provided promising classification results confirming its correctness.

KEYWORDS: Fuzzy Inference System, Fuzzy Classification, Mutual Information, Fusion Operator, Auto Generated FIS.

1. INTRODUCTION

Classification is a process for determining categories of given samples from a set of classes. In the terminology of machine learning, classification is considered an instance of supervised learning, i.e. learning which provides a training set of correctly identified observations (Alpaydin, 2014). Indeed, so many problems in the real world can be expressed as classification problems. Therefore, numerous algorithms with many different approaches have been proposed for classification, so far.

One of the most famous approaches to classification, is to engage fuzzy inference systems (FIS) as a classifier. Because most of the real world classification problems have classes with fuzzy boundaries (Sarkar & Yegnanarayan, 1997), fuzzy classification algorithms have proper capabilities to tackle these problems. The fuzzy classification is an extension of the traditional classification, the same way that the fuzzy sets extend the classical sets (Werro, 2015). Since Zadeh (1965) introduced fuzzy sets, many attempts have been made for solving problems using fuzzy sets and fuzzy inference systems. For example, many algorithms have been proposed only for combination of fuzzy inference system and neural networks to solve function approximation, classification, clustering and other problems.

Moreover, many researches have been conducted to learn fuzzy rules and tune fuzzy inference system parameters in order to reach high accuracy for solving problems. In the following sections, we will refer to some of these investigations. For example, the Adaptive Network-Based Fuzzy Inference System (ANFIS), introduced by Jang (1993), is a well-known and efficient tool for implementing the Takagi-Sugeno fuzzy inference system. ANFIS is able to learn fuzzy rules, their weights, and tunes fuzzy membership function

[*] *Corresponding Author*

RECEIVED: 21 DECEMBER, 2019; **ACCEPTED**: 1 FEBRUARY, 2020; **PUBLISHED ONLINE**: 1 FEBRUARY, 2020

parameters, simultaneously. This tool has been developed for function approximation problems, inherently, but it also can be used to solve classification problems.

In addition, numerous studies have been conducted to tune fuzzy inference system parameters in an attempt to achieve high accuracy and efficiency. For example, a supervised learning technique to fine-tune the fuzzy sets of an existing Sugeno type fuzzy system was proposed by Nomura et al. (1992). In the research, membership function parameters were fine-tuned using a gradient descent procedure. Fuzzy Inference Environment Software with Tuning (FINEST) is another tool for tuning the fuzzy inference system parameters (Tano et al., 1996). FINEST is able to tune fuzzy predicates and implication functions in separate processes.

Another tool that implements a Takagi-Sugeno type fuzzy inference system is the Self-Constructing Neural Fuzzy Inference Network (SONFIN) (Feng & Teng, 1998). This tool uses an aligned clustering-based algorithm to partition input space in a flexible way. For the consequent part of the fuzzy rule, only a singleton value is selected using a clustering method assigned to each rule initially. SONFIN has many details about parameter identification and knowledge representation enhancement, etc. that can be found by Feng & Teng (1998). Also, the metacognitive neuro-fuzzy inference system (McFIS), with a metacognitive sequential learning algorithm for a neuro-fuzzy inference system for classification tasks, uses a human-like learning strategy for tackling these problems (Subramanian et al., 2013). Furthermore, an overview of the adaptation of a fuzzy inference system using neural learning is proposed by Abraham (2005) that can be used to achieve more information in this field.

Shihabudheen et al. (2018) applied an extreme learning machines concept and particle swarm optimization to improve the performance of adaptive neuro-fuzzy inference system. This has led to reduced randomness, reduced computational complexity, and improved generalization. Mutlu et al. (2018) investigated the effect of the issues arrising in modeling and reasoning of Hierarchical Fuzzy System(HFS) on the accuracy and stability of the resultant system. Also, a framework for end-to-end HFS is developed. For fault detection and classification in photovoltaic array, Belaout et al. (2018) developed a Multiclass Adaptive Neuro-Fuzzy Classifier. In this paper, a novel classification system based on ANFIS is developed to improve the generalization performance of the Fuzzy Logic classifiers. Zhang et al. (2019) presented a fuzzy rule-based classification system combined by a multi-population quantum evolutionary algorithm with contradictory rule reconstruction. In this method, a fuzzy C-means clustering is used to heuristically produce representative initial rules. Also, a contradictory rule reconstruction is presented to modify misclassification rules.

One of the most popular approaches to identify and tune fuzzy inference system parameters is to find optimal or near-optimal parameters using evolutionary algorithms. This approach has many adherents and so many researches have been carried out in this field. For example, a learning fuzzy classification rule from labeled data was proposed by Roubos et al. (2003), and in this research genetic algorithm is used for parameter optimization. Also, the genetic algorithm is employed to tune fuzzy membership function parameters (Ilbeygi & Shah-Hosseini, 2012), which led to boosting the overall accuracy of the fuzzy emotion recognition system. In addition, multi-objective genetic algorithms to obtain fuzzy rule-based systems with a better trade-off between interpretability and accuracy in linguistic fuzzy modeling problems was proposed in (Alcalá et al. 2007). A Boosted genetic fuzzy classifier (BGFC) is another research that is used as a genetic algorithm to increase classification performance in land cover classification from multispectral images (Stavrakoudis et al., 2011). In this classifier, a genetic tuning stage is employed after the rule generation stage, aiming at improving the cooperation among the fuzzy rules, thus increasing the classification performance. An adaptive framework based on evolutionary computation and neural learning was proposed by Abraham (2002) and was named EvoNF. In EvoNF, the membership functions, rule base, and fuzzy operators are adapted according to the problem.

In addition, multi-objective genetic algorithms was proposed by Alcalá et al. (2007) in order to obtain Fuzzy Rule-Based Systems with a better trade-off between interpretability and accuracy in linguistic fuzzy modeling problems. A Boosted genetic fuzzy classifier (BGFC) is another research used as a genetic algorithm to increase classification performance in land cover classification from multispectral images (Stavrakoudis et al., 2011). In this classifier, a genetic tuning stage is employed after the rule generation stage, aiming at improving the cooperation among the fuzzy rules, thus increasing the classification performance. An adaptive framework based on evolutionary computation and neural learning was proposed by Abraham (2002) and was named EvoNF. In EvoNF, the membership functions, rule base, and fuzzy operators are adapted according to the problem.

Recently, many investigations have been conducted on feature and attribute selection for fuzzy inference systems using information theory. Also, the most important variables for a fuzzy rule-based system was discovered by computing the mutual information between the fuzzified variables by Sánchez et al. (2008). As mentioned before, many algorithms and tools with very diverse approaches have been proposed for fuzzy rule learning and fuzzy inference system parameters identification and tuning. But based on our knowledge, calculating the impact factor of linguistic variables (in the antecedent part of fuzzy rules) and using them for boosting the final fuzzy reasoning accuracy, has been neglected in previous works. So in this paper, we proposed a new approach for employing the impact factor of linguistic variables based on information theory concepts to promote the accuracy of fuzzy inference systems. Indeed we engaged the mutual information for calculating these impact factors and proposed a new formula for applying them in order to boost the final fuzzy reasoning.

The rest of this paper is organized as follows. In section 2, the basic idea for using mutual information for increasing the accuracy of fuzzy classification is explained. The new equation and method for employing mutual information in the fuzzy inference system are expressed in section 3. Section 4 deals with the implementation and results of employing the proposed method in six different fuzzy classification problems. Finally, conclusions are mentioned in Section 5.

2. Basic idea and Motivation

In a fuzzy reasoning system, final reasoning is done by applying a fuzzy inference system, like Mamdani-Type fuzzy inference (Mamdani & Assilian, 1975), on the fuzzy rule base. Each one of the fuzzy rules has some input parameters that form the antecedent part of this rule. Fuzzy inference system then applies some fuzzy operators on the antecedent part of the fuzzy rules for computing the firing rule strength and then clips the output membership function at the firing rule strength. Finally, the outputs of all of the fuzzy rules must be combined to obtain one fuzzy output distribution.

In the above process of the fuzzy inference system, there is no difference and preference between input parameters of the fuzzy rules. But in many cases, some input parameters play a more effective role in the final reasoning than other parameters. For example, in fuzzy emotion recognition system (Ilbeygi & Shah-Hosseini, 2012) such a rule was employed to recognize happy emotions:

IF (Mouth-Corner-Displacement is Very High) AND (Eye-Opening is Moderate) AND (Mouth-Opening is Very Low) AND (Eyebrow-Constriction is Very Low) THEN (Emotion is Very Happy). In this rule, Mouth-Corner-Displacement is the amount of displacement of mouth corners relative to the natural position of mouth corners with neutral emotion. So it is obvious that Mouth-Corner-Displacement has a very important role to determine the happy emotions from the perspective of human inference. However, in the original fuzzy inference systems, there is no difference between the above three input parameters and indeed, the level of importance of each of them has been neglected. In fact, this is our motivation to discover a method that considers the importance level and the impact factor of each input parameter in the final reasoning. So we decided to find a method for computing the impact factor of each input parameter of the antecedent part of fuzzy rules and then employing these impact factors to perform the final fuzzy reasoning.

To tackle this challenge, we consider classification problems and choose the mutual information for computing the impact factor of each input parameter relevant to a given output class. The mutual information is a measure of the amount of information that one random variable contains about another random variable leading to the reduction of the uncertainty (Cover & Thomas, 1991).

Based on the definition of the mutual information and by assuming that input parameters and classes (in a classification problem) are random variables, then we can use the mutual information for computing the impact factor of each input parameter in the classification process. Consider two random variables X and Y with a joint probability mass function $p(x, y)$ and marginal probability mass function $p(x)$ and $p(y)$. Mutual information $I(X; Y)$ is the relative entropy between the joint distribution and the product distribution $p(x)\, p(y)$ (Cover & Thomas, 1991). So the mutual information is computed using Eq. (1) shown below:

$$I\left(X;Y\right)=\sum_{x\in X}\sum_{y\in Y}p(x,y)\,\log\frac{p(x,y)}{p(x)\,p(y)} \tag{1}$$

In this paper, we consider the input parameters and the classes as the random variables. Because we compute the mutual information between each input parameter and each class, separately, so we use Eq. (2) for calculating the mutual information:

$$I(x_i;y_i)=p(x_i,y_i)\log\frac{p(x_i,y_i)}{p(x_i)\,p(y_i)} \tag{2}$$

In Eq. (2), x_i is one of the input parameters and y_i is one of the output classes. Since in this study the input parameters are linguistic variables, then the main question is: how can we calculate the mutual information between a linguistic variable and a given class?

To answer this question we should focus on the values of the linguistic variables. A linguistic variable is a variable whose values are linguistic terms in a natural or artificial language. For example, the size of an object is a linguistic variable, whose value can be small, medium, and big (Wu et al., 2011). So we can calculate the mutual information between each value of the linguistic variable and a given class. To clarify this process, an example is explained below.

Suppose in a fuzzy emotion recognition system we have six classes of emotions. The emotion is inferred by fuzzy rules that have four linguistic variables in the antecedent part of themselves. The following rule is an example of such a fuzzy rule:

IF (Mouth-Corner-Displacement is high) AND (Eye-Opening is Moderate) AND (Mouth-Opening is Low) AND (Eyebrow-Constriction is Low) THEN (Emotion is Happy)

Assume the Mouth-Corner-Displacement has five values that are very low, low, moderate, high, and very high, as a linguistic variable. So as mentioned before we can calculate the mutual information between each one of the five values and the happy emotions, separately. For example, for computing the mutual information between the high value of the Mouth-Corner-Displacement and the happy emotions, the high value of the Mouth-Corner-Displacement is x_i and the happy emotions is y_i in Eq. (2). So at first, we must calculate $p(x_i)$, $p(y_i)$ and $p(x_i\ ,y_i)$, and then replace these values in Eq. (2).

As previously mentioned, $p(x_i)$ and $p(y_i)$ are marginal probability mass functions and $p(x_i\ ,y_i)$ is joint probability mass function. So for computing *p*(*high value of Mouth-Corner-Displacement*), in our example, we must count all samples of training sets that have the high value for the Mouth-Corner-Displacement linguistic variable and then divide this number by counting of all the samples in the training set. Also, *p* (*happy emotions*) is calculated by counting all of the samples that expose the happy emotions and then dividing this count by counting all the samples in the training set. Finally, for computing *p*(*high value of Mouth-Corner-Displacement, happy emotions*), we must enumerate all of the samples that expose the happy emotions and have the high value for the Mouth-Corner-Displacement and then divide this number by counting all the samples that expose the happy emotions.

Now we can calculate the mutual information between the happy emotions and the Mouth-Corner-Displacement using Eq. (2). So based on the definition of the mutual information, we found a measure that shows the amount of information about a given class that the value of a linguistic variable contains. Therefore, we consider the mutual information as the impact factor of a value of the linguistic variable for inferencing a given class by the fuzzy inference system. In the following section, we will explain a new method and a new equation for employing these impact factors in the fuzzy reasoning process.

3. BOOSTING FUZZY INFERENCE SYSTEM USING MUTUAL INFORMATION

In section 2, we described a method to calculate the mutual information between a value of the linguistic variable and a given class. So in the classification process, we can take into account the obtained mutual information as the impact factor of a given value of a linguistic variable for recognizing a given class. But now, the main question is: how can we use these impact factors of input parameters for boosting the fuzzy classification accuracy?

To answer this question, we should concentrate on the inference method of the fuzzy inference systems. Takagi-Sugeno and Mamdani models are two famous and popular FIS (Takagi & Sugeno, 1985; Mamdani & Assilian, 1975). Each one of these FISs combines the antecedent part of fuzzy rules by different operators such as AND or OR computed with *t*-norms or *t*-conorms, respectively. For example, the most popular *t*-norm operator is the minimum (min) and the most popular *t*-conorm operator is the maximum (max). So when we combine linguistic variables in the antecedent part of fuzzy rules using each *t*-norm or *t*-conorm operator, it is intuitive that the final output of these operators will have greater value if each one of the antecedent part variables has greater value than their previous value.

Consequently, if we promote the value of each part of the antecedent part of fuzzy rules, corresponding to it is calculated impact factor, then the output value of the combination of antecedent part terms will be greater and more effective than before. Now we must respond to this question that how we can promote each part of the antecedent part of fuzzy rules corresponding to its calculated impact factor. Since the computed mutual information as the impact factor is a crisp value then we must propose a formula to fuse this value into the calculated belief value of each linguistic term in the fuzzy rule. Undoubtedly this formula should be able to increase the value of each part of the antecedent parts of fuzzy rules, corresponding to their impact factors. So we can use some aggregation operators for this aim. A comprehensive survey on aggregation operators can be found in the works of (Calvo et al., 2012; Vaníček et al., 2009; Detyniecki, 2001). For example, an aggregation operator that belongs to the fusion group of aggregation operators is shown in Eq. (3).

$$\pi' = (\pi * \alpha) + (1 - \alpha) \tag{3}$$

In Eq. (3), * stands for minimum, product or Łukasiewicz *t*-norm($max(0, \pi + \alpha - 1)$), π is a possibility distribution (for more information refer to Carlsson et al. (2003), and α is degree of certainty of a given source (Yager, 1981). So if we assume π is the grade of membership for a given value of a linguistic term and α is the calculated mutual information for this value of the linguistic term then we can use such an equation for promoting the value of π. But this equation does not satisfy our goals for improving fuzzy classification. For example, if π and α are zero then $\pi'$is be equal to 1! It means that Eq. (3) converts the value of π from zero to one in some situations. So this conversion leads to a wrong inference for fuzzy classification, and as a result, Eq. (3) is not suitable for our problem.

Therefore, we must propose a new aggregation operator that covers all of our goals for improving the fuzzy classification accuracy. After a comprehensive probe, we proposed a new fusion operator that is expressed in Eq. (4).

$$\pi' = \min\left[1, ((\pi) + k * (\alpha * (1 - \pi) * \pi))\right] \tag{4}$$

In Eq. (4), * stands for simple product and π, α, and k are real numbers. In our problem, we considered π as the grade of membership for a given value of a linguistic term and α as the calculated mutual information for this value of the linguistic term. Anyhow, each part of Eq. (4) plays an important role in the computation of π'. For example, α is promotion factor, (1-π) is the amount of possible promotion, π relativity factor (for adopting amount of promotion based-on π value), and k is impact factor for tuning the impact of the promotion. Some properties of Eq. (4) have been listed below:

1. If π and $\alpha \in [0,1]$then $\pi' \in [0, 1]$
2. If π=*1* then $\pi' = 1$
3. If then $\pi' = 0$ (It means that π has more important role in calculating π', than α)
4. If α=0 then $\pi' = \pi$
5. π' monotonic in α
6. $\pi' \geq \pi$

Figure 1 depicts three-dimensional plot of Eq. (4) for k=1.6 and corresponding to different values of π and $\alpha \in [0,1]$. According to Fig. 1, some properties of Eq. (4) such as monotonicity in α, $\pi' \geq \pi$, $0 \leq \pi' \leq 1$, and so on, are obvious.

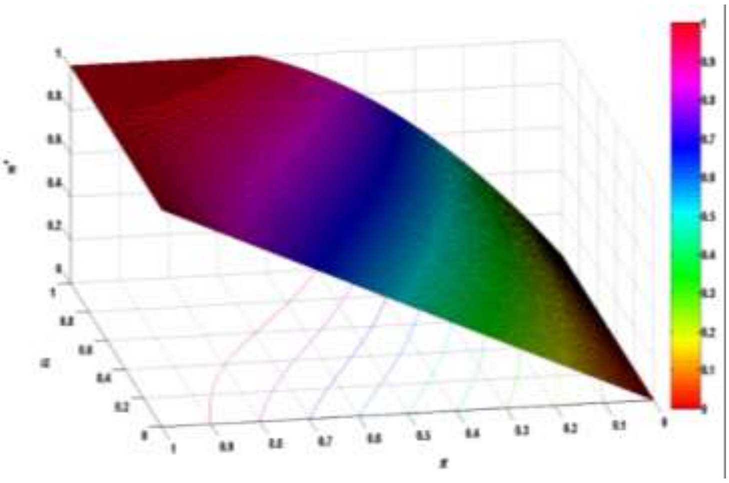

Fig. 1. Plotting the Eq. (4) with k=1.6 and based-on different values of π and $\alpha\epsilon[0,1]$

The point that should be noted here is the fact that the mutual information value between two random variables is not normalized and does not belong to [0,1], but in Eq. (4) we considered α as the mutual information where $\alpha\epsilon[0,1]$. So we must normalize the calculated mutual information values. Therefore, for normalizing the Mutual Information (MI) values we used feature scaling or unity-based normalization formula using Eq. (5).

$$\begin{cases} Normalized_{MI(i,j,k)} = \dfrac{MI(i,j,k) - Min(MI(i,j,:))}{Max(MI(i,j,:)) - Min(MI(i,j,:))}, & \text{Max(MI(i,j,:))} \neq \text{Min(MI(i,j,:))} \\ Normalized_{MI(i,j,k)} = 0, & \text{Max(MI(i,j,:))=Min(MI(i,j,:))} \end{cases} \tag{5}$$

In Eq. (5), $MI(i,j,k)$ is the calculated mutual information between i^{th} class and k^{th} value of j^{th} linguistic term $Min(MI(i,j,:))$ is the minimum of all the mutual information between i^{th} class and all values of j^{th} linguistic term (such as Low, Moderate, and High). So we can normalize each one of the values of mutual information using Eq. (5) and then engage this value in Eq. (4).

Note that, in case all the values of $MI(i,j,:)$ i.e. mutual information between i^{th} class and all values of j^{th} linguistic term are equal, then $Max(MI(i,j,:))$ and $Min(MI(i,j,:))$ are equal. Therefore, we consider the value of the normalized mutual information equal to zero to neutralize the effect of MI in the classification process. It is because, when all values of $MI(i,j,:)$ are equal then these MIs have no useful information for promoting discrimination in classifying input samples.

Now, all parts of the puzzle are ready and we can propose our algorithm for improving fuzzy inference system in the field of classification using the mutual information. The major steps of our algorithm are illustrated in Fig. 2.

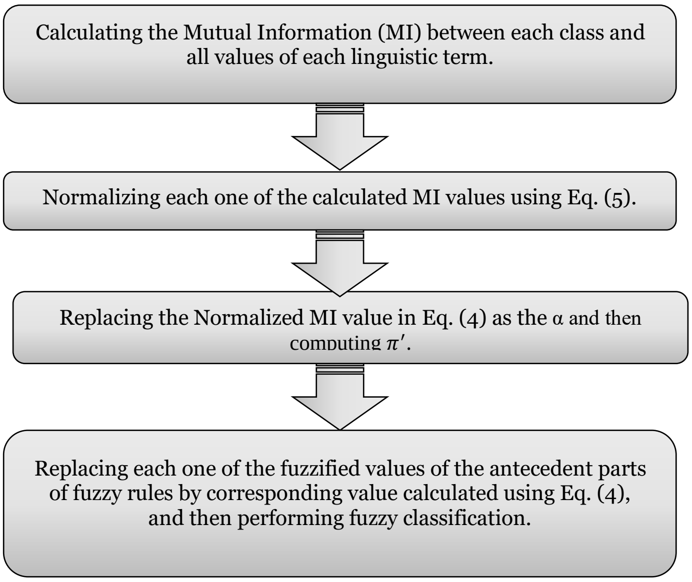

Fig. 2. Major steps of boosting algorithm for fuzzy inference system using the mutual information in the field of classification problems.

There is an intuitive proof for the correctness of our proposed algorithm, but for this proof, we must consider three important assumptions that have been listed below:

1. The selected features that form the antecedent part of all fuzzy rules, must not be dummy and irrelevant to the consequent part of fuzzy rules. Because these features are not only unhelpful to improving classification accuracy but they also can lead to wrong classification results.
2. A classification with moderate accuracy is possible by the current rule base of the fuzzy classification system. It means that we can improve the accuracy of a given fuzzy classification system if this system at least has the acceptable capability for classifying a given problem.
3. The training set must be good enough for calculating correct mutual information values. It means that the training set must contain good samples from all classes with appropriate diversity.

Considering the above assumptions, we have a fuzzy classification system that is able to perform classification with acceptable accuracy and this system has a well-defined rule base. So such a fuzzy classification system has the potential for improving final classification accuracy using our proposed algorithm. Assuming such a system, we tackle the proof of correctness of our proposed algorithm capability for boosting the mentioned system in the following paragraphs.

Whereas we basically use *t*-norm operators for combining each one of the items of the antecedent part of fuzzy rules and because the monotonicity is one of the basic properties of these operators, the firing strength of this rule will be greater than or equal to its previous value. It is because the monotonicity property ensures that the value of firing strength of the fuzzy rule does not decrease if the truth value of each part of the antecedent part is to increase. For more information about *t*-norm operators and their properties, one can refer to Klement et al.(2004).

So because we considered π as the grade of membership for a given value of a linguistic term and also we computed π' based on this value, and whereas $\pi' \geq \pi$ (in Eq. (4)), then we can increase truth value of each part of the antecedent part of fuzzy rules, corresponding to its calculated impact factor, using Eq. (4). Finally, this improvement can lead to promoting firing strength of fuzzy rules and because this promotion is based on reliable mutual information values and also taking into account the above triple assumptions, then the final classification results will be improved at least in terms of belief state of the inferred classes.

4. Implementation and Results

At first, we have chosen emotion recognition as a complex classification problem for proving the correctness of our proposed method. Up to now, many algorithms with different approaches have been proposed for tackling this problem. One of the recent algorithms that have been proposed by Ilbeygi & Shah-

Hosseini (2012), classifies displayed emotions in facial images to six different classes: Happy, Sad, Angry, Disgust, Fear, and Surprise. Also, RaFD Langner et al. (2010), facial images database was chosen by Ilbeygi & Shah-Hosseini (2012), to evaluate the performance and accuracy of the fuzzy emotion recognition system. In Addition, 573 fuzzy rules have been defined for emotion recognition as a knowledgebase. Finally, Ilbeygi & Shah-Hosseini's (2012) experimental results report an average precision rate of 93.96% for emotion recognition of six basic emotions.

We selected this classification problem because of its complexity and intricacy. Eventually, we performed the steps of our proposed algorithm (that have been shown in Fig. (2) on this fuzzy emotion recognition system. Experimental results obtained by applying our algorithm to the same facial image set, as the image set selected by Ilbeygi & Shah-Hosseini (2012), indicate that we could increase the average precision rate for emotion recognition from 93.96% to 94.68%. Table 1 represents a promotion for recognition rate for Happy, Sad and Disgust emotions using our proposed algorithm, separately.

One of the most important features of our proposed algorithm is to boost the firing strength of rules that lead to selecting a class for a given sample with greater confidence. So we calculated the mean value of recognized emotion beliefs by our algorithm in six mentioned emotions and by examination on 69 facial images for each emotion, separately. Then, we compared these results with the mean value of the recognized emotion beliefs that have been calculated by Ilbeygi & Shah-Hosseini (2012). As shown in Table 2, our algorithm's outputs show promotion in the mean value of the recognized emotion beliefs for every one of the six basic emotions.

Table 1. Comparing the recognition rate for *Happy*, *Sad* and *Disgust* emotion

Method	Happy	Sad	Disgust
Ilbeygi & Shah-Hosseini (2012)	98.55%	92.75%	91.30%
Our method	100%	94.20%	92.75%

Table 2. Comparison between the mean values of the recognized emotion beliefs that inferred by our method and the ones that have been calculated by Ilbeygi & Shah-Hosseini (2012).

Method	Happy	Sad	Fear	Surprise	Angry	Disgust
Ilbeygi & Shah-Hosseini	0.6879	0.7902	0.748	0.7943	0.7943	0.6731
Our method	0.8863	0.9071	0.891	0.9289	0.9289	0.8821

In addition, to test our proposed algorithm more precisely, we decided to examine our method for boosting fuzzy classification in the other five different classification problems. So we selected five classification problems whose datasets have been presented in the UCI Machine Learning Repository (Lichman, 2013). These datasets are Iris, Balance-Scale, Haberman, User Knowledge Modeling, and Banknote Authentication. Then we must have implemented a fuzzy classifier for each problem. It means that, first of all, we must have generated a rule base for classifying each problem and next, we must have selected type and count of fuzzy membership functions for each dataset sample attribute for the purpose of fuzzification and finally we must have defuzzified the inferred class.

So, we must have generated a fuzzy rule-based system for classifying each problem at first. For reaching this aim, we have engaged FuzzyWeka machine learning tool (Hall et al., 2009). Using FuzzyWeka we could generate a fuzzy rule base for the classification of each one of the mentioned problems. Note that, only for classification of Iris dataset we used the fuzzy rule set proposed by Castro et al., 1999; Chen et al. (2006). Afterward, we implemented one fuzzy classification application for each problem in MATLAB programming environment, using generated fuzzy rules. Finally, we classified each one of the five selected problems using corresponding classification application and reached to appropriate classification precision.

Now everything is ready to apply our proposed algorithm for boosting fuzzy classification. So we performed all steps of our algorithm, presented in Fig. 2, on each one of the above problems and then computed classification precision. The experiment results show the promotion of classification rate for every one of the mentioned classification problems. Table 3 made the comparison between classification precisions of our algorithm and fuzzy classification based on fuzzy rules that have been generated by FuzzyWeka or have been presented by Ilbeygi & Shah-Hosseini (2012) and Castro et al. (1999), which we call other methods/algorithms.

Table 3. Comparison between classification precisions of our proposed algorithm and other classical fuzzy classification methods.

Dataset name	Number of classes	Number of attributes	Number of membership functions	Number of rules	*K* value	α-cut value	Other methods precisions (%)	Our method precision (%)
Emotion recognition	6	4	5	573	1.6	0.1	93.96	94.68
Iris	3	4	7	72	1.6	0.3	85.71	90.48
Balance-Scale	3	4	3	68	1	0.4	86.78	90.02
Haberman	2	3	3	18	1.8	0.2	73.47	75.00
User knowledge modeling	4	5	3	95	1.4	0.2	71.05	75.56
Banknote authentication	2	5	3	44	1.1	0.3	79.18	87.53

Note that k value in Table 3 signals the impact factor (k) in Eq. (4). We must be careful with selecting the value for k because the high value of k leads to an increase in classification error rate. So we select k value experimentally, but we can employ evolutionary algorithms to find optimum value for k to maximize classification precision and the average of inferred class beliefs. Also α-cut value is a threshold for considering an inferred class as a correct inferred class if the inferred class belief is greater than or equal to this threshold. This value is selected experimentally as well as k value, to better show the boosting capability of our proposed algorithm for classification.

Finally, we computed the average of all the inferred class beliefs (that have been made by other methods and our proposed boosting algorithm) for all problems and found our proposed algorithm could improve the inferred class beliefs, significantly. An overall comparison between mean values of the inferred class beliefs by other methods and our proposed algorithm is depicted in Fig. 3. Note that, the inferred class belief is a value that belongs to [0,1] interval.

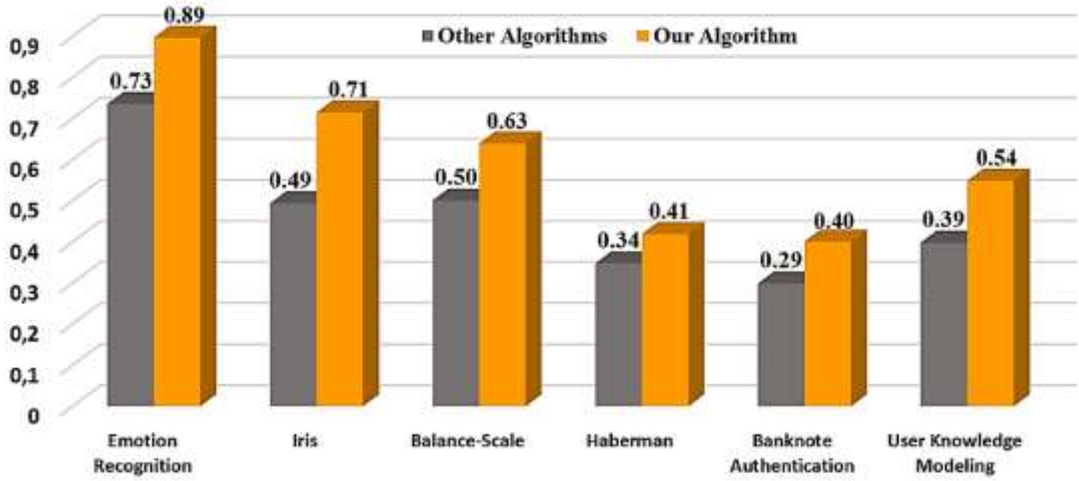

Fig. 3. Comparison between mean values of the inferred class beliefs by other methods and our proposed algorithm for all six classification problems.

5. Conclusions and Future Works

In this paper, we proposed a new boosting algorithm that is able to perform on every fuzzy classification problem. The basic idea of our algorithm is to promote the value of each part of the antecedent parts of fuzzy rules using Mutual Information (MI). Indeed, we calculate the mutual information between each class and all values of linguistic terms and then normalize them. Afterward, we replace the calculated mutual information and the fuzzified values of sample features in a new aggregation formula. Finally, we used the calculated value in the proposed aggregation operator, instead of the fuzzified values of the features of the samples in the antecedent part of the fuzzy rules.

One of the most notable features of our proposed algorithm is to boost firing strength of fuzzy rules that lead to selecting a class for a given sample with greater confidence and finally can lead to boosting classification accuracy. For proving the correctness and effectiveness of our proposed algorithm in the field of fuzzy classification, we examined our algorithm in six different classification problems. The implementation results showed that we can improve the classification accuracy of these problems using our proposed algorithm. Also, these results proved that we can use the proposed algorithm in the field of fuzzy classification problems whose fuzzy rule-base is defined manually (based on human experiences) or is generated automatically by learning from experimental data (like using FuzzyWeka for fuzzy rule-base generation).

For the future, we decided to investigate other fusion operators and to make a comparison between them and our proposed operator. Also, we can learn fuzzy rules for classification using ANFIS and then apply our algorithm on these learned fuzzy rules. Finally, we will be able to generate a new classifier based on ANFIS and the mutual information and name it as ANFIS+. This classifier can train an ANFIS and also calculate the mutual information automatically. Finally, the trained classifier can perform classification by the proposed algorithm. So we only need to feed training data to ANFIS+ and set learning parameters. Subsequently, learning fuzzy rules, tuning membership functions, calculating and normalizing the mutual information, and finally classifying samples using this FIS according to our proposed algorithm, will be performed automatically.

REFERENCES

Abraham, A. (2002). EvoNF: A framework for optimization of fuzzy inference systems using neural network learning and evolutionary computation. *The 17th IEEE International Symposium on Intelligent Control, ISIC'02, IEEE Press*, Vancouver, Canada.

Abraham, A. (2005). *Adaptation of Fuzzy Inference System Using Neural Learning*. In Fuzzy systems engineering, Springer, Berlin, Heidelberg, 53-83.

Alcalá, R., Gacto, M.j., Herrera, F., & Alcalá-Fdez, j. (2007). A multi-objective genetic algorithm for tuning and rule selection to obtain accurate and compact linguistic fuzzy rule-based systems. *World Scientific, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 15(5), 539-557.

Alpaydin, E. (2014). *Introduction to Machine Learning*. London: MIT Press.

Belaout, A., Krim, F., Mellit, A., Talbi, B., & Arabi, A. (2018). Multiclass adaptive neuro-fuzzy classifier and feature selection techniques for photovoltaic array fault detection and classification. *Renewable Energy*, 127, 548-558.

Calvo, T., Mayor, G., & Mesiar. R. (2012). *Aggregation Operators: New Trends and Applications*. Amsterdam: Physica.

Carlsson, C., Fullér, R.,& Majlender, P. (2003). Possibility distributions, a normative view. *Proceedings of Slovakian-Hungarian Joint Symposium on Applied Machine*.

Castro, J.L., Castro-Schez, J.J., & Zurita, J.M. (1999). Learning maximal structure rules in fuzzy logic for knowledge acquisition in expert systems. *Fuzzy Sets and Systems*, 101(3), 331-345.

Cover, T. M., & Thomas, J., A. (1991). Entropy, relative entropy and mutual information. John Wiley & Sons, Inc. *Hoboken, NJ, Elements of Information Theory*.

Detyniecki, M. (2001). Fundamentals on aggregation operators. Ph.D. thesis. *Berkeley University*.

Feng, J.C., & Teng, L.C. (1998). An online self constructing neural fuzzy inference network and its applications. *IEEE Transactions on Fuzzy Systems*, 6(1), 12-33.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, p., & Witten, I. H. (2009). The WEKA data mining software: An update. *ACM SIGKDD Explorations*, 11(1), 10-18.

Ilbeygi, M., & Shah-Hosseini, H. (2012). A novel fuzzy facial expression recognition system based on facial feature extraction from color face images. *Engineering Applications of Artificial Intelligence*, 25(1), 130-146.

Jang, J.S. (1993). ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems. Man and Cybernetics*, 23(3), 665-685.

Klement, E. P., Mesiar, R., & Pap, E. (2004).Triangular norms. Position paper I: basic analytical and algebraic properties. *Fuzzy Sets and System*, 143(1), 5-26.

Langner, O., Dotsch, R.G., Bijlstra, G., Wigboldus, D.H.J., Hawk, S.T., & Van Knippen-berg, A.D. (2010). Presentation and validation of the Radboud Faces Database. *Cognition and Emotion*, 24(8), 1377–1388.

Lichman, M., (2013). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California.

Mamdani, E. H., & Assilian, S. (1975). An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal Man-Machine Studies*, 7(1). 1-13.

Mutlu, B., Sezer, E. A., & Akcayol, M. A. (2018). End-to-End hierarchical fuzzy inference solution. *In 2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, Rio de Janeiro, Brazil.

Nomura, H., Hayashi, I., & Wakami, N. (1992). A Learning Method of Fuzzy Inference Systems by Descent Method. *In Proceedings of the First IEEE International Conference on Fuzzy Systems*, San Diego, US.

Roubos, J. A., Setnes, M., & Abonyi, J. (2003). Learning fuzzy classification rules from labeled data. *Information Sciences*, 150(1), 77-93.

Sánchez, L., Suárez, M., Villar, J.R., & Couso, I. (2008). Mutual information-based feature selection and partition design in fuzzy rule-based classifiers from vague data. *International Journal of Approximate Reasoning*, 49(3), 607-622.

Sarkar, M., & Yegnanarayana, B. (1997). Incorporation of fuzzy classification properties into backpropagation learning algorithm. *IEEE International Conference on Fuzzy Systems*, Barcelona. Spain.

Shihabudheen, K. V., Mahesh, M., & Pillai, G. N. (2018). Particle swarm optimization based extreme learning neuro-fuzzy system for regression and classification. *Expert Systems with Applications*, *92*, 474-484.

Stavrakoudis, D.G., Theocharis, J.B., & Zalidis, G.C., (2011). A boosted genetic fuzzy classifier for land cover classification of remote sensing imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(4), 529-544.

Subramanian, K., Suresh, S. & Sundararajan, N. (2013). A metacognitive neuro-fuzzy inference system (McFIS) for sequential classification problems. *IEEE Transactions on Fuzzy Systems*, 21(6), 1080-1095.

Takagi, T., & Sugeno, M. (1985). Fuzzy identification of systems and its applications to modelling and control. *IEEE Transactions on Systems, Man, and Cybernetics*, 15(1), 116-132.

Tano, S., Oyama, T., & Arnould, T. (1996). Deep combination of fuzzy inference and neural network in fuzzy inference. *Fuzzy Sets and Systems*, 82(2), 151-160.

Vaníček, J., Vrana, I. & Aly, S. (2009). Fuzzy aggregation and averaging for group decision making: A generalization and survey. *Knowledge-Based Systems*, 22(1), 79-84.

Werro, N. (2015). *Fuzzy classification of online customers* (Volume148). Heidelberg: Springer.

Yager, R.R. (1981). A new methodology for ordinal multiple aspect decisions based on fuzzy sets. *Decision Sciences*, 12(4), 589-600.

Wu, Y., Zhang, B., Lu, J., & Du, K.L. (2011). Fuzzy logic and neuro-fuzzy systems: A systematic introduction. *International Journal of Artificial Intelligence and Expert Systems*, 2(2), 47-80.

Chen, Y.S., Wang, L.H., & Chen, S.M. (2006). Generating weighted fuzzy rules from training data for dealing with the Iris data classification problem. *International Journal of Applied Science and Engineering*, 4(1), 41-52.

Zadeh, L.A. (1965). Fuzzy sets. *World Scientific, Information and Control*, 8(3), 338-353.

Zhang, Y., Qian, X., Wang, J., & Gendeel, M. (2019). Fuzzy rule-based classification system using multi-population quantum evolutionary algorithm with contradictory rule reconstruction. *Applied Intelligence*, 49(11), 4007-4021.