

A semantic image classifier based on hierarchical fuzzy association rule mining

Abolfazl Tazaree · Amir-Masud Eftekhari-Moghadam ·
Saeedeh Sajjadi-Ghaem-Maghami

Published online: 22 June 2012
© Springer Science+Business Media, LLC 2012

Abstract One of the major challenges in the content-based information retrieval and machine learning techniques is to-build-the-so-called “semantic classifier” which is able to effectively and efficiently classify semantic concepts in a large database. This paper dealt with semantic image classification based on hierarchical Fuzzy Association Rules (FARs) mining in the image database. Intuitively, an association rule is a unique and significant combination of image features and a semantic concept, which determines the degree of correlation between features and concept. The main idea behind this approach is that any image visual concept has some associated features, so that, there are strong correlations between the concepts and their corresponding features. Regardless of the semantic gap, an image concept appears when the corresponding features emerge in an image and vice versa. Specially, this paper’s contribution was to propose a novel Fuzzy Association Rule for improving traditional association rules. Moreover, it was concerned with establishing a hierarchical fuzzy rule base in the training phase and setup corresponding fuzzy inference engine in order to classify images in the testing phase. The presented approach was independent from image segmentation and can be applied on multi-label images. Experimental results on a database of 6000 general-purpose images demonstrated the superiority of the proposed algorithm.

Keywords Semantic image classifier · Hierarchical fuzzy classification · Fuzzy Association Rule · Fuzzy expert systems

1 Introduction

Rapid growth of multimedia information leads to the appearance of content management as a strategic requirement. Content based information retrieval is one approach for content

A. Tazaree · A.-M. Eftekhari-Moghadam · S. Sajjadi-Ghaem-Maghami (✉)
Faculty of Electrical, Computer and IT Engineering, Islamic Azad University, Qazvin Branch, Qazvin, Iran
e-mail: s.ghaemmaghami@qiau.ac.ir

A. Tazaree
e-mail: tazaree@gmail.com

A.-M. Eftekhari-Moghadam
e-mail: eftekhari@qiau.ac.ir

management which contains a set of techniques for classifying and retrieving information [9]. The focus of the paper was on image classification as an important field in information retrieval and machine vision extensive research efforts have been devoted in the field of image classification in the recent years [10, 11, 13]. Most famous techniques in image classification include support vector machine (SVM), Decision Tree, Naïve Bayes, etc. Association rule based classification is a new approach in image classification [1, 3, 5, 19]. An association rule is a combination of low-level visual features and semantic concept which are unique and significant for describing the semantic concept. In [20], association classification was proven to have higher classification accuracy than other approaches. The main goal of image classification using association rules is to find a set of rules in order to identify the classes of undetermined image feature vectors. Generally, designing an image classifier based on association rules includes the following steps: (a) extracting image features, (b) mining frequently repetitive pairs of features and concepts to generate association rules, and (c) selecting rules with the best evaluation measures. In order to enhance the predictability of the association rules, different papers have been published which presented classification systems based on Fuzzy Association Rules (FAR)[1, 5]. The FAR can be considered to be a classification rule if the antecedent part contains fuzzy feature sets and the consequent part contains a class label [1]. The FAR is an implication, 'If X is A then Y is B', to deal with quantitative attributes. X and Y are sets of attributes and A, B are fuzzy sets which describe X and Y, respectively [17]. Chan Man Kuok et al. proposed an algorithm which used a fuzzy technique for mining association rules and the algorithm solved the problem of sharp boundary [8]. Jing et al. used a cluster technique to extract fuzzy association rules and proposed CBFAR algorithm so that this approach outperformed a known Apriori-based fuzzy association rules mining algorithm [6]. In 2008, P. Santhi Thilagam et al. proposed an algorithm to extract and optimize fuzzy association rules using a multi-objective genetic algorithm. This method was efficient in many scenarios [16].

Class hierarchies are a common way for reducing the complexity of the classification problem, especially when dealing with a large number of classes because some classes are more closely related than others [12]. While flat classification techniques might reach the minimum performance in dealing with a large number of classes, the hierarchical classification techniques are able to overcome the problem by training in a hierarchical structure in a stepwise manner [4, 12]. Given a hierarchy of classes, standard machine learning approaches may find it harder to discern similar classes than the classes that are unrelated according to the classification system. Thus, it would be beneficial to apply a recursive top-down approach to hierarchical classification: first, to discriminate the subsets of classes at the top level of the hierarchy and then to recursively separate the classes (or sets of classes) in those subsets [21]. Figure 1 illustrates an example of hierarchical image classification. The problem consists of five classes including *sea*, *waterfall*, *sunset*, *mountain* and *snow*. The given hierarchy depicts some classes as strongly related to each other. For instance, the *Sea* class is related to *Waterfall* class but not to *Sunset*. So, in the first step, images are classified in two classes: *Waterscape* and *Mainland*. Then, in the second step, the *Waterscape* class is divided into two sub-classes. Furthermore, the *Mainland* class is divided to three sub-classes. Zimek [21] investigated why predictive performance can be improved by exploiting the class hierarchy. Recently, several systems have successfully applied hierarchical classification on large databases [4, 12, 18, 21].

1.1 Paper approach

Many approaches have proposed association rule mining based on image content [19]. In this paper, the major contribution was to present a novel Fuzzy Association Rule (FAR) for

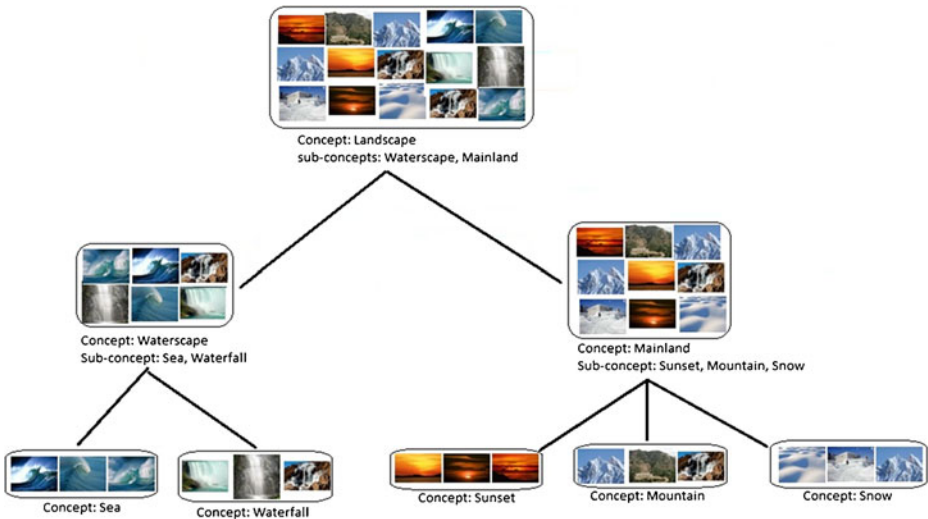


Fig. 1 An example of hierarchical image classification

improving traditional association rules. Intuitively, an association rule is a unique and significant combination of image features and a semantic concept which determines the degree of correlation between features and concepts. The main idea behind this approach is that any image visual concept has some associated features and there are strong correlations between the image concept and its corresponding features. Regardless of the semantic gap, an image concept appears when the corresponding features emerge in an image and vice versa. FAR was used to build a semantic classifier for demonstrating the effectiveness of the proposed approach of FAR mining. In order to build the semantic classifier, a fuzzy rule base was established in the training phase, which contained the best FAR with the highest evaluation measure. Classification was performed in the testing phase using fuzzy inference engines. The experimental results showed the significant improvement of FAR against traditional Association Rules.

Moreover, the minor contribution was a handmade hierarchical database structure IN ORDER to reach better performance in image classification. The fuzzy rule base and fuzzy inference engines were organized according to the database hierarchical structure. All hierarchical structures (database, fuzzy rule base and fuzzy inference engines) were handmade and no algorithm was used for their set up. The test bed of this study was a part of COREL image database. No prior image segmentation was required for feature extraction.

1.2 Outline of the paper

The rest of the paper is organized as follows: in Section 2, the system framework of the proposed approach is presented. Then, feature extraction and fuzzification of low-level visual features were described in the next section. The innovation of this paper is given in Sections 4 and 5. Fundamental concepts of FAR mining are presented in Section 4 and the fuzzy rule base is introduced in Section 5. Hierarchical fuzzy classification is described in the next section. The experimental results were presented and the advantages of this approach over other state-of-the-art papers were demonstrated in Section 7; concluding remarks were given in Section 8.

2 The system framework

Figure 2 demonstrates the framework of the presented approach in two phases: training phase and testing phase. In the former, after feature extraction, low level features were extracted. Fuzzification of feature vectors was done; then, FAR was mined for each concept in all layers and a fuzzy rule base was made up. In the latter, for a new image, a semantic image classifier was built using a hierarchical fuzzy inference system after feature extraction.

Association rules determine the degree of correlation between a feature and a semantic concept [10]. This paper presented a novel FAR that led to more effectiveness than non-fuzzy association rules. Figure 3 shows the value of a feature versus 1000 images with different concepts. Clearly, the feature value for the image subset of 500-600 was very different from other images. If the image subset of 500-600 included the concept C_k , then there was a very strong correlation between feature F_i and concept C_k which could be presented by $F_i \rightarrow C_k$. The features like F_i were looked for in order to establish a fuzzy rule base. However, if no such predictive features could be found because of its challenges, building a fuzzy rule base would be continued by fewer predictive features. Evaluation measures of FARs determined the productivity power of a rule which was investigated in more details later.

3 Feature extraction and fuzzification

In this paper, the feature extraction was performed based on feature availability in the training phase [2] and feature correlation in the testing phase. Feature extraction was performed in HSV and La*b* color space. In fact, each descriptor was extracted from six color matrixes: H, S, V, L, a* and b*. Seven types of descriptors were extracted from each

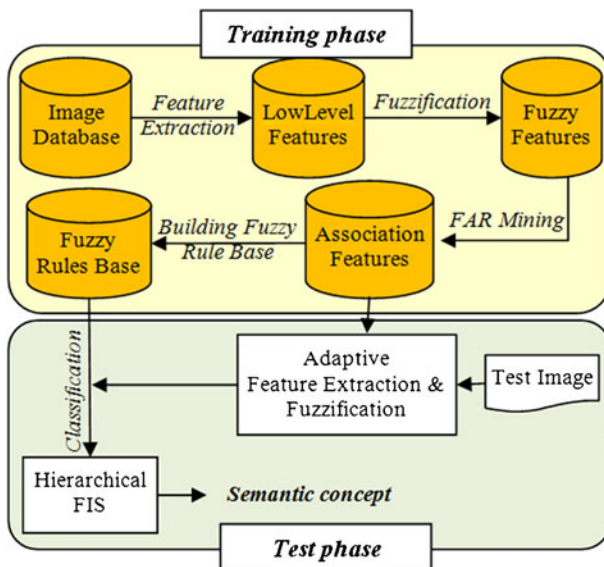


Fig. 2 System framework

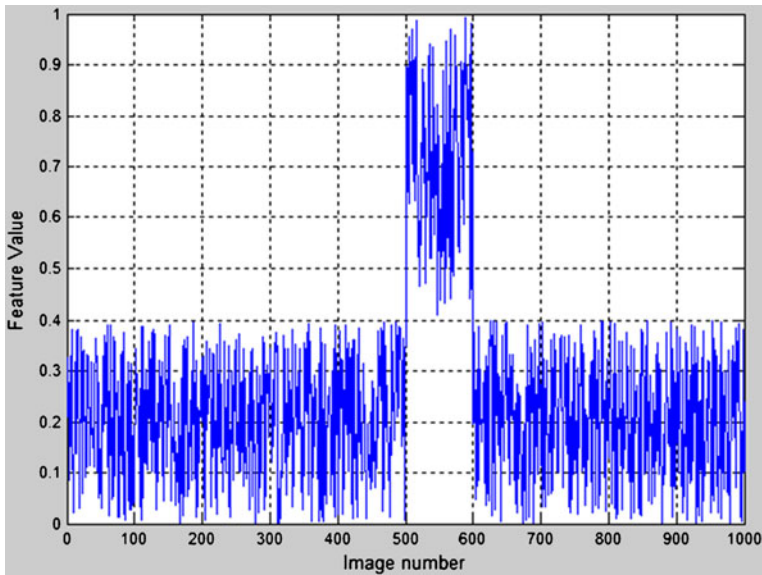
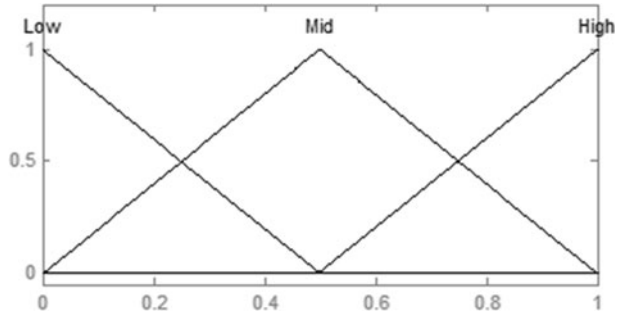


Fig. 3 Images by the number of 500-600 have high correlation with the feature

color matrix: two types of color feature, four types of texture feature and one type of shape feature. The color moments and fuzzy color histogram [7] were color descriptors of feature vector, the former of which took six places (the fifth moment until the tenth moment) and the latter took ten places in the feature vector. Generally the fuzzy features lead to better results [14]. The fuzzy color histogram extracted ten colors from the image: red, blue, green, yellow, brown, cyan, magenta, white, dark and black. Thus, 16 color descriptors were extracted from each color matrix. Four types of exploited texture descriptors were contrast, correlation, energy and homogeneity. Co-occurrence matrix, which is necessary for computing texture descriptors, were extracted in four directions of 0, 45, 90 and 135 degrees by two distances of 5 and 10 pixels; thereby, eight co-occurrence matrixes were obtained from each color matrix. In this way, 32 texture descriptors were extracted from each color matrix because four texture descriptors were extracted from each co-occurrence matrix. Edge detector descriptor was used to extract shape features. Lines in four directions (0, 45, 90 and 135) were extracted using sobel filter. Lines with less than 10 pixels and more than 100 pixels were omitted. The sum length of horizontal lines was one descriptor and the sum length of vertical lines was other descriptor; also two other descriptor values were computed in this way. So, four places of feature vector were dedicated for shape features. After performing feature extraction in the spatial domain, the color matrixes were transformed to five frequency domains (*Fourier*, *wavelet*, *Radon*, *Z* and *cos* domain). Then, the mentioned descriptors were extracted again.

The features with strong correlation with concepts were searched for since the features with such property were not known; thus, many features was inevitably extracted in the training phase and the value of correlation of each feature was computed; finally, the ones with the most correlation were selected in order to build the rule base. It appeared that, in the testing phase, just the most correlated features, which were used to build the rule base, were extracted. Normalization and fuzzification of feature vector were done after feature extraction. Figure 4 shows the triangular membership functions used in the fuzzification. The

Fig. 4 Fuzzy sets for fuzzification



original feature values were replaced with the relative degree of features in all MFs (Low, Mid, High); thus, fuzzy feature vector tripled the dimension of the original feature vector. Note that, dimension of fuzzy feature vector decreased after FAR mining. Table 1 gives an example of original feature value consisting of three low level features, named F1, F2 and F3, which were extracted from 10 images including three concepts of Beach, Sky and Grass. Table 2 represents the fuzzy feature vectors based on the value of Table 1. For more explanation, the original value of the feature F1 of the first image was 0.1 and the relative degrees of 0.1 to Low, Mid, and High membership functions were 0.8, 0.2 and 0 respectively ($\mu_{Low}(F_1) = 0.8$, $\mu_{Mid}(F_1) = 0.2$ and $\mu_{High}(F_1) = 0$). Hence, the original feature value (0.1) was replaced with three values in the fuzzy feature vector (0.8, 0.2, and 0). This process was done for all the low level original feature values.

4 Novel fuzzy association rule

As mentioned before, association rules address the value of correlation between a feature and a concept [5]. In this section, the paper contribution of FAR mining was presented. The presented FAR was more predictive than non-fuzzy Association Rule. For the beginning, given the sample set $U = \{S_1, \dots, S_n\}$ and the semantic concepts $C = \{c_1, \dots, c_k\}$, a concept in layer l was shown as C_i^l . An image is demonstrated as $I = (c, F)$ where c is the image concept and F is the feature vector. For each concept $c \in C$, the image set can be split in two categories: positive samples (S_c^+) and negative samples (S_c^-). The former contained samples with the concept c and the latter contained samples without concept c . In this paper, rules were shown as:

$$f \text{ is } MF_i^l \rightarrow \text{concept is } C_k^l \text{ with } W_j \text{ confidence.} \tag{1}$$

Table 3 describes these variables. As follows, the primary concepts of paper contribution are presented.

FuzzyCount of a rule in the image set D , specified the total degree of membership of the feature f to the MF_i^l in images containing the concept C_k^l .

$$FuzzyCount_D(f \text{ is } MF \rightarrow C_k^l) = \sum_{f \in (S_c^+ \subseteq D)} \mu_{MF}(f_i) \tag{2}$$

where the term f_i refers to a feature of i^{th} image sample S_c^+ in D .

Table 1 Original feature vectors

Image ID	Feature Vectors			Concept
	F1	F2	F3	
1	0.1	0.1	0.8	Beach
2	0.15	0.5	0.95	Beach
3	0.2	0.3	0.9	Beach
4	0.9	0.8	0.6	Sky
5	0.75	0.6	1	Sky
6	0.55	0.9	0.55	Sky
7	0.7	0.15	0.2	Grass
8	0.5	0.1	0.5	Grass
9	0.95	0.2	0.15	Grass
10	0.8	0.5	0.1	Grass

- *FuzzySupport* of a rule in the image set D addresses fuzzy percentage of images including both the feature f_i and concept C_k^l .

$$FuzzySupport_D(f \text{ is } MF \rightarrow c) = \frac{FuzzyCount_D(f \text{ is } MF \rightarrow c)}{|D|} = \frac{\sum_{f \in (S_c^+ \subseteq D)} \mu_{MF}(f_i)}{|D|} \tag{3}$$

where $|D|$ is the number of images in D .

For example, *FuzzySupport* of a rule, $f \text{ is } MF_i^l \rightarrow C_k^l$, equaled 0.9, meaning that 90 % of images had both the feature f and concept C_k^l .

- *FuzzyConfidence* of a rule addressed the fuzzy percentage of images containing both the feature f and concept C_k^l toward the total of images containing the feature f . In fact, it calculated the probability of finding concept c in the image set.

$$FuzzyConfidence_D(f \text{ is } MF \rightarrow c) = \frac{FuzzySupport_D(f \text{ is } MF \rightarrow c)}{FuzzySupport_{\{D' | D' \subseteq D, f \in D'\}}(f \text{ is } MF \rightarrow c)} \\ = \frac{\sum_{f \in (S_c^+ \subseteq D)} \mu_{MF}(f_i)}{\sum_{f \in D'} \mu_{MF}(f_i)} = \frac{\sum_{f \in (S_c^+ \subseteq D)} \mu_{MF}(f_i)}{\sum_{f \in D'} \mu_{MF}(f_i)} \tag{4}$$

FuzzyConfidence presented the implication of a rule and addresses the accuracy of its prediction. A given rule was called *confident* if the value of *FuzzyConfidence* exceeds a minimum threshold. For example, supposing two image sets, one with 1000 images and the other with 100 images, both would have five images with the feature f and concept c . The confidence of the given rule, $f \text{ is } MF_i^l \rightarrow C_k^l$, was 1 for both data sets; so, it can classify five images of these classes. It should be noted that five images formed 5 % of the first class while forming 0.5 % of the second class. Although the confident rules had enough predictive power to build the semantic classifier, to reach more predictive power, the stronger rules had to be extracted.

Table 2 Fuzzy feature vectors based on Table 1

Image ID	Fuzzy(F1)			Fuzzy(F2)			Fuzzy(F3)			Concept
	$\mu_{Low}(f_1)$	$\mu_{Mid}(f_1)$	$\mu_{High}(f_1)$	$\mu_{Low}(f_2)$	$\mu_{Mid}(f_2)$	$\mu_{High}(f_2)$	$\mu_{Low}(f_3)$	$\mu_{Mid}(f_3)$	$\mu_{High}(f_3)$	
1	0.8	0.2	0	0.8	0.2	0	0	0.4	0.6	Beach
2	0.7	0.3	0	0	1	0	0	0.1	0.9	Beach
3	0.6	0.4	0	0.4	0.6	0	0	0.2	0.8	Beach
4	0	0.2	0.8	0	0.4	0.6	0	0.8	0.2	Sky
5	0	0.5	0.5	0	0.8	0.2	0.2	0.8	0	Sky
6	0	0.9	0.1	0	0.2	0.8	0	0.9	0.1	Sky
7	0	0.6	0.4	0.7	0.3	0	0.6	0.4	0	Grass
8	0	1	0	0.8	0.2	0	0	1	0	Grass
9	0	0.1	0.9	0.6	0.4	0	0.7	0.3	0	Grass
10	0	0.4	0.6	0	1	0	0.8	0.2	0	Grass

Table 3 Describing the variables of an FAR (f is $MF_i^l \rightarrow C_k^l$) in the rule base

Variable	Description
C_k^l	k^{th} output fuzzy set in l^{th} layer
W_j	weight of rule in the rule base
f	The original feature
MF_i^l	i^{th} input fuzzy set in layer l

- *FuzzyLift* calculates the prediction power of a rule. It is the ratio of *FuzzyConfidence* to Expected Confidence, which is equal to the distribution of concept c in image set.

$$FuzzyLift_D(f \text{ is } MF \rightarrow c) = \frac{FuzzyConfidence_D(f \text{ is } MF \rightarrow c)}{|S_c^+ \subseteq D|} \tag{5}$$

FuzzyLift addressed the importance of a rule. If the value of *FuzzyLift* was greater than one, then it would indicate that there was an association between the image feature and image concept. On the other hand, if the value of *FuzzyLift* was less than one, it would indicate a negative relationship between the feature and concept. If the value of *FuzzyLift* was equal to 1, then the rule had no effect [17].

For example, 20 percent of images in a dataset can be supposed to contain the concept c . If an image was selected randomly by the probability of 20 percent, the image concept would be c (confidence of 0.2). In cases resembling the one explained, if the confidence was 0.1, the probability of selection would be weaker than the random selection. Generally, the rule confidence must be more than the concept distribution in the dataset.

Table 4 shows all FARs and their objective measures (*FuzzySupport*, *FuzzyConfidence* and *FuzzyLift*) which were extracted from Table 2 based on the proposed algorithm. For further explanation, in the first row, $f1 \text{ is } Low \rightarrow beach$ expresses that, if the feature $f1$ was a member of the low fuzzy set, then the image concept would be the *beach*. The rule confidence was one because the value of $\mu_{Low}(f_1)$ was zero for non-beach concepts. Therefore, $\mu_{Low}(f_1)$ did not emerge for any non-beach concepts. So, if the concept of input image was the beach, the $f1$ feature was sufficient for its classification. Some rules such as 2, 3 had weak *FuzzySupport* and *FuzzyConfidence*, so they were not used in the Fuzzy Rule base. There were several methods for developing association rule mining. One of the famous methods called Apriori, which has been applied in a wide range of tasks [5, 8], was used here to extract association rules.

5 Building fuzzy rule base

Obviously, prediction power of one FAR is not enough to classify large image databases. So, a number of the most qualitative FARs should be exploited for building a semantic classifier. The prediction power of a rule depends on three factors [19]: 1) covering the rate of positive samples, S_c^+ , 2) capability for separating the positive sample from the negative sample (S_c^+ from S_c^-), and 3) the value of rule functionality improvement versus random

Table 4 Fuzzy Association Rules (FAR) of Table 2

#	Fuzzy rules	Fuzzy support	Fuzzy confidence	Fuzzy lift
1	f1 is Low→beach	0.21	1.00	3.33
2	f1 is Low→sky	0.00	0.00	0.00
3	f1 is Low→grass	0.00	0.00	0.00
4	f2 is Low→beach	0.12	0.36	1.21
5	f2 is Low→sky	0.00	0.00	0.00
6	f2 is Low→grass	0.21	0.64	1.59
7	f3 is Low→beach	0.00	0.00	0.00
8	f3 is Low→sky	0.02	0.09	0.29
9	f3 is Low→grass	0.21	0.91	2.28
10	f1 is Mid→beach	0.09	0.20	0.65
11	f1 is Mid→sky	0.16	0.35	1.16
12	f1 is Mid→grass	0.21	0.46	1.14
13	f2 is Mid→beach	0.18	0.35	1.18
14	f2 is Mid→sky	0.14	0.27	0.92
15	f2 is Mid→grass	0.19	0.37	0.93
16	f3 is Mid→beach	0.07	0.14	0.46
17	f3 is Mid→sky	0.25	0.49	1.63
18	f3 is Mid→grass	0.19	0.37	0.93
19	f1 is High→beach	0.00	0.00	0.00
20	f1 is High→sky	0.14	0.42	1.41
21	f1 is High→grass	0.19	0.58	1.44
22	f2 is High→beach	0.00	0.00	0.00
23	f2 is High→sky	0.16	1.00	3.33
24	f2 is High→grass	0.00	0.00	0.00
25	f3 is High→beach	0.23	0.88	2.95
26	f3 is High→sky	0.03	0.12	0.38
27	f3 is High→grass	0.00	0.00	0.00

classification. The first and third factors were explained by *FuzzySupport* and *FuzzyLift*, respectively. The second factor was covered by *GrowthRate*, as defined below:

$$\begin{aligned}
 GrowthRate(f, c, MF, S_C^+, S_C^-) &= \frac{FuzzySupport_{S_C^+}(f \text{ is } MF \rightarrow c)}{FuzzySupport_{S_C^-}(f \text{ is } MF \rightarrow c)} \\
 &= \frac{\frac{FuzzyCount_{S_C^+}(f \text{ is } MF \rightarrow c)}{|S_C^+|}}{\frac{FuzzyCount_{S_C^-}(f \text{ is } MF \rightarrow c)}{|S_C^-|}} = \frac{\sum_{f \in S_C^+} \mu_{MF}(f_i)}{|S_C^+|} \quad (6) \\
 &= \frac{\sum_{f \in S_C^-} \mu_{MF}(f_i)}{|S_C^-|}
 \end{aligned}$$

The *GrowthRate* is the ratio of positive to negative samples of a rule, which evaluates the separating power S_C^+ from S_C^- .

The overall quality of a rule, called rule weight, is measured by combining mentioned triple factors:

$$W_D(R : f \text{ is } MF \rightarrow c) = \text{FuzzyConfidence}_D(R) * \text{GrowthRate}(R) * \text{FuzzyLift}(R) \quad (7)$$

The approach for weighting rules was somewhat similar to [19]. Table 5 summarizes all the concepts about FARs. The weight of a rule specifies the effectiveness of the rule in classification. Since the rule base must be composed of the most effective rules, in this paper, the rules with the highest weight were inserted to the rule base.

6 Hierarchical fuzzy classification

In recent years, hierarchical classification has been successfully used for the classification of large databases [4, 13]. Hierarchical classification has some advantages compared with flat classification such as using posterior probability of each concept, exploiting association between concepts (ontological information), capability of performing parallel classification and improving the accuracy of classification [4, 13]. In this paper, the image database was organized in four layers as a hierarchical structure. Table 6 shows the hierarchy of images. A number of Mamdani fuzzy inference engines were exploited to perform the classification [17]. They were organized in a hierarchical structure corresponding to the hierarchical structure of image database. Figure 5 shows a part of fuzzy inference engines. Figure 6 demonstrates the trapezoidal output fuzzy sets organized corresponding to the hierarchical structure of image database. Defuzzification was performed using the central average method, which has three advantages of simplicity, understandably and continuously [17].

Training phase and testing phase of the proposed algorithm are summarized in Figure 7. The first step of the training phase was feature extraction and fuzzification. Apriori algorithm was used for rule mining in each layer. The most qualitative FARs with the highest *weight* were used to build the fuzzy rule base. The output of the training phase was the fuzzy rule base. The next step was classification. Adaptive feature extraction and fuzzification of test image were the first part of classification. Adaptive feature extraction meant that the feature vector of input test image was composed of the features that built the rule base in the training phase. Afterwards, the hierarchical fuzzy inference engines in partnership with the fuzzy rule base determined the concepts of test image.

Table 5 Basic concepts of FAR mining

Concept	Description
FuzzyCount	The total degree of membership f over MF_i^1 in the image set D .
FuzzySupport	The percentage of <i>FuzzyCount</i> in the image set D .
FuzzyConfidence	The percentage of images contain both feature f and concept c toward total of images contain feature f .
FuzzyLift	The ratio of <i>FuzzyConfidence</i> to <i>Expected Confidence</i> .
GrowthRate	The capability of separating positive sample from negative sample
Weight	The overall importance of a rule.

Table 6 Hierarchical structure of (concept, sub-concept) pair

Layer 1	Layer 2	Layer 3	Layer 4	Concept ID			
Artificial	Indoor	Kitchen	Desk	1			
			Plate	2			
			Cup	3			
			Cabinet	4			
			Fork	5			
			Spoon	6			
			Glass	7			
		Computer	Mouse	8			
			Keyboard	9			
			Monitor	10			
		Toy	Car	11			
			House	12			
			Dolly	13			
		Seat	Sofa	14			
			Chair	15			
		Out door	House	Ancient	16		
				Modern	17		
				Jackal	18		
			Vehicle	Bicycle	19		
				Airplane	20		
	Bus			21			
	Train			22			
	Nature			Animal	Bird	Duck	23
						Butterfly	24
						Stork	25
		Non-bird	Eagle		26		
			Bear		27		
			Fox		28		
			Fish	29			
		Plant		Fruit	Dinosaur	30	
					Monkey	31	
					Cat	32	
					Cow	33	
	Camel				34		
	Horse				35		
	Elephant				36		
	Date				37		
	Apple				38		
	Tomato				39		
	Pear				40		
	Flower				Rose	41	
					Kaktos	42	
Mushroom					43		

Table 6 (continued)

Layer 1	Layer 2	Layer 3	Layer 4	Concept ID
	Landscape	Waterscape	Sea	44
			Waterfall	45
		Mainland	Sunset	46
			Mountain	47
			Snow	48

7 Evaluation and experimental results

The proposed algorithm was implemented and run in a system of Pentium IV 2.8 GHZ CPU with 512 M memory. The image database contained 6000 general-purpose images from COREL including 48 semantic concepts (class), which were manually organized in four layers (<http://imagedb.persiangu.com>). Each class contained about 100 images 60 images of which were dedicated to training and 40 for testing phases. To certify the characteristic of *generality*, 1000 unknown concept images were injected into the database. The unknown concept’s class was not used in the training phase. In other words, the unknown concepts were not learned by the system. It was utilized in the testing phase for evaluating the generality of the algorithm against unknown concepts. MATLAB was used for programming. In this section, the following cases were studied:

- A brief review of evaluation measures of classifiers
- Performance analysis
 - Fuzzy rule base
 - Weight threshold
 - Effectiveness
 - Efficiency
 - Hierarchical performance
- Discussing the effectiveness of fuzzification
 - The quality and quantity of FAR
 - Fuzzy association rules against non-fuzzy association rules

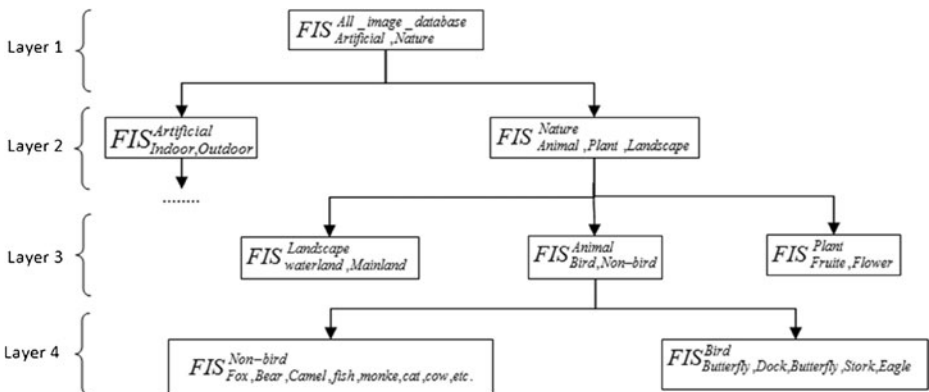


Fig. 5 A part of hierarchical classification structure using four layers of fuzzy inference engines

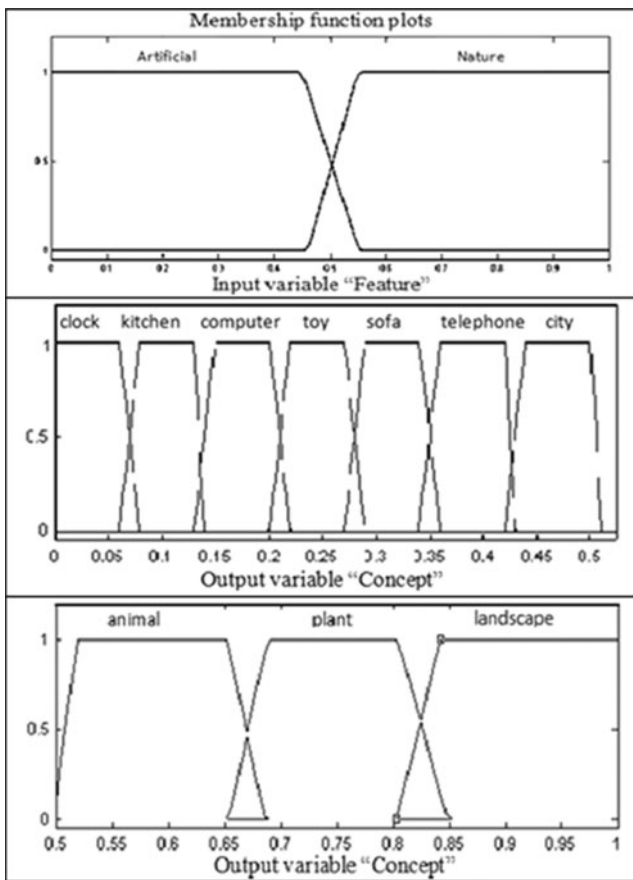


Fig. 6 Output fuzzy sets in two different layers

- System performance in dealing with multi-label images
- Comparison with the related works

Finally, in order to evaluate the performance objectively, the results of the approach were compared with the two works in the literature: CFAR [13] and FARCHD [1], both of which performed classification using Fuzzy Association Rules. To equalize the test bed for comparison, seven real-world datasets were developed in the experiments.

7.1 A brief review of evaluation measures of classifiers

There are some common measures to evaluate systems in information retrieval. In this part, a brief review of some is presented and some others will be reviewed using these criteria in order to evaluate the suggested algorithm. Before anything, some primary concepts must be introduced. According to Figure 8:

True Positive (TP_i): number of images correctly assigned to class i

False Positive (FP_i): number of images incorrectly assigned to class i but actually did not belong to that

```

1  procedure Train
2  input: ImageDatabase, 4-layer concepts structure (L)
3  output: FuzzyRuleBase
4  for each images I in ImageDatabase
5    FV = FeatureExtraction & Fuzzification (I)
6  for each layer l in L
7    for each Concept c in l
8      FARs = FAR_mining(FV, l, c)
9      Add FARs to FuzzyRuleBase
10   endfor
11 endfor
12 return FuzzyRuleBase
13 end procedure

14 procedure FAR_mining
15 input: FuzzyFeatureVector(FV), Layer l, Concept c
16 output: FuzzyAssociationRules of Concept c
17 rules = Apriori (FV, c, other concept of layer l)
18 for each rule r in rules
19   if r.FuzzySup > MinSup &
20     r.FuzzyConfidence > MinConfidence &
21     r.FuzzyLift > MinLift then
22     r.weight = r.FuzzyConfidence * r.growthRate * r.FuzzyLift
23   endif
24 return rules
25 end procedure

26 procedure Classification
27 input: 4-layer FIS structure(FIS), FuzzyRuleBase, test image
28 output: semantic concepts of test image
29 AFV = AdaptiveFeatureExtraction & Fuzzification(test image)
30 for each FIS  $\in$  4-layer FIS structure
31   Fuzzy concepts = FIS(AFV, FuzzyRuleBase)
32   Concept of image = Defuzzifier(Fuzzy concepts, trapezoidal)
33 endfor
34 return concept of image
35 end procedure

```

Fig. 7 Pseudo code of training and testing in presented fuzzy system

False Negative (FN_i): number of images did not assign to class i but actually belonged to that

True Negative (TN_i): number of images did not assign to class i but actually did not belong to that

Fig. 8 Basic concepts of computing evaluation measures

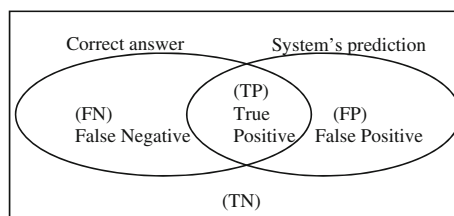


Table 7 Evaluating the performance of classifier in detail (MinSup=0.2)

ID	Concepts	TP	FN	FP	Precision	Recall	F1-measure
1	Desk	92	8	5	0.95	0.92	0.93
2	Plate	85	15	9	0.90	0.85	0.87
3	Cup	84	16	15	0.85	0.84	0.84
4	Cabinet	93	7	25	0.79	0.93	0.85
5	Fork	74	26	13	0.85	0.74	0.79
6	Spoon	86	14	6	0.93	0.86	0.89
7	Glass	92	8	10	0.90	0.92	0.90
8	Mouse	65	35	16	0.80	0.65	0.71
9	Keyboard	78	22	13	0.86	0.78	0.81
10	Monitor	69	31	16	0.81	0.69	0.74
11	Car	75	25	25	0.75	0.75	0.75
12	House	90	10	23	0.80	0.90	0.84
13	Dolly	75	25	8	0.90	0.75	0.81
14	Sofa	61	39	5	0.93	0.61	0.73
15	Chair	69	31	16	0.81	0.69	0.74
16	Ancient	78	22	14	0.85	0.78	0.81
17	Modern	96	4	16	0.86	0.96	0.90
18	Jackal	78	22	17	0.82	0.78	0.79
19	Bicycle	65	35	7	0.90	0.65	0.75
20	Airplane	84	16	5	0.94	0.84	0.88
21	Bus	79	21	7	0.92	0.79	0.85
22	Train	86	14	14	0.86	0.86	0.86
23	Duck	81	19	12	0.87	0.81	0.83
24	Butterfly	74	26	17	0.81	0.74	0.77
25	Stork	67	33	17	0.80	0.67	0.72
26	Eagle	81	19	23	0.78	0.81	0.79
27	Bear	68	32	4	0.94	0.68	0.78
28	Fox	69	31	6	0.92	0.69	0.78
29	Fish	78	22	14	0.85	0.78	0.81
30	Dinosaur	85	15	23	0.79	0.85	0.81
31	Monkey	87	13	15	0.85	0.87	0.85
32	Cat	97	3	13	0.88	0.97	0.92
33	Cow	95	5	32	0.75	0.95	0.83
34	Camel	96	4	24	0.80	0.96	0.87
35	Horse	93	7	18	0.84	0.93	0.88
36	Elephant	80	20	4	0.95	0.80	0.86
37	Date	97	3	11	0.90	0.97	0.93
38	Apple	89	11	13	0.87	0.89	0.87
39	Tomato	93	7	22	0.81	0.93	0.86
40	Pear	78	22	20	0.80	0.78	0.78
41	Rose	68	32	19	0.78	0.68	0.72
42	Kaktos	77	23	19	0.80	0.77	0.78
43	Mushroom	83	17	15	0.85	0.83	0.83
44	Sea	69	31	18	0.79	0.69	0.73

Table 7 (continued)

ID	Concepts	TP	FN	FP	Precision	Recall	F1-measure
45	Waterfall	73	27	13	0.85	0.73	0.78
46	Sunset	76	24	11	0.87	0.76	0.81
47	Mountain	82	18	19	0.81	0.82	0.81
48	Snow	96	4	24	0.80	0.96	0.87

Now, familiarity with evaluation measures can be done. The most commonly used performance measure in classification is precision and recall. $Precision_i$ means the percentage of correctly classified images among all the images assigned to class i . $Recall_i$ means the percentage of correctly classified images among all the images, which must be assigned to class i [15]:

$$\begin{aligned}
 Precision_i &= \frac{TP_i}{TP_i+FP_i} \\
 Recall_i &= \frac{TP_i}{TP_i+FN_i}
 \end{aligned}
 \tag{8}$$

F-measure is another popular criterion for evaluating the classification quality. Since it makes a trade-off between recall and precision, it is widely used for evaluating the classifiers. F-measure expresses the harmonic mean of recall and precision. It is defined for class i as:

$$F - measure_i = \frac{2 * precision_i * recall_i}{recall_i + precision_i}
 \tag{9}$$

This is also known as the $F1$ -measure because recall and precision have equal weight.

While the above measures are used for evaluating the classification quality of each class, there are other measures to evaluate the overall quality of the classifier in multi-label classification. Micro-average and macro-average are two classifier evaluation measures, which are derived from F-measure. They make more sense of the strength and weakness

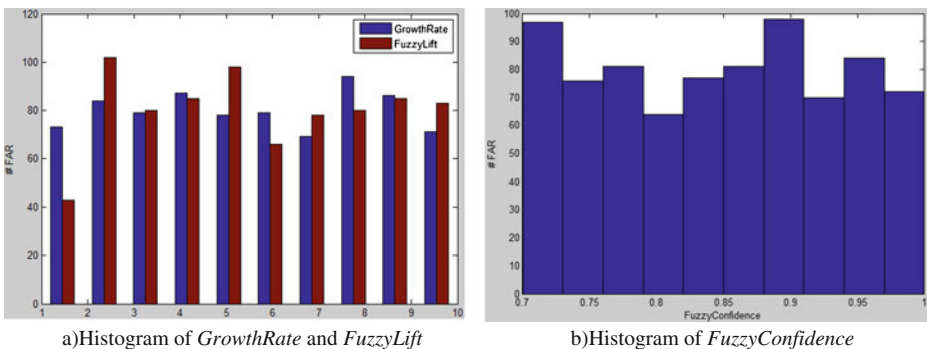


Fig. 9 The histogram of *GrowthRate*, *FuzzyLift* and *FuzzyConfidence* of the rule base

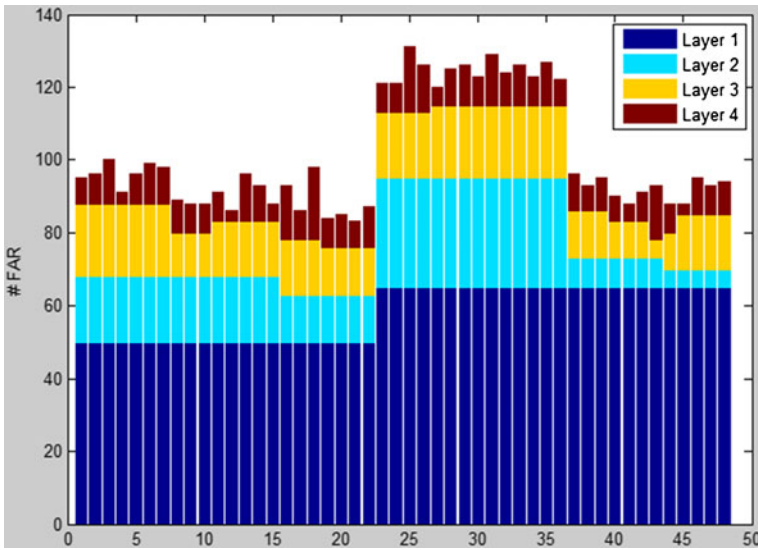
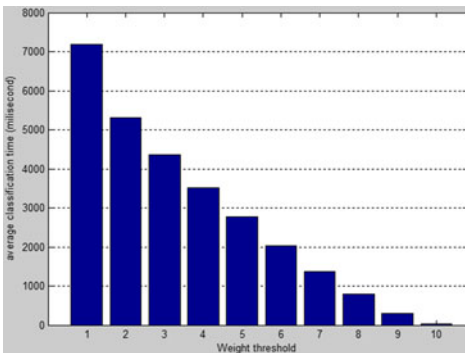


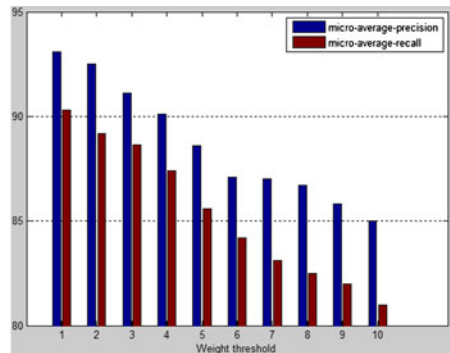
Fig. 10 The number of rules in the rule base over each layer

of systems under evaluation. The micro-average gives equal weight to all images so it is considered as the average over all other classes [15]:

$$\begin{aligned}
 \text{micro - averagePrecision} &= \frac{\sum_{i=1}^M TP_i}{\sum_{i=1}^M (TP_i + FP_i)} \\
 \text{micro - averageRecall} &= \frac{\sum_{i=1}^M TP_i}{\sum_{i=1}^M (TP_i + FN_i)}
 \end{aligned}
 \tag{10}$$



a) Average time of classification over different values of weight threshold



b) System performance over different values of weight threshold (*micro-average_{precision}* and *micro-average_{recall}*)

Fig 11 Effectiveness and efficiency over different values of weight threshold

where M is the number of classes. In the same way, it is possible to compute the overall micro-average of the classifier. It is the harmonic mean of $micro\text{-average}_{precision}$ and $micro\text{-average}_{recall}$:

$$micro - average = \frac{2 * precision * recall}{precision + recall} \tag{11}$$

where precision and recall are $micro\text{-average}_{precision}$ and $micro\text{-average}_{recall}$, respectively. The macro-average gives equal weight to each class [15]:

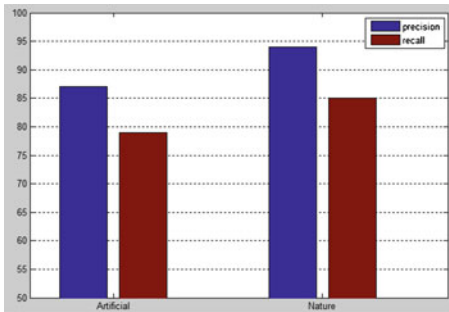
$$macro - average_{Precision} = \frac{\sum_{i=1}^M Precision_i}{M}$$

$$macro - average_{Recall} = \frac{\sum_{i=1}^M Recall_i}{M} \tag{12}$$

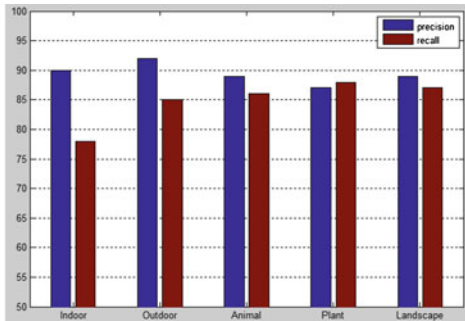
where M is the number of classes. In the same way, it is possible to compute the overall macro-average of classifier. It is the average of F-measure over all the classes:

$$F - measure_i = \frac{2 * precision_i * recall_i}{recall_i + precision_i}$$

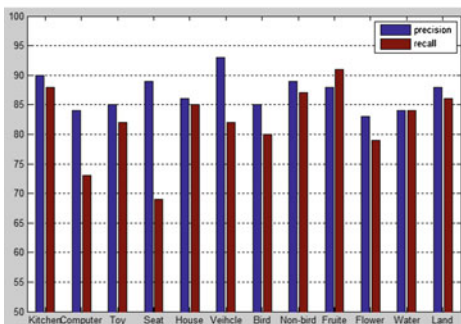
$$macro - average = \frac{\sum_{i=1}^M F - measure_i}{M} \tag{13}$$



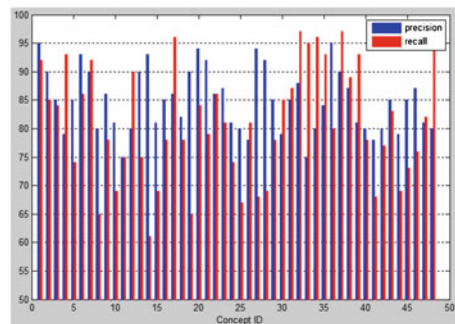
a) Performance of layer 1; percision average is 90; recall average is 81



b) Performance of layer 2; percision average is 90; recall average is 85



c) Performance of layer 3; percision average is 87; recall average is 82



d) Performance of layer 4; percision average is 85; recall average is 81

Fig. 12 Recall and precision behavior in the hierarchical structure

where $precision_i$ and $recall_i$ are the value of precision and recall for the class I , respectively

Table 7 shows the result of classification in detail. For more clarity, the measures of *spoon* class are explained. As appeared in the *spoon* row, the classifier assigns 86 images to this class correctly (TP) but it assigned 14 spoon images to other classes incorrectly (FN). After classification, there were six images in the spoon class which actually did not belong to this class (FP).

According to the formula 8, 9, the evaluation measures were computed as:

$$\begin{aligned}
 Precision_i &= \frac{TP_i}{TP_i+FP_i} = \frac{86}{86+6} = \frac{86}{92} = 0.93 \\
 Recall_i &= \frac{TP_i}{TP_i+FN_i} = \frac{86}{86+14} = \frac{86}{100} = 0.86 \\
 F1 - measure &= \frac{2 * precision_i * recall_i}{recall_i + precision_i} = \frac{2 * 0.93 * 0.86}{0.93 + 0.86} = \frac{1.60}{1.80} = 0.89
 \end{aligned} \tag{14}$$

7.2 Performance analysis

7.2.1 Fuzzy rule base

The fuzzy rule base should be composed of the most qualitative rules which have a big *weight* value. In this paper, the rules with the weight of more than ten were selected to build the rule base (*weight threshold*=10). Regarding Formula 7, the weight of a rule is a combination of three factors: *FuzzyConfidence*, *GrowthRate* and *FuzzyLift*. Therefore, more value of these measures leads to the more value of rules *weight*. So, most qualitative rules have the greatest value of these measures. Figure 9 shows the histogram of these measures for rules in the fuzzy rule base. The maximum value of *FuzzyConfidence* is one while the values of *FuzzyLift* and *GrowthRate* should be more than one to be beneficial for classification. Figure 10 shows the number of rules in each layer. Classes of layer n had a smaller number of rules against the classes of layer $n-1$ because the number of images in each layer was less than the number of images in the previous layer. Classification of a few images requires less number of rules. For example, in layer one, 6000 images should be classified while, in layer three, the fuzzy inference engine of *seat* concept should classify only 200 images. Classification of 6000 images requires a greater number of rules than that of 200 images.

7.2.2 Weight threshold

In this part, the discussion of weight threshold tuning and its effectiveness on system performance and time complexity is presented. If the weight threshold is a small number, then the rule base contains a great number of rules; subsequently, the time of classification increases because the fuzzy inference engines have to process a great number of rules. Figure 11a shows the average time of classification over different values of weight threshold. Decrease of the weight threshold leads to a greater number of rules inserted to the rule base. Thus, the system effectiveness increases because a greater number of rules can classify images more precisely. Figure 11b shows the system performance over different values of weight threshold. If the weight threshold is equal to n , then all rules with the weight of more than or equal to n will

Table 8 Overall performance of the classifier in each layer

	Layer 1	Layer 2	Layer 3	Layer 4
Micro-average _{Precision}	0.9435	0.894	0.8700	0.8453
Micro-average _{Recall}	0.875	0.854	0.8211	0.8096
Macro-average _{Precision}	0.9037	0.8845	0.8598	0.8490
Macro-average _{Recall}	0.8647	0.8465	0.8236	0.8096

be inserted to the rule base. The selected weight threshold is only one possible choice and authors have no claims about the optimal value of weight threshold.

7.2.3 Effectiveness

The overall quality of rule base plays the most important role in system effectiveness.

Fuzzy inference engines with the partnership of fuzzy rule base assign each image to an appropriate class. Figure 12 shows the precision and recall rate over each layer. The overall precision of layer n was less than the performance of layer $n-1$ because, if an image was classified incorrectly, then it would be incorrectly classified in the next layer, inevitably. Table 8 shows the overall performance in each layer. All images would be classified correctly if all the rules had absolute confidence, meaning that: $FuzzyConfidence=1$. Since the rules were somewhat confident, $FuzzyConfidence < 1$, some images were incorrectly classified. Figure 13 shows the number of misclassified images in all the layers. The misclassified images of concept *Artificial* were those that must be assigned to the *Nature* class but they were wrongly assigned to *Artificial* class.

Due to evaluating the generality of the algorithm against new concepts, 1000 unknown concept images were inserted in the image database. The unknown concepts were not learned by the system. Figure 14 shows the number of misclassified unknown concept images. About

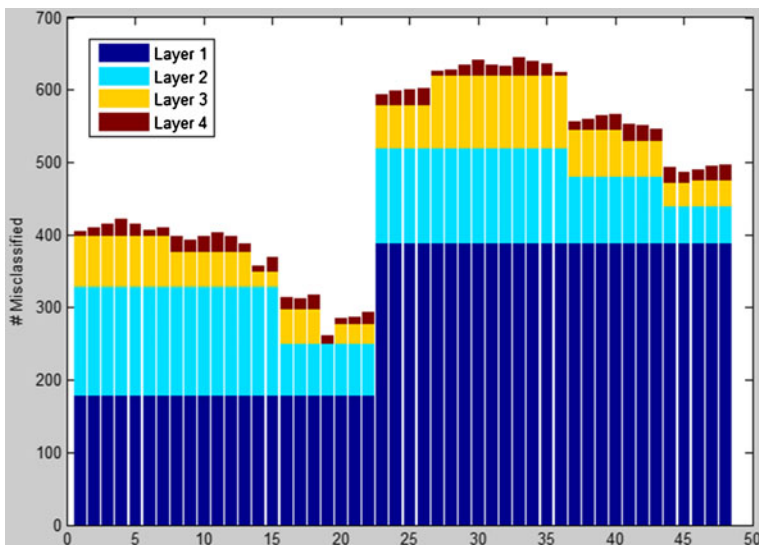


Fig. 13 The number of misclassified images

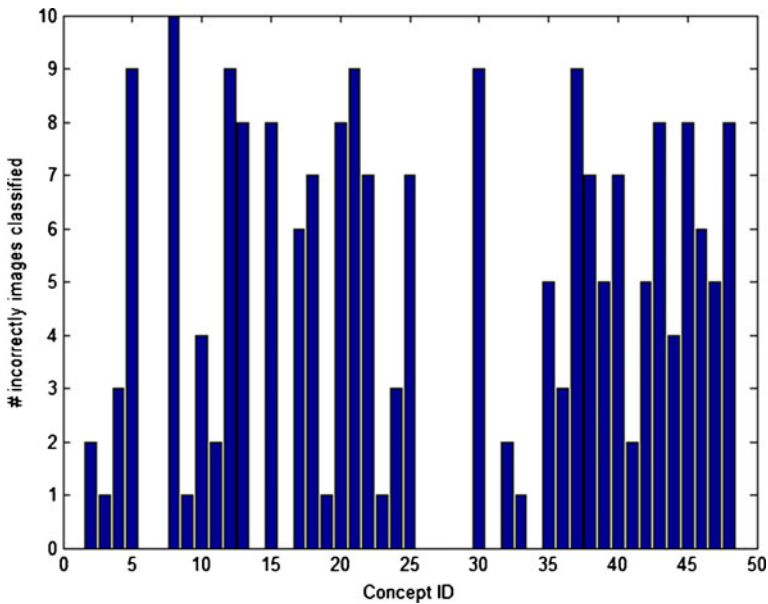


Fig. 14 The number of misclassification of unknown concept images

20 % of unknown concept images were wrongly assigned to other concepts and 80 % were correctly classified as unknown concepts. Table 9 shows the system performance after inserting 1000 unknown concept images to the database. As can be observed, the classifier had good persistence against unknown concepts. In order to evaluate the rate of performance reduction against unknown concepts, the unknown concept images were injected into the image database step by step. Figure 15 shows the micro-average_{Precision} and micro-average_{Recall} in ten steps. In each step, 100 images were inserted to the database. System performance decreased gradually when unknown concept images were inserted to the database.

7.2.4 Efficiency

The time of classification was in correlation with the number of rules in fuzzy rule base because the fuzzy inference engines had to process the rules and rule processing was time consuming. Figure 16 shows the average time of classification over each layer. This figure had meaningful correlation with Figure 10 which showed the number of rules in each layer because a greater number of rules led to greater classification time.

Fuzzy inference engines in layer *n* classified images more quickly than inference engines of layer *n-1* because the number of rules in layer *n* was less than that in

Table 9 Overall performance of the classifier with and without unknown concept

	Without unknown concept (5000 images)	With unknown concept (6000 images)
Micro-average _{Precision}	0.8453	0.8185
Micro-average _{Recall}	0.8096	0.7584
Macro-average _{precision}	0.8490	0.8105
Macro-average _{Recall}	0.8096	0.7511

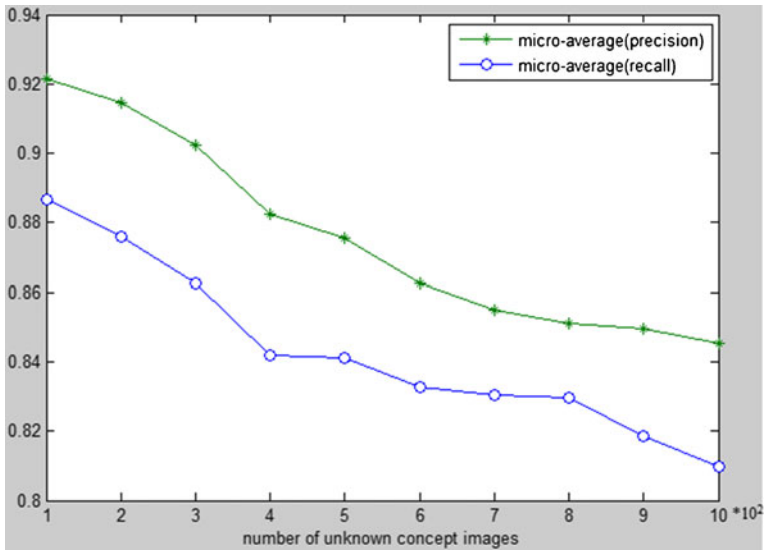


Fig. 15 System performance in dealing with injecting unknown concept

layer $n-1$. For example, the fuzzy inference engine in layer one should classify 6000 images into two classes of *Artificial* and *Nature* using 105 rules while, in layer 3, the fuzzy inference engine of *seat* concept classified only 200 images in two classes using 15 rules. Obviously, processing 15 rules would take less time that processing 105 images.

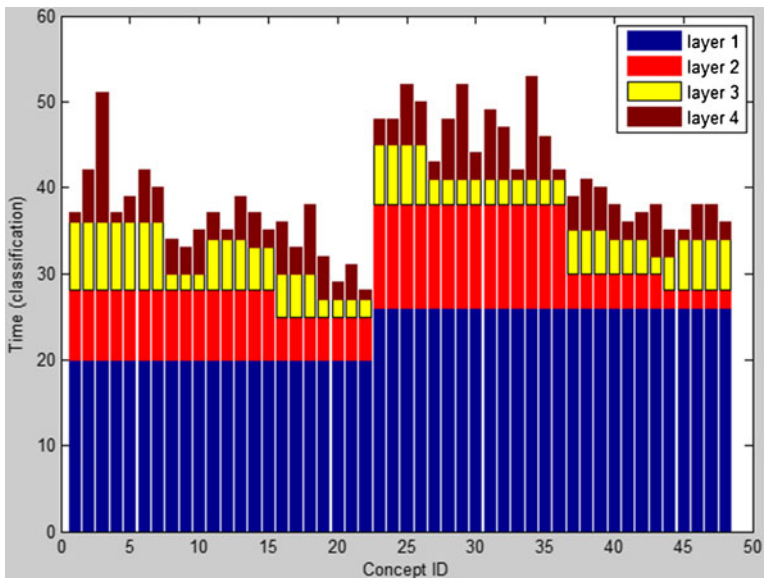


Fig. 16 Time of classification in each layer

Table 10 Overall performance of hierarchical classification versus flat classification

	Hierarchical	Flat
Micro-average _{Precision}	0.8453	0.6950
Micro-average _{Recall}	0.8096	0.6781
Macro-average _{Precision}	0.8490	0.7014
Macro-average _{Recall}	0.8096	0.6751

7.2.5 Hierarchical performance

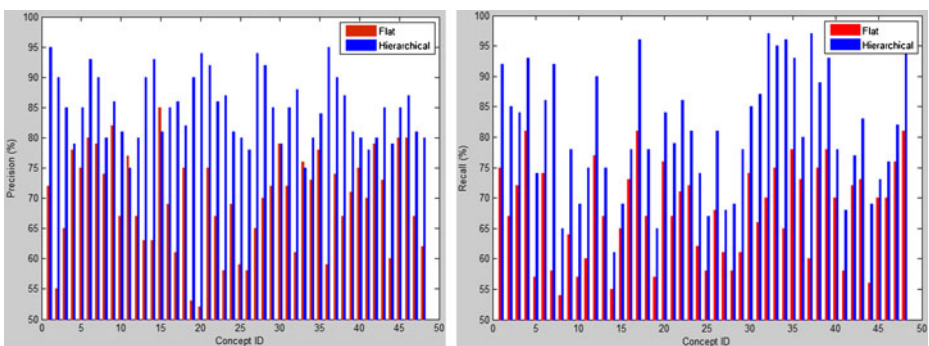
To demonstrate the performance of hierarchical classification versus flat classification, a flat classifier was developed, which classified images of layer 4 without considering the three first layers. In other words, it tried to classify the test image in 48 classes. Table 10 shows the overall performance of hierarchical and flat classifier. The hierarchical performance was more than flat classification because it could use the similarity between the classes [12]. Figure 17 shows the efficiency and effectiveness of flat classifier versus hierarchical classifier. Generally, hierarchical classifiers are more time consuming because an image should be classified several times in layers. The result of time spending was to reach better performance in recall and precision.

7.3 Discussing the effectiveness of fuzzification

Fuzzification of feature vector plays an important role in the prediction power of FAR. In other words, the rules extracted from fuzzy feature vector had higher values of weight versus the rules extracted from non-fuzzy feature vector. Fuzzification of feature vector was applied after feature extraction. In this part, different aspects of fuzzification are discussed.

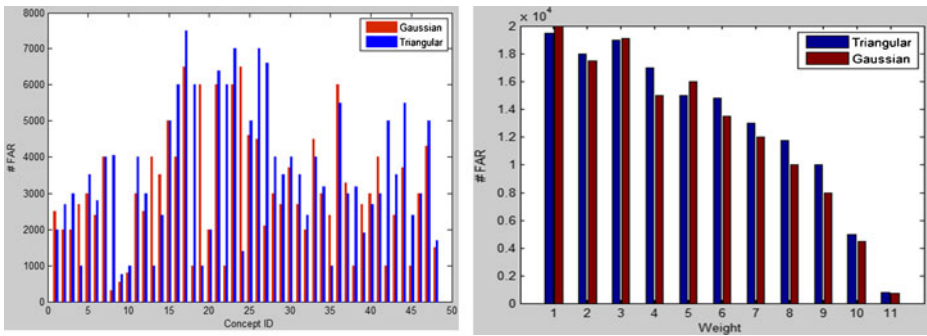
- The quantity and quality of FAR

In order to evaluate the effect of membership function, two different fuzzy sets were tested: triangular and Gaussian fuzzy sets. Figure 18 compares the quantity and quality of extracted rules with two fuzzy sets. The triangular fuzzy set generated a greater number of



a) Precision rate; average of hierarchy=85% , average of flat=68%
 b) Recall rate; average of hierarchy=81% , average of flat=65%;

Fig. 17 A comparison of effectiveness and efficiency of hierarchical classification versus flat classification



a) The quantity of extracted rules (average of Gaussian=3000; average of Triangular=3600)

b) The quality of extracted rules

Fig. 18 Comparisons of the quality and quantity of extracted rules with two different fuzzy sets

rules against Gaussian fuzzy set. The average rules generated by Triangular fuzzy set was 3600 rules while the average rules generated by Gaussian fuzzy set was 3000 rules.

The quality of rules is the most important thing for classifier performance. Figure 18.b shows that the triangular fuzzy set generates more qualitative rules because of producing a greater number of rules with the weight threshold of six. Table 11 compares the system performance using triangular and Gaussian fuzzy sets. As was expected, the triangular fuzzy set built more qualitative classifier than Gaussian fuzzy set. In this paper, all the experiments were performed using the triangular membership function.

- Fuzzy association rules against non-fuzzy association rules

Non-fuzzy association rules were extracted from quantized feature data. Quantization led to the change of the real value of data and replaced unrealistic valued in the dataset. This problem led to the reduction in the effectiveness of non-fuzzy association rules. To evaluate the suggested fuzzy association rule against non-fuzzy association rules, the feature vectors were discretized; then the association rules were extracted like in [19]. Table 12 shows the overall performance of the suggested algorithm and non-fuzzy association rules.

7.4 System performance in dealing with multi-label images

No special experiment was developed for evaluating the system performance in dealing with multi-label images because most of the images in the database were multi-label; thus, all the results implicated the system behavior in dealing with multi-label images. However, the

Table 11 Effectiveness of *membership function* for overall performance of the classifier

	Triangular fuzzy set	Gaussian fuzzy set
Micro-average _{Precision}	0.8453	0.8327
Micro-average _{Recall}	0.8096	0.8140
Macro-average _{Precision}	0.8490	0.8355
Macro-average _{Recall}	0.8096	0.8059

Table 12 Overall performance of the proposed algorithm and non-fuzzy association rules

	Proposed algorithm	Non-fuzzy association rules
Micro-average _{Precision}	0.8453	0.7891
Micro-average _{Recall}	0.8096	0.7681
Macro-average _{Precision}	0.8490	0.7904
Macro-average _{Recall}	0.8096	0.7745

process of handling multi-label images was given here. Consider a multi-label image with two concepts, for example, sunset and mountain. So, it belonged to these two classes (sunset and mountain) altogether. *In the training phase*, the sunset rules were extracted from the image when the image was placed in the sunset class; also the mountain rules were extracted from the image when it was placed in the mountain class. Therefore, the rules of each image concept were extracted separately from image when the image belonged to each class. *In the testing phase*, adaptive feature extraction was applied on the test image and the feature vector was constructed; subsequently, the fuzzy inference engines started the matching process between feature vector and fuzzy rule base. In this process, perhaps some rules with different concepts were matched with the feature vector; meaning that the input image contained more than one concept. For example, if some sunset rules and some mountain rules were matched with the feature vector, then fuzzy inference engine would classify the test image in two classes of sunset and mountain.

7.5 Comparison with the related works

To evaluate the proposed algorithm, these results were compared with the findings of some other state-of-the-art similar paper: CFAR [13] and FARCHD [1], both of which dealt with fuzzy association rule in order to establish a semantic classifier. Table 13 indicates that the proposed approach reached 0 more precision with a less numbers of rules compared with two other methods. Due to equalizing the test bed of the comparison, six real-world datasets were considered in the experiments. All the datasets were multivariate, real attributes and dedicated for the classification task.

Table 13 Comparison of the proposed algorithm with other similar methods

# Dataset	#Instance	# Attribute	#Classes	FARC-HD		CFAR		Our approach	
				#RULE	Precision (%)	#RULE	Precision (%)	#RULE	Precision (%)
1 Sonar	208	60	2	18	80.19	37	72.48	13	83.65
2 Pima	768	8	2	22	75.66	2	65.11	14	86.70
3 Page-blocks	5473	10	5	19	95.01	24	89.19	16	93.14
4 Glass	214	10	7	22	70.24	3	44.66	17	80.52
5 Magic	19020	11	2	43	84.51	2	64.84	8	83.27
6 Yeast	1484	8	10	38	58.50	40	50.90	20	72.10
	Mean			26.67	75.21	14.33	61.44	14	80.74

Available at: <http://archive.ics.uci.edu/ml/datasets.html>

8 Conclusion

In this paper, a new approach was proposed to improve image classification using fuzzy association rules (FAR). The FAR was a combination of low-level visual features and semantic concept, which was unique and significant for describing the semantic concept. Fuzzy rule base was constructed using the most qualitative FARs after feature extraction and fuzzification of feature vector. A hierarchical fuzzy inference engine was developed for building the semantic classifier.

In this study, a part of COREL image database was used. It had about 6000 images with 48 categories. To investigate the generality characteristic of the database, it was enriched by injecting 1000 unknown concept images, which were not used in the training phase but participated in the testing phase. The experimental results demonstrated that the proposed classifier had the factors of an outstanding classifier. Despite the impressive advantages of the suggested classifier such as the independence of image segmentation, solving dimensionality problem and efficiency in run-time, one of its limitations was being time consuming in FAR mining.

For future works, the methodology might be improved in the following directions:

The application of the developed methodology could be changed to a dynamic image database environment (i.e. support insertion and deletion of images and concepts in the database).

Besides the semantic classification, this methodology may be also applied to solving other problems such as automatic image annotation, image retrieval, object recognition and image database summarization.

Fuzzifier and defuzzifier are static fuzzy sets and can be dynamic to different concepts because some semantic concepts are more consistent with some special fuzzy sets.

References

1. Alcal J, Alcal A, Herrera F (2011) A fuzzy association rule based classification model for high-dimensional problems with genetic rule selection and lateral tuning. *IEEE Trans Fuzzy Syst.* doi:10.1109/TFUZZ.2011.2147794
2. Chang B, Hang H, Huang H (2003) Research friendly MPEG-7 software testbed. *IS&T/SPIE Symposium on Electronic Imaging Science and Technology*, pp 890–901
3. Chen Z, Chen G (2008) Building an associative classifier based on fuzzy association rules. *Int J Comput Intell Syst* 1:262–273
4. Fan J, Gao Y, Luo H, Jain R (2008) Mining multilevel image semantics via hierarchical classification. *IEEE Trans Multimed* 10:167–187
5. Hua Y, Chena S, Tzengb G (2002) Mining fuzzy association rules for classification problems. *Comput Ind Eng* 43:735–750
6. Jing W, Huang L, Luo Y et al (2006) An algorithm for privacy-preserving quantitative association rules mining. *Proceedings of the 2nd IEEE International Symposium*, pp 315–324
7. Konstantinidis K, Gasteratos A, Andreadis I (2005) Image retrieval based on fuzzy color histogram processing. *Opt Commun* 248:375–386
8. Kuok CM, Fu A, Wong MH (2008) Mining fuzzy association rules in databases. *ACM SIGMOD* 27:159–168
9. Liu Y, Zhang D, Lu G, Ma W (2007) A survey of content-based image retrieval with high-level semantics. *Pattern Recogn* 40:262–282
10. Lu D, Weng Q (2007) A survey of image classification methods and techniques for improving classification performance. *Rem Sen* 28:823–870
11. Malathi A, Shanthi V (2011) Statistical measurement of ultrasound placenta images complicated by gestational diabetes mellitus using segmentation approach. *J. of information Hiding and Multimedia Signal Process* 2:332–343
12. Marsza M, Schmid C (2008) Constructing category hierarchies for visual recognition. *European Conference on Computer Vision (ECCV '08)*, pp 479–491
13. Silaa C, Freitas A (2011) A survey of hierarchical classification across different application domains. *Data min. and knowl. Discovery* 22:31–72

14. Singh C (2011) Fuzzy rule based median filter for gray-scale images. *J. of information Hiding and Multimedia Signal Process* 2:108–122
15. Sun A, Lim E (2001) Hierarchical text classification and evaluation. In *proc. ICDM*, pp 521–528
16. Thilagam S, Ananthanarayana P et al (2008) Extraction and optimization of fuzzy association rules using multi-objective genetic algorithm. *Pattern Anal Appl* 11:315–324
17. Wang LX (1996) *A course in fuzzy systems and control*. Prentice-Hall, New Jersey
18. Wang F, Kan M (2006) NPIC: Hierarchical synthetic image classification using image search and generic features. In *Proc. CIVR*, pp 473–482
19. Wang W, Zhang A (2006) Extracting semantic concepts from images: a decisive feature pattern mining approach. *Multimedia Syst*. doi:10.1007/s00530-006-0029-x
20. Xiaoyun C, Wei C (2006) Text categorization based on classification rules tree by frequent patterns. *Softw* 17:1017–1025
21. Zimek A, Buchwald F, Frank E, Kramer S (2008) A study of hierarchical and flat classification of proteins. *IEEE Trans Comput Biol Bioinformatics* 7:563–571



Abolfazl Tazaree was born in Iran in 1982. He entered Pnu University as an undergraduate student in 2000, received the B.S. degree in computer engineering in 2005. He received the M.Sc. in 2008 in Artificial Intelligence from the Qazvin branch, Islamic Azad University. His research interests are in softcomputing, machine learning, data mining and knowledge discovery, and their application to software engineering.



Amir-Masud Eftekhari-Moghadam was born in Tehran, Iran in 1964. He received B.E. and M.Sc. degree in computer engineering from Iran University of Science and Technology in 1992 and 1995, respectively and Ph.D. degree in computer engineering in 2002 from Science and Research branch, Islamic Azad University, Tehran, Iran. In 1996 he was a lecturer of computer engineering and since 2002, he has been assistant professor at the Qazvin branch, Islamic Azad University, Qazvin, Iran. His research interests include image processing, image retrieval, image/video coding, semantic video analysis, data mining and image security.



Saeedeh Sajjadi Ghaem Maghami was born in Tehran, Iran in 1984. She received B.E. and M.Sc. degree in computer engineering from Qazvin Azad University, Qazvin, Iran in 2008 and 2012, respectively. In 2008 she was a lecturer of computer engineering and since at the Qazvin branch, Islamic Azad University, Qazvin, Iran. Her research interests include image processing, image retrieval, image mining, knowledge discovery, speech processing, natural language processing and machine learning.