Twelfth Int. Conf. on IFSs, Sofia, 17-18 May 2008
NIFS Vol. 14 (2008) 2, 91-99

Bacillus Colonies Recognition Using Intuitionistic Fuzzy Sets

Vahid Khatibi, Gh.A. Montazer*

Information Technology Dept., School of Engineering, Tarbiat Modares University

Address: P.O. Box 14115-179, Tehran, Iran

Abstract
The pattern recognition problem in many practical domains such as medical diagnosis needs to encounter uncertain and imprecise hypotheses. Instead of solely relying on statistical inference for classification-type recognition, we need to quantify uncertainty first, and then, classify the unknown samples according to uncertainty quantification result. On the other hand, the Intuitionistic Fuzzy Sets provide a convenient framework for uncertainty quantification. Instead of measuring the similarity between certain pattern feature vectors and samples, the similarity between the Intuitionistic Fuzzy Sets of uncertain pattern feature vectors and samples are represented. In this paper, an IFS similarity measure is exploited in the practical problem of bacillus colonies classification. Our experiment showed that the classifier built on this approach could satisfactorily classify the unknown imprecise samples of bacillus colonies correctly, so as the final results were rational and acceptable.

Keywords: Intuitionistic Fuzzy Sets, IFS Similarity Measure, Pattern Recognition, Medical Diagnosis, Bacillus Colony Recognition

1. Introduction

Pattern recognition (PR) [7, 8] is an activity which processes received unknown information to identify the source of the information. The goal of pattern recognition research is to devise ways and means of automating decision-making processes that lead to classification and recognition. Pattern classification is concerned with the assignment of unknown patterns, represented by feature vectors, to predefined categories or classes. In fact, pattern recognition problems typically involve the classification of an unknown pattern $\boldsymbol{L}$ to a given set of $\boldsymbol{K}$ prototypes $\boldsymbol{P_k}$, $k \in \{1,2,...,\boldsymbol{K}\}$ [12, 13]. Each prototype $\boldsymbol{P_k}$ belongs to a given class $\boldsymbol{C_m}$, $m \in \{1,2,...,\boldsymbol{M}\}$, which is specified by the indicator function $\boldsymbol{A_k}$:

$$\boldsymbol{A_k} = \boldsymbol{l_m} \text{ if } \boldsymbol{P_k} \text{ belongs to the } m\text{th class } \boldsymbol{l_m} \tag{1}$$

Let $\boldsymbol{S(L,P_k)}$ be a similarity measure which measures the degree of similarity, or compatibility, between the unknown pattern $\boldsymbol{L}$ and the kth prototype $\boldsymbol{P_k}$. Then , formally, we

* Corresponding Author, Telfax: +98(21)82883990

Email Address: *montazer@modares.ac.ir* (GH.A. Montazer)

may write the process of classifying, or assigning the unknown pattern $\boldsymbol{L}$ to the class $\boldsymbol{l}^* = \boldsymbol{A}_{k^*}$, where

$$k^* = \arg\max_k \left(\boldsymbol{S}(\boldsymbol{L}, \boldsymbol{P}_k)\right). \tag{2}$$

We typically encounter imprecise classes and samples and are not able to confront uncertainty and impreciseness using certain hypotheses, whereas, in the real life classification problems, we envisage numerous ambiguous recognition problems in which we could not classify the unknown samples with certainty. To deal with these problems, we use *Soft Computing* [19] as a comprehensive collection of methods which works synergistically and provides flexible information processing capabilities to exploit the tolerance for imprecision, uncertainty, approximate reasoning, and partial truth. In this paper, we use Intuitionistic Fuzzy Sets as a tool of Soft Computing to solve a practical classification problem in medical diagnosis domain which deals with uncertainty quantification. Bacillus Colonies recognition involves imprecise and uncertain feature vectors which are interpreted through Intuitionistic Fuzzy Sets and then classified using a specified similarity measure.

This paper is organized as follows. At the next section, the fundamental notions of Intuitionistic Fuzzy Sets are presented. Then, its application in pattern recognition and respective measures based on similarity and distance measures are argued. At the last section, the bacillus colonies recognition using Intuitionistic Fuzzy Sets is proposed and the experimental results are investigated.

2. Intuitionistic Fuzzy Sets

The Fuzzy Sets Theory proposed by Zadeh [18] in 1965 has showed successful applications in various fields. In this theory, the membership of an element to a fuzzy set is a single value between zero and one. But in reality, it may not always be certain that the degree of non-membership of an element in a fuzzy set is just equal to 1 minus the degree of membership. That is to say, there may be some hesitation degree. So, as a generalization of fuzzy sets, the concept of *Intuitionistic Fuzzy Sets* (*IFS*) was introduced by Atanassov [1] in 1985 . Bustince and Burillo [4] showed that this notion coincides with the notion of Vague Sets (VSs).

The intuitionistic fuzzy sets were defined as an extension of the ordinary fuzzy sets [3]. As opposed to a fuzzy set in X, given by

$$A = \{(x, \mu_A(x)) \mid x \in X\} \tag{3}$$

where $\mu_A : X \to [0,1]$ is the membership function of the fuzzy set *A*, an intuitionistic fuzzy set *B* is given by

$$B = \{(x, \mu_B(x), \nu_B(x)) \mid x \in X\} \tag{4}$$

where $\mu_B : X \to [0,1]$ and $\nu_B : X \to [0,1]$ are such that

$$0 \le \mu_B + \nu_B \le 1 \tag{5}$$

and $\mu_B, \nu_B \in [0,1]$ denote a degree of membership and a degree of non-membership of $x \in A$, respectively. For each intuitionistic fuzzy set *B* in *X*, we will call

$$\pi_B = 1 - \mu_B - \nu_B \tag{6}$$

''hesitation margin'' (or ''intuitionistic fuzzy index'') of $x \in B$ and, it expresses a hesitation degree of whether *x* belongs to *B* or not. It is obvious that $0 \le \pi_B(x) \le 1$, for each $x \in X$.

Therefore, if we want to fully describe an intuitionistic fuzzy set, we must use any two functions from the triplet [2, 3]:

- Membership function;
- Non-membership function;
- Hesitation margin.

In other words, the application of intuitionistic fuzzy sets instead of fuzzy sets means the introduction of another degree of freedom into a set description (i.e. in addition to μ_B we also have ν_B or π_B). Since the intuitionistic fuzzy sets being a generalization of fuzzy sets give us an additional possibility to represent imperfect knowledge, they can make it possible to describe many real problems in a more adequate way.

3. IFS application in Pattern Recognition

Similarity and distance measures as pattern recognition mechanisms could be considered as dual concepts [17]. On the other hand, as important contents in intuitionistic fuzzy mathematics, similarity measure and distance measure between IFSs have attracted many researchers. Li and Cheng [10] proposed similarity measures of IFSs and applied these measures to pattern recognition. But Liang and Shi [11] pointed out that Li and Cheng's measures are not always effective in some cases, and made some modifications respectively. At the same time, Szmidt and Kacprzyk [16] proposed four distance measures between IFSs, which were in some extent based on the geometric interpretation of intuitionistic fuzzy sets, and have some good geometric properties. Let $A = \{\langle x, \mu_A(x), \nu_A(x)\rangle | x \in X\}$, $B = \{\langle x, \mu_B(x), \nu_B(x)\rangle | x \in X\}$ be two IFSs in X = {x_1,x_2, . . . ,x_n}. Based on the geometric interpretation of IFS, Szmidt and Kacprzyk proposed the following distance measures between *A* and *B*:

- *Hamming distance:*

$$d^1_{IFS}(A,B) = \frac{1}{2}\sum_{i=1}^{n}(|\mu_A(x_i) - \mu_B(x_i)| + |\nu_A(x_i) - \nu_B(x_i)| + |\pi_A(x_i) - \pi_B(x_i)|) \tag{7}$$

- *Euclidean distance:*

$$e^1_{IFS}(A,B) = \sqrt{\frac{1}{2}\sum_{i=1}^{n}(\mu_A(x_i) - \mu_B(x_i))^2 + (\nu_A(x_i) - \nu_B(x_i))^2 + (\pi_A(x_i) - \pi_B(x_i))^2} \tag{8}$$

As mentioned above, several similarity measures between two IFSs have been proposed so far. Li and Cheng [10] defined the similarity between the two IFSs, *A* and *B* as follows:

$$S_d^p(A,B) = 1 - \frac{1}{\sqrt[p]{n}}\sqrt[p]{\sum_{i=1}^{n}|m_A(i) - m_B(i)|^p} \tag{9}$$

where $m_A(i) = (\mu_A(x_i) + 1 - \nu_A(x_i))/2$, $m_B(i) = (\mu_B(x_i) + 1 - \nu_B(x_i))/2$, and $1 \le p < \infty$.

This measure could be assumed as a right classifier for many applications of pattern recognition in different domains such as medical diagnosis. Nevertheless, some examples could be mentioned in which the Li and Cheng similarity measure does not lead to reasonable results and the acquired classification is not fully correct. On this basis, several similarity measures are proposed [6, 11, 17].

4. Bacillus Colonies Recognition Using IFS

One of the fields which observed many applications of PR is medical diagnosis [5]. There are numerous diagnostic problems which need to identify unknown samples through assigning them to specified patterns (classes). Bacillus colony recognition is an instance of medical diagnostic problems which involves pattern recognition. Recognition in this problem could not be performed with certainty. Hence, we need a PR approach which has uncertainty handling capability.

4.1 Bacillus colony recognition

Microbiologists broadly classify bacteria according to their shape [9, 14]. Most bacteria come in one of three shapes: rod, sphere, or spiral. Rod-shaped bacteria are called *bacilli*. Spherical bacteria are called *cocci*, and spiral or cockscrew-shaped bacteria are called *spirilla*. Some bacteria come in more complex shapes, e.g. plemorphic bacteria can assume a variety of shapes. It must be noted that the bacterial cell wall generally determines the shape of the bacterial cell. Bacteria may be further classified according to whether they require oxygen (aerobic or anaerobic) and how they react to a test with Gram's stain. Bacteria in which alcohol washes away Gram's stain are called *Gram-negative*, while bacteria in which alcohol cause the bacteria's wall to absorb the stain, are called *Gram-positive*. Another system for bacteria classification depends on source of carbon which the bacteria are fed on (autotrophic or heterotrophic)[9, 14].

As noted before, bacillus is a rod-shaped bacterium which is active only in the presence of the oxygen (aerobic bacterium). Bacilli occur mainly in chains, produce spores, and include many saprophytes, some parasites, and the bacterium that causes anthrax [15]. In this research, four intestinal bacilli named *Shigella*, *Salmonella*, *Bacillus Coli* and *Klebsiella*, exhibited in Fig. 1, are considered which have some similarity in culture medium and are Gram-negative.

Shigella *Klebsiella* *Bacillus Coli* *Salmonella*

Fig. 1. Four intestinal bacilli considered in the recognition problem

Pattern recognition of these bacilli is a problem for microbiologists. They use medical experiments such as Gram stain, to classify unknown samples [9, 14]. Hence, they spend much time and energy to identify bacillus type whereas this repetitive task could be done by a medical expert system. Instead of human medical expert, this system exploits medical expertise in pattern recognition. Our target is to use IFS similarity measure as a classifier, to recognize the bacillus type of the samples. It could be used as a classifier in medical diagnosis or generally in a pattern recognition problem which is in need of uncertainty handling.

Shigella [15] is a rod-shaped Gram-negative bacterium (bacillus) that lives in the intestinal tracts of human beings and animals and causes bacillary dysentery. There are four species, all causing dysentery but with varying degrees of severity. Salmonella [15] is also a rod-shaped bacterium found in the intestine that can cause food poisoning, gastroenteritis, and typhoid fever. Other two bacilli considered herein, Klebsiella and Bacillus Coli are also Gram-negative and found in the intestine. Primal features used for these bacilli classification comprise of *macro shape* which is domical in these four bacillus colonies, *single microscopic shape*, *double microscopic shape*, and *existence of flagellum*.

In bacillus colonies recognition, we confront uncertainty in the unknown samples classification and could not state their classes with certainty. Cause of this phenomenon is imprecision and uncertainty in sample's feature existence. To model feature existence uncertainty, each sample's feature could be represented in intuitionistic fuzzy sets with two degrees of membership and non-membership functions. A dataset comprised of 127 different samples were gathered from the Resalat Laboratory, Microbiology Section, located in Tehran, Iran, so as each of sample's features is represented with two degrees of membership and non-membership functions. Some part of it is exhibited in Table 1.

Table 1. Some part of the bacilli data set

Samples	Domical Shape		Single Micro. Shape		Double Micro. Shape		Flagellum	
	$\mu(A_i)$	$v(A_i)$	$\mu(B_i)$	$v(B_i)$	$\mu(C_i)$	$v(C_i)$	$\mu(D_i)$	$v(D_i)$
S_1	**0.85674**	**0**	**0.91253**	**0.028768**	**0.001125**	**0.93549**	**0.9535**	**0.037096**
S_2	**0.81677**	**0.10944**	**0.76638**	**0.12143**	**0.10512**	**0.81002**	**0**	**0.8559**
S_3	**0.79802**	**0.084328**	**0.63959**	**0.12573**	**0**	**0.99151**	**0.92964**	**0.028088**
S_4	**0.69909**	**0.14805**	**0.69518**	**0.15**	**0.1861**	**0.63366**	**0.053622**	**0.92836**
S_5	**0.91184**	**0.081481**	**0.89523**	**0.037934**	**0.03075**	**0.94277**	**0.80829**	**0.046994**
S_6	**0.80501**	**0.15812**	**0.9569**	**0**	**0.023439**	**0.80122**	**0.037573**	**0.86546**

Gathered data set was investigated by microbiologists and its samples were interpreted, so as four classes were extracted which their feature vectors are represented in Table 2. Each of features in class feature vector is also represented through IFS structure.

Table 2. Imprecise feature values of four classes (patterns)

Classes	Domical Shape		Single Micro. Shape		Double Micro. Shape		Flagellum	
	$\mu(A_i)$	$v(A_i)$	$\mu(B_i)$	$v(B_i)$	$\mu(C_i)$	$v(C_i)$	$\mu(D_i)$	$v(D_i)$
Bacillus Coli	0.9	0.05	0.9	0	0	1	0.9	0.06
Shigella	0.9	0.08	0.9	0.05	0.05	0.92	0.08	0.9
Salmonella	0.8	0.1	0.8	0.1	0.1	0.85	0.9	0.01
Klebsiella	0.8	0.15	0.7	0.15	0.2	0.75	0.1	0.85

4.2 Experimental Results

To recognize the appropriate class for each unknown sample, the similarity measure between bacillus colonies' intuitionistic fuzzy sets was selected, so as the Eq. (9) is used in which the *p* parameter is set to one. So we have:

$$Simlarity\ (S,P)=1-\frac{1}{n}\sum_{i=1}^{n}\left|m_S(i)-m_P(i)\right| \tag{10}$$

where $m_S(i)=(\mu_S(x_i)+1-v_S(x_i))/2, m_P(i)=(\mu_P(x_i)+1-v_P(x_i))/2$ and S and P represent sample and pattern feature vectors, respectively. Taking into account the membership and non-membership degrees of feature existence, this similarity measure could represent a rational uncertainty modeling which leads to approximate reasoning in PR. To review the recognition process, an example is denoted. According to the four bacillus colonies patterns exhibited in Table 2, the following classes are depicted with IFSs:

Bacillus Coli = {(0.9,0.05),(0.9,0),(0,1),(0.9,0.06)},

Shigella = {(0.9,0.08),(0.9,0.05),(0.05,0.92),(0.08,0.9)},

Salmonella= {(0.8,0.1),(0.8,0.1),(0.1,0.85),(0.9,0.01)}, and

Klebsiella = {(0.8,0.15),(0.7,0.15),(0.2,0.75),(0.1,0.85)}

If we consider a sample *A* = {(0.69909, 0.14805),(0.69518, 0.15),(0.1861, 0.63366), (0.053622, 0.92836)}, based on the Eq. (10), *Similarity*(*Bacillus Coli*,*A*)= 0.63488, *Similarity*(*Shigella*,*A*)= 0.86863, *Similarity*(*Salmonella*,*A*)= 0.70363, and *Similarity*(*Klebsiella*,*A*)= 0.95863. From the above results, it is evident that sample *A* has the most similarity to Klebsiella pattern and therefore could be classified as a Klebsiella bacillus colony.

After performing the aforementioned process and calculations completion, the similarity degree of the sample against the whole feature vector of a class would be obtained. This process was iterated to acquire the similarity of the sample versus other classes. Having compared the final results, the class which has obtained the biggest similarity value or in other words, has the most similarity to the sample would be recognized as the appropriate pattern for the sample. After testing phase completion, all the unknown samples were classified. The final results of the classification have been checked with microbiologists' opinions. Table 3 summarizes the classification result of ten samples which are compared with the microbiologists' opinions.

Table 3. Classification result based on the similarity measure versus microbiologists opinions

S. No.	***Bacillus Coli***	***Shigella***	***Salmonella***	***Klebsiella***	***Microbiologists***
S_1	0.964676	0.755116	0.939104	0.684104	*Bacillus Coli*
S_2	0.725392	0.933597	0.701097	0.868597	*Shigella*
S_3	0.96043	0.779048	0.945679	0.715547	*Salmonella*
S_4	0.669741	0.903491	0.736889	0.979389	*Klebsiella*
S_5	0.949687	0.739669	0.928161	0.692169	*Bacillus Coli*
S_6	0.731529	0.957779	0.725873	0.880873	*Shigella*
S_7	0.910958	0.729708	0.948352	0.752326	*Klebsiella*
S_8	0.701731	0.928405	0.733405	0.92783	*Klebsiella*
S_9	0.929233	0.747983	0.973862	0.718862	*Salmonella*
S_{10}	0.684561	0.905685	0.717416	0.959916	*Salmonella*

Having compared all the classified samples with medical experts' ideas, we observed that 96.7 percent of the unknown samples have been classified correctly. The bacillus recognition based on the similarity measure for a number of samples is represented in Fig. 2, in which the most similar class has obtained a higher rank, meanwhile compared with the microbiologists' opinions. Also, the compatibility between microbiologist' opinions and IFS classifier is represented in Fig. 3. Since the classifier built on the IFS similarity measure, satisfactorily classified the unknown uncertain and imprecise samples and confronted with the problem uncertainty, this result is acceptable and rational for the first step.

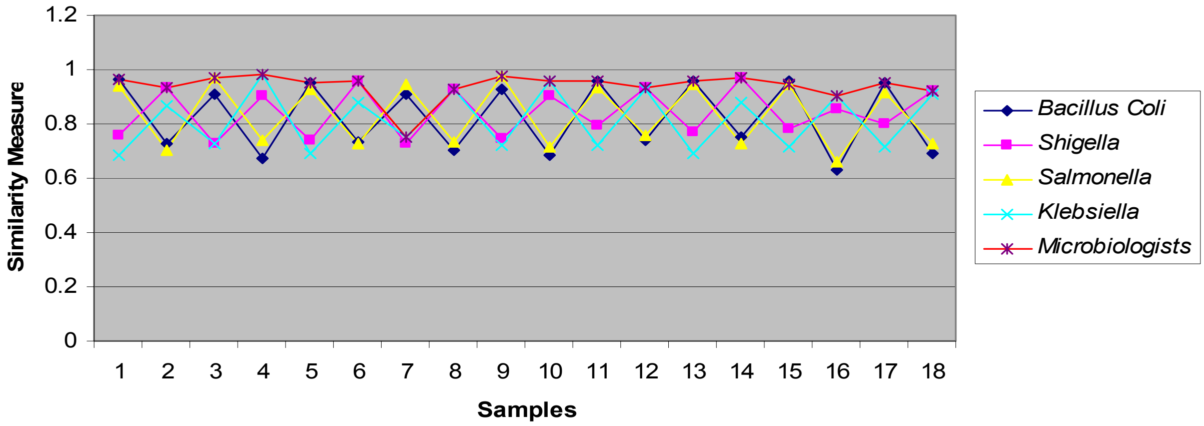

Fig. 2. Bacillus colonies recognition based on the similarity compared with microbiologists' opinions

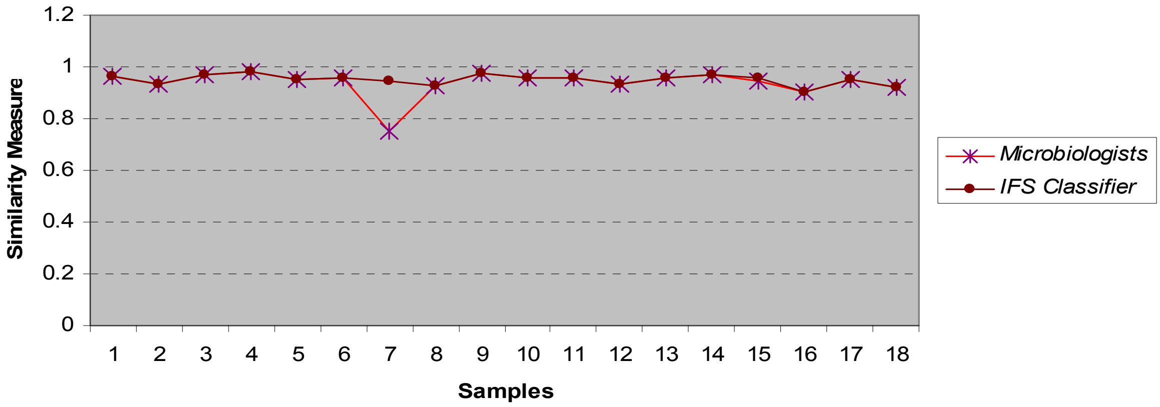

Fig. 3. Bacillus colonies' IFS classifier performance compatibility with microbiologists' opinions

As a main advantage of this approach, it could be mentioned that experts (such as microbiologists in our study) have a great reasonable insight into the uncertainty quantification process and how the expert system works. In fact, this approach looks at the problem and its information realistically, so that first, models the existed information through quantifying the problem uncertainty which leads to the membership and non-membership functions and then classifies the unknown samples according to the similarity to predefined classes.

Conclusion

In many practical situations, we encounter imprecise and uncertain hypotheses which could not be expressed through certain values. Hence, it is more convenient and reasonable to represent them in Fuzzy Sets. Intuitionistic Fuzzy Sets as an extension of regular Fuzzy Sets could represent the uncertainty more accurately. In this paper, a similarity measure based on Intuitionistic Fuzzy model is used to quantify a pattern recognition uncertainty. Having obtained the membership and non-membership functions for feature vectors, the similarity between the uncertain pattern feature vectors and samples are computed.

This mechanism could be used as a classifier for imprecise and uncertain hypotheses with a realistic look at problem uncertainty. In this approach, we first quantify problem uncertainty and then examine the similarity between unknown pattern and the classes using the appropriate membership and non-membership functions. As a practical application of this similarity measure in classification, it was applied in bacillus colony recognition. Our experimental results showed that the classifier built on this similarity measure could satisfactorily classify the unknown uncertain and imprecise samples correctly.

Acknowledgements

The authors would like to thank Ms. Hoda Davarzani, Ph.D. student at the Tarbiat Modares University for her precious assistance in data analysis of our research.

References

[1] K. Atanassov, Intuitionistic fuzzy sets, *Fuzzy Sets and Systems* **20**(1) (1986) 87-96.
[2] K. Atanassov, More on intuitionistic fuzzy sets, *Fuzzy Sets and Systems* **33**(1) (1989) 37-46.
[3] K.T Atanassov, Intuitionistic Fuzzy Sets, Physica-Verlag, Heidelberg, 1999.
[4] H. Bustince, P. Burillo, Vague sets are intuitionistic fuzzy sets, *Fuzzy Sets and Systems* **79**(3) (1996) 403-405.
[5] S.K. De, A. Biswas, R. Roy, An application of intuitionistic fuzzy sets in medical diagnosis, *Fuzzy Sets and Systems* **117**(2) (2001) 209-213.
[6] L. Dengfeng, C. Chuntian, New similarity measures of intuitionistic fuzzy sets and applications to pattern recognitions, *Pattern Recognition Letters* **23** (2002) 221-225.
[7] P.A. Devijer, J. Kittler, Pattern Recognition, A Statistical Approach, Prentice-Hall, London, England, 1982.
[8] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, 2nd ed, Wiley-Inetscience, New York, 2000.
[9] H. Facklam, M. Facklam, Bacteria, Twenty-First Century, 1995.
[10] D. Li, C. Cheng, New similarity measures of intuitionistic fuzzy sets and application to pattern recognition, *Pattern Recognition Letters* **23**(1-3) (2002) 221-225.
[11] Z. Liang, P. Shi, Similarity measures for intuitionistic fuzzy sets, *Pattern Recognition Letters* **24** (2003) 2687–2693.
[12] H.B. Mitchell, Pattern recognition using type-II fuzzy sets, *Information Sciences* **170** (2005) 409–418.
[13] S.K. Pal, P. Mitra, Pattern Recognition Algorithms and Data Minig, Chapman and Hall/CRC, Florida, 2004.
[14] E.R. Ricciuti, Microorganisms: The Unseen World, Blackbirch, 1993.
[15] N. Sankaran, Microbes and People: An A-Z of Microorganisms in Our Lives, Oryx, 2001.
[16] E. Szmidt, J. Kacprzyk, Distances between intuitionistic fuzzy sets, *Fuzzy Sets and Systems* **114**(3) (2000) 505-518.
[17] W. Wang, X. Xin, Distance measure between intuitionistic fuzzy sets, *Pattern Recognition Letters* **26** (2005) 2063-2069.
[18] L.A. Zadeh, Fuzzy sets, *Inform. Control* **8** (1965) 338-353.
[19] L.A. Zadeh, Fuzzy Logic, neural networks, and soft computing, *Comm. ACM* **37** (1994) 77-84