NIFS 4 (1998) 4, 62-67

# INTUITIONISTIC FUZZY VERSIONS OF K - NN METHOD AND THEIR APPLICATION TO RESPIRATORY DISTRESS SYNDROME DETECTION

S. T. Hadjitodorov

Department of Biomedical Informatics, Central Laboratory of Biomedical Engineering, Bulgarian Academy of Sciences, Acad. G. Bonchev str. bl.105, 1113 Sofia, Bulgaria
E-mail: STHADJ@ARGO.BAS.BG

**KEYWORDS**: Pattern recognition; Fuzzy classification; Intuitionistic fuzzy sets; Fuzzy K - NN; Respiratory Distress Syndrome (RDS) detection

**ABSTRACT:** Intuitionistic fuzzy versions of one of the basic statistical nonparametrical methods, the K - NN method, are proposed. The inclusion of fuzzy information is made through modification of the distances by means of the pattern degrees of membership and nonmembership to the classes to which the reference pattern belongs. Thus for each of the labeled samples its typicalness and nontypicalness are taken into consideration. The versions are applied to Respiratory Distress Syndrome (RDS) detection.

## 1. INTRODUCTION

The K - nearest neighbors ( K- NN ) method is one of the most often applied and most accurate methods in pattern recognition [1,3,5,6,9,10,11,12,]. The basic assumption of this statistical and nonparametrical method is that inside of the small hypersphere around a given pattern the probability density function is approximately constant [5]. For that reason the classifiers have to be restricted to a small value of K. However in cases of small sample size and overlapping classes this requirement causes accuracy decreasing. A lot of modifications of the method have been published which improve the method with respect to increasing the classification accuracy and minimization of the calculations. Several fuzzy versions have been also developed. In [11,12], for example, the degree of membership of a given pattern (vector) to a given class is determined by weighted averaging of the degree of membership of the K-NNs. As weights, the reciprocal values of the distances between the vector and its K - NNs are used. In [10] the degree of membership of a pattern with unknown classification to a class is calculated as a mean of the class membership of the K - NNs of the pattern. Some generalizations of this method are developed in [1,3]. In [14] an attempt to combine the advantages of the fuzzy description of the pattern and classes is made. The idea is to include expert experience and knowledge in pattern recognition in a new way. In this way the pattern distribution in the space is implicitly taken into consideration. The goal is to cope with cases of small sample size and overlapping classes. A transformation of the distances using fuzzy description (degree of membership) is proposed and it is proved that the probability of error in that case is less than or equal to the probability of error in case of classical (nontransformed distances) K - NN. Some fuzzy decision rules were combined with these modified distances. The combination of statistical and fuzzy approaches in pattern recognition intuitively leads to an improvement of the recognition results. In [17] an improvement of this method by means of application of degree of nonmembership besides the degree of membership is described but not tested. Here that version and a new one are tested and applied to the Respiratory Distress Syndrome (RDS) detection problem.

## 2. THEORETICAL BACKGROUND

Let the sample ( training, reference ) set of pattern is:

$$X = \{x_1, x_2, \dots x_N\};\ x_l \in X;\ x_l \in \Re^n,\ l = 1, 2, \dots, N \tag{1}$$

Each pattern $x_l \in \Re^n$ is described by ***n*** features. The feature set Z is :

$$Z = \{Z_1, Z_2, \dots, Z_n\};\ Z_j \in Z;\ j = 1,2,\dots,n, \tag{2}$$

The set of the classes defined over $\Re^n$ is:

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_M\};\ \omega_i \in \Omega;\quad i = 1, 2, \dots, M \tag{3}$$

The fuzzy description can be used not only in cases of insufficient and small training(reference) set but it can be used also in case of large training set in order to improve the classification accuracy. In [17] is proposed the use of the intuitionistic fuzzy sets (IFSs) for classification purposes, i.e. of the degrees of membership $\mu(x)$ and nonmembership $\nu(x)$. It is known that the IFS $A^*$ in E is an object having the form [15,16]

$$A^* = \{< x, \mu(x), \nu(x) >, x \in E\} \tag{4},$$

where $\mu_A : E \to [0,1]$, $\nu_A : E \to [0,1]$ define the degree of membership and nonmembership of the element $x \in E$ to the set A which is a subset of E, respectively, and for every $x \in E$

$$0 \le \mu_A(x) + \nu_A(x) \le 1 \tag{5}$$

We consider the degree of membership as typicalness and confidence for membership to the given class and the degree of nonmembership as nontypicalness and diffidence for membership.

### 2.1. Evaluation of the degrees of membership and nonmembership

The implementation of IFS in a K-NN rule is possible if the $\mu(x)$ and $\nu(x)$ for each pattern from the reference set are known. These memberships can be determined from the sample X in several different ways: on the basis of the geometrical properties of the classes[1,7,8,12]; on the basis of the probabilistic measures and expert knowledge [1,4,5,13,14].

In [17] the following way for evaluating of the degree of membership and nonmembership to the class $\omega_i$ is proposed, to which x (a pattern from the reference set) belongs:

$$\mu_{\omega_i}(x) = e^{-d_m}, \tag{6}$$

where $d_m$ is the Euclidean distance between the pattern x and the mean vector for class $\omega_i$ to which x belongs, $x_{i_m} = \frac{1}{N_i} \sum x_i$ ; $N_i$, is the number of the training patterns belonging to class $\omega_i$;

$$\nu_{\omega_i}(x) = e^{-d_{m_N}}, \tag{7}$$

where $d_{m_N}$ is the Euclidean distance between the pattern x and the mean vector of the class $\omega_{m_N}$ which is the next nearest mean vector among the other classes ( to which x doesn't belong, i.e. the competitive class),

$$d_{m_N} = \min\{d_{m_j}\}; j = 1,2,\dots, M; i \neq j \tag{8}$$

Here we propose an additional way of determining of $\mu(x)$ and $\nu(x)$:

$$\mu_{\omega_i}(x) = \frac{\left(\left\|x - x_{im}\right\|^2\right)^{\left(-\frac{1}{q-1}\right)}}{\sum_{k=1}^{M}\left(\left\|x - x_{km}\right\|^2\right)^{\left(-\frac{1}{q-1}\right)}}, \tag{9}$$

$$\nu_{\omega_i}(x) = \frac{\left(\left\|x - x_{im_N}\right\|^2\right)^{\left(-\frac{1}{q-1}\right)}}{\sum_{k=1}^{M}\left(\left\|x - x_{km}\right\|^2\right)^{\left(-\frac{1}{q-1}\right)}}, \tag{10}$$

where q>1 is a degree of fuzziness.

In such way each class $\omega_i$; $i = 1,2,\dots, M$ is presented as an IFS. The value of $\mu_{\omega_i}(x)$ determine the subset at level to which the pattern belongs, i.e. the typicalness of x for class $\omega_i$ and $\nu_{\omega_i}(x)$ is the nontypicalness. If (5) is not satisfied then $\nu_{\omega_i}(x)$ is evaluated according to:

$$\nu_{\omega_i}(x) = \min\{(1 - \mu_{\omega_i}(x)), \nu_{\omega_i}(x)\} \tag{11}$$

## 2.2 Transformation of the distances

Using the idea of K - NN method, for the classification of a pattern $x_u$ with unknown class-membership , the distances $d_l$ between $x_u$ and the reference patterns from the sample set $x_l$; $l = 1,2,\ldots,N$ are determined. Then the smallest K of them are chosen. No restrictions are imposed on the type of the metrics, in which the distances are calculated. The basic idea of implementing IFS is to transform these distances and to take into account the degree of membership (that fix the typicalness of $x_l$ for the class) and the degree of nonmembership. These degrees are determined according to (6), (7) or (9), (10).

Taking into consideration the fact that the distance $D_l$ is calculated to each reference $x_l$ which belongs to fixed class $\omega_i$ we propose the transformation to be performed according to:

$$D_{l_i} = \frac{d_{l_i}}{\mu_{l_i}} \quad \text{and} \quad D_{l_i}' = d_{l_i} \, . \, \nu_{l_i} \tag{12}$$

After the transformation we obtain two sets $S_k$ and $S_k^{'}$ of K-distances. These transformed distances are proposed to be used for classification. In [14] the following proposition for application of transformed distances $D_l$ is proved:

*PROPOSITION:* If the transformed distances are used as distances in K - NN method then the probability of error is less than or equal to the same in case of nontransformed distances.

Similar proposition could be easily generalized for both distances $D_{l_i}$ and $D^{'}_{l_i}$.

## 2.3. A two-level classification scheme

We propose a simple two level classification scheme. On the first level the classical K-NN rule for the two sets of NN $S_k$ and $S_k^{'}$ is applied, i.e. if $k_1, k_2, \ldots, k_M$ are the number of the patterns (among K) which respectively belong to $\omega_1, \omega_2, \ldots, \omega_M$, then

$$x_u \in \omega_i \text{ if } k_i = \max\{k_1, k_2, \ldots, k_M\}. \tag{13}$$

Let with respect to $S_k$ - $x_u \in \omega_i$, and with respect to $S_k^{'}$ - $x_u \in \omega_{i'}$. Then the decision rule on the second level is:

if $i = i \Rightarrow x \in \omega_i$,

if $i \neq i \Rightarrow$ refusal from classification, (14)

i.e. if the decisions from $S_k$ and $S_k^{'}$ coincide we have simple classification, otherwise we have ambiguous classification and refusal from decision.

## 3. EXPERIMENTAL RESULTS AND CONCLUSIONS

The proposed modifications are applied to Respiratory Distress Syndrome (RDS) detection. The clinical data are collected in the Institute of Obstetrics and Gynecology, Sofia and consists of 85 preterm newborn infants, belonging to two partially overlapping classes:

$\omega_1$ - neonates without RDS - 40 cases

$\omega_2$ - neonates suffering from RDS - 45 cases.

The patterns are described by18 parameters from anamnesis, laboratory tests and clinical observations.. The experiments include testing of the following methods:

- K-NN - classical (crisp) version [5];
- K-NN - fuzzy version [10];
- K-NN - fuzzy version with modified distance [14];
- K-NN - intuitionistic version with modified distances according to (6), (7), (12);
- K-NN - intuitionistic version with modified distances according to (9), (10), (12);

During the testing the data are divided into two sets - reference set consisting of 55 cases and control (test) set consisting of 30 cases. Twenty independent testing sessions have been carried out, i.e. a cross validation testing scheme with 20 different reference and control sets is implemented. The averaged over the 20 independent sessions results are shown on Table 1.

**Table 1**..The averaged accuracy of the experimented K-NN versions

| **K-NN versions** | Number of neighbors | 1 | 3 | 5 | 7 |
|---|---|---|---|---|---|
| Classical version | Accuracy [%] | 93.3 | 93.3 | 94.4 | 95.6 |
| | Mean accuracy [%] | 94.15 | | | |
| Fuzzy version [10] | Accuracy [%] | 94.4 | 94.4 | 94.4 | 94.4 |
| | Mean accuracy [%] | 94.4 | | | |
| Fuzzy version with modified distances [14] | Accuracy [%] | 95.6 | 95.6 | 95.6 | 95.6 |
| | Mean accuracy [%] | 95.6 | | | |
| Intuitionistic version with modified distances (6), (7), (12); | Accuracy [%] | 96.7 | 96.7 | 95.6 | 95.6 |
| | Mean accuracy [%] | 96.15 | | | |
| Intuitionistic version with modified distances (9), (10), (12); | Accuracy [%] | 96.7 | 96.7 | 95.6 | 95.6 |
| | Mean accuracy [%] | 96.15 | | | |

The modified versions with distances' transformation show an improved classification accuracy compared with the classical (crisp) K-NN and classical fuzzy version of K-NN rule. It could be seen that using fuzzy information in this rule, decreases the variance of the accuracy with varying the number of neighbors. The intuitionistic fuzzy versions have the highest accuracy. There is no difference between the proposed two intuitionistic fuzzy versions. The results show that the combination of statistical and fuzzy approaches in pattern recognition leads to an improvement of the recognition accuracy. Moreover using both pattern degree of membership and nonmembership to the class, i.e. the typicalness and confidence for membership and the nontypicalness and diffidence for membership, reduces additionally the classification error. Taking into account the experimental results reported here, it seems promising to apply such classification rules to small consulting systems for clinical practice in the field of RDS detection.

REFERENCES


1. J.C. Bezdek, S.K. Chuali and D.Leep, Generalized K - Nearest Neighbor Rules, Fuzzy Sets and Systems 18 (1986) 237-256.
2. A.F. Blishun, Comparative Analysis of Methods for Fuzziness Measuring, Ann. USSR Acad. Sci. Techn. Cyber. 5 (1988) 152-175, (in Russian).
3. D.Cabello, S.Barro, J.M. Salceda, R. Ruiz and J. Mira, Fuzzy K- Nearest Neighbor Classifiers for Ventricular Arrhythmia Detection, Int. J. Bio-Medical Computing 27 (1991) 77-93.
4. W.J. Dixon, BMDP: Biomedical Computer Programs. P-series (Univ. of Calif. Press, Los Angeles, 1977).
5. K. Fukunaga, Introduction to Statistical Pattern Recognition, (Nauka, Moscow, 1979), ( in Russian).
6. K. Fukunaga and L.D. Hostetler, K- Nearest Neighbor Pattern Classification, IEEE Trans. Inf. Theory, 21 (1975) 285-293.
7. S.T. Hadjitodorov, A Fuzzy Method for Pattern Classification, Avtomatika, Izchislitelna Technika i Avtomatizirany Systemy 4 (1987) 8-11, (in Bulgarian).
8. B. Hussien, R. McLaren and S. Bleha, An Application of Fuzzy Algorithms in a Computer Access Security System, Pattern Recognition Letters 9 (1989) 39-43.
9. A. Jousselin and B. Dubuisson, A Link between K-NN rules and Knowledge Based Systems by Sequence Analysis, Pattern Recognition Letters 6 (1987) 287-295.
10. A.A. Jozwik, A Learning Scheme for a Fuzzy K - NN Rule, Pattern Recognition Letters 1 (1983) 287-289.
11. J.M. Keller, M.R. Gray and J.A. Givens , Fuzzy K -Nearest Neighbor Algorithm, IEEE Trans. on Systems, Man and Cyber , SMC 15 (1985) 580-585.
12. J.M. Keller and J.A. Givens, Membership Function Issues in Fuzzy Pattern Recognition, in: Proc Int.Conf.on Cybernetics and Society, (USA, Arizona, 1985) 210 - 214.
13. V.T. Kissiov and S.T. Hadjitodorov, The K - NN Method by Fuzzy Description of the Classes, Avtomatika, Izchislitelna Technika i Avtomatizirany Systemy 3 (1988) 5 - 7, ( in Bulgarian ).
14. V.T. Kissiov and S. Hadjitodorov, A Fuzzy version of the K-NN Method, Fuzzy Sets and Systems 49, (1992) 323-329.
15. K. Atanassov, Intuitionistic Fuzzy Sets, Fuzzy Sets and systems 20,(1986), 87-96.
16. K. Atanassov, More on Intuitionistic Fuzzy Sets, Fuzzy Sets and systems 33,(1989), 37-46.
17. Hadjitodorov, S. An intuitionistic fuzzy sets application of the K-NN method. *Notes of Intuitionistic Fuzzy Sets*, vol 1, No 1, 1995, 66-69.