Application of K-Nearest Neighbor Rule in the Case of Intuitionistic Fuzzy Sets for Pattern Recognition

Todorova L., Vassilev P.*

Centre of Biomedical Engineering
Bulgarian Academy of Sciences
105 Acad. G. Bonchev Str., 1113 Sofia, Bulgaria
E-mail: peter.vassilev@gmail.com

Summary: In the present paper an algorithm based on the k-nearest neighbors (KNN) rule modified for the case of intuitionistic fuzziness is proposed. The algorithm calculates the degrees of membership, non-membership and indeterminacy for each new element that needs to be classified. The choice of the KNN rule is due to the high precision of the method in decision making for pattern recognition problems, while the apparatus of the intuitionistic fuzzy sets is used to describe more adequately the considered objects and allows for pattern recognition with non-strict membership of the patterns.

Keywords: Intuitionistic fuzzy sets, k-nearest neighbors rule, Pattern recognition

1. INTRODUCTION

Intuitionistic fuzzy sets (IFS) were introduced by Atanassov in 1983 [1]. They are an extension and generalization of the fuzzy sets (FS) which Zadeh introduced in [2]. FS and subsequently IFS are mainly used to address problems with imprecise or incomplete data. An IFS is usually denoted as an ordered triple, e.g. \( \langle x, \mu_A(x), \nu_A(x) \rangle \), where \( x \) is an element of some universal set \( E \), \( A \) is a subset of \( E \) and \( \mu_A(x) \), \( \nu_A(x) \) denote the degree of membership and non-membership of the element \( x \) to the set \( A \). The mappings

\[
\mu : E \to [0, 1], \quad \nu : E \to [0, 1]
\]

are chosen so that for all it is fulfilled that:

\[
\mu_A(x) + \nu_A(x) \leq 1
\]

By choosing \( \nu_A(x) \stackrel{\text{def}}{=} 1 - \mu_A(x) \), we obtain the FS.

* Corresponding author
In the case of IFS the difference \( \pi_A(x) \overset{\text{def}}{=} 1 - \mu_A(x) - \nu_A(x) \) is called degree of indeterminacy and represents the level of uncertainty associated with the information. One of the most used geometric interpretations is the interpretation triangle introduced in [3]. It is shown on Fig. 1.

Fig. 1 The interpretation triangle

IFS have been successfully used to improve pattern recognition (see e.g. [4, 5, 6, 7]).

The KNN classifier has been used in different areas with great success [8, 9, 10]. KNN classification algorithm tries to find the K nearest neighbors of a given unclassified pattern \( x_0 \) and uses a majority vote to determine the class label of \( x_0 \). An example is shown on Fig.2.

Fig. 2 In the inner circle the case \( k=3 \) has been shown, the bigger circle corresponds to \( k=5 \)

This simple and easy-to-implement method can still yield competitive results even compared to the most sophisticated machine learning methods. Because of this the KNN rule is one of the most used and precise classification method based on distance functions.
The high performance of the two approaches led to the idea of their combination to develop an improved recognition rule.

2. ALGORITHM

In the present work we propose an algorithm which determines the degree of membership, degree of non-membership, degree of indeterminacy in the sense of IFS using the KNN rule. For clarity we will consider the case of two classes – class $\omega_1$ and class $\omega_2$. The patterns belonging to class $\omega_1$ we will denote by $x_i^1$, the patterns belonging to class $\omega_2$ we will denote by $x_i^2$.

Description of the algorithm:

Step 1: We identify etalons (by etalon here and further we denote a typical or representative pattern for the given class) for each of the considered classes. Let them be $\varepsilon_1$ and $\varepsilon_2$ for class $\omega_1$ and class $\omega_2$, respectively.

Step 2: We find the closest to $\varepsilon_1$ pattern $x_i^2$ from class $\omega_2$. We will denote the distance between $\varepsilon_1$ and $x_i^2$ by $r_1$. We construct a circle $C_{\min}^1$ with center $\varepsilon_1$ and radius $r_1$. We make an analogous construction for the class $\omega_2$. Obviously, the disks bounded by the so-constructed circles contain only images from the respective class (e.g. – in the disk bounded by $C_{\min}^1$ there are only images from class $\omega_1$). The degrees of membership for the patterns in these areas are $\mu_i^1 \overset{\text{def}}{=} 1$ and $\nu_i^2 \overset{\text{def}}{=} 0$ (respectively $\mu_i^2 \overset{\text{def}}{=} 1$ and $\nu_i^2 \overset{\text{def}}{=} 0$ for class $\omega_2$).

Step 3: We find the furthest from $\varepsilon_1$ pattern $x_i^1$ from class $\omega_1$. We will denote the distance between $\varepsilon_1$ and $x_i^1$ by $R_1$. We construct a circle $C_{\max}^1$ with center $\varepsilon_1$ and radius $R_1$. We make an analogous construction for class $\omega_2$. Obviously, each of the disks bounded by the so-constructed circles contains all patterns (and not only them) of the respective class (e.g. in the disk bounded by $C_{\max}^1$, there are all images from class $\omega_1$, as well as images from $\omega_2$).

Step 4: Let $\varepsilon_1 \varepsilon_2 \cap C_{\min}^1 = p^1$ and $\varepsilon_1 \varepsilon_2 \cap C_{\min}^2 = p^2$. Let $m$ be (preliminary selected) natural number $s = \frac{\text{dist}(p^1, p^2)}{m}$. We construct
the concentric circles \( C_i^1(\varepsilon_1, r^1 + i\varepsilon) \) and \( C_i^2(\varepsilon_2, r^2 + i\varepsilon) \) (see Fig. 3.)

![Fig. 3 The result of Step 4](image)

**Step 5:** In all so-generated annuli \( V \) from both classes we find the number of contained patterns and let

\[
n_{\text{min}} = \begin{cases} 
2 & \text{if } \min n_i < 2; \\
\min n_i & \text{if } \min n_i \geq 2 \geq 2.
\end{cases}
\]

We choose an odd \( k < n_{\text{min}} \). For every pattern \( x_i^\delta \) (where \( \delta \) denotes the class) we determine the degree of membership by the formula:

\[
\mu(x_i^\delta) = \frac{n_{\delta_i}}{n_{\delta_i} + n_{\overline{\delta_i}} + 1}
\]

where \( \overline{\delta} \) denotes the other class and \( \delta_i, \overline{\delta_i} \) refer to the \( j \)-th annuli for the class \( \delta \).

**Step 6:** For every \( x_i^\delta \) we determine the degree of non-membership by the formula:

\[
\nu(x_i^\delta) = \frac{n_{\overline{\delta}}}{n_{\delta} + n_{\overline{\delta}} + 1}
\]

**Step 7:** For every \( x_i^\delta \) we determine the degree of indeterminacy by the formula:

\[
\pi(x_i^\delta) = 1 - \mu(x_i^\delta) - \nu(x_i^\delta)
\]

**Step 8:** For every annulus we determine the aggregated values for the degrees following the formulas:

\[
\mu(V_i) = \left(1 + \frac{i\varepsilon}{R_{\delta}}\right)^{-1} \frac{1}{t} \sum_{i=1}^{t} \mu(x_i^\delta)
\]

\[
\pi(V_i) = \frac{1}{t} \sum_{i=1}^{t} \pi(x_i^\delta)
\]
\[ \nu(V_i) = 1 - \mu(V_i) - \pi(V_i) \]

**Step 9:** We check whether the desired recognition accuracy has been reached. If "yes" – we go to Step 10, if "no" – return to Step 4 and select another value for \( m \).

**Step 10:** We determine the degrees of membership, non-membership and indeterminacy of a pattern with unknown classification. To each pattern with unknown classification we assign the degrees of the annulus in which it is contained.

*Flowchart of the algorithm*

3. CONCLUSION

The proposed algorithm is appropriate for solving pattern recognition problems in medicine, since it combines the precision of the \( KNN \)
rule with the apparatus of the IFS, which reflects the fact that one and the same illness develops differently in different patients. This particularity of the data obtained in medical research is accounted for by the different degrees of membership of the patient to a given class corresponding to the specific disease.

The accuracy of the algorithm depends on two parameters – \( k \) (the number of nearest neighbors in the KNN rule) and \( s \) (the width of the annuli). For every particular problem these parameters are adjusted until a maximum precision is achieved, which makes the algorithm flexible and suitable for a wide range of applications.

REFERENCES

1. Atanassov K., Intuitionistic fuzzy sets, VII ITKR's Session, Sofia, June, 1983