Classification with Fuzzification Optimization Combining Fuzzy Information Systems and Type-2 Fuzzy Inference

Martin Tabakov *, Adrian Chlopowiec , Adam Chlopowiec and Adam Dłubak

Department of Computational Intelligence, Wroclaw University of Science and Technology, 50-370 Wroclaw, Poland; 254517@student.pwr.edu.pl (A.C.); 254518@student.pwr.edu.pl (A.C.); adam.dlubak@gmail.com (A.D.)

Abstract: In this research, we introduce a classification procedure based on rule induction and fuzzy reasoning. The classifier generalizes attribute information to handle uncertainty, which often occurs in real data. To induce fuzzy rules, we define the corresponding fuzzy information system. A transformation of the derived rules into interval type-2 fuzzy rules is provided as well. The fuzzification applied is optimized with respect to the footprint of uncertainty of the corresponding type-2 fuzzy sets. The classification process is related to a Mamdani type fuzzy inference. The method proposed was evaluated by the F-score measure on benchmark data.

Keywords: fuzzy sets; interval type-2 fuzzy sets; fuzzy information systems; classification; fuzzification optimization

1. Introduction

Nowadays, machine learning and its applications are growing rapidly. Classification is one of the major machine learning problems. As there are various kinds of data, the need for robust techniques has become essential, especially in real-world applications [1,2]. What is more, often the data are not well-defined, and are vague or imbalanced, which is an additional obstacle. Therefore, fuzzy techniques became the right solution. We may divide these techniques into two primary groups: those related to the classical type-1 fuzzy sets, introduced by Zadeh [3] and those related to new concepts concerning type-2 fuzzy sets [4,5]. In practice, the interval type-2 fuzzy sets [6], a particular case of type-2 fuzzy sets, are commonly used for their reduced computational cost and easy implementation. What is more, researchers have proved that interval type-2 fuzzy concepts are better able to handle uncertainties than type-1 fuzzy approaches [6–11]. Interval type-2 fuzzy sets are very effective, as they provide better generalisations when it is difficult to determine the exact membership functions of the fuzzy sets applied.

On the other hand, to deal with information issues, data discovery techniques, assumed as the computational process of discovering patterns in data, are widely used to induce knowledge. Information systems and rough sets, introduced by Pawlak [12–14], are applied to represent knowledge. Therefore, it seems that the combination of fuzzy techniques and data discovery is very appropriate for the analysis of complex data such as medical data [15,16].

That is how the concept of fuzzy information systems came out. If each attribute of an information system is related to a fuzzy set, a fuzzy information system is defined [17]. Fuzzy information systems find their applications in the field of decision-making [18,19]. In recent research, a novel multi-criteria approach with application in investments was proposed [20]. Many other applications also can be found concerning rule extraction [19,21], feature selection and classification [22–25]. Mathematical properties of fuzzy information systems are under consideration as well [17,26]. The concept of rough sets was extended by analyzing the properties of lower and upper approximations of fuzzy sets [27–31] with...
corresponding practicable applications. For example, in the significant field of machine learning [32] or even medical image analyses, as an image segmentation approach [33].

In our research, we introduce a fuzzy information system which induces type-1 fuzzy rules and transforms them into interval type-2 fuzzy rules. We assume the information function values to be fuzzy sets, but still applying the original definitions of indistinguishability relation and rough sets defined by Pawlak. Our research experiments consist of investigating different fuzzification procedures, which determine the optimal fuzzy classifier.

The rest of the paper is organized as follows: in Section 2, the benchmark data used and the classification methodology proposed are explained, in Section 3 the classification evaluation is presented, and finally, Sections 4 and 5 draw discussion and conclusions.

2. Materials and Methods

2.1. Materials

In our research, we have applied our classification proposal on well-known benchmark data [34] for binary classification. The benchmark details are given in Table 1.

Table 1. The data sets used.

| Benchmark                      | Number of Attributes | Number of Samples | Class Proportions |
|--------------------------------|----------------------|-------------------|-------------------|
| Blood Transfusion [35]         | 7                    | 748               | 76%/24%           |
| Breast Cancer Data             | 4                    | 569               | 63%/37%           |
| Breast Cancer Wisconsin [36]   | 9                    | 683               | 65%/35%           |
| Data Banknote Authentication   | 4                    | 1372              | 56%/44%           |
| Haberman                       | 3                    | 305               | 73%/27%           |
| Heart                          | 13                   | 303               | 46%/54%           |
| HTRU 2 [37]                    | 8                    | 17898             | 9%/91%            |
| Immunotherapy [38]             | 7                    | 90                | 79%/21%           |
| Indian Liver Patient           | 10                   | 583               | 71%/29%           |
| Ionosphere                     | 34                   | 351               | 64%/36%           |
| Parkinson [39]                 | 22                   | 187               | 74%/26%           |
| Pima Indians Diabetes          | 8                    | 768               | 65%/35%           |
| Vertebral                      | 6                    | 310               | 68%/32%           |

2.2. Methods

This section explains the preliminaries of the type-2 Mamdani fuzzy model [4,40] and information systems rule induction procedure [41,42]. Next, we introduce our proposal of combination information systems with type-2 fuzzy inference to introduce a classification procedure.

2.2.1. Fuzzy Sets and Interval Type-2 Fuzzy Sets

A fuzzy set or type-1 fuzzy set F consists of a domain X of real numbers together with a function \( \mu_F: X \rightarrow [0, 1] \), [3] i.e.,:

\[
F = \mu_F \int_X / x, x \in X
\]  

Here, the integral denotes the collection of all points \( x \in X \) with associated membership grade \( \mu_F(x) \in [0, 1] \). The function \( \mu_F \) is also known as the membership function of the fuzzy set F, as its value represents the grade of membership of the elements of X to the fuzzy set F. The idea is to use membership functions to describe imprecise or vague information.
By expanding this concept with the assumption that membership function values can be fuzzified themselves, the idea of type-2 fuzzy sets was introduced. A type-2 fuzzy set, denoted as \( \tilde{F} \), is defined as follows [4]:

\[
\tilde{F} = df \int_{x \in X} \int_{u \in J_x} \mu_{\tilde{F}}(x, u)/(x, u), J_x \subseteq [0, 1]
\]  

(2)

where \( \bigcup \) denotes union over all admissible \( x \) and \( u \). Interval type-2 (IT2) fuzzy sets [4], a special case of type-2 fuzzy sets, are the most widely used because of their acceptable computational complexity and easy interpretation. Uncertainty about \( \tilde{F} \) is conveyed by the so-called footprint of uncertainty (FOU) of \( \tilde{F} \):

\[
\text{FOU}(\tilde{F}) = df \bigcup_{x \in X} J_x, J_x \subseteq [0, 1]
\]  

(3)

The size of an FOU (the corresponding surface) is directly related to the uncertainty that is conveyed by an interval type-2 fuzzy set, and what follows, an FOU with more area is more uncertain than the one with less area. The upper membership function and lower membership function of \( \tilde{F} \) are two type-1 membership functions \( \tilde{F} \) and \( \tilde{F} \) that bound the FOU, which might be used to describe \( J_x \), i.e.:

\[
J_x = df \ [\mu_{\tilde{F}}(x), \mu_{\tilde{F}}(x)]
\]  

(4)

which leads to the following:

\[
\text{FOU}(\tilde{F}) = df \bigcup_{x \in X} [\mu_{\tilde{F}}(x), \mu_{\tilde{F}}(x)]
\]  

(5)

Figure 1 clearly illustrates the above definitions.

Figure 1. An interval type-2 fuzzy set \( \tilde{F} \).

2.2.2. Mamdani Type-2 Fuzzy System

In Figure 2, we show the information flow within an IT2 fuzzy system. It is very similar to its type-1 analogue. The major difference is that because of the IT2 fuzzy sets used in the rulebase, the outputs of the inference engine are IT2 fuzzy sets too. Therefore, a type reducer [6, 43] must be applied to convert them into type-1 fuzzy sets to enable the defuzzification procedure.

Figure 2. Information flow in a type-2 fuzzy system.
Below, we give a brief description of the basic steps of the computations in an IT2 fuzzy system [40]. Let consider the rule base of an IT2 fuzzy system consisting of \( N \) rules taking the following form:

\[
\bar{R}^n \triangleq \text{IF } (x_1 \text{ is } \tilde{X}_1^n) \circ \ldots \circ (x_i \text{ is } \tilde{X}_i^n) \text{ THEN } y \text{ is } \bar{Y}_j^n
\]

where \( \tilde{X}_i^n \) (\( i = 1, \ldots, I; n = 1, \ldots, N \)) are IT2 fuzzy sets defined over corresponding domains, and \( \bar{Y}_j^n \) (\( j = 1, \ldots, M \)) is an IT2 fuzzy set, assuming a Mamdani type-2 system, which represents the corresponding rule conclusion. The logical operator ‘\( \circ \)’ is defined as a fuzzy conjunction or disjunction, i.e., \( \circ \in \{ \otimes, \oplus \} \), where \( \otimes \) and \( \oplus \) are binary operators (\( t\)-norm and \( s\)-norm respectively) defined over \([0, 1]\) (\( \otimes, \oplus : [0, 1]^2 \rightarrow [0, 1] \)). The \( t\)-norm operator provides the characterization of the AND fuzzy operator, while the \( s\)-norm provides the characterization of the OR fuzzy operator [44]. In our research, we applied the Zadeh’s \( t\)- and \( s\)-norms, which correspond to the \( \min \) and \( \max \) operators, respectively.

Assuming an input vector \( \mathbf{x} = \{x_1, x_2, x_3, \ldots, x_I\} \), typical computations of an IT2 fuzzy system consist of the following steps:

1. Compute the membership intervals of \( x_i \) for each \( \tilde{X}_i^n \), \( [\mu_{\tilde{X}_i^n}(x_i), \mu_{\tilde{X}_i^n}(x_i')] \), \( i = 1, \ldots, I; \ n = 1, \ldots, N \)
2. Compute the firing interval of the \( n\)-th rule:

\[
F^n(\mathbf{x}) = \text{df} \left[ \mu_{\tilde{X}_1^n}(x_1) \circ \ldots \circ \mu_{\tilde{X}_I^n}(x_I) \circ \ldots \circ \mu_{\tilde{X}_I^n}(x_I) \right] \equiv [\bar{F}^n, \bar{F}^n]
\]

3. Apply type reduction to combine \( F^n(\mathbf{x}) \) with corresponding rule consequents.

There are some methods of type reduction [6,45,46], but the most commonly used one is the \textit{center-of-set} (COS) type reducer [6] using the Karnik–Mendel algorithms [6,43] or their variants [47,48]. As the whole procedure concerning Mamdani type-2 fuzzy system is well described in the literature, we will omit further details. For more clarity, in Figures 3 and 4, we show the idea of the introduced calculations.

![Figure 3. Firing interval calculation along with footprint of uncertainty (FOU) rule output calculation for a sample rule: IF (\( x_1 \) is \( \tilde{X}_1 \)) AND (\( x_2 \) is \( \tilde{X}_2 \)) THEN \( \bar{Y}_k \).]
A rule induction process with respect to all considered decisions. The procedure consists of the following steps:

1. Introduce an information system with decision attribute,
2. Eliminate information system inconsistency, i.e., objects with the same information function values, but different decision values, applying lower or upper approximation precision analysis,
3. Provide attribute reduct,
4. Apply rule induction and define rules for all considered decisions, which cover the decision problem.

Below, we explain the main algorithms applied in the procedure above with examples. The reduct generation is omitted, as any other attribute selection method can be applied.

### Inconsistency Elimination

Let consider the following information system with decision attribute: $(U, A \cup \{a^*\}, V \cup V_{a^*}, f)$. If inconsistency is found, then Algorithm 1, which uses the lower approximation precision as a decisive factor, should be applied. The lower approximation precision is defined as follows:

$$\gamma_B(X) = \frac{|B \downarrow X|}{|U|}, \quad B \subseteq A$$
Algorithm 1 (inconsistency elimination algorithm):

**Input:** Inconsistent information system  
**Output:** Consistent information system

Let \( x_i \) and \( x_j \) are casing inconsistency \((x_i, x_j \in U)\) and let \( f(x_i, a^*) = d_1, f(x_j, a^*) = d_2 \), then:

1. Define \( X_1 = \{ x \in U : f(x, a^*) = d_1 \} \), \( X_2 = \{ x \in U : f(x, a^*) = d_2 \} \), \((X_1, X_2) \subseteq U\).
2. Calculate \( \gamma_B(X_1) \) and \( \gamma_B(X_2) \) \((B \subseteq A)\).

If \( \gamma_B(X_1) < \gamma_B(X_2) \) then remove from \( U \) \( x_j \) Else remove from \( U \) \( x_i \).

**Example 1.** Let \( U = \{ x_1, x_2, x_3, x_4, x_5, x_6 \} \), \( A = \{a_1, a_2, a_3\} \subseteq A \), \( V_{a_1}, V_{a_2}, V_{a_3}, V_a = \{0, 1, 2, 3, 4\} \) with information function values represented in the matrix below.

\[
\begin{array}{cccc}
  & a_1 & a_2 & a_3 & a^* \\
x_1 & 1 & 4 & 1 & 1 \\
x_2 & 0 & 3 & 2 & 2 \\
x_3 & 1 & 2 & 1 & 1 \\
x_4 & 0 & 2 & 2 & 1 \\
x_5 & 0 & 2 & 0 & 2 \\
x_6 & 1 & 3 & 1 & 1 \\
\end{array}
\]

In the above matrix, we can discover inconsistency as objects \( x_2 \) and \( x_4 \) have the same information function values with respect to \( B \), but different decision values. Therefore, we can apply Algorithm 1:

1. \( X_1 = \{ x_1, x_3, x_4, x_6 \} \), \( X_2 = \{ x_2, x_5 \} \).
2. \( B \downarrow X_1 = \{ x_1, x_3, x_4 \} \), \( \gamma_B(X_1) = 3/6 = 0.5 \), \( B \downarrow X_2 = \{ x_5 \} \), \( \gamma_B(X_2) = 1/6 \approx 0.17 \), which implies \( U = U \backslash \{ x_2 \} \).

Of course, it is only our assumption that different decision values for the same attribute values cause inconsistency. In the decision making process introduced, such decision values could be considered as a set of acceptable values. If that is the case, then the \( d_a^B \) sets defined below wouldn’t be singletons.

Rule Induction

The aim of this step is to induce rules for each decision regarding Algorithm 2 below.

Algorithm 2 (rule induction algorithm):

**Input:** Information system with decision attribute  
**Output:** Corresponding decision rules

**Step 1:** Generate \( M_k \), \( k = 1, 2, \ldots, |U| \) matrixes using the so-called discernibility matrix of the information system, denoted below as \( M(IS) \) and defined as:

\[
M(IS) = \text{df}[m_{ij}]_{i,j=1,\ldots,n} = \{ a \in A : f(x_i, a) \neq f(x_j, a), n = |U| \}
\]

**Step 2:** Define the object implicants,

**Step 3:** Define the set of decision rules.

**Step 1:**
Let \( c_{ij} \) are the elements of \( M(IS) \), \( \hat{c}_i \) are the elements of \( M_k \) (with respect to \( k \), \( k = 1, \ldots, n \); \( n = |U| \)) and \( a^* \) is the decision attribute, then:

For each \( k = 1, \ldots, n \):

1. If \( i \neq k \) then \( \hat{c}_i = \emptyset \),
2. If \( (c_{kj} \neq \emptyset) \) and \( (d_{a^*}^B(x_j) \neq d_{a^*}^B(x_i)) \) then \( \hat{c}_j = \text{df} \text{ } c_{kj} \cap B \) else \( \hat{c}_j = \text{df} \emptyset \).
where $B \subseteq A$ and

$$d^B_{a^*}(x_k) = df \{ v \in V_{a^*} : \exists y \in U (x_k) \text{IND}(B) y \land f(y, a^*) = v \}$$

(10)

For simplicity, let consider the example below that illustrates the generation of matrix $M_1$

**Example 2.** Let assume an information system with decision attribute extended by the corresponding $d^B_{a^*}$ values $(U = \{x_1, x_2, x_3, x_4, x_5\}, A = B = \{a, b, c\}, V_a = V_b = V_c = V_{a^*} = \{0, 1, 2\})$:

|     | a  | b  | c  | a* | $d^x_{a^*}$ |
|-----|----|----|----|----|-------------|
| $x_1$ | 1  | 0  | 1  | 0  | {0}         |
| $x_2$ | 0  | 0  | 0  | 1  | {1}         |
| $x_3$ | 2  | 0  | 1  | 0  | {0}         |
| $x_4$ | 0  | 0  | 1  | 2  | {2}         |
| $x_5$ | 1  | 1  | 1  | 0  | {0}         |

The corresponding discernibility matrix takes the form:

$$
\begin{array}{cccc}
\emptyset & [a, c] & \emptyset & \ldots \\
[a] & [a, c] & \emptyset & \\
[a] & [c] & [a] & \emptyset \\
[b] & [a, b, c] & [a, b] & \emptyset \\
\end{array}
$$

The matrix $M(IS)$ is a symmetric matrix. Next, we can apply Algorithm 2—step 1 and therefore, define the matrix $M_1$ ($k = 1$):

$$
\begin{array}{cccc}
\emptyset & [a, c] & \emptyset & \emptyset \\
[a, c] & \emptyset & \ldots & \emptyset \\
[a] & \emptyset & \emptyset \\
[b] & \emptyset & \emptyset \\
\end{array}
$$

because: $d^B_{a^*}(x_2), d^B_{a^*}(x_4) \neq \{0\}$, therefore $c_{ij} = c_{1j} \cap B (j = 2, 4)$, and so: $c_{12} = \{a, c\} \cap B = \{a\}, c_{14} = \{a\} \cap B = \{a\}$.

**Step 2:**

The object implicants indicate which attributes and objects are strongly related. The first implicant is obtained from $M_1$ applying the corresponding Boolean algebra reduction rules: Implicant$_1$ (M$_1$): $x_1 \Rightarrow a$. The rest of the implicants are presented below:

- Implicant$_2$ (M$_2$): $x_2 \Rightarrow c$,
- Implicant$_3$ (M$_3$): $x_3 \Rightarrow a$,
- Implicant$_4$ (M$_4$): $x_4 \Rightarrow a \land c$,
- Implicant$_5$ (M$_5$): $x_5 \Rightarrow a \lor b$. 

The corresponding discernibility matrix takes the form:
Step 3:
Finally, using the above implicants, we can generate the target set of decision rules regarding the decision attribute values. Each rule represents one decision and it is obtained as a sum of the object implicants related to that decision, i.e., concerning decision attribute value ‘0’, we have: \( f(x_1,a^*) = f(x_3,a^*) = f(x_5,a^*) = 0 \), and therefore: Rule1 \( \equiv df, f(x_1) \lor f(x_3) \lor (f(x_5) \lor f(x_5,b)) \Rightarrow (\text{decision: 0}) \)
by analogy:
Rule2 \( \equiv df, f(x_2,c) \Rightarrow (\text{decision: 1}) \),
Rule3 \( \equiv df, f(x_4,a) \land f(x_4,c) \Rightarrow (\text{decision: 2}) \).

The main disadvantage of the above decision making approach, in terms of vague or imprecise data, is that the induced rules are crisp, i.e., Rule2 has the following interpretation: if an object of \( x_2 \) type has an information function value for the attribute ‘c’ exactly equal to ‘0’ then make the decision ‘1’. We solve this disadvantage by extending information systems to our assumption of a type-1 fuzzy information system.

### 2.2.4. Type-1 Fuzzy Information System

The major assumption of our research is the interpretation of a fuzzy information system. First, we define the values of the information function as type-1 fuzzy sets. This assumption solves the problem of crisp values by providing a generalisation. Therefore, we propose better data granulation and thus more robustness to vague or imprecise data. Next, by assuming Gaussian distribution for each attribute, we are able to define any information function value as one of the very basic fuzzy sets: low, medium, and high. Therefore, the value of any pair object-attribute is generalised by a fuzzy set. Below in Table 2, we give an illustration of a type-1 fuzzy information system with a decision attribute, according to our assumptions.

| Attribute1       | Attribute2       | Attribute3       | ... | Decision Attribute |
|------------------|------------------|------------------|-----|-------------------|
| Object1 [low, medium, high]  | [low, medium, high] | [low, medium, high] | ... | D1 |
| Object2 [low, medium, high]  | [low, medium, high] | [low, medium, high] | ... | D2 |
| Object3 [low, medium, high]  | [low, medium, high] | [low, medium, high] | ... | D3 |

where:
\[
\mu_{\text{medium}}(a) = df \left[ 1 - e^{-\frac{(a-a_0)^2}{2\sigma^2}} : a < a_0 \right] \\
\mu_{\text{low}}(a) = df \left\{ \begin{array}{ll} 1 - e^{-\frac{(a-a_0)^2}{2\sigma^2}} & : a < a_0 \\ 0 & : a \geq a_0 \end{array} \right. \\
\mu_{\text{high}}(a) = df \left\{ \begin{array}{ll} 1 - e^{-\frac{(a-a_0)^2}{2\sigma^2}} & : a \geq a_0 \\ 0 & : a < a_0 \end{array} \right. 
\]

If a pair object\(_i\)-attribute\(_j\) has a value defined as low, for example, it means that:
\[
\mu_{\text{low}}(f(\text{object}_i, \text{attribute}_j)) \geq \max \{ \mu_{\text{medium}}(f(\text{object}_i, \text{attribute}_j)), \mu_{\text{high}}(f(\text{object}_i, \text{attribute}_j)) \}.
\]

It is important to note that our assumption was also the direct application of the original Skowron and Suraj algorithm [41,42] for rule induction. As a result, we did not have to modify the introduced indistinguishability relation (IND) to define fuzzy rough
sets or modify the quality of the approximation introduced by Equation (8). We directly used the fuzzified values of the information function as if they were crisp values. This means that object\(_i\) is related to object\(_j\) with respect to IND in accordance with our model if the values of the corresponding attributes define the same fuzzy sets. For example, below we show the partition \(P\) of the set of objects \{Object\(_1\), \ldots, Object\(_5\)\} implied by IND for a sample information system given in Table 3.

**Table 3.** Sample information system with five objects and two attributes.

|       | Attribute\(_1\) | Attribute\(_2\) |
|-------|-----------------|-----------------|
| Object\(_1\) | low             | high            |
| Object\(_2\) | medium          | low             |
| Object\(_3\) | medium          | low             |
| Object\(_4\) | high            | high            |
| Object\(_5\) | low             | high            |

The corresponding partition with respect to IND has three equivalence classes:

\[
P_{/IND(Attribute1, Attribute2)} = \{\{Object1, Object5\}, \{Object2, Object3\}, \{Object4\}\}.
\]

The above assumption made possible the use of classical algorithms introduced in information systems theory, while at the same time generalizing the information given by the attributes. The use of membership degrees inferred from corresponding membership functions is performed after the induction of rules and their transformation into fuzzy rules, in accordance with the assumptions of the fuzzy model used. Of course, such an assumption may be too strong a generalization. We solve this problem by introducing various possibilities of fuzzification. Defining additional fuzzy sets and transforming them into interval type-2 fuzzy sets with the corresponding optimization, depending on the examined dataset (see Table 4 below).

Next, we define the decision attribute values as fuzzy sets, which affect the degree of decision. Any induced rules from such an information system can be easily transformed into fuzzy rules for a type-1 Mamdani fuzzy system. For example, if we have pairs: (object\(_1\), attribute\(_1\)): low, (object\(_2\), attribute\(_1\)): high, (object\(_3\), attribute\(_2\)): medium and a rule which defines decision \(D\) as:

\[
(f(object1, attribute1) \land f(object2, attribute1)) \lor f(object3, attribute2),
\]

then we can transform it into the fuzzy rule:

If \(((f(object1, attribute1) \text{ is } low) \odot (f(object2, attribute1) \text{ is } high)) \oplus (f(object3, attribute2) \text{ is } medium)\) Then \(D\).

2.2.5. Transformation into Type-2 Fuzzy Sets

To apply type-2 fuzzy sets in the fuzzy rule induction procedure described, we propose a simple extension. After induction of type-1 fuzzy rules, we modify the membership functions applied by changing the standard deviation of the ‘medium’ fuzzy set. By doing so, we can define the bounds of the FOU for the corresponding type-2 fuzzy set medium. This also makes the possibility to define low and high type-2 fuzzy sets as well. The idea is given in Figure 5.
This also makes the possibility to define various possibilities of fuzzification. Defining additional fuzzy sets and transforming assumption may be too strong a generalization. We solve this problem by introducing figures to the FOU for the corresponding type-2 fuzzy set with the corresponding optimization, depending on various possibilities of fuzzification. Defining additional fuzzy sets and transforming assumption may be too strong a generalization. We solve this problem by introducing.

The above extension gives us the possibility to apply the type-2 Mamdani fuzzy inference procedure. In our research, we have applied the procedure proposed to solve classification problems.

2.2.6. Fuzzy Rules Optimization

In our experiments, the basic assumption was to analyze the influence of the shape and the FOU of the type-2 fuzzy sets used in the rule premises, derived from the corresponding fuzzy information system, on the classification accuracy. For this purpose, we have used the type-2 fuzzy sets shown in Table 4. Each of these type-2 membership functions was derived for every attribute of all benchmark data analyzed. Below, we explain the descriptions used in our research:

1. The number of Gaussian functions for fuzzification: \{3, 5, 7, 9, 11\}. The fuzzification procedure is as follows: generate the medium membership function and, next, the rest of the membership functions based on it. All Gaussians are transformed into type-2 fuzzy sets by changing the standard deviation value (called \texttt{sigma\_offset} parameter). The \texttt{sigma\_offset} defines the FOU of the corresponding IT2 fuzzy sets. The same value was applied for each Gaussian function.

2. If the standard deviations applied to the Gaussians are the same or not: \{equal, progressive\}—progressive assumes the concentration of membership functions around the mean or central values. For 3 Gaussians, equal and progressive, give the same fuzzification result.

3. If the mean of the medium basic function is derived directly from the corresponding data set or if a fixed value is used: \{mean, center\}—center assumes Gaussian distribution with fixed mean 0.5.

For example, <3 Gausses, Equal, Mean> is considered as 3 Gaussian membership functions with equal standard deviation values used to define 3 type-2 fuzzy sets, and the medium function is defined by its corresponding mean value derived from the considered data. The <5 Gausses, Equal, Center> is considered as 5 Gaussian membership functions and the medium function is defined by a fixed mean value 0.5. The <7 Gausses, Progressive, Mean> is considered as 7 Gaussian membership functions with different standard deviation values, assuming concentration of membership functions around the mean value of the medium function.

![Figure 5. Type-1 and type-2 fuzzification proposal.](image-url)
Table 4. The IT2 fuzzy sets used for a sample attribute of the blood transfusion data set. For all attributes, we normalized the corresponding domain values.
Table 4. Cont.

| Data Set                        | k-Value |
|---------------------------------|---------|
| Blood Transfusion               | 10      |
| Breast Cancer Data              | 10      |
| Breast Cancer Wisconsin         | 10      |
| Data Banknote Authentication    | 10      |
| Haberman                        | 6       |
| Heart                           | 5       |

For the type-2 sets used in the conclusions, we define s-shaped functions representing each class in the classification problem considered. We optimize the FOU of each function by changing the cross-points. We called the corresponding parameter under optimization `center_offset`—see Figure 6. The final classification decision value was fixed = 0.5, assuming normalized domain [0, 1]. Changes of the classification threshold did not show a positive effect on the experiment.

![Figure 6. IT2 fuzzy sets used in the rule conclusions.](image)

For each benchmark data, we optimized the `sigma_offset` and `center_offset` parameters for the type-2 fuzzy sets considered, using the grid search optimization algorithm applied in k-folded cross-validation process with evaluation on a held-out validation set. The `sigma_offset` takes values in the range [0.01, 0.03] with step 0.005 and the `center_offset` takes values in the range [0.01, 0.21] with step 0.02. The k value was chosen for any data set assuming a fair number of samples in the corresponding validation sets, presented in Table 5. Additionally, for any data partition, we have used stratified sampling to ensure balanced validation and test sets. For better clarity, we show the experiment information flow in Figure 7.
Table 5. Cont.

| Data Set                | k-Value |
|-------------------------|---------|
| HTRU 2                  | 10      |
| Immunotherapy           | 4       |
| Indian Liver Patient    | 10      |
| Ionosphere              | 6       |
| Parkinson               | 4       |
| Pima Indians Diabetes   | 10      |
| Vertebral               | 5       |

3. Results

In Table 6 we present the results achieved in our experiments. The number of rules generated for each data set is constant and equal to the cardinality of the decision attribute. In the case of binary classification, two fuzzy rules will be generated, but with very extensive and complex rule premises.

Figure 7. Experiment information flow.

3. Results

In Table 6 we present the results achieved in our experiments. The number of rules generated for each data set is constant and equal to the cardinality of the decision attribute. In the case of binary classification, two fuzzy rules will be generated, but with very extensive and complex rule premises.
Table 6. Evaluation of the method proposed.

| Data Set                  | Classification Accuracy | F-Score | Classifier                        |
|---------------------------|-------------------------|---------|-----------------------------------|
| Blood Transfusion         | 0.742                   | 0.541   | <11 Gausses, Equal, Mean>         |
| Breast Cancer Data        | 0.918                   | 0.912   | <7 Gausses, Progressive, Mean>    |
| Breast Cancer Wisconsin   | 0.961                   | 0.957   | <3 Gausses, Equal, Center>        |
|                           |                         |         | or <3 Gausses, Progressive, Center>|
| Data Banknote Authentication| 0.99                   | 0.993   | <9 Gausses, Progressive, Mean>    |
| Haberman                  | 0.783                   | 0.655   | <9 Gausses, Progressive, Mean>    |
| Heart                     | 0.802                   | 0.802   | <5 Gausses, Equal, Mean>          |
| HTRU 2                    | 0.968                   | 0.89    | <9 Gausses, Progressive, Center>  |
| Immunotherapy             | 0.889                   | 0.8     | <9 Gausses, Progressive, Mean>    |
| Indian Liver Patient      | 0.686                   | 0.592   | <7 Gausses, Progressive, Mean>    |
| Ionosphere                | 0.896                   | 0.888   | <7 Gausses, Equal, Mean>          |
| Parkinson                 | 0.881                   | 0.847   | <7 Gausses, Progressive, Center>  |
| Pima Indians Diabetes     | 0.732                   | 0.677   | <11 Gausses, Progressive, Center> |
| Vertebral                 | 0.85                    | 0.821   | <9 Gausses, Equal, Center>        |

We achieved quite high classification results for seven data sets, which proves the purposefulness of the methodology used.

Below in Table 7, we show a comparison regarding high performing classifiers for some of the data sets with respect to the F-score measure.

Table 7. Comparison regarding other classifiers with respect to the F-score measure.

| Data Set                  | Our Approach                                      | Other Classifiers                                                                 |
|---------------------------|---------------------------------------------------|-----------------------------------------------------------------------------------|
| Breast Cancer Data        | 0.912                                             | Immune-inspired semi-supervised Algorithm, introduced in [49]: 0.973              |
| Breast Cancer Wisconsin   | 0.957                                             | Extreme Learning Machine Neural Networks, introduced in [50]: 0.978              |
|                           |                                                   | Immune-inspired semi-supervised algorithm, introduced in [49]: 0.965             |
|                           |                                                   | Support vector machines combined with Feature Selection, introduced in [51]: 0.997|
| Data Banknote Authentication| 0.993                                             | Deep Neural Network with PCA and LDA, introduced in [52]: 0.99                  |
|                           |                                                   | Decision tree approach, introduced in [53]: 0.994                               |
|                           |                                                   | Random Forest approach, introduced in [54]: 0.948                               |
|                           |                                                   | Neural Network-Genetic Algorithm, introduced in [55]: 1                         |
| HTRU 2                    | 0.89                                              | Classical classifiers: C4.5: 0.74; MLP: 0.752; NB: 0.692; SVM: 0.789               |
|                           |                                                   | GH-VFDT Algorithm, introduced in [37]: 0.862                                    |
|                           |                                                   | A hybrid ensemble method, introduced in [56]: 0.918 (with respect to a voting threshold parameter) |
| Immunotherapy             | 0.8                                               | A hybrid classification approach, introduced in [57]: 0.851                      |
|                           |                                                   | Considering evaluation of machine learning techniques [58]: 0.91                |

As we can see above, we achieved quite good results for the datasets data banknote authentication and HTRU 2, despite the very general fuzzy approach proposed.
Classification accuracy comparison with fuzzy classification techniques for chosen data sets is shown in Table 8 as well.

**Table 8. Classification accuracy comparison with fuzzy classifiers.**

| Data Set            | Our Approach | Fuzzy Classifiers                                      |
|---------------------|--------------|--------------------------------------------------------|
| Breast Cancer Wisconsin | 0.961        | EM-PCA-CART-Fuzzy Rule-Based classification, introduced in [59]: 0.932 |
| HTRU 2              | 0.968        | Fuzzy neural network models: SORAF model [60]: 0.9597; models introduced in [61,62]: 0.961 and 0.952 respectively. |
| Immunotherapy       | 0.889        | Fuzzy neural network approach, introduced in [63]: 84.32 |

**4. Discussion**

We distinguish several important conclusions and advantages of our research. First of all, it is possible to aggregate information by interpreting attributes describing objects as fuzzy sets, using the advantages of fuzzy logic for flexible modelling of real data. This enables a large generalization of information, which positively affects unbalanced or inaccurately defined data. Additionally, the introduced interpretation of fuzzy information systems allows direct work on fuzzy sets and the induction of rules that can be easily changed into type-2 fuzzy rules. By appropriate fuzzification, the values of the information function can be derived directly from the data. Removal of inconsistencies, which may appear because of the information generalization applied, is performed by appropriate use of rough sets, i.e., by analyzing the quality of the corresponding approximations. The fuzzification applied is very general. It allows even better adjustment to special cases by optimization of the fuzzy sets in the premises and the conclusions of the rules applied. In this research, we have used only one of the many possible fuzzifications of the data presented. Applying fuzzification adapted to a specific problem, after an in-depth analysis, we should be able to adjust the fuzzy sets in the premises and the conclusions of the rules to achieve the best possible result. The methodology we introduced in our work is a general approach, aiming to show the advantage of fuzzy inference using a simple and homogeneous model. Despite that our algorithm was not strictly adjusted to any data set, we managed to obtain good results for some of them. It shows practical potential and creates the basis for generating robust classifiers.

**5. Conclusions**

In this research, we proposed a new fuzzy classification method. Our concept assumes the induction of fuzzy rules from the corresponding fuzzy information system. The applied fuzzification of data is done in a very simple manner by Gaussian type membership functions. Moreover, we are able to extend the induced rules into interval type-2 fuzzy rules. The experiments provided with benchmark data proved the method usefulness. Additionally, by further optimization of the interval type-2 fuzzy sets applied, we were able to improve our classification results. More suitable fuzzification, adjusted to a specific data set, is possible as well. Our concept should be considered as a general approach for unbalanced or inaccurately defined data.

**Author Contributions:** M.T. conceived of the presented idea, developed the theory and supervised the project. A.C. (Adrian Chlopowiec), A.C. (Adam Chlopowiec) and A.D. performed the computations and the method evaluation. All authors discussed the results and contributed to the final manuscript. All authors have read and agreed to the published version of the manuscript.
References

1. Bertsimas, D.; Dunn, J.; Pawlowski, C.; Zhuo, Y.D. Robust Classification. *Inf. J. Optim.* 2019, 1, 2–34. [CrossRef]

2. Sarkar, S.K.; Oshiba, K.; Giebisch, D.; Singer, Y. Robust Classification of Financial Risk, NIPS 2018 Workshop on Challenges and Opportunities for AI in Financial Services: The Impact of Fairness, Explainability, Accuracy, and Privacy. *arXiv* 2018, arXiv:1811.11079.

3. Zadeh, L. Fuzzy sets. *Inf. Control* 1965, 8, 338–353. [CrossRef]

4. Mendel, J.M.; John, R.I. Type-2 fuzzy sets made simple. *IEEE Trans. Fuzzy Syst.* 2002, 10, 117–127. [CrossRef]

5. Mendel, J.M. Type-2 Fuzzy Sets and Systems: How to Learn About Them. *IEEE Smc Envelope* 2009, 27, 1–7.

6. Mendel, J.M. Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions; Prentice-Hall: Upper Saddle River, NJ, USA, 2001.

7. Castillo, O.; Melin, P. Type-2 Fuzzy Logic Theory and Applications; Springer: Berlin, Germany, 2008.

8. Hagras, H. A hierarchical type-2 fuzzy logic control architecture for autonomous mobile robots. *Ieee Trans. Fuzzy Syst.* 2004, 12, 524–539. [CrossRef]

9. Hagras, H. Type-2 FLCS: A new generation of fuzzy controllers. *Ieee Comput. Intel. Mag.* 2007, 2, 30–43. [CrossRef]

10. Wu, D.; Tan, W.W. Genetic learning and performance evaluation of type-2 fuzzy logic controllers. *Eng. Appl. Artif. Intell.* 2006, 19, 829–841. [CrossRef]

11. Wu, D.; Tan, W.W. A simplified type-2 fuzzy controller for real-time control. *ISA Trans.* 2006, 15, 503–516.

12. Pawlak, Z. Information systems—Theoretical foundations. *Inf. Syst.* 1981, 6, 205–218. [CrossRef]

13. Pawlak, Z. Rough Sets: Theoretical Aspects of Reasoning about Data, System Theory, Knowledge Engineering and Problem Solving; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1991; Volume 9.

14. Pawlak, Z. Rough Sets; Basic notions, Report no 431; Institute of Computer Science, Polish Academy of Sciences: Warsaw, Poland, 1981.

15. Halder, A.; Kumar, A. Active learning using rough fuzzy classifier for cancer prediction from microarray gene expression data. *J. Biomed. Inform.* 2019, 92, 1–13. [CrossRef] [PubMed]

16. Onan, A. A fuzzy-rough nearest neighbor classifier combined with consistency-based subset evaluation and instance selection for automated diagnosis of breast cancer. *Expert Syst. Appl.* 2015, 42, 6844–6852. [CrossRef]

17. Wang, C.; Chen, D.; Hu, Q. Fuzzy information systems and their homomorphisms. *Fuzzy Sets Syst.* 2014, 249, 128–138. [CrossRef]

18. Dubois, D. The role of fuzzy sets in decision sciences: Old techniques and new directions. *Fuzzy Sets Syst.* 2011, 184, 3–28. [CrossRef]

19. Lee, M.-C.; Chang, T. Rule Extraction Based on Rough Fuzzy Sets in Fuzzy Information Systems. *Trans. Cei lli Lucs* 2011, 6560, 115–127.

20. Ye, J.; Zhan, J.; Xu, Z. A novel decision-making approach based on three-way decisions in fuzzy information systems. *Inf. Sci.* 2020, 541, 362–390. [CrossRef]

21. Cheruku, R.; Edla, D.R.; Kuppili, V.; Dharavath, R. RST-Bat-Miner: A fuzzy rule miner integrating rough set feature selection and bat optimization for detection of diabetes disease. *Appl. Soft. Comput.* 2018, 67, 764–780. [CrossRef]

22. Dai, J.H.; Chen, J.L. Feature selection via normative fuzzy information weight with application into tumor classification. *Appl. Soft Comput.* 2020, 92, 106299. [CrossRef]

23. Jensen, R.; Shen, Q. Semantics-preserving dimensionality reduction: Rough and fuzzy-rough-based approaches. *Ieee Trans. Knowl. Data Eng.* 2004, 16, 1457–1471. [CrossRef]

24. Lin, Y.; Li, Y.; Wang, C.; Chen, J. Attribute reduction for multi-label learning with fuzzy rough set. *Knowl. Based Syst.* 2018, 152, 51–61. [CrossRef]

25. Maji, P.; Garai, P. Fuzzy–Rough Simultaneous Attribute Selection and Feature Extraction Algorithm. *Ieee Trans. Cybern.* 2013, 43, 1166–1177. [CrossRef]

26. Yu, G. Characterizations and uncertainty measurement of a fuzzy information system and related results. *Soft Comput.* 2020, 24, 12753–12771. [CrossRef]

27. de Cock, M.; Cornelis, C.; Kerre, E.E. Fuzzy rough sets: The forgotten step. *Ieee Trans. Fuzzy Syst.* 2007, 15, 121–130. [CrossRef]

28. Dubois, D.; Prade, H. Rough fuzzy sets and fuzzy rough sets. *Int. J. Gen. Syst.* 1990, 17, 191–209. [CrossRef]

29. Mieszkowicz-Rolka, A.; Rolka, L. Fuzziness in information systems. In *Electronic Notes in Theoretical Computer Science*; Elsevier: Amsterdam, The Netherlands, 2003; Volume 82.

30. Wu, W.; Mi, J.; Zhang, W. Generalized fuzzy rough sets. *Inf. Sci.* 2003, 151, 263–282. [CrossRef]

31. Yeung, D.S.; Chen, D.G.; Tsang, E.; Lee, J.; Wang, X.Z. On the generalization of fuzzy rough sets. *Ieee Trans. Fuzzy Syst.* 2005, 13, 343–361. [CrossRef]
62. De Campos Souza, P.V. Pruning fuzzy neural networks based on unineuron for problems of classification of patterns. *J. Intell. Fuzzy Syst.* 2018, 35, 1–9.

63. Guimarães, A.J.; Souza, P.V.d.; Araújo, V.J.S.; Rezende, T.S.; Araújo, V.S. Pruning Fuzzy Neural Network Applied to the Construction of Expert Systems to Aid in the Diagnosis of the Treatment of Cryotherapy and Immunotherapy. *Big Data Cogn. Comput.* 2019, 3, 22. [CrossRef]