An efficient way of material classification using uncertainty assessment

Arvind Kumar Sinha* and Pradeep Shende

Department of Mathematics, National Institute of Technology, Raipur, Chhattisgarh India

*Corresponding author’s e-mail address: aksinha.maths@nitrr.ac.in

Abstract. Analysis of material properties using the available data is a compulsory part of the material classification and is considered as a challenging task in industrial processes. Classification refers to the segregation of objects using known data or knowledge. The uncertainties in the data increase the complexities of material classification. In this paper, we present a methodology to measure uncertainties in the attributes-values. An example illustrates the efficiency, validity, and effectiveness of the proposed method. We find the most significant features for the classification of glass materials using uncertainty evaluation and optimization.

1. Introduction

Knowledge-based systems use available data to find useful information out of it. It is a very challenging task, as incomplete or inexact data may yield misleading information. Such data leading to wrong information is termed as uncertainties in the data set [1]. In today's time, knowledge is accumulating at the exponential pace, and with this hoard of information, data evaluation complexities surge intensely. Finding useful patterns out of the available information becomes obscured due to the presence of uncertainties in it. Many approaches are studied in the literature to deal with uncertain values in the data-set. One of the very effective ways to deal with vagueness is fuzzy sets introduced by Zadeh [2]. In 1982, Pawlak [3] introduced the rough set theory (RST) as a new approach to deal with uncertainties. Both the Fuzzy set theory and RST finds vast scientific applications in almost all fields. RST requires only available information and is very useful in reasoning about data and to find helpful information out of it. The material information stored using various attributes of the material is called an information system. We use the data of elements in the form of the decision table. The collection of elements or instances or examples whose information is stored in the information system using various attributes is the universe of discourse or universal set. The set of all features is called the feature space or the set of attributes. These features lead to the labeling of objects using the decision feature. The values of the decision attributes are termed as decision class, and each decision class denotes the concept in the information system. The information system whose instances are labeled using the decision attribute is term as a decision system. The learning from the decision system is supervised learning, and learning from examples without the decision parameter is termed as unsupervised learning. The feature values add knowledge to the information system and are called as granules of knowledge [4]. Material classification is the segregation of distinct elements using the various attributes. Due to missing information, different objects seem indistinguishable. In such cases, the classification of elements using the most significant parameters becomes mandatory. Elements
classification using RST is studied in [5]. Uncertainty assessment is useful in finding the significance of the attribute [6]. Also, it helps to find the most suitable combination of features that keeps the classification of the elements intact. Finding features subset out of the given set of features is termed as feature subset selection (FSS).

[7] proposed a wrapper method for FSS based on the evaluation of subsets of the feature set for classification. FSS using a genetic algorithm is studied in [8]. Optimization of the feature space is studied in [9]. Decision-rule is another way of representing data. These rules help tremendously in recognition of new instances. These rules also find useful in the categorization of new elements [10]. The decision rules generation in case of missing information is studied in [11,12]. Statistical inference based decision rules generation is studied in [13]. The variable precision rough set model is discussed in [14], which is the generalization of RST using optimization of uncertainties allowing small errors in the rule set. The data reduction and feature optimization processes mainly evaluate uncertainties in the information system. We assess the uncertainties in the feature values by using the range of decision classes.

2. Data pre-processing

Data pre-processing is the process of cleaning the raw data employing the removal of missing values, if any, reducing noise and non-linearity, discretization processes, etc. We remove those elements whose attribute values outdistance from the remaining values. In the case of continuous attribute values, we get a vast number of rule-set. Discretization is the representation of continuous attribute values to its categorical counterparts. The complete data preprocessing method is represented in Figure 1.

![Material data pre-processing](image)

**Figure 1.** Material data pre-processing

2.1. Rough set reducts

The uncertain features severely escalate the noise in the information system. Finding the combination of features among the given feature set, which keeping the classification of objects intact, is termed as feature subset selection (FSS). The rough set provides skilful algorithms for FSS. Let \( I = (U, \{ \alpha \}) \) is the information system where \( U \) is the universal set of elements, \( \mathcal{A} \) is the features space and \( d \) is the decision feature. Let any \( \alpha \in \mathcal{A} \) is the feature whose uncertain region is to be determine, \( \alpha(o) \) is the feature value of object \( o \) for feature \( \alpha \) and \( V\alpha \) is the set of feature values of feature \( \alpha \) then \( V\alpha \) is given by the function \( f : \alpha \rightarrow V\alpha \) defined as \( f(o) = V\alpha(o) \), \( o \in U \).

Let \( B \subseteq \mathcal{A} \), define an equivalence relation \( R(B) \) on \( U \) as follows:

\[
\text{for } o_i, o_j \in U, \quad o_i \sim_{R(B)} o_j \quad \text{iff} \quad V\alpha(o_i) = V\alpha(o_j), \quad \forall \alpha \in B
\]  

(1)

The cells of the relation \( R(B) \) are

\[
[o_i]_{R(B)} = \{ o_j : o_i R(B) o_j \}
\]  

(2)
Each feature-value represents a concept in the decision system $I$. Let $\delta$ be a concept in $I$, approximation of $\delta$ using feature-values of $B \subseteq A$ by creating the $B$-lower approximation and $B$-upper approximation of $\delta$ denoted by $B_\uparrow(\delta)$ and $B_\downarrow(\delta)$ respectively where,

$$B_\downarrow(\delta) = \{o : [o]_{R(B)} \subseteq \delta\} \quad \text{and} \quad B_\uparrow(\delta) = \{o : [o]_{R(B)} \cap \delta \neq \phi\}$$  \hspace{1cm} (3)

**B-Boundary Region** of $\delta$ $BN_\downarrow(\delta) = B_\uparrow(\delta) - B_\downarrow(\delta)$ amount to the elements whose precise categorization into the concept $\delta$ is not possible using the feature $B$. The rough set incorporates the concepts with an imprecise boundary region. This boundary region embodies the uncertain-region. We call the feature $\delta$ as dispensable in $B \subseteq A$ if $[o]_{R(B)} = [o]_{R(B-\{\alpha\})}$, $\forall o \in U$. The non-dispensable features are termed as indispensable. The indispensable elements set are described as independent set.

If $B \subseteq A$ is independent and $[o]_{R(B)} = [o]_{R(A)}$, $\forall o \in U$ then $B$ is described as reduct of $A$.

### 3. Uncertainties assessment methodology

The equivalence relation $R(\alpha)$ on $U$ using the feature $\alpha \in A$ is defined as

$$o_1 R(\alpha) o_2 \quad \text{iff} \quad V_{\alpha(o_1)} = V_{\alpha(o_2)}$$  \hspace{1cm} (4)

The cells of the relation $R(\alpha)$ are obtained as

$$[o]_{R(\alpha)} = \{\xi : o R(\alpha) \xi\}, \quad o, \xi \in A$$  \hspace{1cm} (5)

For the concepts $\delta_i, i = 1, 2,..., k$ in the decision system, we designate the membership of any $o \in U$ in the concept $\delta_i$ as

$$\mu^\alpha_{\delta_i}(o) = \frac{|\delta_i \cap [o]_{R(\alpha)}|}{|[o]_{R(\alpha)}|}$$  \hspace{1cm} (6)

where, $\cdot$ denotes the cardinality. We designate the feature-range with corresponding to decision-class $\delta_i$ as

$$\alpha_{\delta_i} = \{o \in U : \mu^\alpha_{\delta_i}(o) > 0\}$$  \hspace{1cm} (7)

If $0 < \mu^\alpha_{\delta_i}(o) < 1$ then $o$ is the imprecise case and its categorization into the concept $\delta_i$ is not possible. The uncertain-region $Unc(\alpha)$ of the feature $\alpha$ is

$$Unc(\alpha) = \bigcup_{i,j} (\alpha_{\delta_i} \cap \alpha_{\delta_j}), \quad i, j = 1, 2, ..., k$$  \hspace{1cm} (8)

The certain-space $Cer(\alpha)$ of the feature $\alpha$ is

$$Cer(\alpha) = U \setminus Unc(\alpha)$$  \hspace{1cm} (9)

Let $R(\alpha), R(\alpha)_1, R(\alpha)_2, ..., R(\alpha)_n$ be the classes of the relation $R(\alpha)$ on the feature $\alpha$ , then we prescribe the significance-factor $\sigma^\alpha_{\alpha}$ of the feature $\alpha$ as:

$$\sigma^\alpha_{\alpha} = \sum_{i=1}^{n} \frac{|Cer(\alpha) \cap R(\alpha)_i|}{|R(\alpha)_i|}$$  \hspace{1cm} (10)
Figure 2 is the flowchart of the feature subset selection (FSS) using the uncertain-region optimization of the features.

![Flowchart](image)

**Figure 2.** Flowchart of FSS using uncertainty optimisation

Following example demonstrate the methodology of FSS using certainty region of features. The Table 1 is the decision table consisting of information of seven elements using four features $\alpha_1$, $\alpha_2$, $\alpha_3$, and $\alpha_4$. Using the information given in Table 1, the classification of the examples is given by $\text{Clfs} = \{E1, E3, E6\}, \{E2\}, \{E5\}$. Examples $E4$ and $E7$ are conflicting cases, and its precise classification is not possible.
Table 1. Example of the Decision table.

| Example | $\alpha_1$ | $\alpha_2$ | $\alpha_3$ | $\alpha_4$ | Decision |
|---------|------------|------------|------------|------------|----------|
| $E1$    | 2          | 1          | 0          | 1          | 1        |
| $E2$    | 0          | 2          | 0          | 1          | 0        |
| $E3$    | 0          | 1          | 1          | 2          | 1        |
| $E4$    | 1          | 2          | 0          | 1          | 1        |
| $E5$    | 0          | 0          | 1          | 2          | 2        |
| $E6$    | 1          | 0          | 2          | 2          | 1        |
| $E7$    | 1          | 2          | 0          | 1          | 2        |

For Table 1, feature space $A = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$ and universal set of examples $U = \{E1, E2, E3, E4, E5, E6, E7\}$. There are three decision classes $\delta_1 = 0, \delta_2 = 1, \delta_3 = 2$. Using the equivalence relation defined in equation (4), for feature $\alpha_1$ we get the classes as $[E1]_{\alpha_1} = \{E1\}, [E2]_{\alpha_1} = [E3]_{\alpha_1} = [E5]_{\alpha_1} = \{E2, E3, E5\}, [E4]_{\alpha_1} = [E6]_{\alpha_1} = [E7]_{\alpha_1} = \{E4, E6, E7\}$.

For feature $\alpha_2$, we get the classes as $[E1]_{\alpha_2} = [E3]_{\alpha_2} = \{E1, E3\}, [E5]_{\alpha_2} = [E6]_{\alpha_2}$. $[E2]_{\alpha_2} = [E4]_{\alpha_2} = [E7]_{\alpha_2} = \{E2, E4, E7\}$. For feature $\alpha_3$, the classes are $[E1]_{\alpha_3} = [E2]_{\alpha_3} = [E4]_{\alpha_3} = [E7]_{\alpha_3} = \{E1, E2, E4, E7\}, [E3]_{\alpha_3} = [E5]_{\alpha_3} = \{E3, E5\}, [E4]_{\alpha_3} = [E7]_{\alpha_3} = \{E4, E7\}$. Also the classes for the feature $\alpha_4$ are given by $[E1]_{\alpha_4} = [E2]_{\alpha_4} = [E4]_{\alpha_4} = [E7]_{\alpha_4} = \{E1, E2, E4, E7\}$ and $[E3]_{\alpha_4} = [E5]_{\alpha_4} = [E6]_{\alpha_4} = \{E3, E5, E6\}$.

The certainty region for the features is obtained as $Cer(\alpha_1) = \{E1\}, Cer(\alpha_2) = \{E1, E3\}, Cer(\alpha_3) = \{E6\}, Cer(\alpha_4) = \phi$. Using equation (10), we get the significance of the features as $\sigma_{\alpha_1} = 1, \sigma_{\alpha_2} = 2, \sigma_{\alpha_3} = 1, \sigma_{\alpha_4} = 0$. The feature with maximum significance factor is $\alpha_2$. Hence adding feature $\alpha_2$ in the set $Re\, d$ and removing the elements which belongs to $Cer(\alpha_2)$, the reduced decision table obtained is specified in

Table 2. Reduced decision table.

| Example | $\alpha_1$ | $\alpha_3$ | $\alpha_4$ | Decision |
|---------|------------|------------|------------|----------|
| $E2$    | 0          | 0          | 1          | 0        |
| $E4$    | 1          | 0          | 1          | 1        |
| $E5$    | 0          | 1          | 2          | 2        |
| $E6$    | 1          | 2          | 2          | 1        |
| $E7$    | 1          | 0          | 1          | 2        |
We have \( \text{Red} = \{\alpha_2\} \). Note that \([o]_{R(\text{Red})} \neq [o]_{\text{Red}}\). Proceeding in the same manner, here we get the classes as \([E2]_{\alpha_1} = [E5]_{\alpha_1} = \{E2, E5\}\), \([E4]_{\alpha_1} = [E6]_{\alpha_1} = [E7]_{\alpha_1} = \{E4, E6, E7\}\), \([E2]_{\alpha_1} = [E4]_{\alpha_1} = [E7]_{\alpha_1} = \{E2, E4, E7\}\), \([E5]_{\alpha_1} = (E5)\), \([E6]_{\alpha_1} = (E6)\), and for feature we get \([E2]_{\alpha_2} = [E4]_{\alpha_2} = [E7]_{\alpha_2} = \{E2, E4, E7\}\), \([E5]_{\alpha_2} = [E6]_{\alpha_2} = \{E5, E6\}\). The certainty region is given by \(\text{Cer}(\alpha_1) = \emptyset\), \(\text{Cer}(\alpha_2) = \{E5, E6\}\), \(\text{Cer}(\alpha_3) = \emptyset\). The significance factor is \(\sigma_{\alpha_1} = 0, \sigma_{\alpha_2} = 2, \sigma_{\alpha_3} = 0\). Therefore the feature added to set \(\text{Red}\) is \(\alpha_1\). After eliminating the examples belonging to \(\text{Cer}(\alpha_1)\) we get the reduced decision table shown in Table 3.

| Example | \(\alpha_1\) | \(\alpha_2\) | Decision |
|---------|-------------|-------------|----------|
| \(E2\)  | 0           | 1           | 0        |
| \(E4\)  | 1           | 1           | 1        |
| \(E7\)  | 1           | 1           | 2        |

Here \([o]_{R(\text{Red})} \neq [o]_{\text{Red}}\). Here we get \([E2]_{\alpha_1} = \{E2\}\), \([E4]_{\alpha_1} = [E7]_{\alpha_1} = \{E4, E7\}\), \([E2]_{\alpha_1} = [E4]_{\alpha_1} = [E7]_{\alpha_1} = \{E2, E4, E7\}\). The certainty region is given by \(\text{Cer}(\alpha_1) = \{E2\}\), and \(\text{Cer}(\alpha_2) = \emptyset\). The significance factor obtained as \(\sigma_{\alpha_1} = 1\) and \(\sigma_{\alpha_2} = 0\). Thus the feature added to the set \(\text{Red}\) is \(\alpha_1\). At this stage we get \([o]_{R(\text{Red})} = [o]_{\text{Red}}\). Therefore the reduct set obtained is \(\text{Red} = \{\alpha_1, \alpha_2, \alpha_3\}\). The feature \(\alpha_4\) is dispensable.

4. Research Data

In this section, we present the data used in the study. We use the data of materials concentration in the glass given in the UCI machine learning repository [15]. The glass classification utilizing the information of its nine constituent elements is given in the data. The data set comprises information of a total of 214 objects. The elements are classified into six classes using the nine features. All the features are continuous-valued. Table 4 gives detailed of the data-set used in the study. The feature noted as \(R1\) shows the refractive index of the glass. Other features notation is the standard notation of elements. Among the 214 objects, 163 objects are the example of windows glass, and 51 objects are the example of the non-windows glass. Among the 163 examples of the glass of the windows, 87 objects are float-processed, and 56 objects are non-float processed. We applied the maximum discernibility discretization algorithm for feature-values discretization. The algorithm is implemented in ROSETTA toolkit [16] which provides a platform for rough set methods. Using the discretized features, we applied the FSS using uncertainty assessment; we found the reduct of the feature space as \(\{R1, Na, Mg, Si, Ba\}\). To compare the classification accuracy we used the SVM classification algorithm [17] with 10-fold validation. The comparison of the classification accuracy using SVM to the feature set and its reduct is given in Table 5.
### Table 4. Description of the data set used in the study.

| Sr.No | Feature | Min   | Max   | Average | SD    | Correlation With Class |
|-------|---------|-------|-------|---------|-------|------------------------|
| 1.    | $RI$    | 1.5112| 1.5339| 1.5184  | 0.0030| -0.1642                |
| 2.    | $Na$    | 10.73 | 17.38 | 13.407  | 0.8166| 0.5030                 |
| 3.    | $Mg$    | 0     | 4.49  | 2.684   | 1.4424| -0.745                 |
| 4.    | $Al$    | 0.29  | 3.5   | 1.445   | 0.4993| 0.5988                 |
| 5.    | $Si$    | 69.81 | 75.41 | 72.6509 | 0.7745| 0.1515                 |
| 6.    | $K$     | 0     | 6.21  | 0.4971  | 0.6522| -0.0100                |
| 7.    | $Ca$    | 5.43  | 16.19 | 8.9570  | 1.4232| 0.0007                 |
| 8.    | $Ba$    | 0     | 3.15  | 0.1750  | 0.4972| 0.5751                 |
| 9.    | $Fe$    | 0     | 0.51  | 0.0570  | 0.0974| -0.1879                |

There are 70 examples of float-processed windows building window glass and denoted by class 1. Decision class 3 comprises of Seventeen examples are the float processed vehicle windows glass. Seventy-six examples of the non-float glass of building windows are denoted by class 2. Decision class 5 are examples of container glass. Decision classes 6 and 7 represent the examples of tableware glass and headlamps, respectively.

### Table 5. Accuracy obtained using feature space and its reduct.

| Set       | TP   | FP   | Precision | Recall | F-measure | Accuracy |
|-----------|------|------|-----------|--------|-----------|----------|
| Feature Space | 0.67 | 0.15 | 0.63      | 0.67   | 0.65      | 67.29%   |
| Reduct    | 0.72 | 0.11 | 0.71      | 0.73   | 0.7       | 72.89%   |

Table 5 shows that the accuracy obtained in the case of reduct is significantly higher than the accuracy of full feature space. The features $RI$ and $Na$ found the most significant for the classification of the glass of class 1. Feature Barium $Ba$ found most notable for the categorization of headlamp glass. Silicon $Si$ found most significant for the classification of tableware glass. Calcium $Ca$ obtained as a most significant feature for the classification of non-float processed building window glass.

### 5. Conclusion

Imprecise or inexact data may yield misleading information. In this work, we presented a novel method for assessing the uncertainties in the feature space. A flowchart for finding the combination of features among the feature space, which keeps the entire information about the dataset as the feature space is specified. We introduced an innovative methodology for feature selection using uncertainty optimization. Using the proposed method, we evaluated the glass material dataset and obtained the accuracy of classification significantly more than the accuracy of the entire feature space.
References

[1] Alchian A A 1950 Uncertainty, evolution, and economic theory J. Polit. Econ. 58 211–21
[2] Zadeh L A 1965 Fuzzy sets Inf. Control 8 338–53
[3] Pawlak Z 1982 Rough sets Int. J. Comput. Inf. Sci. 11 341–56
[4] Pawlak Z 1998 Granularity of knowledge, indiscernibility and rough sets 1998 IEEE International Conference on Fuzzy Systems Proceedings. IEEE World Congress on Computational Intelligence (Cat. No. 98CH36228) vol 1 pp 106–10
[5] Pawlak Z 1984 Rough classification Int. J. Man. Mach. Stud. 20 469–83
[6] Jurado K, Ludvigson S C and Ng S 2015 Measuring uncertainty Am. Econ. Rev. 105 1177–216
[7] Kohavi R and John G H 1997 Wrappers for feature subset selection Artif. Intell. 97 273–324
[8] Yang J and Honavar V 1998 Feature subset selection using a genetic algorithm Feature extraction, construction and selection (Springer) pp 117–36
[9] Koller D and Sahami M 1996 Toward optimal feature selection
[10] Apté C, Damerau F and Weiss S M 1994 Automated learning of decision rules for text categorization ACM Trans. Inf. Syst. 12 233–51
[11] Grzymala-Busse J W and Werbrouck P 1998 On the best search method in the LEM1 and LEM2 algorithms Incomplete Information: Rough Set Analysis (Springer) pp 75–91
[12] Holmström B and Myerson R B 1983 Efficient and durable decision rules with incomplete information Econom. J. Econom. Soc. 1799–819
[13] Cencov N N 2000 Statistical decision rules and optimal inference (American Mathematical Soc.)
[14] Ziarko W 1993 Variable precision rough set model J. Comput. Syst. Sci. 46 39–59
[15] Dua D and Graff C 2017 (UCI) Machine Learning Repository
[16] Nguyen H S 1998 Discretization problem for rough sets methods International Conference on Rough Sets and Current Trends in Computing pp 545–52
[17] Cortes C and Vapnik V 1995 Support-vector networks Mach. Learn. 20 273–97