Research on Classification Model based on Neighborhood Rough Set and Evidence Theory

Yuzhuo Zhang* and Yunqiong Wang
School of Information Science, Yunnan Normal University, Kunming, China,

*Corresponding author email: yuzhuo2013@sina.com

Abstract. The forward greedy numerical attribute reduction algorithm based on the neighborhood rough set is used to reduce the continuous numerical evaluation metrics of the health condition of complex equipment. This eliminates the risk of data loss and the additional processing time, due to avoiding discretization of continuous numerical evaluation metrics. Furthermore, the reduced evaluation decision table is processed to construct the basic probability assignment function (BBAS). Finally, multiple evaluation metrics are fused by Dempster’s rule of combination to get the health condition grade, and the relationship between evaluation metrics and health condition grade is mined further. The theoretical analysis and experimental results show that the proposed model is effective and efficient for classification.

Keywords: Neighborhood rough set; attribute reduction; D-S evidence theory; classification model.

1. Introduction
Since Polish mathematician Pawlak proposed rough set theory in 1981, it has been proved to be an effective technique for selecting feature in classified data, which is commonly applied in pattern recognition, machine learning, knowledge discovery, process control and other fields[1]. However, classical rough set theory has some limitation[2][3], it is based on equivalence relationship[4], and is only suitable for processing data with nominal attributes (e.g., 0-1). When continuous data is processed, it must be discretized first. Through the discretization process, some information will inevitably be lost, and the processing time overhead will increase. The neighborhood rough set is expanded in accordance with the classic rough set theory. By introducing the concept of neighborhood, it not only inherits the advantages of the classical rough set theory, using the information of the data itself without any prior knowledge[5], but also uses the domain space of neighborhood granulation theory, which can directly reduce the continuous numerical evaluation metrics. Because rough set theory fails to have a way to treat with inaccurate or uncertain raw data, it is strongly recommended to be complemented with the evidence theory and other theories handling uncertainty.

D-S evidence theory was formally built by Shafer in 1976 [6], as a decision-level fusion method. This theory is broadly applied to the field of information fusion, due to its advantages of realizing multi-evidence data integration and dealing with uncertainty[7]. The fundamental principle is to complete independent judgments on frame of discernment through multiple irrelevant evidence, and apply the combination rules to fuse those results to make the final decision.

In the article, we present a classification model, combining both rough set theory and D-S evidence theory. The model does not need to discretize continuous quantitative data, but directly follows the forward greedy numerical attribute reduction algorithm to eliminate the irrelevant and redundant attribute, allowing the classification ability to remain unchanged[8]. Then, the basic belief assignments
function (BBAS) is constructed according to the reduced evaluation feature metrics, and multiple reduced feature metrics are synthesized by Dempster’s rule of combination, and classify the combined results. The proposed usage of this model is to evaluate the health status of the equipment examples in the literature [14], to obtain the health category of the equipment, which illustrates the applicability and effectiveness of the method.

2. Classification Model Combining Neighborhood Rough Set and Evidence Theory

First, attributes of the decision table are reduced, through the application of the neighborhood rough set reduction algorithm. Then, Dempster’s rule of combination is employed to synthesize the basic belief assignments (BBAS), and make evaluation decisions. The classification process by neighborhood rough set and D-S evidence theory is illustrated in Figure 1.

![Figure 1. Classification flow through neighbourhood rough set and evidence theory.](image)

2.1. Basic Concepts of Neighborhood Rough Sets

2.1.1. Neighborhood. Formally, structured data offered for categorization learning can be described as an information system\( < U, A > \), if \( A = C \cup D \), \( < U, A > \) is called the decision table[9], where \( U \) is a nonempty and finite set of samples\{\( x_1, x_2, x_3 \ldots x_n \}\), called a universe, and \( C \) is the conditional attribute set (feature), and \( D \) is the decisional attribute (tag). \( \forall x_i \in U \) and \( B \subseteq C \), the neighborhood \( \delta_B (x_i) \) of \( x_i \) in feature space \( B \) is defined as:

\[
\delta_B (x_i) = \{x_j|x_j \in U, \Delta_B (x_i, x_j) \leq \delta)\}
\]

\( \delta_B (x_i) \) is the neighborhood information granule, centered around sample \( x_i \) [9]. It represents samples set similar to \( x_i \) under feature space \( B \). Let \( x \) be any sample of the universe \( U \), and \( f(x, c) \) represents the value of object \( x \) on attribute \( c \). \( \Delta_B (x_i, x_j) = \left( \sum_{c \in B} |f(x_i, c) - f(x_j, c)| \right)^{1/2} \) is the distance function and represents similarity between the sample \( x_i \) and the sample \( x_j \) in the feature space \( B \). \( \delta \) plays a role in controlling the size of the sample neighborhood information granule, otherwise known as the neighborhood threshold which is real number between 0 and 1. The smaller the number is, the higher similarity between sample \( x_i \) and sample \( x_j \) and fewer samples fall into the neighborhood set of \( x_i \) in the feature space \( B \). So \( \Delta_B \) determines the shape of the neighborhood and \( \delta \) control the size of neighborhood granule[10].

2.1.2. Upper and lower approximation. For a decision system \( NDT = < U, C \cup D, N > \), the universe \( U \) is divided into \( N \) equivalence classes\( \{X_1, X_2, \ldots, X_N\} \) according to decisional attribute \( D \) and \( X_i (i = \ldots) \)
1, \ldots, N) is a subset of the samples with the decisional tag. For any \( X_i \), there are two subsets of objects, called the upper and lower approximations of decisional attribute \( D \) on conditional attribute subset \( B \), are defined as

\[
\overline{N}_B = \bigcup_{i=1}^{N} \overline{N}_B X_i
\]
\[
\underline{N}_B = \bigcup_{i=1}^{N} \underline{N}_B X_i
\]

\( \overline{N}_B X_i = \{x_i | \delta(x_i) \subseteq X, x_i \in U \}, \underline{N}_B X_i = \{x_i | \delta(x_i) \cap X \neq \emptyset \} \). The lower approximation set is also called the positive region of the decision table, denoted by \( \text{pos}_B(D) \), equal to \( \overline{N}_B D \). The positive region is the subset of objects whose neighborhood granules consistently belong to one of the decision[11]. The more samples in the positive region, the stronger the classification ability is to define that feature set and differentiate it from other feature sets.

2.1.3. Neighborhood dependence. For neighbourhood a decision system \( NDT = \langle U, C \cup D, N \rangle \), the distance function is \( \Delta \) and the neighborhood threshold is \( \delta \), the neighborhood dependence of decisional attribute \( D \) on the conditional attribute \( B \) is defined as

\[
\gamma_B(D) = \frac{|\text{pos}_B(D)|}{|U|}
\]

Due to the \( \text{pos}_B(D) \subseteq U \), the neighborhood dependence meets \( 0 \leq \gamma_B(D) \leq 1 \). If \( \gamma_B(D) = 1 \), \( D \) is completely dependent on \( B \) and the decision system is consistent in terms of \( \Delta_B \) and \( \delta \), otherwise we say \( D \) depends on \( B \) in the degree of \( \gamma \).[11]

Neighborhood dependence is the proportion of approximate samples under the decision class in the neighborhood decision system. It is also a measure to describe the degree of relationship between conditional attribute and decisional attribute set. Therefore, the measure can be used as a heuristic function to perform attribute reduction of neighborhood decision system.

2.1.4. Reduction. \( NDT = \langle U, C \cup D, N \rangle \), for a neighborhood decision system \( NDT \), if \( B \) satisfies:

\[
\forall a \in B, \gamma_B(D) > \gamma_{B-\{a\}}(D)
\]
\[
\gamma_B(D) = \gamma_C(D)
\]

Called \( B(B \subseteq C) \) is a reduction of \( C \). Equation(5) indicates that there are no redundant attributes in the reduction, and equation(6) guarantees \( \text{POS}_B(D) = \text{POS}_C(D) \). Therefore, reduction is the minimum subset in the conditional attribute set in the neighborhood decision system, which has the same approximation ability as the whole attribute set.

2.2. A Forward Greedy Search Algorithm based on Neighborhood Rough Set

Attribute reduction within information systems is an important method to reduce the dimension of complex data. Attribute \( a_i \) importance \( Sig(a_i, B, D) = \gamma_{B\cup\{a_i\}}(D) - \gamma_B(D) \) measure the impact of conditional attributes \( a_i \) on decisional attributes \( D \). When \( Sig(a_i, B, D) = 0 \), the conditional attribute \( a_i \) is unnecessary with respect to \( D \), but if greater than 0, then \( a_i \) is necessary. The purpose of rough set under attribute reduction is to get a subset of attributes which has the same discriminating power as the original data and has not any redundant. The forward greedy search algorithm for attribute reduction can be formulated as follows.
2.3. State Assessment Model based on Evidence Theory

D-S evidence theory is often used to handle indeterminate information. It uses sets to represent propositions and can directly express “uncertainty” by assigning probability to a subset of a set composed of single object or multiple objects. In addition, it can also combine various kinds of evidence to produce new evidence. The basic concept is as follows.

2.3.1. Frame of discernment.

Set $\Omega$ is a set of mutually exclusive and collectively exhaustive events, which can be represented as $\Omega = \{E_1, E_2, \ldots, E_n\}$. This is called frame of discernment. The power set of $\Omega$ is denoted by $2^\Omega$, which includes the empty set, in addition to all other sets. $\emptyset \in 2^\Omega$ is called a hypothesis.

2.3.2. The basic probability assignment function (mass function).

In a frame of discernment $\Omega$, a mass function, denoted by $m: 2^\Omega \rightarrow [0, 1]$, and the following conditions are satisfied:

$$m(\emptyset) = 0$$  

$$\sum_{A \subseteq \Omega} m(A) = 1$$  

In D-S theory, a mass function is also defined as a BBAS. If $m(A) > 0$, $A$ is called focal element.

2.3.3. Dempster’s rule of combination.

Let $m_1, m_2, \ldots, m_n$ be many independent BBAS and $A_1, A_2, \ldots, A_N$ be many focal element in the frame of discernment $\Omega$. Dempster’s rule of combination can be defined as follows[12].

$$m(A) = K^{-1} \times \sum_{\prod_i \neq \emptyset} \prod_{i=1}^{N} m_i(A_i)$$  

Where $K = \sum_{\prod_i \neq \emptyset} \prod_{i=1}^{N} m_i(A_i)$, $K$ is the coefficient of conflict of $m_1, m_2, \ldots, m_n$. $K \neq \emptyset$, the equation (9) is represented in the form $m_1 \oplus m_2 \oplus \ldots \oplus m_n[13]$. The reasoning steps of multi-evidence synthesis by D-S evidence theory are as follows:

- Form the frame of discernment $\Omega$ according to the actual evaluated problem.
- Construct the mass function by combining relevant data and expert experience.
- Determine BBAS value of each focal element according to different evidence sources.
- Obtain the final decision result by fusing multi-source evidence through equation (9)
3. Experimental Analysis

The data in literature [14] are used for experimental analysis. Take, for example, a positioning and orientation device. Its health status assessment decision information table for continuation-oriented tasks is shown in Table 1. Health status of the system is divided into five grades ----for $\Omega = \{\text{super ideal(1)}, \text{ideal(2)}, \text{average (3)}, \text{below average (4)}, \text{poor (5)}\}$. The health assessment metrics are also seen in Table 1, as quick start time $c_1$, system positioning accuracy $c_2$, maximum elevation angle $c_3$, azimuth accuracy $c_4$, alignment time $c_5$, and continuous working time $c_6$[14], which are used to obtain the health condition grade.

Table 1. Original evaluation decision table.

|   | $c_1$ | $c_2$ | $c_3$ | $c_4$ | $c_5$ | $c_6$ | $D$ |
|---|-------|-------|-------|-------|-------|-------|-----|
| 1 | 2.5   | 6.8   | 24.4  | 0.4   | 7.3   | 14.9  | 1   |
| 2 | 3.1   | 7.6   | 25.6  | 0.7   | 6.5   | 13.7  | 1   |
| 3 | 2.8   | 6.5   | 25.2  | 0.8   | 8.2   | 14.3  | 1   |
| 4 | 6.8   | 8.1   | 21.8  | 0.9   | 13.6  | 12.1  | 2   |
| 5 | 7.5   | 13.2  | 22.1  | 13.6  | 12.1  | 11.4  | 2   |
| 6 | 7.8   | 12.8  | 22.8  | 1.6   | 11.9  | 9.9   | 2   |
| 7 | 8.1   | 16.3  | 20.1  | 1.9   | 18.5  | 9.3   | 3   |
| 8 | 4.2   | 15.9  | 17.8  | 2.4   | 15.8  | 8.1   | 3   |
| 9 | 5.6   | 17.4  | 18.4  | 2.5   | 17.3  | 7.7   | 3   |
| 10| 8.6   | 19.2  | 13.7  | 3.1   | 22.6  | 6.9   | 4   |
| 11| 9.4   | 20.1  | 12.2  | 3.3   | 23.4  | 7.3   | 4   |
| 12| 8.8   | 18.9  | 14.1  | 2.9   | 25.1  | 5.8   | 4   |
| 13| 14.4  | 23.1  | 11.5  | 4.2   | 26.3  | 5.4   | 5   |
| 14| 12.6  | 21.9  | 9.9   | 4.9   | 25.6  | 6.1   | 5   |
| 15| 13.7  | 22.6  | 10.6  | 5.2   | 27.7  | 5.5   | 5   |

3.1. Reduction Processing of Conditional Attributes

The forward greedy search algorithm of the neighborhood rough set (see Algorithm 1) was followed to reduce the evaluation metrics in Table 1. Due to the important role the threshold function plays in the rough concentration of neighborhoods rough set, it can be utilized as a parameter to limit the granularity of data analysis. The importance of feature metrics will change with respect to the granularity level, while the classification accuracy will also change with the threshold. The number of selected feature metrics positively correlates with the threshold, and will then decrease after reaching a peak. This trend is very similar to that of altering classification accuracy based on CART and SVM algorithms[15]. Roughly speaking, a candidate interval of threshold function delta is $[0.1, 0.3]$. When the interval exceeds 0.4, the attribute reduction algorithm based on neighborhood rough set cannot obtain enough features to distinguish samples[15]. Therefore, we use 0.3 for the experiment. The attribute importance obtained from the first cycle of the algorithm 1 are $\text{var}(c_1, \phi, D) = 14/15$, $\text{var}(c_2, \phi, D) = 15/15, \text{var}(c_3, \phi, D) = 15/15, \text{var}(c_4, \phi, D) = 10/15, \text{var}(c_5, \phi, D) = 15/15, \text{var}(c_6, \phi, D) = 12/15$. We select three the most important feature $c_2, c_3, c_5$ as attribute importance are the bigger. The second cycle calculates $\text{var}(c_1, \text{red}, D)$, where $c_2, c_3, c_5$ combined with other attributes. The result of $\text{var}(c_1, \text{red}, D)$ are all 0, which depicts that the device's health assessment for continuous tasking is only related to the characteristic metrics, $c_2, c_3, c_5$, and has nothing to do with other metrics. Therefore, $c_2, c_3, c_5$ are selected and synthesized by Dempster’s rule of combination, and the results are compared and analyzed.

3.2. Multi-evidence Synthesis

As mentioned above, we determine the frame of discernment as $\Omega = \{\text{super ideal(1)}, \text{ideal(2)}, \text{average (3)}, \text{below average (4)}, \text{poor (5)}\}$ . The formula (9) combines the three evaluation metrics $c_2, c_3, c_5$ to produce composite evidence. One with the most credibility could be regarded as the result of health condition grade. The mass functions of device health level are $m(1), m(2), m(3), m(4), m(5)$, and $m(\Omega)$ which is the mass function of uncertainty hypothesis. The evaluation decision information in Table 1 is used to construct the mass function of evaluation metrics for each health grade. Metrics $c_3,$
for example, the values of the five health grades are \{24.4, 25.6, 25.2\}, \{21.8, 22.1, 22.8\}, \{20.1, 17.8, 18.4\}, \{13.7, 12.2, 14.1\}, \{11.5,9.9,10.6\}, so their mean values are 25.07, 22.23, 18.77, 13.33, 10.67, respectively. The mass function of each health level are constructed as Table2 below.

**Table2.** The construction of mass function about c3.

| c3 (> 25.07) | 22.3 ≤ c3 ≤ 25.07 | 18.77 ≤ c3 ≤ 22.23 |
|---------------|---------------------|---------------------|
| m(1) = 0.9    | m(1) = \((c3 - 22.23)/(25.07 - 22.23)\) × 0.9 | m(1) = \((c3 - 18.77)/(22.23 - 18.77)\) × 0.9 |
| m(2) = m(3) = m(4) = m(5) = 0 | m(2) = \((25.07 - c3)/(25.07 - 22.23)\) × 0.9 | m(2) = \((25.07 - c3)/(25.07 - 22.23)\) × 0.9 |
| m(Ω) = 0.1    | m(3) = m(4) = m(5) = 0; m(Ω) = 0.1 | m(1) = m(2) = m(3) = m(4) = 0 |

m(1) = m(2) = m(3) = m(4) = 0; m(Ω) = 0.1  

m(1) = m(2) = m(3) = m(4) = 0; m(Ω) = 0.1

The same method is true for constructing the mass function of other attribute metrics c2, c5. At a certain time, attribute metrics in the existing equipment positioning and orientation system, c2, c3, and c5, are measured and obtain the following results, and it is E1 = \{17.5, 17.3, 19.2\}, E2 = \{13.7, 18.7\}. This is shown below in Table 3.

**Table 3.** Basic belief function distribution of E1 for different health status.

| sample | metrics | m(1) | m(2) | m(3) | m(4) | m(5) | m(Ω) |
|--------|---------|------|------|------|------|------|------|
| E1     | c2      | 0    | 0    | 0.3042 | 0.5958 | 0    | 0.1  |
|        | c3      | 0    | 0    | 0.6568 | 0.2432 | 0    | 0.1  |
|        | c5      | 0    | 0    | 0.6231 | 0.2769 | 0    | 0.1  |
| E2     | c2      | 0.9  | 0    | 0    | 0    | 0    | 0.1  |
|        | c3      | 0    | 0    | 0.6568 | 0.2432 | 0    | 0.1  |
|        | c5      | 0    | 0    | 0.6932 | 0.2077 | 0    | 0.1  |

The evaluation metrics set in Table 4, c2, c3, c5, is synthesized by using the formula (9) and results are shown in Table 4.

**Table 4.** Multi-attribute composition result of samples.

| sample | results | Super ideal | Ideal | Average | Below average | poor | Ω |
|--------|---------|-------------|-------|---------|---------------|------|----|
| E1     | c2 ⊕ c3 | 0           | 0     | 0.5334  | 0.4279        | 0    | 0.0187 |
|        | c2 ⊕ c1 ⊕ c5 | 0    | 0    | 0.7099  | 0.2870        | 0    | 0.0031 |
| E2     | c2 ⊕ c3 | 0           | 0     | 0.3458  | 0.1279        | 0    | 0.5263 |
|        | c2 ⊕ c1 ⊕ c5 | 0    | 0    | 0.7613  | 0.1771        | 0    | 0.0620 |

3.3. Result Analysis

According to the composite results, shown in Tables 3 and 4, when using a single evidence for determination, the basic credibility of the uncertainty of each evidence is greater than or equal to 0.1. After the synthesis, the uncertainty greatly decreases from 0.1 to 0.0031. The algorithm1 can obtain the reduction directly from the neighborhood decision table, and there is no need to further discretize the decision table, reducing average computing time. The algorithm1 can effectively eliminate redundant attributes, which reduces the computational time complexity, when applying the D-S evidence theory to fuse the feature metrics.

Second, the evidence theory is utilized to assess the three groups of evaluation metrics set, \{c2\}, \{c3\}, \{c5\}, in the evaluation evidence E1, E2. Comparing E1 with E2, both have the same or similar general state of health, except in the attribute metrics c2, where a significant difference can be seen. If the three attributes are used separately for judgment, the uncertainty is as high as 0.1 or 0.5263. By following the D-S evidence combination rule, the results of three attributes reduction are combined to
obtain that maximum reliability of 0.7099 and 0.7613, for E1, E2 respectively, which were regarded as the results of health status. Therefore, the assessment of health status is average according to data of E1, E2, and the uncertainty after fusion is greatly reduced.

4. Conclusion
In this article, a classification model is proposed, combining both the neighborhood rough set and evidence theory, and applied to the health condition grade assessment of a kind of device. Considering the evaluation metrics of health condition grade measures continuous number values, the forward greedy search algorithm of the neighborhood rough set can be employed to directly reduce the continuous attribute values in the decision table. This will not only help avoid data loss, but also reduce processing time for the attribute metrics in the decision table. After reduction, the mass function is constructed for the evaluation decision table, and the basic credibility distribution value of each focal element is collected. As the last step, the final evaluation result is obtained through Dempster’s rule of combination. This analysis shows that the method offered can effectively determine the main evaluation metrics under the condition of small samples, greatly reduce the uncertainty in the assessment process, and produce more reliable evaluation results.

Acknowledgments
Our research work is supported by National Natural Science Foundation of China(61741219).

References
[1] Geng J B Qiu W, Kong X C et al. Technical condition evaluation for devices based on rough set theory and D-S evidence theory[J].Systems Engineering and Electronics,2008,30(1):113-115.
[2] Duan J, Hu Q H, Zhang L J. Feature Selection for Multi-Label Classification Based on Neighborhood Rough Sets[J].Journal of Computer Research and Development,2015,52(1):56-65.
[3] Hu Q H,Yu D R, Liu J F et al. Neighborhood rough set based heterogeneous feature subset selection[J].Information Sciences, 2008, 178(18):3577-3594.
[4] Zhang W X, Wu Z W, Liang J F et al. Rough Set Theory and Method[M].Peking: Science Press, 2001: 1-19.
[5] An R M,,Suo M L. Application of attributes reduction and weights calculation through neighborhood rough set[J].Computer Engineering and Applications,2016,52(7):160-165.
[6] Shafer G. A mathematical theory of evidence[M]. Princenton University Press,1976,78-103.
[7] Li F J, Qian Y H, Wang J T et al. Multi granulation information fusion: a Dempster-Shafer evidence theory-based clustering ensemble method[J].Information Sciences,2017,378:389-409.
[8] Li W H, Wang Y P, Shang Y S et al. Research on health assessment based on hidden Markov model[C]//Proceedings of International Conference on Electronic &Mechanical Engineering and Information Technology. Washington D C USA: IEEE Computer Society,2011:4330-4332.
[9] Hu Q H, Zhao H, Yu D R. Efficient Symbolic and Numerical Attribute Reduction with Neighborhood Rough Sets[J]. Patten Regnization and Artificial Intelegence,2008,21(6):732-738.
[10] Weiwei Li, Zhiqiu Huang,Xinyi Jia,xinye Cai. Neighborhood based decision-theoretic rough set models[J].International journal of Approximate Reasoning 69(2016)1-17.
[11] Li B Y, Xiao J M, Wang X H. Power System Transient Stability Assessment Based on Hybrid Neighborhood Rough Reduction and Deep Forest [J].Transactions of China Electrotechnical Society.2020,35(15):3245-3257.
[12] Li Y,Cai Y Z, Y R P .Support Vector Machine Ensemble Based on Evidence Theory for Multi-Class Classification [J].Journal of Computer Research and Development.2008,45(2):571-578.
[13] Fuyuan Xiao. A new divergence measure for belief function in D-S evidence theory for multisensory data fusion[J].Information science,514(2020)462-483.
[14] Wang L, Lu Z Y, Li Z Y. Condition assessment method of equipment based on rough sets and evidence theory[J].Systems Engineering and Electronics, 2020,42(1):141-146.
[15] Qinghua Hu, Daren Yu, Jinfu Liu et al. Neighborhood Rough set based Heterogeneous Feature Subset Selection [J]. Information Science. 2008, 178: 3577-3594.