Attribute Reduction of Coal Mine Fire Incidents Based on Finding Maximum Mutual Information

Manting Zhang*, Weibing Feng,
College of science, Xi’an University of Science and Technology, Xi’an 710054, China
1564462093@qq.com

Abstract. There are many redundant attributes for the original data of the emergency system with uncertainty and data loss, and the existing incomplete information system attribute reduction algorithms have the disadvantages of high time complexity and completeness of the reduction result, an attribute reduction method for incomplete emergencies based on finding the maximum mutual information is proposed. Firstly, the original data of the emergency was preprocessed, and the situation and the consequences were regarded respectively as conditional attributes and decision attributes, thereby an incomplete decision information table was constructed; then, according to the characteristic that the mutual information of the conditional attribute set and the decision attribute is equal to the information entropy of the decision attribute in the consistency incomplete information system, the attribute reduction algorithm based on finding the maximum mutual information was obtained; finally, this method was used to reduce the attribute of coal mine fire data and compared it with other three methods. The results indicate that this method can effectively reduce the redundant attributes of coal mine fire emergencies, and can ensure that the reduction results have strong completeness when the algorithm time is low, thus decision-making for the emergency department is provided.

1. Introduction

In the field of emergency management, emergency is a small-probability complex event with insufficient precursors before the occurrence of the event, and it is highly destructive and potentially secondary hazard after the event. It is difficult to take traditional solutions to deal with it. In recent years, the frequency of emergencies has become higher and higher, and in particular coal mine fires have caused certain economic losses and casualties. The analysis and decision of emergency information is instructive to timely rescue and dispatch of manpower and material resources after the event. A lot of potential and meaningful knowledge and rules are contained in the raw data information of various emergencies. Each data item is composed of attributes that reflect the characteristics of the incident, but the criteria for dividing the attributes are ambiguous and the attributes are related to each other, so that there are many redundant attributes in the incident information table, which masks the knowledge in the information table, to some extent, it leads to errors in decision-making for emergencies. Therefore, the attribute information table is effectively reduced, and the redundant attributes are deleted, which is of great significance in assisting emergency decision-making.

In an emergency information system, information has uncertainty and data is missing, which is consistent with the characteristics of incomplete information systems. In the rough set model, the heuristic information about the attribute reduction algorithm of incomplete information systems mainly includes positive domain, difference matrix, and information entropy. At present, many scholars have applied the attribute reduction in rough set theory to emergencies in the field of...
emergency management. Gao Tian and Li Feng [2,3] respectively applied the attribute reduction algorithm based on attribute significance and positive domain to the emergency response of tourist emergencies; Chen Wu [4] constructed the risk assessment of mountain tunnel collapse as a decision information table, and used the attribute reduction method based on conditional information entropy to predict the risk level; Li Yinxu [5] applied the attribute reduction algorithm based on similarity matrix to the reasoning of emergency case examples, thereby optimizing the emergency case library. However, the literature [2–5] is only applicable to complete information systems. For incomplete emergency information systems, Zhong Qiuyan [6] proposed a heuristic attribute reduction algorithm based on the importance of attributes, which is based on the ordered weighted operator to calculate the importance of attributes, but there are some limitations for mixed attribute information systems. Based on the relationship between the information entropy of decision attributes and the mutual information of condition attributes in incomplete information systems, this paper proposes an attribute reduction algorithm to find the maximum mutual information and applies it to the attribute reduction of incomplete emergency information.

Aiming at the shortcomings of the existing algorithms, such as lack of completeness, high time complexity, and inability to deal with the redundant attributes of emergencies, because the knowledge entropy defined by the concept of information has better knowledge classification ability, a method for reducing the emergency event attributes based on mutual information is proposed. This method takes respectively scenario factors and event consequences as condition attributes and decision attributes, then establishes corresponding decision tables. It uses an incomplete decision-making information system knowledge reduction algorithm based on finding the maximum mutual information of condition attribute sets relative to decision attributes. The attribute reduction of the emergency decision information table is implemented. Finally, the method is used to implement the attribute reduction of the fire emergency information, and compared with other methods, which reflects the practicability and effectiveness of the algorithm.

2. Rough set and attribute reduction

For the processing of incomplete and inaccurate information systems, the use of rough set theory is an effective method. Its main advantage is that it can directly analyze and process the original data without any prior information, remove redundant attributes, and find the smallest related set, so get the core rules.

**Definition 1** [7]. $S = (U, A = C \cup D, V, f)$ is an information system. $U$ is non-empty finite set of objects, $A = C \cup D$ is non-empty finite set of attributes, $C$ is condition attributes set, $D$ is decision attributes set, and $C \cap D = \emptyset$; $V = \bigcup_{a \in A} V_a$, $V_a$ is the range of values of the attribute $a$; $f$ expresses an information function of $U \times A \rightarrow V$, which is assigned to each object an information value on each attribute, that is $a \in A, x \in U, f(x, a) \in V_a$, if $D = \emptyset$, the information system is called data table, otherwise it is called decision table. If there exists $x \in U, a \in C, f(x, a)$ is unknown (denoted as: $f(x, a) = \ast$), it is called incomplete information system; otherwise it is called complete information system.

**Definition 2** [8]. The tolerance relationship $T$ is defined as:

$$T(x, y) = (x, y) | x \in U \land y \in U \land \forall c_j (c_j \in B) \Rightarrow (c_j (x) = c_j (y) \lor c_j (x) = \ast \lor c_j (y) = \ast)$$  \hspace{1cm} (1)

$I_B(x)$ is expressed the set of objects $y$ that satisfy tolerance relationship with $x$ on the attribute subset $B$, namely $I_B(x) = \{ y | y \in T(x, y) \}$.

For tolerance relationship $T$, $B$ is a subset of attributes, note: $U/SM(B) = \{ T_B(x) | x \in U \}$. Then $T_B(x)$ is expressed the tolerance class of object $x$ on attribute set $B$. 
Definition 3[9]. Let \( S = (U, C \cup D, V, f) \) be an incomplete information system, where \( R \subseteq C \), 
\[
U/\mathcal{T}_R = \{T_R(x_1), T_R(x_2), \ldots, T_R(x_l)\}; \quad U/D = \{Z_1, Z_2, \ldots, Z_m\}.
\]

The entropy of \( R \) is defined as:
\[
H(R) = -\sum_{i=1}^{l} p(T_R(x_i)) \log_2 p(T_R(x_i))
\]
(2)

The conditional entropy of \( R \) relative to \( D \) is defined as:
\[
H(D|R) = -\sum_{i=1}^{l} p(T_R(x_i)) \sum_{j=1}^{m} p(Z_j | T_R(x_i)) \log_2 p(Z_j | T_R(x_i))
\]
(3)

where \( p(T_R(x_i)) = |T_R(x_i)|/|U| \), \( p(Z_j | X_i) = |Z_j \cap T_R(x_i)|/|T_R(x_i)| \).

The mutual information of \( R \) and \( D \) is defined as:
\[
I(R; D) = H(D) - H(D|R)
\]
(4)

According to mutual information the importance degree of attribute is defined as:
\[
SIG(a, R, D) = I(R \cup \{a\}; D) - I(R; D)
\]
(5)

Theorem 1. Let \( S = (U, C \cup D, V, f) \) be a consistent incomplete decision information system, Where \( U \) is universe, \( C \) is condition attribute set, \( D \) is decision attribute set, then \( I(C; D) = H(D) \).

Proof. Let \( \partial_a : U \rightarrow V \) be the generalized decision function, defined as:
\[
\partial_a(x_i) = \{i \mid f(d, y) \land y \in T_a(x_i)\},
\]

since \( S \) is consistent, it can be obtained for any \( x_i \in U \), there is \( \partial_a(x_i) = 1 \) [12], then there exist \( Z_j \in U/D, T_c x_i \subseteq \subseteq Z_j \). Then \( T_c x_i \cap Z_j = \phi \), which \( Z_r \in U/D, Z_r \neq Z_j \), that is
\[
\sum_{j=1}^{m} \frac{|T_c(x_i) \cap Z_j|}{|T_c(x_i)|} = 0 + 0 + \cdots + \log_2 1 + 0 + \cdots = 0,
\]

\[
H(D|C) = -\sum_{i=1}^{l} p(T_R(x_i)) \sum_{j=1}^{m} \left[\frac{|T_c(x_i) \cap Z_j|}{|T_c(x_i)|}\right] = 0,
\]

therefore \( I(C; D) = H(D) - H(D|C) = H(D) - 0 = H(D) \).

Theorem 1 shows that in the incomplete information system, only mutual information of condition attribute set and decision attribute is equal to information entropy of decision attribute, and then can obtain the relative reduction of the information system. Therefore, according to the calculation formula of mutual information, in the process of attribute reduction, only the condition attribute set corresponding to the maximum mutual information can be found. This avoids calculating mutual information of the redundant condition attribute set and decision attribute, thereby the amount of calculation is reduced, and the calculation efficiency is improved.

3. Mutual information-based method for reducing event attributes

In the emergency information system of emergency management, most of the emergency information with data missing and uncertainty is an incomplete information system in nature. The redundant attributes in the emergency system may lead to errors in emergency decisions. Therefore, it is necessary to reduce the attributes of incomplete emergency information systems. For the complete information system, Huang Weihua [10] proposed an attribute reduction algorithm that uses mutual information as the heuristic information which only needs to satisfy \( I(R; D) = I(C, D) \), so \( R \) is the attribute reduction of the information system. This algorithm verifies that mutual information can be
used to measure the classification ability of attribute sets, but it is necessary to calculate the mutual information of the decision attributes and all condition attribute sets which are composed of all condition attributes. The calculation amount is relatively large and the efficiency is relatively low.

3.1. Improved reduction algorithm of incomplete information system based on finding maximum mutual information

On the basis of attribute reduction algorithm based on mutual information in complete information system [10], the equivalence relation is extended to the tolerance relationship. According to the definition of information entropy in incomplete information system, firstly, calculating the mutual information of each condition attribute $c_i$ and decision attribute $D$, then finding the corresponding condition attribute when the mutual information is the maximum value, which is selecting the condition attribute with the greatest importance measurement, and the mutual information is used as the iterative criterion, the corresponding condition attribute is superimposed to satisfy $I(R; D) = H(D)$, then $R$ is the relative reduction of the information system. The flowchart is shown in Figure 1:

![Flow chart of an incomplete information system reduction algorithm based on finding the maximum mutual information](image)

The incomplete information system reduction algorithm based on finding the maximum mutual information

1. Obtain relative reduction algorithm in incomplete information system.
2. Input: incomplete information system $S = (U, A, V, f)$
3. Output: relative reduction $R$ in $S$
4. a) Let $R = \emptyset$, $U/D = \{Y_1, Y_2, \cdots, Y_m\}$
5. b) Solve the information entropy $H(D)$ of the decision attribute $D$
6. c) Let $R = R \cup \{c_i\}$, calculate tolerance class $T_x(x_i)$, where $x_i \in U, i = 1, \cdots, |U|$, calculate the conditional entropy $H(D|R)$ of $R$ relative to the decision attribute $D$, according to the formula $I(R; D) = H(D) - H(D|R)$, the mutual information $I(R; D)$ of $R$ and $D$ is obtained;
7. d) Select the maximum value of mutual information, then $I(R; D) = \max I(R \cup \{c_i\}; D)$, which is selecting the attribute with the greatest importance measure $SIG(a, R, D)$;
8. e) $R = R \cup \{c_i\}$, if $I(R; D) = H(D)$ is satisfied, then $R$ is the relative reduction of $C$, and $R$ is directly output, otherwise turn c).
3.2. Emergency attribute reduction algorithm

According to the characteristics of fire emergency data, combined with the incomplete decision-making information system attribute reduction algorithm based on finding maximum mutual information, a new fire emergency attribute reduction algorithm is proposed on this basis. As shown in Figure 2:

![Flow chart of fire emergency attribute reduction algorithm](image_url)

Figure 2. Flow chart of fire emergency attribute reduction algorithm

At present, the emergency attribute reduction methods proposed in most literature are only applicable to complete emergency information, but in reality, there are a large number of incomplete emergency information, and the existing incomplete emergency attribute reduction algorithm has high time complexity and incomplete reduction results. From the perspective of information, this paper applies the improved reduction algorithm based on finding the maximum mutual information to the fire emergency attribute reduction. The basic idea is: firstly, preprocessing the data on the original information table; then classify the attributes, consider respectively the scenario factors and consequences of the incident as conditional attribute sets and decision attributes, so as to establish a consistent incomplete information decision table; finally, an attribute reduction algorithm based on finding maximum mutual information is used for attribute reduction of incomplete information decision tables.

4. Fire emergency case analysis

In the emergency information, the scenario information at the time of the disaster and the information about the casualties and property damage caused by the disaster are recorded. There is a causal relationship between the former and the latter, and there is a correlation between many scenarios. In fire emergencies, different levels of fire are caused by different fire scenario information. The correlation between fire scenario factors such as time, wind, temperature, and relative humidity obscures the causal relationship between fire scenarios and fire consequences. If the redundant attributes can be removed, the correlation between fire information and response levels can be found, and the understanding of the fire situation can be improved. Through the decision-making correlation, a new emergency response level for fire incidents can be obtained, so as to determine the severity of fire incidents for relevant departments and make corresponding emergency rescue dispatch, then can provide decision support. According to the actual situation of a certain coal mine, the fire information data provided in [6] was partially revised to obtain the initial fire information table shown in Table 1. Among them, the Fire rating indicates the fire situation and the size of manpower and materials that should be arranged for rescue.
Table 1. Fire initial information table

| U  | Time | Temperature | Relative humidity | Wind power | Population density | Fire rating |
|----|------|-------------|-------------------|------------|--------------------|-------------|
| 1  | unknown | unknown | unknown | 12 | unknown | 1 |
| 2  | day | 10~15 | 28.3 | unknown | 85.2 | 3 |
| 3  | unknown | 4~10 | 76.1 | 2 | 634.8 | 2 |
| 4  | night | 17~23 | unknown | 7 | 476.6 | 4 |
| 5  | night | 2~8 | 42.9 | 3 | 584.3 | 2 |
| 6  | day | 12~18 | 18.5 | 5 | 162.9 | 3 |
| 7  | night | 5~9 | unknown | 2 | 693.3 | 2 |
| 8  | night | unknown | 64.7 | 4 | unknown | 4 |
| 9  | unknown | 27~23 | 53.3 | 7 | 218.4 | 4 |
| 10 | day | 16~19 | 30.4 | 3 | 118.9 | 3 |
| 11 | day | unknown | 22.6 | unknown | 27.4 | 3 |
| 12 | day | 30~33 | 44.5 | 4 | 246.5 | 2 |
| 13 | day | 29~34 | 59 | 5 | 387.1 | 2 |

First preprocess the data and replace the unknown values in the data with "*"; determine discrete breakpoints based on the data characteristics of each attribute to discretize the data; because there are multiple unknown attribute values in object 1, object 1 is regarded as noise data and deleted. At the same time, the Fire rating is used as the decision attribute, and the remaining fire scenario factor attributes are used as the condition attributes. Establish a corresponding fire information decision table, as shown in Table 2.

Table 2. Fire information decision table

| U  | $c_1$ | $c_2$ | $c_3$ | $c_4$ | $c_5$ | $D$ |
|----|-------|-------|-------|-------|-------|-----|
| 1  | day   | middle | lower | *     | lower | 3   |
| 2  | *     | lower  | high  | I     | high  | 2   |
| 3  | night | high   | *     | III   | middle| 4   |
| 4  | night | lower  | middle | I   | high  | 2   |
| 5  | day   | middle | lower | II    | lower | 3   |
| 6  | night | lower  | *     | I     | high  | 2   |
| 7  | night | *      | high  | II    | *     | 4   |
| 8  | *     | high   | middle | III  | middle| 4   |
| 9  | day   | middle | lower | I     | lower | 3   |
| 10 | day   | *      | lower  | *     | lower | 3   |
| 11 | day   | high   | middle | II   | middle| 2   |
| 12 | day   | high   | middle | II   | middle| 2   |

Next, use the algorithm based on finding the maximum mutual information proposed in this paper to perform attribute reduction on the consistent incomplete decision information table 2 obtained above. Use $c_1, c_2, c_3, c_4, c_5$ to represent the five condition attributes, and use $D$ to represent the decision attribute fire level.

$U/D = \{Y_1, Y_2, Y_3\}$, where $Y_1 = \{1, 2, 9, 10\}$, $Y_2 = \{2, 4, 6, 11, 12\}$, $Y_3 = \{3, 7, 8\}$, $H(D) = 0.468$

When $R = \phi$, calculating mutual information:

$I(\{c_1\}; D) = -2.762$; $I(\{c_2\}; D) = -2.134$; $I(\{c_3\}; D) = -1.975$; $I(\{c_4\}; D) = -2.149$; $I(\{c_5\}; D) = -1.092$.

From this $I(R; D) = I(\{c_1\}; D)$;

When $R = \{c_1\}$.

$I(\{c_1, c_2\}; D) = -0.186$; $I(\{c_1, c_3\}; D) = -1.092$; $I(\{c_1, c_4\}; D) = -0.205$; $I(\{c_1, c_5\}; D) = -0.151$.

Therefore $I(R; D) = I(\{c_1, c_5\}; D)$.

When $R = \{c_4, c_5\}$.
Through three iterations, the relative reductions of the fire information table are \( R = \{c_1, c_4, c_5\} \) and \( R = \{c_1, c_4, c_5\} \).

The reduction results indicate that temperature is a redundant attribute, and it can be deleted from the scenario factors to obtain two reduction decision tables. When a fire event occurs, correct decisions and judgments can be made based on relative humidity, wind and population density conditions or time, and wind and population density conditions, so as to improve the understanding of the extent of the disaster and determine the size of the rescue force dispatched. It avoids the situation that the rescue force is insufficient due to improper estimation of the disaster situation and the precious rescue opportunity is missed.

Compare the algorithm proposed in this paper with the algorithms in [6], [8], and [11] in terms of completeness of reduction results and algorithm time complexity in the fire emergency problem, as shown in Table 3 and Table 4.

| Algorithms | Reduction results |
|------------|-------------------|
| Literature [6] | \( \{c_1, c_4, c_5\} \) |
| Literature [8] | \( \{c_1, c_4, c_5\} \) |
| Literature [11] | \( \{c_1, c_4, c_5\} \) and \( \{c_1, c_4, c_5\} \) |
| The algorithm in the paper | \( \{c_1, c_4, c_5\} \) and \( \{c_1, c_4, c_5\} \) |

| Algorithms | Inspirational information | Time complexity |
|------------|--------------------------|-----------------|
| Literature [6] | Attribute significance | \( O(|\mathcal{C}||\mathcal{U}|) \) |
| Literature [8] | Positive domain | \( O(|\mathcal{C}||\mathcal{U}|) \) |
| Literature [11] | Conditional entropy | \( O(|\mathcal{C}||\mathcal{U}|) \) |
| The algorithm in the paper | Maximum mutual information | \( O(|\mathcal{C}|(|\mathcal{C}|-1)||\mathcal{U}|) \) |

Through a comparative analysis of algorithm reduction results and algorithm time complexity, we can get that: in the algorithm proposed in this paper and literature [11], the attribute reduction both from the perspective of information theory, so the algorithm in this paper and literature [11] have stronger completeness than the literature [6] and [8] in the reduction results. In terms of algorithm time complexity, the time complexity of the algorithm proposed in this paper is obviously lower than the algorithm time complexity of the literature [6], [8], [11]. Reference [11] adopts global operations with high time complexity, however, in this paper, in order to satisfy \( I(R; D) = H(D) \) when calculating the mutual information of the condition attribute set and the decision attribute, only the attribute set superimposed on the condition attribute corresponding to the maximum value of the mutual information value is selected, which avoids global operations and improves the calculation efficiency, thereby It shows the effectiveness of the algorithm based on finding the maximum mutual information.

5. Conclusions

The emergency system has uncertainty, many redundant attributes, and missing data. Based on the rough set theory, this paper proposes a method of incomplete emergency information attribute reduction based on finding the maximum mutual information. The method is applied to the attribute reduction of coal mine fire emergency events, which verifies the practicability and effectiveness of the algorithm in attribute reduction of emergency information, and solves the problem of redundant attributes of emergency events in the absence of information. The problem is eliminated, the information table is simplified, the accuracy of the awareness of the disaster level is improved, and it
plays a certain auxiliary role in the rescue dispatch decision. The next step is to study how to use the rough set theory to efficiently extract emergency rules.

Acknowledgment
This work was financially supported by National Natural Science Foundation of China (51504186)

References
[1] Pawlak Z. Rough set theory and its application to data analysis[J].Cybernetics and Systems,1998,9(4):661-668.
[2] GAO Tian, DU Jun-ping, WANG Su. Attribute reduction of tourism emergencies based on rough set[J].Journal of Southeast University (Natural Science Edition),2009,39(1): 164-167.
[3] LI Feng, LIU Xin-ran. Application of rough set theory in emergency response of tourism emergencies[J].Technical Economy,2019,38(3): 78-87.
[4] CHEN Wu, ZHANG Guo-hua, WANG Hao, et al. Evaluation of collapse risk of mountain tunnel based on rough set condition information entropy[J].Rock and Soil Mechanics,2019, 40(9):1-10.
[5] LI Yin-xu. Application of rough sets in case reasoning of sudden crisis events[J].Journal of Shaanxi University of Science and Technology (Natural Science Edition),2011,29(5): 107-112.
[6] ZHONG Qiu-yan, WANG Ran, QU Yi. A method of contingency event reduction based on rough set[J].Operations Research and Management,2018,27(1):89-95.
[7] HUANG Bing,ZHOU Xianzhong . Attribute Reduction of Incomplete Information System Based on Information Quantity[J]. System Engineering Theory and Practice,2005,25(4): 55-60.
[8] YANG Cheng-fu, SHU Lan. Attribute reduction algorithm for incomplete decision systems based on tolerance relation[J]. Computer Technology and Development,2006,16(9):68-69.
[9] LIU Gui-long, HUA Zheng, ZOU Ji-yang. Local attribute reductions for decision tables[J]. Information Sciences,2018,422:204-217.
[10] HUANG Wei-hua. Attribute reduction algorithm based on information theory[J]. Journal of Hubei Institute for Nationalities(Natural Science Edition),2018,36(3):289-292.
[11] DAI Jian-hua. Attribute selection based on a new conditional entropy for incomplete decision systems[J].Knowledge-Based Systems,2013,39: 207-213.
[12] HU Feng. Knowledge reduction method of incomplete information system based on decision entropy[J]. Computer engineering and design,2013,34(1):289-292.
[13] YANG Jilin . Importance and reduction algorithm of attributes in incomplete information systems[J]. Computer Engineering and Applications,2010,46(1):99-102.
[14] FU Ang,WANG Guoyin . Attribute Reduction Algorithm in Incomplete Information System Based on Information Entropy[J]. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition),2008,20(5): 586-592.
[15] H. Zhao, K.Y. Qin, Mixed feature selection in incomplete decision table[J]. Knowledge-Based Systems, 2014,57: 181-190.
[16] SHU Wen-hao, QIAN Wen-bin. A fast approach to attribute reduction from perspective of attribute measures in incomplete decision systems[J].Knowledge-Based Systems,2014,72: 60-71.
[17] ZHANG Yan-qiong, JIANG Xun, XU Xu-kan. Study on emergency retrieval model based on hierarchical rough set[J].Information Science, 2018,36(10):30-47.
[18] CHAO Ying, WANG Qing-rong. Emergency material demand forecast based on fuzzy rough set case reasoning[J].Transportation Technology and Economy,2015,17(5):1-4.
[19] FENG Wei-bing, ZHANG Man-ting. Reduction algorithm based on finding the maximum mutual information in incomplete information systems [C]//Journal of Physics:Conference Series.IOP Publishing,2019,1237(2):022020.