Acquisition of missile fault diagnosis knowledge based on incomplete information of flow graph

Zhaozheng Liu*, Mingqing Xiao, Haizhen Zhu and Jianfeng Li
Aeronautics Engineering School of Air Force Engineering University, Xi’an, China

*Corresponding author e-mail: zhaozhengliu@foxmail.com

Abstract. Aiming at the problem of missing information and incomplete information in missile fault diagnosis, a method of acquiring missile fault diagnosis knowledge based on flow graph is proposed. Firstly, the fault feature information is extracted, and the fault diagnosis knowledge is classified by using the feature relationship to obtain a set of fault diagnosis examples under the same feature relationship. Then, the flow graph of incomplete fault diagnosis is established, and the confidence level and coverage level between each node are expressed. The knowledge acquisition framework of missile fault diagnosis is constructed. Finally, combining with the case analysis, it is verified that the fault diagnosis knowledge acquisition method has a good intuition, which provides an effective reference for missile fault diagnosis.

1. Introduction
With the continuous development of science and technology and the continuous improvement of the level of mechanical automation, the satisfaction of mechanical functions needs to increase the complexity of the system to achieve. In recent years, mechanical failures caused by complex machinery have also led to frequent occurrence of accident cases, and ensuring the safe and stable operation of construction machinery equipment has become an urgent problem to be solved.

The flow graph theory was proposed by the Polish scholar Z. at the 2003 International Conference on Rough Set Theory [1]. It is a new method for graphically expressing the fault symptom attribute value and fault type. It is applied to machine learning, pattern recognition, and systems research in various professional fields. Huang Wentao applied the flow graph to data mining and decision rule extraction [2]. Yu Jun used the flow graph to diagnose the knowledge of gearbox fault diagnosis [3]. The extensive use of flow graphs has good prospects for fault diagnosis under incomplete information.

Traditional missile fault diagnosis mainly uses limited diagnosis knowledge and combined with the experience of domain experts to obtain the diagnosis results, and the incompleteness of diagnosis knowledge brings great problems to the diagnosis process. For the specific problems of fault diagnosis, the flow graph of diagnostic reasoning can be expressed formally, which provides a clearer picture of the relationship between the cause of the fault and the symptoms of the fault under incomplete conditions. Firstly, this paper analyses the definition of an incomplete information system and specifies the main components of an incomplete information system. Secondly, it processes the incomplete fault diagnosis information according to the characteristic relationship to improve the accuracy of the classification of missile fault diagnosis examples. Finally, the use of incomplete
information flow graph is used to acquire knowledge, calculate the confidence and coverage of the path, and obtain the final complete path, which provides an effective reference for the fault diagnosis of the missile.

2. Incomplete information system

Incomplete information systems are compared with traditional information systems. Traditional information systems are usually complete. And it is that the value of an object under certain attributes exists and is unique. The actual system is often caused by various problems such as data deviation, inaccurate definition of data, etc. For example, in the gear box system, because of the various types of failures and the different reasons for the failure, the fault diagnosis information of gearbox is incomplete information with deviation [4].

Compared with the traditional complete information system, the definition of an incomplete fault diagnosis information system can be obtained due to the occurrence of mechanical failures in the process of information acquisition. Or, in the process of information transmission, erroneous diagnosis information appears, which leads to a missing attribute value and form an incomplete fault diagnosis information system. By analogy with the definition of a complete information system, the definition of an incomplete information system can be obtained as follows:

Establish a complete incomplete information system, a quadruple \( S = \langle U, AT, V, f \rangle \) can be described as an information system \( S \), \( U \) is a finite non-empty set representing the object, and it is called the universe of discourses; \( AT \) is a finite non-empty set representing all the attributes describing the object, and it is called the attribute set; \( V_a \) represents all possible values of the attribute \( a \), and it is called the attribute value range; \( V \) represents all possible values of an attribute, and it is called an attribute value range; we define an information function \( f \) as a mapping which is from \( U \times AT \) to \( V \).

For example, in a typical example of fault diagnosis of incomplete information in a missile guidance system, in order to determine whether the gyro in the system is intact, we take test points \( TP_1, TP_2, TP_3 \) and measure the frequency of the gyro to obtain the system fault diagnosis decision table as shown in the following figure:

| \( U \) | \( TP_1 \) | \( TP_2 \) | \( TP_3 \) | \( d \) |
|---|---|---|---|---|
| \( u_1 \) | VH | L | L | Y |
| \( u_2 \) | H | H | * | Y |
| \( u_3 \) | VH | H | H | Y |
| \( u_4 \) | H | L | N | Y |
| \( u_5 \) | VH | N | * | Y |
| \( u_6 \) | H | N | N | N |

Among them, the object set \( U = \{ u_1, u_2, u_3, u_4, u_5, u_6 \} \) represents the six health states of the gyro. \( AT = \{ T_1, T_2, T_3 \} \) are the three symptom attributes of the system \( V_{T_1} = \{ L, N, H, VH \} \), \( V_{T_2} = \{ L, N, H \} \), \( V_{T_3} = \{ *, L, N, H \} \), where \( L \) means low, \( N \) means normal, \( H \) means high, \( VH \) means very high, and * means null.

For an incomplete information system \( S = \langle U, AT, V, f \rangle \), let the attribute set \( B \subseteq AT \). For \( a \in AT \) and \( v \in V \), \( (a, v) \) is an attribute value pair; if the attribute value is known, then the set of all \( x \in U \) satisfied instances \( a(x) = v \) is the attribute value pair block of attribute value pairs, which is recorded as \( [(a, v)] \) [5].

Define the unknown attribute values as the following two types. The first type is called the missing attribute value. It is represented by the symbol "?" and is used to define the information lost during the gyro frequency measurement. Attribute values of this category cannot be replaced with known
attribute values. The second type is called unknown attribute value, which is indicated by the symbol "*" and is used to define unknown information, but it can be replaced by a known attribute value. Then, the attribute pair block with missing attribute value is \([a,?]\) and the attribute pair block with unknown attribute value is the intersection of all contained attribute \(a\) pair blocks.

Set an incomplete information system \(S = (U, A, V, f)\). Define the attribute set \(B \subseteq A T\) and for \(\forall x \in U\), the feature set \(B\) of the instance \(x\) under the attribute set \(K(x)\) is the intersection of all attribute value pairs.

\[
K_x = \bigcap_{a \in B, a \in V} \{ y \in U | y \in [(a, v)] \}
\]

(1)

The characteristic relationship \(R(B)\) is defined as follows:

\[
R(B) = \{(x, y) \in U \times U | y \in K_x \}
\]

(2)

The feature relationship is used to process the two unknown attribute values, which is based on its performance superior to the tolerance relationship and the asymmetric similarity relationship, and can deal with the incomplete faults of two different attribute categories [6].

3. Incomplete fault diagnosis flow graphs

Define the fault diagnosis information system flow graph under incomplete conditions as a directed, acyclic graph \(IG = (N, M, D, B, \varphi)\); \(N\) is a collection of nodes; \(M\) is a collection of symptom attribute nodes, and \(M = \{m_1, m_2, \ldots, m_m\}\) is that the value range of the symptom attribute \(m_i\) refers to the set of all states of the system. \(D\) is a set of decision attribute nodes. The value range of decision attributes refers to the set of decision attributes \(d\) of all states of the system. \(B \subseteq N \times N\) is a set of directed branches; \(\varphi\) is defined as flow, and the input and output of any attribute node \(x\) are represented by \(I(x)\) and \(O(x)\).

Define any symptom attribute node of incomplete fault diagnosis flow graph, and define the input flow as:

\[
\varphi_i(m_i(x)) = \sum_{m_i(x) \in \varphi(R_{B}(m_i(x)))} \varphi(m_i(x), m_{-i}(x))
\]

(3)

The output flow is defined as:

\[
\varphi_i(m_i(x)) = \sum_{m_i(x) \in \varphi(R_{B}(m_{-i}(x)))} \varphi(m_{-i}(x), m_i(x))
\]

(4)

The input flow of any decision attribute node is defined as:

\[
\varphi_i(d_j(x)) = \sum_{d_j(x) \in \varphi(R_{B}(m_i(x)))} \varphi(d_j(x), m_i(x))
\]

(5)

The directed path from the symptom attribute node \(m_i(x)\) to the decision attribute node \(d_j\) is defined as \(m_i(x), \ldots, d_j\), and its complete path can be expressed as \([m_i(x), \ldots, d_j]\).

Through the incomplete flow graph, each instance in the decision table can be represented with a complete path. In order to quantitatively analyse the reliability of each path in the incomplete flow graph, the concepts of confidence and coverage are introduced, which are defined as follows.

The confidence level of the incomplete flow graph is defined as:

\[
\text{cer}[m_i(x), d_j] = \frac{\text{instance}\{[m_i(x), \ldots, m_i(x)]\}}{\text{instance}\{[m_i(x), \ldots, m_i(x), d_j]\}}
\]

(6)
Where the \( \text{instance}[[m_1(x), \cdots, m_l(x), d_j]] \) is the number of instances representing known attribute nodes in the path, and \( \text{instance}[[m_1(x), \cdots, m_l(x)]] \) is the number of instances representing known attribute nodes in the path condition only.

The coverage level of the incomplete flow graph is defined as:

\[
\text{cov}[m_i(x), d_j] = \frac{\text{instance}[[d_j]]}{\text{instance}[[m_1(x), \cdots, m_l(x), d_j]]}
\]  

Where the \( \text{instance}[[m_1(x), \cdots, m_l(x), d_j]] \) is the number of instances representing known attribute nodes in the path, and \( \text{instance}[[d_j]] \) is the number of instances satisfying only the attribute nodes of the path flow decision part.

In the incomplete flow graph, the confidence indicates the reliability of the path. The greater the confidence, the higher the reliability of the fault diagnosis decision made according to the path; the coverage indicates that the path accounts for the path that flows through its decision attribute nodes proportion.

4. Case Analysis

4.1. Analysis of the reason for the incomplete fault information of the missile

Missile incomplete fault diagnosis information refers to the fact that certain data or some cases have incomplete or missing unknown attribute values during the missile test process, which in turn affects the accuracy of the missile's fault diagnosis. Many reasons may cause incomplete information on missile fault diagnosis. For example, the source of unknown attributes may be due to the failure or damage of the equipment, or it may be due to the interference of data during the transmission process, and the temporary interruption of communication may cause the information to be ineffective collection. Based on the above viewpoints, the reasons for incomplete missile fault diagnosis information can be divided into the following three points:

1. The key components of the missile have failed. In the process of missile fault diagnosis, due to the influence of equipment working environment and complicated working conditions, it needs to perform fault detection.

2. It is difficult to obtain the attribute value of the fault symptom. Because the missile has the characteristics of long-term storage and one-time use, it is difficult to obtain sufficient fault diagnosis information in a short time.

3. The impact of data normalization. In order to extract the effective information in the data, it is generally necessary to normalize the data to obtain the attribute value of the fault symptom at the abstract level.

The above three reasons have caused incomplete missile fault information and increased the difficulty of missile fault diagnosis.

4.2. Establish incomplete fault diagnosis flow graph

Due to the strong connectivity between the various subsystems and components in the missile system, the missile's failure is often at a certain level of the subsystem or a smaller unit. When determining the fault diagnosis knowledge of the missile, it is necessary that the failure diagnosis knowledge acquisition framework is established for the subsystems of key components. The final overall diagnosis knowledge acquisition result is obtained, which is used to improve the accuracy of missile fault diagnosis.

The gyro is a key component of the missile guidance system. It is mainly composed of a gyro rotor, a gimbal support, and a motor. It uses its fixed axis to maintain the normal operation of the seeker and
uses precession to track the target. Taking the gyro of the missile guidance system as an example, the steps of acquiring its fault diagnosis knowledge under incomplete conditions are as follows.

First, the collected fault diagnosis information is used to extract fault features, and an incomplete gyro incomplete fault diagnosis system diagnostic table is established. Secondly, the information is processed based on the feature relationship to improve the accuracy of fault classification. Finally, the flow of incomplete fault diagnosis is drawn. It generates a complete fault diagnosis path, and finally achieves the purpose of fault diagnosis knowledge expression and acquisition. The specific process is shown in the following figure:

**Figure 1. Knowledge acquisition process.**

4.3. Fault diagnosis knowledge acquisition example

Study and analyze the knowledge of fault diagnosis. Table 2 is a set consisting of 10 fault diagnosis instances, namely \( U = \{u_1, u_2, u_3, u_4, u_5, u_6, u_7, u_8, u_9, u_{10}\} \), where \( U_1 = \{u_1, u_2, u_3, u_4, u_5, u_6, u_7, u_8\} \) is the faulted instance set and \( U_2 = \{u_9, u_{10}\} \) is the non-faulted instance set. \( AT = \{t_1, t_2, t_3\} \) is the symptom attribute set. The test points \( TP_1, TP_2, TP_3 \) are selected to measure the gyro frequency offset of the missile's subsystem, and the data is discretized. \( k \) is the number of identical instances at the same test point. Out of the confidentiality requirements for data, this paper uses the degree of deviation from the normal data to represent the value of the test node data that is small, normal and large. If it is too small, then \( t(x) = 1 \); if it is normal, then \( t(x) = 2 \); if it is too large, then \( t(x) = 3 \). The decision attribute \( d \) is and its value range is expressed as \( \{Y, N\} \), where the decision attribute value \( Y \) indicates that the gyro has failed, and the decision attribute value \( N \) indicates that the gyro is in good working condition.
Using characteristic relationships to process Table 2, according to the definition of attribute values on the block, all attribute value pairs in the incomplete fault diagnosis information system of missile gyro are obtained as follows:

\[(t_1, 1) = \{u_1, u_3, u_4, u_5, u_6, u_7, u_8\}\)
\[(t_2, 2) = \{u_1, u_2, u_4, u_5, u_6, u_7, u_8, u_9\}\)
\[(t_3, 3) = \{u_2, u_3, u_4, u_5, u_6, u_7, u_8, u_9, u_{10}\}\)

According to the definition of the feature set, the feature set of each instance in the incomplete fault diagnosis information system of the gyroscope is calculated as follows:

\[K_c(u_1) = [(t_1, 1)] \cap [(t_2, 2)] \cap U = \{u_1, u_2\}\]

Similarly, we can get:

\[K_c(u_2) = \{u_2\} , K_c(u_3) = \{u_3\} , K_c(u_4) = \{u_4\} , K_c(u_5) = \{u_5, u_4\} , K_c(u_6) = \{u_2, u_6\} , K_c(u_7) = \{u_7\} , K_c(u_8) = \{u_8\} , K_c(u_9) = \{u_9\} , K_c(u_{10}) = \{u_9, u_{10}\}\]

According to the definition of the characteristic relationship in Section 1, the characteristic relationship of each instance in the incomplete diagnostic information system is as follows:

\[R(C) = \{(u_1, u_1), (u_1, u_7), (u_2, u_2), (u_3, u_3), (u_4, u_4), (u_5, u_5), (u_6, u_6), (u_7, u_7), (u_8, u_8), (u_9, u_9), (u_9, u_{10}), (u_{10}, u_{10})\}\]

A graphical representation of the above-mentioned fault diagnosis decision table, using incomplete fault diagnosis flow graph, the relationship between the attribute value of the fault symptom and the attribute value of the fault decision is clearly and clearly expressed. According to the gyro frequency measurement results of the three test points, the actual symptom attributes of each test point are determined. The decision attribute value indicates whether a fault occurs, and the incomplete attribute

| Table 2. Gyro incomplete fault diagnosis decision table. |
|-----------------|---|---|---|---|---|
| U   | k | t_1 | t_2 | t_3 | d  |
|-----|---|-----|-----|-----|----|
| u_1 | 3 | 1   | 2   | ?   | Y  |
| u_2 | 1 | 3   | *   | 1   | Y  |
| u_3 | 1 | *   | 3   | 1   | Y  |
| u_4 | 2 | 1   | 3   | 1   | Y  |
| u_5 | 3 | 2   | 3   | 1   | Y  |
| u_6 | 2 | 3   | 2   | ?   | Y  |
| u_7 | 2 | 1   | *   | 2   | Y  |
| u_8 | 1 | *   | 1   | *   | Y  |
| u_9 | 3 | 2   | *   | 2   | N  |
| u_{10}| 2 | 2   | 2   | 2   | N  |
values are indicated by "*" and "?". Establish an incomplete fault diagnosis flow graph, as shown in the following figure.

![Flow graph of incomplete fault diagnosis.](image)

The above example is based on the analysis of 20 typical fault samples. According to the relevant theory of section 2 for the processing of incomplete diagnosis flow graphs, the path of fault diagnosis knowledge can be fully expressed. Calculate the confidence and coverage of the path. Some complete paths and their related indicators are shown in the following table:

| Instance | Complete paths | Confidence level | Coverage level |
|----------|----------------|------------------|----------------|
| u1       | \{(t1,1),(t2,2),(t3,?),(d,Y)\} | 8/29             | 5/29           |
| u2       | \{(t1,3),(t2,*),(t3,1),(d,Y)\} | 8/25             | 7/25           |
| u3       | \{(t1,*),(t2,3),(t3,1),(d,Y)\} | 13/28            | 7/28           |
| u4       | \{(t1,1),(t2,3),(t3,1),(d,Y)\} | 3/7              | 1/5            |
| u5       | \{(t1,2),(t2,3),(t3,1),(d,Y)\} | 4/9              | 7/36           |
| u6       | \{(t1,3),(t2,2),(t3,?),(d,Y)\} | 7/25             | 1/5            |
| u7       | \{(t1,1),(t2,*),(t3,2),(d,Y)\} | 4/29             | 2/29           |
| u8       | \{(t1,*),(t2,1),(t3,*),(d,Y)\} | 1/16             | 1/16           |
| u9       | \{(t1,2),(t2,*),(t3,2),(d,N)\} | 2/5              | 1/4            |
| u10      | \{(t1,2),(t2,2),(t3,2),(d,N)\} | 1/3              | 5/27           |

From the above path calculation results, it can be known that this method can quantitatively describe the reliability of the fault diagnosis path, and has achieved good results in the acquisition and utilization of incomplete fault diagnosis knowledge.

5. Conclusion

This paper presents a method for acquiring knowledge of missile fault diagnosis under incomplete information. The feasibility and effectiveness of this knowledge acquisition method are verified by the gyroscopic fault examples of the missile guidance system, and the fault diagnosis examples are classified by feature relationship processing, and reliable fault diagnosis information is obtained using limited data. The incomplete flow graph is used to make the incomplete fault diagnosis decision table concrete and graphical, and to quantitatively describe the relationship between attribute values, which is convenient for reference in the classification and processing of actual missile fault diagnosis data.
Acknowledgments
The author would like to thank all of my teachers and classmates who help me solve my problems in the difficult process of my paper.

References
[1] Z. Pawlak. Rough sets, decision algorithms and Bayes' theorem [J]. European Journal of Operational Research. 2002, 136 (1): 181-189.
[2] W. T. Huang, P. L. Niu, Y. F. Liu. Spur Bevel Gearbox Fault Diagnosis Based on Wavelet Packet Transform for Feature Extraction and Flow Graph Data Mining [C]. In: Conference on Advanced Engineering Materials and Technology, 2013: 2297-2302.
[3] Yu J, Huang W, Zhao X. Knowledge discovery for gearbox fault diagnosis using flow graph [C]// International Conference on Computer Science & Network Technology. IEEE, 2016.
[4] M. Wang, N. Hu, G. Qin. A method for rule extraction based on granular computing: Application in the fault diagnosis of a helicopter transmission system [J]. Journal of Intelligent and Robotic Systems: Theory and Applications. 2013, 71 (3-4): 445-455
[5] J. W. Grzymala-Busse, P. G. Clark, M. Kuehnhausen. Generalized probabilistic approximations of incomplete data [J]. International Journal of Approximate Reasoning. 2013, 55 (1): 180-196.
[6] J. W. Grzymala-Busse. Rough set strategies to data with missing attribute values [C]. Proceedings of the 2nd Workshop on Foundations and New Directions of Data Mining, Melbourne, 2003: 197-212