Complexity metrics for auto fault diagnosis based on information entropy

Ying CHENG, ZhiWei GUAN, HongLin ZHAO

School of Automotive & Transportation, Tianjin University of Technology & Education, Tianjin 300222, China

Email: chengying1650@aliyun.com

Abstract: Due to plenty of advanced and electronic technology used in the modern auto, the modern auto maintenance workload should consider the influence of the fault diagnosis complexity on maintenance man hour, so it is necessary and valuable to evaluate the complexity of the auto fault diagnosis. In this paper, the method of complexity assessment of auto fault diagnosis is discussed by using the information entropy. Firstly, the theory of complexity measurement based on information entropy is introduced, and then the complexity assessment index and the calculation process are shown. Finally, take the engine overheating fault for example, the hierarchy structure of this fault diagnosis is analyzed, and the complexity value of this fault diagnosis based on information entropy is obtained. The results show that this assessment method has a certain practicality and feasibility, which could be utilized to optimize the auto fault diagnosis strategy and maintenance work arrangement.

1. Introduction

The complexity of fault diagnosis can directly affect whether the maintenance job can be effectively implemented, the level of the maintenance work will obviously affected by the complexity of fault diagnosis, in order to draw up a reasonable and effective maintenance project, the complexity of the fault diagnosis should be determined by experiments in the process of establishing plans. However, the experimental tests takes a lot of time, manpower and material resources due to the diversity and complexity of fault types, and thus the completion of this work is very difficult.

The complexity is based on the diversity and difference. Filiz (2010) proposed an approach to the measurement of complexity in supply chains based on Shannon’s information entropy. Liu et al. (2008) applied an information-theoretic approach to study the complexity measurement of equipment management. Li et al. (2010) proposed a MTC (Maintenance Task Complexity) measurement model to quantify the complexity of maintenance task. Rao et al. (2006) developed the entropy-based measurement model to study the complexity of manufacturing systems and applied this model to investigate the effectiveness of schedules for manufacturing systems. Kong et al.(2011)evaluated the complexity of electrical equipment fault diagnosis using information entropy, and provided a new theoretical support for the complexity measurement of electrical equipment maintenance. However, little is known about the complexity measurement of auto fault diagnosis based on the information entropy. So, considering the characteristics of diversity and complexity in process of auto fault diagnosis, in this paper, we aim at applying an information-theoretic approach to study the complexity of auto fault diagnosis, and the quantification of complexity can provide a reference for drawing up maintenance plan and maintenance decision-making.
2. Evaluation indexes of complexity for auto fault diagnosis system

Auto fault diagnosis system is a multi-hierarchy, multi-factor complex structure system. The complexity of the system structure is reflected by the hierarchy structure relationship, factors correlation relationship and information transfer relationship, etc. In this paper, the complexity of auto fault diagnosis system is evaluated from three aspects: information transfer relationship, factors correlation relationship and hierarchies structure relationship(Kahneman et al.1973; Luiza et al. 2016). Evaluation indexes of complexity of auto fault diagnosis system are shown in table.1.

| Evaluation domain                          | Complexity parameter                                    | Information entropy |
|-------------------------------------------|---------------------------------------------------------|---------------------|
| Information transformation complexity      | Fault information feedback loop(X₁)                     | H(X₁)               |
| among fault factors (X)                   | Fault information communication network(X₂)             | H(X₂)               |
|                                           | Diagnose process path(X₃)                               | H(X₃)               |
| Correlation complexity among              | Fault factors coupling (Y₁)                             | H(Y₁)               |
| fault factors (Y)                         | Cohesion degree of fault factors (Y₂)                   | H(Y₂)               |
|                                           | Correlation degree of fault factors (Y₃)                 | H(Y₃)               |
| Structure complexity of factors hierarchies (Z) | Fault factor hierarchies (Z₁)                         | H(Z₁)               |
|                                           | Fault hierarchy span(Z₂)                               | H(Z₂)               |
|                                           | Correlation level (Z₃)                                 | H(Z₃)               |

3. Entropy as a measurement of complexity

3.1 Complexity measurement model

Definition 1: a complex vector space E contains three dimensions of X, Y and Z. The elements of E are defined as the complex vector \( E_{mi} = (e_{m,xi}, e_{m,yi}, e_{m,zi}) \), \( e_{m,xi} \), \( e_{m,yi} \) and \( e_{m,zi} \) is the information entropy in the subordinate 3 dimensions of \( E_{mi} \), thereinto: \( m \) represents the fault diagnosis hierarchy; \( i \) represents the judging index of fault factors complexity. \( E_{mi} \in H^3 \), here, the \( H^3 \) is the 3D space in H, a set of information entropy. \( \|E_{mi}\| \) represents the information of fault factors or the length of complexity vector. The distance between two vectors in \( H \) is defined as follow.

\[
\|E_{mi} - E_{m-1,i}\| = \|e_{m,xi} - e_{m-1,xi}, e_{m,yi} - e_{m-1,yi}, e_{m,zi} - e_{m-1,zi}\|
\]

(1)

The matrix form of the vector is

\[
B_m = \begin{bmatrix}
E_{m1} \\
E_{m2} \\
E_{m3}
\end{bmatrix} = \begin{bmatrix}
e_{m,xi} & e_{m,yi} & e_{m,zi} \\
e_{m,xi} & e_{m,yi} & e_{m,zi} \\
e_{m,xi} & e_{m,yi} & e_{m,zi}
\end{bmatrix}
\]

\(B_m\) is the complexity information entropy matrix for the fault diagnosis of the \( m \) hierarchy.

Definition 2: \( E_x, E_y \) and \( E_z \) are information transformation vector space, function vector space and structured vector space in \( H^3 \) respectively. In order to get a scale of judging fault diagnosis system complexity, a unified complexity space is established by combining three space complexity and then a new complexity scale is defined based on the new combining space, thus \( E_x \times E_y \times E_z \) is called 3D entropy space. Assuming \( \Phi \) is a reflection, defined by \( \Phi: E_x \times E_y \times E_z \rightarrow H \). \( \Phi \) is
three-dimensional linear form or three-dimensional covariance tensor in $H$, this tensor is defined as $T_i(H)$, also a vector space in $H$.

If a $3 \times 3$ matrix $B_e$ is defined that $B_e$ is formed by the three-dimensional complexity vector of the fault $B_i$ on $E_x$, $E_y$ and $E_z$, as a matrix row, then $\|\phi\| = \|\phi(E_{x, i}, E_{y, i}, E_{z, i})\| = \det(B_{m, i})$ is the form of the tensor, representing the unified entropy of $B_i$ hierarchy fault, and $\|\phi\|$ is the information content value (tensor complexity value). The set $\Phi_i$ is the set of multi linear space. Therefore, the distance from $E_{i-1}$ to $E_i$ is

$$d(E_{i-1} - E_i) = \sum_{m=1}^{m_i} \det(B_{m, i} - B_{m, i-1})$$

This is the total amount of information generated by fault $B_i$ after the $B_{i-1}$ hierarchy fault. So,

$$\|\beta\|_1 = \sum_{i=1}^{m} \Phi_i = \sum_{i=1}^{m} \|\Phi(E_{x, i}, E_{y, i}, E_{z, i})\| = \sum_{i=1}^{m} \det(B_{m, i})$$

and all the tensor entropy is,

$$\|\beta\|_E = \sum_{i=1}^{m} \|\Phi(E_{x, i}, E_{y, i}, E_{z, i}) - \Phi(E_{x, i-1}, E_{y, i-1}, E_{z, i-1})\| = \sum_{i=1}^{m} \det(B_{m, i} - B_{m, i-1})$$

This is the evaluation result of all information entropy for auto fault diagnosis system B.

3.2 Application example

In practical applications, the phenomena and causes of auto faults are needed to be classified, some common auto fault diagnosis process are selected, and the hierarchical structure of these auto fault diagnosis is examined in order to study the complexity measurement of auto fault diagnosis. Take the engine overheating for an example, the fault diagnosis hierarchy structure is determined according to the maintenance manual of auto manufacturers and related experience. The fault diagnosis hierarchy structure is shown in Fig 1.

![Fig.1 The hierarchy structure of engine overheating fault diagnosis](image)

According to complexity measurement of information transformation mentioned above, the information feedback loop is summarized by decision of Boolean function, then feedback to next fault hierarchy. The calculating process of the information transformation complexity among fault factors is shown in Table 2.
Table 2 The calculating process of the information transformation complexity among fault factors

| Fault hierarchy | Fault factors | Fault feedback $X_i$ | Communication type $X_y$ | Diagnose path $X_z$ | $H(X_i)$ | $e_i$ | $e_{di}$ | $H(X_y)$ | $H(X_z)$ |
|-----------------|--------------|---------------------|-------------------------|-------------------|----------|------|--------|----------|----------|
| 1st Hierarchy   | Engine overheating | 2 | 1 | 1 | | | | | | |
| 2nd Hierarchy   | engine body overheating | 4 | Straight line | 1 | 1 | 0.0899 | 0.0899 | 0.0504 | 0.0276 |
| 3rd Hierarchy   | Ignition timing faults | 0 | Straight line | 1 | 2 | | | | |
|                 | poor lubrication | 2 | Straight line | 1 | 2 | | | | |
|                 | engine overload | 1 | Straight line | 1 | 2 | | | | |
|                 | Combustion chamber caused by carbon | 0 | Straight line | 1 | 2 | | | | |
|                 | Thermostat faults | 3 | Straight line | 1 | 2 | | | | |
|                 | Radiator problems | 3 | Straight line | 1 | 2 | | | | |
|                 | Cooling water problems | 3 | Straight line | 1 | 2 | | | | |
| Total of $M_2$  | 11 | 7 | 14 | 0.1553 | 0.4013 | 0.1490 | 0.1428 |
| 4th Hierarchy   | Low oil pressure | 0 | Straight line | 1 | 3 | | | | |
|                 | lack of lubricating oil | 0 | Straight line | 1 | 3 | | | | |
|                 | Motor overload | 0 | Straight line | 1 | 3 | | | | |
|                 | Thermostat Damage | 0 | Straight line | 1 | 3 | | | | |
|                 | Control valve not working | 0 | Straight line | 1 | 3 | | | | |
|                 | Radiator blocking | 0 | Straight line | 1 | 3 | | | | |
|                 | Fan poor working | 2 | Straight line | 1 | 3 | | | | |
|                 | Having water scale | 0 | Straight line | 1 | 3 | | | | |
|                 | Dirty | 0 | Straight line | 1 | 3 | | | | |
|                 | Insufficiency | 2 | Straight line | 1 | 3 | | | | |
|                 | Water pump poor working | 0 | Straight line | 1 | 3 | | | | |
| Total of $M_4$  | 4 | 11 | 33 | 0.1297 | 0.5310 | 0.1597 | 0.1505 |
| 5th Hierarchy   | Belt loose | 0 | 2 linear type | 2 | 4 | | | | |
|                 | Clutch fault | 0 | 2 linear type | 2 | 4 | | | | |
|                 | Water tank leaking | 0 | 2 linear type | 2 | 4 | | | | |
|                 | Cooling water evaporation | 0 | 2 linear type | 2 | 4 | | | | |
| Total of $M_5$  | 0 | 8 | 16 | 0.0000 | 0.5310 | 0.1187 | 0.1492 |
| Total           | 24 | 29 | 66 | | | | | |

Note: $X_n = H(X_{n-1}) + X_{di(n-1)}$ means fault feedback is a cumulative process.

The calculation principle of other complexity measurement is the same as above, the calculation process is omitted, and results are shown in table 3.

Table 3 Results of total information entropy

| Structure hierarchy | $e_{x1}$ | $e_{y1}$ | $e_{x1}$ | $e_{y1}$ | $e_{x2}$ | $e_{y2}$ | $e_{x3}$ | $e_{y3}$ | $e_{x1}$ |
|---------------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| 1                   | 0.0899   | 0.0504   | 0.0276   | 0        | 0.0610   | 0.2000   | 0.1398   | 0.0899   | 0.1000   |
| 2                   | 0.2460   | 0.0801   | 0.0460   | 0.0465   | 0.1180   | 0.3695   | 0.2796   | 0.1561   | 0.3197   |
| 3                   | 0.4013   | 0.1490   | 0.1428   | 0.1901   | 0.1596   | 0.4571   | 0.4194   | 0.1553   | 0.4627   |
According to the corresponding calculation results, the three-dimensional vector space determinant of each fault structure hierarchy can be obtained by using complexity calculation formula mentioned above, so only the 1st hierarchy vector space determinant calculation is shown as follow. Therefore, the result of complexity measurement for engine overheating fault diagnosis is shown in table 4.

Table 4 The unified complexity entropy value of the engine overheating fault

| Structure hierarchy | $\text{det}(B_{m1})$ | $\text{det}(B_{m1-m(i-1)})$ |
|---------------------|----------------------|-----------------------------|
| m1                  | 0.0001               | ----                        |
| m2                  | 0.0010               | 0.0026                      |
| m3                  | 0.0113               | 0.0002                      |
| m4                  | 0.0158               | 0.0003                      |
| m5                  | 0.0011               | 0.0007                      |
| Total               | 0.0293               | 0.0014                      |

From table 4, $\|B\|_F = 0.0293$, $\|B\|_E = 0.0014$, this is a unified complexity entropy value of engine overheating fault diagnosis, the complexity entropy value of other auto fault diagnosis can also be obtained in the same way. Therefore, the complexity of various auto fault diagnosis can be compared with and evaluated according to these values, and the higher the value is, the more difficult the fault diagnosis becomes, which means that the fault diagnosis need longer hours, and the complexity of auto fault diagnosis is greater. As a result, these values can provide a reference for determining the corresponding diagnosis working hours of various auto faults, and can further improve the calculation of modern maintenance working hours.

3.3 Validation tests

In order to verify the validity of the complexity measurement method of auto fault diagnosis, the complexity metrics of 10 common auto fault diagnosis are calculated respectively using the above methods. According to data collected from the maintenance workers and maintenance records, ten auto fault diagnosis in table 4 are arrayed according to their fault diagnosis time sequence, and the horizontal axis represents auto fault diagnosis order (horizontal axis equidistant intervals does not represent the specific duration), the vertical axis represents complexity metrics values (See Fig 2).

Fig 2 Relationship between the entropy values and the complexity of auto fault diagnosis
From Fig2, it can be seen that both the entropy values and the complexity of auto fault diagnosis have a little linear relationship, that is to say, the more complex the auto fault diagnosis is, the bigger the corresponding information entropy values are. Therefore, the information entropy value can reflect the complexity of the auto fault diagnosis and maintenance workers’ workload. The complexity of auto fault diagnosis can be quantitatively evaluated using information entropy, and results can provide guidance for drawing up a reasonable and effective maintenance plan. It shows that the complexity measurement method in this paper can effectively reflect the degree of complexity real auto fault diagnosis, which can contribute to auto fault diagnosis strategy optimization and maintenance work arrangement.

4. Conclusion
Since modern auto applies the module design, the equipment repair and maintenance are not carried out at the hierarchy of small components, but at the higher hierarchy of the larger unit (composite parts, module). Sometimes, the entire assembly unit has to be replaced to complete the repair tasks. Therefore, the effect of fault diagnosis complexity must be taken into consideration in modern car maintenance workload.

In this paper, the author applies the information entropy to evaluate the complexity of auto fault diagnosis. It shows that entropy values calculated by the model given in this paper are consistent with the actual situation. However, the application of this method requires the government and industry management organizations to develop the auto unified classification. The unified classification of complexity of auto fault diagnosis will contribute to the auto fault diagnosis strategy optimization and maintenance work arrangement. Matched with the corresponding calculation software, this method will be conducive not only to the reasonable pricing and development of the auto repair and maintenance industry, but also to the protection of the interests of consumers.

Reference
[1] Filiz, I. 2010. “An entropy-based approach for measuring complexity in supply chains,” International Journal of Production Research, 48:12, 3681-3696, DOI:10.1080/00207540902810593
[2] Liu, J.S, H.N.Tu and H. Zhang et al. 2008. “Research on measurement entropy-based of equipment management complexity and its application in production planning”//Lecture Notes in Computer Science, 5315:604-611.
[3] Li, S.T., Q. Cai, B. Zhu. 2010. “Research on the quantification of emergency maintenance task complexity for marine power facility.” Journal of Wuhan university of technology(Transportation Science& Engineering) , Vol.34 No.4:813-817
[4] Rao, Y.Q., E. Janet. 2006. “Entropy- based of manufacturing system complexity and its application in scheduling.” Chinese Journal of mechanical engineering, Vol.4, No.7:8-13
[5] Kong, F.S., Y.F. Wu, C.Li. 2011. “Assessment of fault diagnosis complexity about electrical fault diagnosis of equipment based on information entropy.” Journal of Jilin University(Engineering and Technology Edition), vol 41 (3) :697-701.
[6] Kahneman, T. Daniel, Versky, Amos. 1973. On the psychology of prediction. Psychological Review, Vol 80(4): 237-251. http://dx.doi.org/10.1037/h0034747
[7] Luiza, Z., E.H. Aghezzaf, V. Limère.2016. “Workload balancing and manufacturing complexity levelling in mixed-model assembly lines,” International Journal of Production Research, DOI: 10.1080/00207543.2016.1213452
[8] Song, H.L. 2010. “Researches of complexity measurement theory, method and case study for enterprise complex systems based on the check space and entropy metrics.” Beijing:Economic management press.