Information Transfer-Based Root Cause Tracing for Information Quality Problems of Complex Electromechanical Systems in Process Industry

(Completed Research Paper)

Rongxi Wang, Jianmin Gao, Zhiyong Gao, Yanjie Liang, Hongquan Jiang
State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an, China

rongxiwang@163.com

Abstract: Functional equipment items of complex electromechanical systems in process industry are coupled. The problems of information quality of any one of the internal components may propagate and pervade to the other equipment due to coupling. In order to fast and effectively identify the root cause to avoid the further propagation of these problems, an information transfer based novel data-driven method for root cause tracing of information quality problems was proposed considering the experience and qualitative analysis of conventional methods. An actual application is used to verify the effectiveness of the proposed method. A unique root cause is obtained regardless of the choice of the initial variable. The proposed method can be flexibly and effectively used in root cause tracing of complex electromechanical systems in process industry, and formulate the foundation of system vulnerability analysis and condition prediction, as well as other engineering applications.

Keywords: coupling analysis, root cause tracing, information quality, complex electromechanical systems, process industry

1 INTRODUCTION

Production systems in modern process industry are typical distributed and complex electromechanical systems which composed with mass electromechanical equipment and connected with control, power and flow signals (Wang et al. 2016). Different functional equipment items in such system are coupled and will be working in collaboration with each other. The problems of information quality of any one of the internal components may propagate and pervade to the other equipment due to coupling, and it may also lead to serious accident. Thus, information quality control has been one of the most important works for complex electromechanical systems in process industry, which has attracted much attention from researchers (dos Santos and Bastos 2017; Guan and Chen 2017; Liu et al. 2017; Myrelid 2017; Senaratne et al. 2017; Wijaya et al. 2017).

There are three major issues in information quality control: the evaluation criterion of information quality, the identification and root cause tracing of information quality problems. However, the emphasis of these issues are distinguishing. The first and second issues focused on the detection of information quality problems, and the last issue highlights on tracing the root cause of these existing problems. The tracing results formulated the foundation of information quality control, and it can help the enterprise to take immediate and corrective measures to prevent any major accident due to information coupling.

In recent years, several methods for fault root cause tracing have been proposed. Among these methods, due to the abilities of characterizing the relationships between different system conditions and carrying amount of information, the signed graph method which has first appeared in a mathematical paper in 1953 (Harary 1953; Hussein 2014) is suited for identifying the root node of abnormal condition and it has been rediscovered many times because it come up naturally in many unrelated areas (Fleiner and Wiener 2016; Liu et al. 2016; Macajova and Skoviera 2016; Meng et al. 2015). M.Iri et al. (Iri et al. 1979; Iri et al. 1980) introduced the definition of signed directed graph (SDG) in the chemical field, and the depth-first search strategy was applied to implement the basic search for incomplete samples. Multi-level SDG was proposed to characterize the causality over time and to analyze the dynamic process by T.Umeda (Umeda et al. 1980). The drawbacks of complexity and high occupancy rate of computation source limit its application. Unlike the previous methods that have limitations to trace the root cause based on SDG, M.A.Kramer et al. (Kramer and Palowitch 1987) proposed the rules for expert system based on forward analysis of SDG, and he applied these rules to online fault diagnosis of the plant. However, it cannot avoid the problem of explosion of rules. C.C.Chang et al. (Chang and Yu 1990) presented a method to simplify the SDG model by deleting the non-potential root nodes based on the system conditions and fault propagation pathway. Huang et al. (Huang et al. 2013a; Huang et al. 2013b) proposed a fault root cause tracing method integrated with SDG and fault graph based on the relationships between multiple-factors. There is no doubt that the derivation of knowledge base rule based on SDG is a big process in the field of fault root cause tracing. However, these rules come from the superficial expert knowledge, and they are difficult to reveal the deep disciplines of faults and hard to achieve the requirements of completeness.

The commonly used method for detailed event analysis and risk assessment is failure mode and effects (FMEA) (Arvanitoyannis and Varzakas 2008; Liu et al. 2013b). Since its appearance, FMEA has become a powerful tool extensively used for safety and reliability analysis of systems, products, processes, and services in many industries (Chang et al. 2012; Kutlu and Ekmekçioğlu 2012; Liu et al. 2014; Liu et al. 2013a; Yang et al. 2016). When also aimed at the prioritization of potential failure modes, FMEA is referred to as failure mode, effects and criticality analysis (FMECA) (Certa et al. 2017). According to ISO 9000 and ISO/TS 16949 standards, the best classified analysis for prevention of failures during the production process is process failure mode and effect analysis (PFMEA) (Banduka et al. 2016). As a top-down deductive method that aims to compute the probability of occurrence of the top event as a function of the probabilities of occurrence of basic events, fault tree analysis (FTA) is the easiest and most used technique in dependability assessment (Talebberrouane et al. 2016). In the literature, hybridized fault trees have been successfully used in some work to assess reliability, availability, and safety of complex systems (Ayav and Sozer 2016; Chen et al. 2017; Talebberrouane et al. 2016; Volkanovski et al. 2009). FTA and FMEA have advantages of describing the fault root cause and fault propagation pathway. However, because of the complexity of electromechanical systems in the process industry, it is difficult to model the physical systems based on the fault disciplines, which limits their application in actual scenes.

Thus, new approaches for root cause tracing are welcome and needed. As outlined above, production systems in process industry are typical complex electromechanical systems, and discrete electromechanical equipment in them are coupled and high corrective. The power, energy and control information are transformed between different functional components due to coupling. Actually, the transformation of information in complex electromechanical systems is one of the descriptions of coupling relationships, and the local and abnormal information quality can be represented by the exceptional relationships of information transfer. Based on this thinking, the propagation pathway and root causes of information quality problems can be obtained without any prior knowledge of system if the information models of complex electromechanical systems are known. In our work, an information transfer-based novel method was proposed to trace the root causes of the information quality problems of complex electromechanical systems in process industry. Symbolic transfer entropy method was improved to obtain the requested information for information modeling. Assuming that information quality problems have been detected, then, the tracing process was performed from an arbitrary initial point along with the bidirectional information transfer in information model by comparing the changes of information transfer between different variables.

The rest of the paper is organized as follows: Section 2 gives a brief presentation of transfer entropy and symbolic transfer entropy. An improved symbolic transfer entropy method is introduced in Section 3. The detail processing procedures of information transfer based root cause tracing of information quality problems are described in Section 4. Section 5 investigates the effectiveness of the proposed method by empirical research and analysis, while the paper is concluded in Section 6.

2 PRELIMINARIES

Based on the analysis introduced above, the information modeling is the foundation and core of root cause tracing of information quality problems. The essence of an information model for a specific complex electromechanical system is a directed-weighted graphic. Information nodes, directed edges and their weights are the basic elements in information model. In order to construct an information model for a complex electromechanical system, the basic elements in model should be obtained firstly. The monitoring variables and the relationships between them can be views as information nodes and directed edges respectively. However, there still are three major issues needed to be addressed: Whether there is a relationship between two information nodes? What are the directions and weights of these relationships? Among which, the determination of the direction of these relationships is the most important one. In information theory, joint entropy, mutual information, condition entropy and other measures can be applied to characterize the relationships between different random variables, but these quantifications are non-directed or directions generated by the different variables' individual entropies and not due to information flow. In 2000, an asymmetric information transfer measure named as transfer entropy (TE) was proposed to satisfy the requirements of reflecting original dynamic and direction of coupling (Hahs and Pethel 2013; Schreiber 2000).

Assuming there are two random processes I and J, the transfer entropy from J to I can be interpreted as the addition information provided by J for the prediction of $I(t+1)$ apart from the information provided by I itself in the past. A mathematical expression of transfer entropy is:

$$\begin{aligned} T_{J\to I} &= \sum p(i_{t+1}, i_t^{(k)}, j_t^{(l)}) \log_2 p(i_{t+1} \mid i_t^{(k)}, j_t^{(l)}) - \sum p(i_{t+1}, i_t^{(k)}) \log_2 p(i_{t+1} \mid i_t^{(k)}) \\ &= \sum p(i_{t+1}, i_t^{(k)}, j_t^{(l)}) \log_2 \frac{p(i_{t+1} \mid i_t^{(k)}, j_t^{(l)})}{p(i_{t+1} \mid i_t^{(k)})} \end{aligned} \tag{1}$$

Here, i_t represents the state of I at time t, $i_t^{(k)}$ is a k-dimensional vector representing the k most recent states of I before i_{t+1}, while $j_t^{(l)}$ and l are the corresponding quantities for J. $T_{I\to J}$ can be obtained by exchanging the variables in Equation (1).

As a measure for information flow, transfer entropy supplies an effective method to obtain the required information for modeling. However, transfer entropy method is sensitive to noise (Wu et al. 2013) which is inescapable in monitoring data of process industry. In order to overcome this drawback, a symbolization-based extension of TE named symbolic transfer entropy (STE) was proposed by Staniek in 2008 (Staniek and Lehnertz 2008) to improve the anti-noise ability of traditional transfer entropy method.

In symbolic transfer entropy method, phase space reconstruction based on embedding theorems (Takens 1981; Wang et al. 2016) is the foundation for symbolization. Let τ and m denote the lag and embedding dimension in phase space reconstruction. For a given but arbitrary phase point, there are m amplitude values $s_t = (i_t, i_{t+\tau}, \cdots, i_{t+(m-1)\tau})$. Then, values in this point are arranged in ascending order $i_{t+\tau(k_{t1}-1)} \le i_{t+\tau(k_{t2}-1)} \le \cdots \le i_{t+\tau(k_{tm}-1)}$ with rank k_{tl} and $l \in \{1, \cdots, m\}$. Equal amplitude values are arranged by their time index, i.e., such that if $i_{t+\tau(k_{t1}-1)} = i_{t+\tau(k_{t2}-1)}$ then $k_{t1} < k_{t2}$. This ensures that every s_t is uniquely mapped onto one of the $m!$ possible permutations, and a symbolic sequence permutation is defined as:

$$\hat{i}_t = (k_{t1}, k_{t2}, \cdots, k_{tm}) \tag{2}$$

The relative frequencies of symbols provide an estimator of joint and conditional probabilities of the sequences of permutation indices. Given two symbol sequences $\hat{i}_t$ and $\hat{j}_t$, based on Equation (2), symbolic transfer entropy is defined as in Equation (3).

$$\hat{T}_{J\to I} = \sum p(\hat{i}_{t+1}, \hat{i}_t, \hat{j}_t) \log_2 \frac{p(\hat{i}_{t+1} \mid \hat{i}_t, \hat{j}_t)}{p(\hat{i}_{t+1} \mid \hat{i}_t)} \tag{3}$$

In (3), the sum runs over all symbols. $\hat{T}_{J\to I}$ is defined in an analogous manner. $\hat{T}_{J\to I}$ is positive and explicitly asymmetric for the exchange between I and J since it measures the flow of information from J to I and not vice versa. The difference $\hat{T}_{J\to I} - \hat{T}_{I\to J}$ indicates the flow of information (Dickten and Lehnertz 2014; Staniek and Lehnertz 2008).

3 IMPROVED SYMBOLIC TRANSFER ENTROPY BASED INFORMATION MODELING

The symbolic transfer entropy approach improves the noise robustness of traditional transfer entropy methods. However, symbolization retains the coarse information of the basic dynamic characters but loses the detailed information. This in turn can seriously influence computation results because of the non-stationarity of the observation sequence (Daw et al. 2003). In this section, an adaptive symbolization technology based on improved symbolic transfer entropy (ISTE) is proposed to increase resistance to noise and reduce information loss.

3.1 Determination of the Optimal Number of Symbols

Logically, the larger of the number of symbol, the less information loss and the weaker of the anti-noise ability. In order to clearly demonstrate the influence of symbolization due to different symbol number, a sequence with 300 samples following a uniform distribution was generated using MATLAB, and it was marked as x_k, where $1 \le k \le N$. We also added 15 dB of white Gaussian noise to the sequence. The signal-to-noise ratio (SNR) and information entropy were applied as measures of noise resistance and information loss of the symbolized sequence. Assuming there are n symbols, the original sequence is symbolized using the following rule.

$$x_k = i-1, min + \frac{(i-1)(max-min)}{n} \le x_k < min + \frac{i(max-min)}{n}, i \in \{1, \cdots, n\} \tag{4}$$

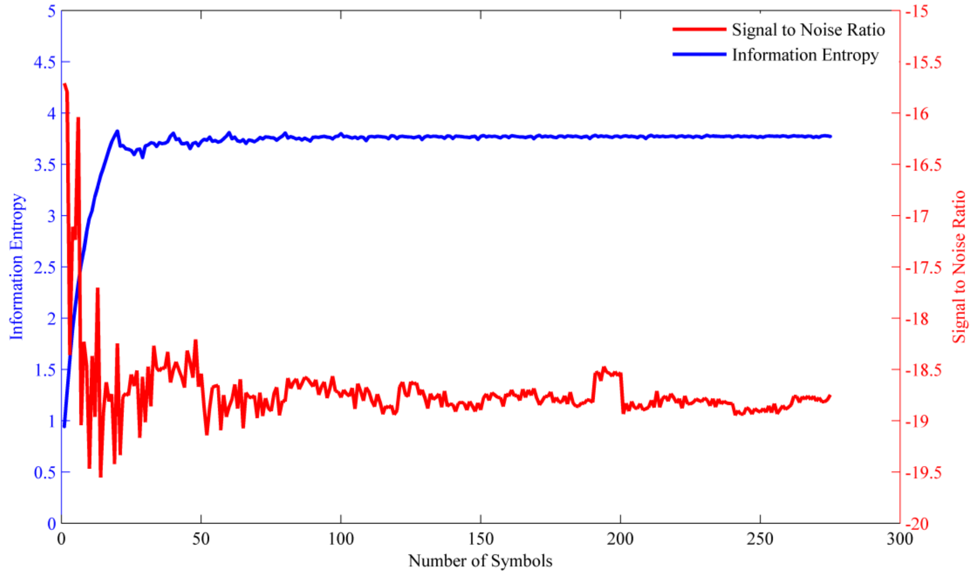

Figure 1 Relationships between the number of symbols, information loss and SNR

As shown in Figure 1, with an increase of symbols, the information entropy of the symbolized sequence increases sharply in the beginning, and gradually stabilizes. The SNR shows a reverse behavior, as it decreases sharply. In other words, with an increase of symbols, the noise suppression property of symbolization becomes weaker and becomes asymptotically stable. The information loss due to symbolization also stabilizes asymptotically. Thus, it is necessary to determine an optimal number of symbols to balance noise suppression and information loss in practical applications. In our work, the optimal number of symbols used was that which achieved equilibrium between noise robustness and information loss.

3.2 Improved Transfer Entropy Calculation

The most important procedure in STE is symbolization. The purpose of STE is to transmit the original continuous sequence into a discrete symbol sequence with specific intervals, and then transfer entropy is computed for the new sequence.

Following the work of Staniek on STE (Staniek and Lehnertz 2008), the first step in symbolization is phase space reconstruction based on the embedding theorem. In 1981, Takens proposed an embedding theorem to obtain more systematic information from single time series. In his theory, $\{x(n)\} = \{x_1, x_2, x_3, ..., x_{N-1}, x_N\}$ is defined as a noise-free scalar time series with a d dimensional chaotic attractor. Embedding theorem shows that one always can reconstruct a m dimensional embedding phase space if the relationship $m \geq 2d+1$ holds (Takens 1981).

Considering the determined optimal number of symbols, each vector in reconstructed phase space can be mapped to a unique symbolic sequence. Assuming there are n symbols and the number of embedding dimensions is m, then, there are n^m possible permutations. In order to gather statistics for each permutation, all the symbolized vectors are further transformed before transfer entropy computations. In our work, this transformation was applied using binary encoding and decimal decoding.

Binary Encoding - A $b = \lceil log_2 n \rceil$-bit binary number was used to encode n symbols, where $\lceil \bullet \rceil$ indicates rounding up to an integer. For example, a 3-bit binary number can represent 8 symbols (1=001, 2=010, 3=011, and so on) and a binary number having at least 4 bits is needed to represent 15 symbols.

Decimal Decoding - It is difficult to gather statistics of the symbolized sequence and the binary encoding sequence. Because of the uniqueness of each set of encoding, it can be interpreted as a unique decimal number, for which it is easy to determine the probability density function and probability. Figure 2 shows an example of this transformation for a 5-dimensional vector point with 8 symbols.

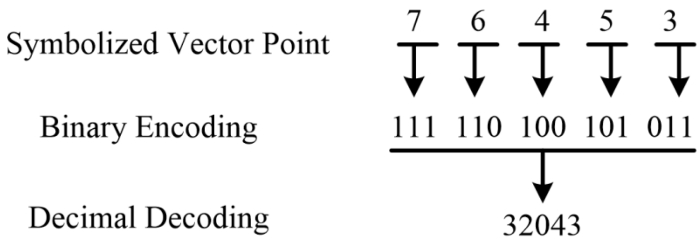

Figure 2 Binary encoding and decimal decoding for a symbolized vector point

Based on the previous analysis, the symbolic transfer entropy can be calculated using Equation (3). To summarize our approach, the improved symbolic transfer entropy proposed is shown in Figure 3.

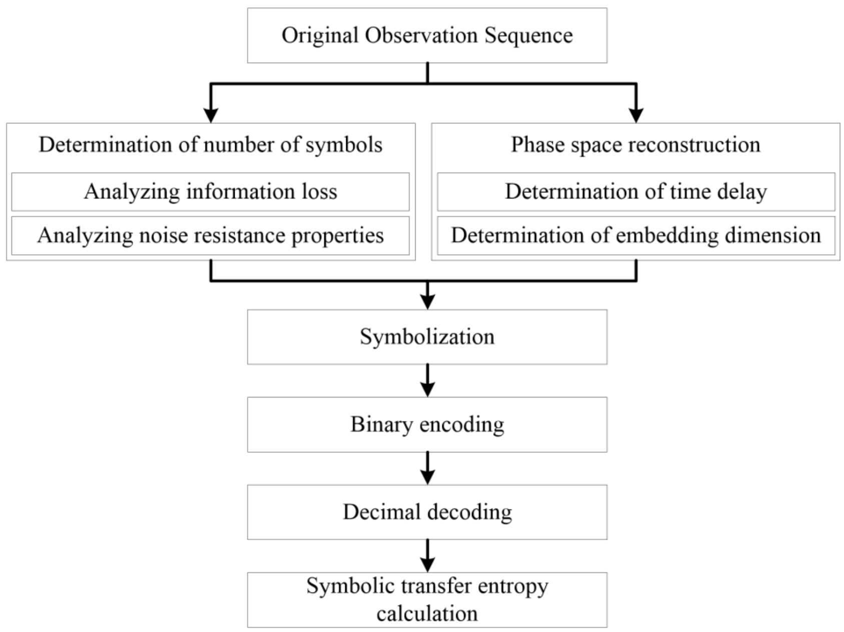

Figure 3 Steps of proposed symbolic transfer entropy calculation

In theory, the improvements of symbolic transfer entropy in our work comprehensively consider the noise robustness and information loss of symbolization. The proposed further processing of symbolized vector point, binary encoding and decimal decoding, simplify the determination of probabilities, and the optimal time delay guarantees the detection of maximal flow of information.

3.3 Coupling Analysis Based on Improved Symbolic Transfer Entropy

The analysis of coupling relationships between monitoring variables plays an important role in root cause identification for information quality problems. Three major issues in the investigation of couplings among the monitoring variables should be addressed: the detection of the couplings, their direction, and their quantification (Papana et al. 2013; Papana et al. 2016). It is no doubt the couplings between different monitoring variables are existing for complex electromechanical systems in process industry. The remaining problems for coupling analysis of complex electromechanical systems are the determinations of directions and their quantifications. Actually, the information transfer between monitoring variables is often asymmetric. That is to say the information transfer is directed, and the amount of information transferred between different variables reflects its quantification. Then, the essence of information transfer computation is the coupling analysis based on information transfer. The results for transfer entropy computation can be utilized as the coupling relationships between different monitoring variables.

4 COUPLING ANALYSIS BASED ROOT CAUSE IDENTIFICATION FOR INFORMATION QUALITY PROBLEMS

In this section, the detail procedures for root cause tracing for information quality problems will be introduced. Firstly, the information model for complex electromechanical system was established based on coupling analysis. Secondly, the information transfer between different monitoring variables is applied as the criterion of root cause tracing for information quality problems.

4.1 Information Modeling for Complex Electromechanical Systems

Information modeling is the foundation of root cause identification based on information theory. The core of information modeling is the analysis of coupling relationships between different information nodes.

Information modeling integrates the decentralized information nodes to an organic whole. The detection of information flow, which is the connection rule for nodes, is the necessary premise in our general framework. However, three issues should be solved before modeling: How to determine whether there is a flow of information or not? What is the direction and what is the strength of this flow of information? Based on the above explanations and introduction, the results of interaction analysis contain the answers to these three questions. Therefore, in our general framework for information modeling, the improved symbolic transfer entropy proposed in our work is used to analyze the interaction relationships between information modes, and the analysis results are used to connect information nodes with directed edges.

Given a specific complex electromechanical system with n sensors, the determination of information relationships and modeling can be realized according to the following procedures.

1) Compute the transfer entropy of each pair of information nodes based on the improved symbolic transfer entropy proposed in Section 3.2 to obtain the weight matrix A_W.
2) Determine the interaction relationship of each pair of nodes. According to A_W, if $A_{Wi,j}, A_{Wj,i} \in A_W$ and $A_{Wi,j} - A_{Wj,i} > 0$, then the direction is $i \to j$, which indicates a positive interaction from i to j. Otherwise, there is a negative interaction. Symbols + and − are used to express a positive and negative interaction respectively. After this procedure, the contents of A_R in the information model can be obtained.
3) Starting from an arbitrary information node i and searching for nodes j who have positive interaction relationship with $i \to j$ in all nodes excepted itself. All the searched nodes were composed to a set S.
4) We connect node i to each node of S with a directed edge, and the direction is $i \to j$. The weight of the directed edge is marked on it.
5) Iterate all the elements of S and repeat steps 3 - 5.

The inputs of this step are the information nodes and monitoring data, and the outputs are the flows of information between the nodes and their extensions, such as A_W and A_R. Then, a weighted and directed information model is constructed, which is unique and independent of the starting node.

4.2 Root Cause Identification for Information Quality Problems

In this section, the information transfer relationships between each pair of monitoring variables are determined based on the improved symbolic transfer entropy proposed in section 3, including the direction and strength. As the phase transition sustains a specific period, the strengths of information transfer for each pair of variables are specific values, and the direction is certain.

The information transfer relationships during phase transition and the normal system condition are the inputs for root cause tracing of information quality problems. Figure 7 shows an information model of a specific electromechanical system and this model is constructed using the methods proposed in our work.

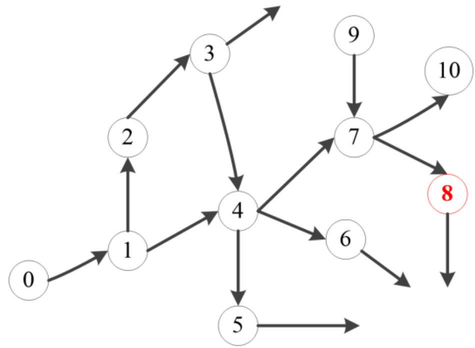

Figure 4 A simple example of information model for a specific electromechanical system

In this model, there are 11 information nodes. Here, it is known that a fault has occurred. However, the fault root cause is unknown. The node with number 8 is selected as the initial point to determine the fault propagation pathway. The detailed procedure is described below.

1) Determine the source node for node eight along with the inverse direction of information transfer and mark it as S . In this example, the result is $S=\{7\}$.

2) Compare the current information transfer relationships between each pair of nodes. If the current strength lies in the weight interval of normal conditions, then this source node is deleted from S . Otherwise, the source node is retained. The new node set is labeled as $S^{'}$.

3) For each element in $S^{'}$, repeat steps 1) and 2) until there is no new source node or $S^{'}$ is empty.

Here, a simple example is given to help readers understand this procedure. Assuming the changes of S and $S^{'}$ in previous iteration is $\{8\}\to\{7\}\to\{4,9,10\}\to\{4\}\to\{1,3\}\to\varnothing$, then, the propagation pathway is $8\to7\to4$, and the root cause for this fault is node 4.

These steps are used to perform all the procedures in our framework for fault root cause tracing of complex electromechanical systems in the process industry. The obtained result can be used for maintenance and safety precautions.

5 EMPIRICAL RESEARCH AND ANALYSIS

In this section, an actual application of root cause tracing in a chemical plant is investigated to explain the detailed treatment processes of the framework proposed in our work. Additionally, the effectiveness of the proposed method is analyzed.

5.1 Selection and Analysis of Monitoring Data

Compressor groups are standard production equipment in chemical plants. Thousands of sensors are deployed to monitor the attributes of them, such as pressure, temperature, flow, vibration, rotating speed and so on. An example of this equipment's connections is shown in Figure 5, which shows the configuration required to realize an air compression function.

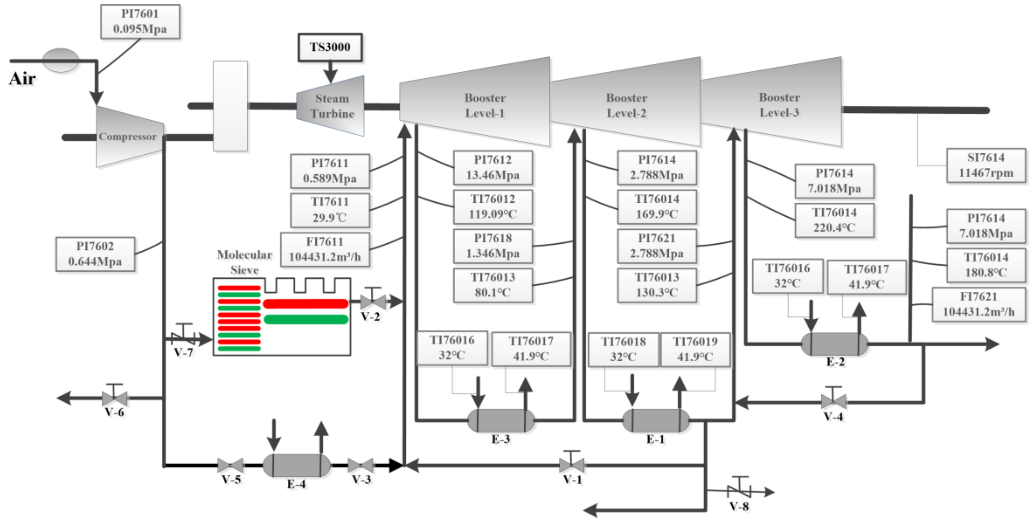

Figure 4 Equipment connections of air compressor group

For the purpose of this experiment, 1189 monitoring variables were allocated to the compressor group. In order to introduce the modeling and tracing procedures clearly and to demonstrate the effectiveness of the proposed methods, ten variables applicable to the steam turbine that have clear interactions were selected as the analysis objects. These variables and their descriptions are listed in Table 1, along with the number used to replace the variable identification code in the following for convenience. The monitoring data in the example was taken from a DCS of a chemical plant. Continuous sampling with a sampling rate of 1/60 Hz for 120h was used. The total number of samples was 7200 and the samples were labeled from 1 to 7200. It can be observed that a tripping fault has occurred at 5:50AM of the fourth day because of the turbine rushing. The average period of each variable of the normal data was calculated based on the Hilbert-Huang transform (Huang et al. 1998; Wang et al. 2015), which takes into consideration the differences in statistical properties for different sampling lengths. The results are listed in Table 1.

Table 1 Variables selected for information modeling example

NO.	Variable	Description	Average Period
1	RXI7650	shaft vibration of steam turbine	975
2	RZI7650	shaft displacement of steam turbine	252
3	AFI7650	initial steam flow of steam turbine	178
4	AFI7651	pumping flow of steam turbine	178
5	API7654	discharge pressure of steam turbine	314
6	API7655	pumping pressure of steam turbine	110
7	API7658	inlet pressure of steam turbine	90
8	ATI7654	inlet temperature of steam turbine	366
9	ATI7655	pumping temperature of steam turbine	302
10	RKI7650	key phase of steam turbine	68

5.2 Information Modeling Based on Coupling Analysis

According to the framework and methods of root cause tracing proposed in our work, the information model of the electromechanical system should be established first. In particular, the phase space reconstruction of each variable, determining the public number of symbols and the computation of transfer entropy based on our improved method for each pair of variables are the main elements of practical applications of information modeling.

Information loss and noise robustness were considered to determine the optimal number of symbols, and then, the procedures shown in Figure 3 and 4 were applied for information modeling. Here, the directions of interactions and transfer entropies between each pair of variables are listed in Tables 2 and 3, respectively. It should be noted that every symbol or value of Table 2 and Table 3 indicates the interaction direction or information transfer from source variables (rows) to destination variables (columns). The symbol "--" in Tables 2 and 3 indicates that no flow of information was detected between the two variables.

Table 2 Directions of interactions between each pair of variables

NO.	1	2	3	4	5	6	7	8	9	10
1	--	-	-	-	-	+	+	-	-	+
2	+	--	-	-	-	+	-	-	-	-
3	+	+	--	+	+	+	-	+	+	-
4	+	+	-	--	+	+	-	+	+	+
5	+	+	-	-	--	-	+	-	+	-
6	-	-	-	-	+	--	-	+	+	-
7	-	+	+	+	-	+	--	-	+	-
8	+	+	-	-	+	-	+	--	+	+
9	+	+	-	-	-	-	-	-	--	-
10	-	+	+	-	+	+	+	-	+	--

Table 3 Transfer entropies between each pair of variables (10^{-2} bits)

NO.	1	2	3	4	5	6	7	8	9	10
1	--	--	--	--	--	3.46	**0.06**	--	--	0.79
2	3.17	--	--	--	--	1.50	--	--	--	--
3	2.26	4.55	--	3.52	5.00	**0.02**	--	3.27	6.84	--
4	3.93	1.98	--	--	1.35	1.34	--	0.83	4.61	0.9
5	2.53	**0.40**	--	-	--	--	0.98	--	**0.13**	--
6	--	--	--	-	1.09	--	--	**0.68**	**0.08**	--
7	--	5.82	4.71	3.31	--	0.72	--	--	2.29	--
8	1.38	3.26	--	--	4.50	--	0.73	--	8.34	0.79
9	3.79	**0.40**	--	--	--	--	--	--	--	--
10	--	**0.60**	4.06	--	**0.63**	4.10	2.50	--	2.76	--

Based on the results listed in Table 2 and Table 3, the distributed variables were connected, and all the information necessary for creating the model was obtained. In Table 3, some values of transfer entropies were clearly smaller than others, which indicate that the flow of information between these variables was also smaller. In order to simplify the information model, a transfer entropy threshold value of 0.0070 was applied, resulting in some directed edges being deleted from the model. These deleted interactions are marked with red and bold fonts in Table 2 and Table 3, respectively. After obtaining the weights of relationships, the information model was constructed.

5.3 Root Cause Tracing for Information Quality Problems

Based on the information model obtained above, the root cause tracing was introduced in this section. The model reflects the information transfer relationships between each pair of variables within the actual electromechanical system. Moreover, the direction and strength are also indicated in the model, and the pathways of fault propagation are the same as the information transfer. Then, tracing along with the opposite directions of fault propagation or information transfer is an effective and direct method to determine the root cause for fault.

Based on the procedure introduced in 4.4.3, the points which are connected with 6 compose a set marked as $S = \{1, 2, 4, 7, 10\}$, then the information transfer between 6 and each element of this set was calculated. The obtained results were compared with the weights of the directed edges based on the direction. The strengths of information transfer for (1, 6), (2, 6), (4, 6), (7, 6) and (10, 6) are 0.005 bits, 0.0206 bits, 0.0306 bits, 0.0449 bits and 0.0358 bits respectively. All the strengths are lying in the range of weights for normal conditions with the exception of 0.005 bits for (1, 6). Next, a new variable set, $S = \{2, 3, 4, 5, 8, 9\}$, was generated by searching the connected relationships with the end node of 1. Then, calculate the strengths of information transfer and compare them with the normal weights until S is an empty set or there is no new set. For instance, the procedure for updating S are presented as following:

$$\begin{aligned} S &= \{1, 2, 7, 10\} \\ &\to \{1\} \to \{2, 3, 4, 5, 8, 9\} \\ &\to \{2\} \to \{3, 4, 7, 8\} \\ &\to \varnothing \end{aligned}.$$

Thus, it can be determined from this example that the variable with number 2, i.e., RZI7650 (shaft displacement of steam turbine) is the root cause of this fault based on the information transfer.

5.4 Results and Discussions

The root cause of a specific information quality problem is certain. In other words, the tracing results should be consistent regardless of the initial point. In order to verify the tracing results and procedure mentioned in 5.3.2, variable AFI7651 with number 4 was considered to trace the root cause again, and the tracing processes are described as follows:

$$\begin{aligned} S &= \{7\} \to \{5, 8, 10\} \\ &\to \{5\} \to \{4, 6, 8\} \\ &\to \{6\} \to \{1, 2, 4, 7, 10\} \\ &\to \{1\} \to \{2, 3, 4, 5, 8, 9\} \\ &\to \{2\} \to \{3, 4, 7, 8\} \\ &\to \varnothing \end{aligned}.$$

Although the tracing process for this case is more complex than the former case, the end results of both of them are the same. Based on the method proposed in our work, the exclusive root cause of any fault can be traced regardless of the initial monitoring variable. The tracing results are acceptable.

Moreover, the analysis report of the same fault obtained from the plant was used to verify the validity of the tracing results further. The comparison shows that the tracing results of our work and the actual fault reasons are consistent, which confirm the effectiveness of the proposed method.

6 CONCLUSIONS

Root cause tracing of information quality problems is a directive method for fault analysis and location, and it is one of the most important ways to explore the propagation pathway and to avoid any major accident. However, effective root cause tracing of complex electromechanical systems in a process industry is still an open issue. In this work, an information transfer based novel framework driven by data analysis and centered on improved symbolic transfer entropy method was proposed for root cause tracing in complex electromechanical systems in the process industry. In this framework, an improved symbolic transfer entropy method with binary encoding and decimal decoding was presented to increase the resistance to noise and reduce information loss when constructing the information model for complex electromechanical systems. An actual root cause tracing application of a complex electromechanical system is applied to verify the effectiveness of the proposed framework. A unique root cause was obtained regardless of the choice of the initial variable. The comparison with the fault analysis report demonstrated the validity of our tracing results. This framework can handle common problems of root cause tracing and overcome some of the drawbacks of other existing methods since a prior knowledge of fault propagation and topological structures of systems are not required. Thus, the proposed framework can be flexible and effectively used in root cause tracing of problems for complex electromechanical systems in the process industry, and formulate the foundation of information modeling, system vulnerability analysis, and condition prediction, as well as other engineering applications.

FUNDING

This work was supported by the National Natural Science Foundation of China (Grant by No. 51375375).

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the valuable cooperation of Weihe Coal Chemical Group in accomplishing this research project.

REFERENCES

Arvanitoyannis, I. S., and Varzakas, T. H. 2008. "Application of Iso 22000 and Failure Mode and Effect Analysis (Fmea) for Industrial Processing of Salmon: A Case Study," *Critical Reviews in Food Science and Nutrition* (48:5), pp. 411-429.

Ayav, T., and Sozer, H. 2016. "Identifying Critical Architectural Components with Spectral Analysis of Fault Trees," *Applied Soft Computing* (49), pp. 1270-1282.

Banduka, N., Veza, I., and Bilic, B. 2016. "An Integrated Lean Approach to Process Failure Mode and Effect Analysis (Pfmea): A Case Study from Automotive Industry," *Advances in Production Engineering & Management* (11:4), pp. 355-365.

Certa, A., Hopps, F., Inghilleri, R., and La Fata, C. M. 2017. "A Dempster-Shafer Theory-Based Approach to the Failure Mode, Effects and Criticality Analysis (Fmeca) under Epistemic Uncertainty: Application to the Propulsion System of a Fishing Vessel," *Reliability Engineering & System Safety* (159), pp. 69-79.

Chang, C. C., and Yu, C. C. 1990. "On-Line Fault Diagnosis Using the Signed Directed Graph," *Industrial & Engineering Chemistry Research* (29:7), pp. 1290-1299.

Chang, D. S., Chung, J. H., Sun, K. L., and Yang, F. C. 2012. "A Novel Approach for Evaluating the Risk of Health Care Failure Modes," *Journal of Medical Systems* (36:6), pp. 3967-3974.

Chen, Y. Y., Zhen, Z. M., Yu, H. H., and Xu, J. 2017. "Application of Fault Tree Analysis and Fuzzy Neural Networks to Fault Diagnosis in the Internet of Things (Iot) for Aquaculture," *Sensors* (17:1).

Daw, C. S., Finney, C. E. A., and Tracy, E. R. 2003. "A Review of Symbolic Analysis of Experimental Data," *Review of Scientific Instruments* (74:2), pp. 915-930.

Dickten, H., and Lehnertz, K. 2014. "Identifying Delayed Directional Couplings with Symbolic Transfer Entropy," *Physical Review E* (90:6), p. 062706.

dos Santos, F., and Bastos, L. C. 2017. "House of Quality and Information Quality: Systematic Review," *Perspectivas Em Ciencia Da Informacao* (22:1), pp. 100-111.

Fleiner, T., and Wiener, G. 2016. "Coloring Signed Graphs Using Dfs," *Optimization Letters* (10:4), pp. 865-869.

Guan, X., and Chen, Y. J. 2017. "The Interplay between Information Acquisition and Quality Disclosure," *Production and Operations Management* (26:3), pp. 389-408.

Hahs, D., and Pethel, S. 2013. "Transfer Entropy for Coupled Autoregressive Processes," *Entropy* (15:3), p. 767.

Harary, F. 1953. "On the Notion of Balance of a Signed Graph," *Michigan Mathematical Journal* (2:2), pp. 143-146.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. L. C., Shih, H. H., Zheng, Q. N., Yen, N. C., Tung, C. C., and Liu, H. H. 1998. "The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-Stationary

Time Series Analysis," *Proceedings of the Royal Society a-Mathematical Physical and Engineering Sciences* (454:1971), pp. 903-995.

Huang, X., Gao, J., and Gao, Z. 2013a. "Precise Signed Digraph Modelling Based on Causal Dependence Identification," *Proceedings of the Institution of Mechanical Engineers Part E-Journal of Process Mechanical Engineering* (227:E1), pp. 3-16.

Huang, X., Gao, J., Jiang, H., Gao, Z., and Chen, F. 2013b. "Fault Root Cause Tracing of Complicated Equipment Based on Fault Graph," *Proceedings of the Institution of Mechanical Engineers Part E-Journal of Process Mechanical Engineering* (227:E1), pp. 17-32.

Hussein, A. 2014. "Sign-Indefinite Second-Order Differential Operators on Finite Metric Graphs," *Reviews in Mathematical Physics* (26:4), pp. 900-914.

Iri, M., Aoki, K., O'Shima, E., and Matsuyama, H. 1979. "An Algorithm for Diagnosis of System Failures in the Chemical Process," *Computers & Chemical Engineering* (3:1), pp. 489-493.

Iri, M., Oshima, E., Aoki, K., and Matsuyama, H. 1980. "A Graphical Approach to the Problem of Locating the Origin of the System Failure," *Journal of the Operations Research Society of Japan* (23:4), pp. 295-312.

Kramer, M. A., and Palowitch, B. L. 1987. "A Rule-Based Approach to Fault-Diagnosis Using the Signed Directed Graph," *Aiche Journal* (33:7), pp. 1067-1078.

Kutlu, A. C., and Ekmekçioğlu, M. 2012. "Fuzzy Failure Modes and Effects Analysis by Using Fuzzy Topsis-Based Fuzzy Ahp," *Expert Systems with Applications* (39:1), pp. 61-67.

Li, J., Liang, C., Zhu, X., Sun, X., and Wu, D. 2013. "Risk Contagion in Chinese Banking Industry: A Transfer Entropy-Based Analysis," *Entropy* (15:12), p. 5549.

Liu, H. C., Liu, L., and Li, P. 2014. "Failure Mode and Effects Analysis Using Intuitionistic Fuzzy Hybrid Weighted Euclidean Distance Operator," *International Journal of Systems Science* (45:10), pp. 2012-2030.

Liu, H. C., Liu, L., and Lin, Q. L. 2013a. "Fuzzy Failure Mode and Effects Analysis Using Fuzzy Evidential Reasoning and Belief Rule-Based Methodology," *IEEE Transactions on Reliability* (62:1), pp. 23-36.

Liu, H. C., Liu, L., and Liu, N. 2013b. "Risk Evaluation Approaches in Failure Mode and Effects Analysis: A Literature Review," *Expert Systems with Applications* (40:2), pp. 828-838.

Liu, J. H., Lin, Y. J., Lin, M. L., Wu, S. X., and Zhang, J. 2017. "Feature Selection Based on Quality of Information," *Neurocomputing* (225), pp. 11-22.

Liu, Y. K., Wu, G. H., Xie, C. L., Duan, Z. Y., Peng, M. J., and Li, M. K. 2016. "A Fault Diagnosis Method Based on Signed Directed Graph and Matrix for Nuclear Power Plants," *Nuclear Engineering and Design* (297), pp. 166-174.

Macajova, E., and Skoviera, M. 2016. "Characteristic Flows on Signed Graphs and Short Circuit Covers," *Electronic Journal of Combinatorics* (23:3).

Meng, D. Y., Jia, Y. M., and Du, J. P. 2015. "Nonlinear Finite-Time Bipartite Consensus Protocol for Multi-Agent Systems Associated with Signed Graphs," *International Journal of Control* (88:10), pp. 2074-2085.

Myrelid, P. 2017. "Information Quality Deficiencies in Delivery Schedules and Their Impact on Production Scheduling," *Production Planning & Control* (28:3), pp. 232-243.

Papana, A., Kyrtsou, C., Kugiumtzis, D., and Diks, C. 2013. "Partial Symbolic Transfer Entropy," Universiteit van Amsterdam, Center for Nonlinear Dynamics in Economics and Finance.

Papana, A., Kyrtsou, C., Kugiumtzis, D., and Diks, C. 2016. "Detecting Causality in Non-Stationary Time Series Using Partial Symbolic Transfer Entropy: Evidence in Financial Data," *Computational Economics* (47:3), pp. 341-365.

Schreiber, T. 2000. "Measuring Information Transfer," *Physical Review Letters* (85:2), pp. 461-464.

Senaratne, H., Mobasheri, A., Ali, A. L., Capineri, C., and Haklay, M. 2017. "A Review of Volunteered Geographic Information Quality Assessment Methods," *International Journal of Geographical Information Science* (31:1), pp. 139-167.

Staniek, M., and Lehnertz, K. 2008. "Symbolic Transfer Entropy," *Physical Review Letters* (100:15), p. 158101.

Takens, F. 1981. *Detecting Strange Attractors in Turbulence*. Springer Berlin Heidelberg.

Talebberrouane, M., Khan, F., and Lounis, Z. 2016. "Availability Analysis of Safety Critical Systems Using Advanced Fault Tree and Stochastic Petri Net Formalisms," *Journal of Loss Prevention in the Process Industries* (44), pp. 193-203.

Umeda, T., Kuriyama, T., Oshima, E., and Matsuyama, H. 1980. "A Graphical Approach to Cause and Effect Analysis of Chemical-Processing Systems," *Chemical Engineering Science* (35:12), pp. 2379-2388.

Volkanovski, A., Čepin, M., and Mavko, B. 2009. "Application of the Fault Tree Analysis for Assessment of Power System Reliability," *Reliability Engineering & System Safety* (94:6), pp. 1116-1127.

Wang, R., Gao, J., Gao, Z., Gao, X., and Jiang, H. 2015. "Hilbert-Huang Transform Based Pseudo-Periodic Feature Extraction of Nonlinear Time Series," *Measuring Technology and Mechatronics Automation (ICMTMA), 2015 Seventh International Conference on*, pp. 532-537.

Wang, R., Gao, J., Gao, Z., Gao, X., and Jiang, H. 2016. "Complex Network Theory-Based Condition Recognition of Electromechanical System in Process Industry," *Science China Technological Sciences* (59:4), pp. 604-617.

Wijaya, D. R., Sarno, R., and Zulaika, E. 2017. "Information Quality Ratio as a Novel Metric for Mother Wavelet Selection," *Chemometrics and Intelligent Laboratory Systems* (160), pp. 59-71.

Wu, S., Li, J., Zhang, M. L., and Wang, J. 2013. "Coupling Analysis of Electrocardiogram and Electroencephalogram Based on Improved Symbolic Transfer Entropy," *Acta Physica Sinica* (62:23), p. 238701.

Yang, Z., Liu, P. X., Zhu, Y. P., and Zhang, Y. M. 2016. "A Comprehensive Reliability Allocation Method for Series Systems Based on Failure Mode and Effects Analysis Transformed Functions," *Proceedings of the Institution of Mechanical Engineers Part B-Journal of Engineering Manufacture* (230:12), pp. 2239-2248.