International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 4 Issue: 3 27 - 29

Text Mining Method to Develop D-Matrix for Fault Diagnosis

Ms. Amruta Kulkarni
Department of Computer Network
KJ College of Engg. & Management
Pune, India
amrutak0307@gmail.com

Prof. Jyoti Nighot
Department of Computer Network
KJ College of Engg. & Management
Pune, India
jyotinighot67@gmail.com

***Abstract*—** The D-matrix is one amongst the quality diagnostic models specified by IEEE Standard. This framework catches underlying connections between symptoms and failure modes in structured fashion. This framework is called as Dependency or Diagnosis framework (D-matrix).Proposed system describes text mining method based on an ontology to develop D-matrix by mining repair verbatim written in unstructured text. Here repair verbatim are collected during fault diagnosis. Then mining algorithms are applied to find dependencies. D-Matrix is constructed for different dataset, then we generate a combined D-matrix by taking common parameters from each D-matrix and then a graph is formed for that D-matrix.

Keywords- *Data mining, fault analysis, fault diagnosis, information retrieval, text processing.*

*****

I. INTRODUCTION

Nowadays digital information is increasing at an incredible rate because of the internet, which causes the problem of information overload. It is not possible to read everything, yet we often have to make critical decisions based on what we are able to understand. Thus, effective management of electronic documents, especially management of complexity and specialization of knowledge expressed in those text documents, is vital to enterprise knowledge management [1]. Each system performs its task of communicating with its adjacent system to maintain its performance within an adequate range of tolerances. One challenge that managers face is how to find a deep knowledge from a collection of documents to solve a problem [1]. The fault detection and diagnosis (FDD) is performed to diagnose the root-causes of the fault and to minimize the downtime of a system [1].

DUE to the rapid growth of digital data in recent years, data mining have attracted a great deal of attention with an imminent need for turning such data into useful information and knowledge [8]. The goal of data mining is to discover or derive new information from data, finding patterns across datasets, and/or separating signal from noise. Text mining is gaining a heavy attention because of its ability to discover the required data assets from unstructured text. In this paper, we tend to propose a text mining method to map the diagnostic info extracted from the unstructured repair verbatim in a very D-matrix [1].

In the past decade, a significant number of data mining techniques have been presented in order to perform different knowledge tasks. The D-matrix is one amongst the quality diagnostic models laid out in IEEE Standard. This framework catches causal connections between symptoms and failure modes in structured fashion. This framework is called as Dependency or Diagnosis framework (D-matrix) [1] [4].

When fault is identified in system technicians follow some diagnosis procedure based on error codes previously stored in system, to identify the nature of faults. Different data types are collected during fault diagnosis is then stored in the database. This data is then mined to construct diagnosis matrix models. Such models can be used by the technicians and stakeholders to find accuracy of fault detection.

II. LITERATURE SURVEY

The rapidly increasing amounts of text data in the context of these large online collections has led to an interest in creating scalable and effective mining algorithms [9].

In existing fault models, the knowledge which was embedded in the unstructured repair data have improved the performance of fault diagnosis by introducing an approach for constructing D-matrices based on an ontology-based text mining method[1].

Ontology includes identifying, defining and entering concept definitions and their relationship [6]. Here an approach is described for ontology extraction from an existing knowledge base of heterogeneous documents. Misinterpretation of pattern derived from data mining techniques lead to the ineffective performance. In [8], proposed technique uses two processes, pattern deploying and pattern evolving to refine the discovered pattern in text documents.

Multi signal flow modeling is one of the particular interests because it is one example where tools have been developed that use D-Matrix [4], [5]. Different diagnostic algorithms used in diagnostic interference model (DIM) are summarized in [4].

In the [2], [7], the limited efforts are done to create a D-matrix by analyzing unstructured repair verbatim.

Traditionally, the D-matrices are constructed by using the knowledge included in the field of failure data. The data includes historical data, engineering data, and sensory data, error codes [2], [7], [5], [10], but the authors have not provided any perception for new symptoms and failure modes which are observed for the first time and their insertion in the D-matrix models.

A maturation approach is proposed in [7] which represent Timed Failure Propagation Graph (TFPG) models. Further in [3] the researcher worked on developing D-matrices from dissimilar information format and data sources. The D-matrices is classified based on their data source and the

imperfectness of symptoms. They have considered for both boolean-value and real-valued [0, 1] D-matrices.

III. EXISTING SYSTEM

The methodology consists of ontological text mining method. It captures the terms and the relations seen in the automobile fault diagnosis domain. The D-matrix is constructed using following steps -document annotation, term extraction, and phrase merging.

Initially, the repair verbatim data points are collected by retrieving them from the database, which are recorded during FD. In the first step, the terms relevant for the D-matrix are annotated from each repair verbatim by developing the document annotation algorithm.

A repair verbatim consists of several parts, symptoms, failure modes and actions and the correct associations must be established between the relevant terms based on their propinquity with each other. A repair verbatim is first split in different sentences using the sentence boundary detection rules and the terms appearing in the same sentence are co-related with each other. Finally, Naive Bayes probability model is developed to disambiguate the abbreviated terms by considering the context in which they are specified.

Next, the phrase merging is used to avoid ambiguous references of the failure mode phrases, where the failure mode phrases that are written by using an inconsistent vocabulary are merged into a single, consistent failure mode phrase. Finally, the newly constructed D-matrix is audited by subject matter experts (SMEs) to identify the discovery of new symptoms and failure modes for in-time FDD.

The existing system [1] creates the D-Matrix for one dataset. It provides accurate d-matrix but graph is not generated from the matrix. So that, every time, the new d-matrix is created for the dataset. Even if the different datasets contains some similar data, the new D-Matrix is developed for each dataset.

IV. MOTIVATION

Fault diagnosis is critical process when performed on unstructured data. D-matrix generation from fault diagnosis on unstructured data makes it easier to identify faults and also shows the relationship between symptoms and fault. This is helpful to take action. This motivates the work in this paper. Implementing this using a different technique (in our case ANN & SVM) will give more accuracy.

V. PROPOSED SYSTEM

Proposed Ontology-Based Comprehensive D-Matrix Using Support Vector Machine and ANN consist mainly three steps.

1. Generating D-Matrix from unstructured repair verbatim
2. Generate a combined D-Matrix
3. Form a combined graph

In this system first a collection of repair verbatim is mined by using the a. Document annotation b. Term extraction and c. Phrase merging. In these steps fault diagnosis ontology is applied on text. The d-matrix is generated for each mined dataset. Then a combined D-Matrix is generated by taking only the common patterns from all the generated various D-Matrices to construct a single, generic D-Matrix. After that a graph is generated for that combined D-Matrix which shows the relationship between all common fault and symptoms from the merged d-matrices.

Our system will create D-Matrix for each dataset of unstructured repair verbatim. Here we are using a dataset containing diseases and its repair actions from a hospital domain. At first, the repair verbatim data is collected by recovering them from the OEM database, which are recorded during fault diagnosis. In the first step, the terms, for example, part, symptom, and failure mode, repair actions which are applicable to generate the D-Matrix are annotated from each repair verbatim by using the document annotation step. In this step, stop word are removed from the dataset and also lexical matching is performed to find a correct meaning of abbreviations.

Here, a boundary of sentence is identified by checking sentence separator. Then identified terms are get checked in repair verbatim. At last, Support Vector Machine and Artificial Neural Network model is developed to perform training on unstructured repair verbatim. Comparison of both the technique will show the accuracy of fault diagnosis.

VI. CONCLUSION

Generating D-Matrix by using text Driven approach have more impact. This framework helps the service technician to detect the faults related to complex system and diagnose it. This dependency model (D-Matrix) contains symptoms, failure modes and their causal relationship. The probabilistic approach helps to solve various problems such for removing ambiguity and to establish better co-relation between the terms. Development of a graph from D-matrix gives better visualization and analysis.

REFERENCES

[1] Dnyanesh G. Rajpathak, "An Ontology-Based Text Mining Method To Develop D-Matrix From Unstructured Text", *Member, IEEE And Satnam Singh, Senior Member, IEEE.*

[2] J. Sheppard, M. Kaufman, And T. Wilmering, "Model Based Standards For Diagnostic And Maintenance Information Integration", *In Proc. IEEE AUTOTESTCON Conf., Pp. 304–310, 2012*

[3] S. Singh, S. W. Holland, And P. Bandyopadhyay, "Trends In The Development Of System-Level Fault Dependency Matrices", *In Proc. IEEE Aerosp. Conf., Pp. 1–9, 2010.*

[4] J. W. SHEPPARD AND S. G. W. BUTCHER,"A Formal Analysis Of Fault Diagnosis With D-Matrices", *Springer Science, JOURNAL OF ELECTRONIC TESTING: Theory And Applications, 2007*

[5] S. Deb, S. K. Pattipati, V. Raghavan, M. Shakeri, And R. Shrestha, "Multi-Signal Flow Graphs: A Novel Approach For System Testability Analysis And Fault Diagnosis", *IEEE Aerosp. Electron. Syst.., Vol. 10,No. 5, Pp. 14–25, May 1995.*

[6] Matteo Gaeta, Member, IEEE, Francesco Orciuoli, Stefano Paolozzi, And Saverio Salerno, "Ontology Extraction For Knowledge Reuse:The E-Learning Perspective, Ieee

Transactions On Systems, Man, And Cybernetics", *Part A: Systems And Humans, Vol. 41, No. 4, July 2011*

[7] S. Strasser, J. Sheppard, M. Schuh, R. Angryk, And C. Izurieta, "Graph Based Ontology-Guided Data Mining For D-Matrix Model Maturation", *In Proc. IEEE Aerosp. Conf., Pp. 1–12, 2011.*

[8] Ning Zhong, Yuefeng Li, And Sheng-Tang Wu, "Effective Pattern Discovery For Text Mining", *IEEE Transactions On Knowledge And Data Engineering, Vol. 24, No. 1, January 2012*

[9] S. Singh, A. Kodali, K. Choi, K. R. Pattipati, S. M. Namburu, S.C. Sean, D. V. Prokhorov, And L. Qiao, "Dynamic Multiple Faul T Diagnosis:Mathematical Formulations And Solution Techniques", *IEEE Trans. Syst., Man Cybern. A, Syst. Humans, Vol. 39, No. 1, Pp. 160–176,Jan. 2009.*

[10] Tinal R. Thombare, Lalit Dole, "D-Matrix: Fault Diagnosis Framework", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 3, Issue 3, March 2015.Margaret H. Dunham, "Data Mining: Introductory and Advanced Topics", Pearson Education

[11] https://en.wikipedia.org/wiki/Data mining

[12] https://en.wikipedia.org/wiki/Support vector machine.

[13] https://en.wikipedia.org/wiki/Artificialneural network