ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

A Survey: on Text Mining Unstructured Data Using D-Matrix

Sneha P. Mendhe[1] , K.N.Hande[2]

PG Student, Dept. of Computer Science & Engg., Priyadarshini Bhagwati College of Engineering, Nagpur, India

Assistant Professor, Dept. of Computer Science & Engg., Priyadarshini Bhagwati College of Engineering, Nagpur, India

ABSTRACT: Fault dependency (D-matrix) is used as a diagnostic model that identifies the fault system data and its causal relationship at the hierarchical system-level. It consists of dependencies and relationship between identified failure modes and symptoms related to a system. Constructing such D-matrix fault detection model is time overwhelming task .A system is proposed that describes associate ontology based text mining on unstructured data using D-matrix for automatically constructing D-matrix by mining many repair verbatim text data (typically written in unstructured text) collected throughout the identification process. And also graphical model generation for each generated D-matrix. Initially we construct fault diagnosis ontology and then text mining techniques are applied to spot dependencies among failure modes and identified symptom. D-matrix is represented in graph so analysis gets easier and faulty parts becomes simply detectable. The proposed methodology are implemented as a prototype tool and validated by using real-life information collected from the automobile domain.

KEYWORDS: Fault diagnosis; fault detection; information retrieval; dependency-matrix; text mining.

I. INTRODUCTION

To maintain consistency in the performance within range of tolerances the system must interacts with its surrounding for execute some set of task. A term fault is a deviation from its normal process behaviour. Fault Detection and Diagnosis (FDD) performance for identifying the fault and diagnose root causes of the system. Data recording book kept diagnosis information comes within the form of unstructured repair verbatim that gives a lot of useful information for data diagnosis purpose. Thousands of repair verbatim are collected and argue that there's an requirement need to mine this information to enhance fault detection (FD).Now a day's Text mining is getting tremendous response for its ability to automatically discover the informational assets within unstructured text. In this paper, we propose a text mining method to map the diagnostic data extracted from the unstructured repair verbatim in a very D-matrix. This is used to set a correlation between symptoms and its failure modes in structured text fashion. This framework is termed as Dependency or diagnosis framework (D-matrix). During fault diagnosis, several data varieties are collected, like error code, scanned values of operating values related to faulty component system, repair verbatim.

The collected data transferred to the database and particularly repair verbatim data collected over a period of time can be extracted to construct the D-matrix diagnostic models. These models are often used by field technicians and different stakeholders for performing correct FDD. Generally, the D-matrix are constructed by utilizing the historical data, engineering data, and sensor data. Even so, a much nothing understanding is given regarding the disclosure of current or recent symptoms furthermore; faulty condition saw firstly and their incorporation within the dependency matrix models the perfect D-matrix.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

II. RELATED WORK

This section primarily concentrates on how fault detection and diagnosis (FDD) is done to identify the faults through D-matrix framework associated with the automobile domain.

[1] Describe about to construct Ontology based text mining method encompassing of dependencies and relationships currently saw in the fault diagnosis. The text mining algorithms makes utilized of this ontology to find the necessary outcome such as parts, symptoms, faulty modes to develop the D-matrix by auto mining the unstructured repair verbatim information collected during fault diagnosis episode.[2] They mainly describe about analyzing of information from the on-board Diagnosis by utilizing the data mining process. Onboard diagnosis model is assumed to be still and proper. But in real world, due to engineering challenges and design, newly vehicle structure and vehicle frame is launching. This approach gives, Faults are removed and provided conception data mining and data transformation, But Discovery is mislaying because result was not in form of matrix.

In the [11], [20], [16], [4], normal efforts are to be taken to create a D-matrix by analyzing unstructured repair verbatim. The [17] tool is proposed which collects the data from the on-board diagnosis by using the ontology-based data mining. The onboard diagnosis is processed by collecting the real time data and integrates onboard ECUs. This model is assumed to be static and complete. However in real world, due to engineering changes and design, new vehicle structure and vehicle design is launching. The vehicles launch new failure symptoms and failure modes a number of the tools suffer some drawbacks associated with the symptoms and fault parts. In past D-matrix constructed manually, to overcome the issues within the real life industry of getting to construct model D-matrix manually. Conventionally, the D-matrices are made by using the data hidden within the field failure data. This type of data having generated historical data, engineering data, and sensory data, generated error codes.

Traditionally, the D-matrices are constructed by using the knowledge buried in the field failure data. The data includes historical data, engineering data, and sensory data, error codes [11], [12], [13], [4], [14], [15], but the authors haven't provided any perception for recent failure symptoms and failure modes that are observed for the first time and their insertion in the D-matrix models. The prior information is useful for constructing D-matrix fault diagnosis model to create it additional correct. The field failure data using spreadsheets that is time consuming and labour-intensive method. So a data-driven framework is develop [2] that automatically find the unusual activity that leads to fault and saves a significant expert's time. This framework is developed so they could completely work on analyzing anomalies and taking correct actions. Further in [6] the researcher worked on developing D-matrices from dissimilar information format and data sources. The fault diagnosis D-matrix models are used successfully in aerospace industry [9], [10] to spot the dependencies among failure modes, symptoms, and repair claims by analyzing the structured service manual data.

III. CONCLUSION AND FUTURE WORK

In previous method using the data sources that view the overall data that is saving all database and firstly parse that data after that scan overall data so it is takes more data base memory and it is very much time consuming. In my proposed system mainly works on ontology text mining method which will perform auto mining construction and updation the dependency matrix (D-Matrix) for optimization of time. Also it will improve the accuracy of Fault Detection and Diagnosis. To overcome these limitations where natural language processing algorithms were proposed to immediately construct the D-matrices from the unregulated repair verbatim.

We compared the testability and diagnosability matrix of the historical data-driven D-matrix and the text-driven D-matrix, where the text-driven D-matrix approach shows higher fault recognition, higher fault reclusion, and lower ambiguity group size due to textual symptoms and the corresponding failure modes included in the text-driven D-matrix. Development of a graph from D-matrix model gives better visualization and analysis. It helps in real world industry to identify the necessary facts.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

REFERENCES

1. Dnyanesh G. Rajpathak, Satnam Singh, "An Ontology-Based Text Mining Method to Develop D-Matrix from Unstructured Text", IEEE trasactions on system, man and cybernetics : systems, vol. 44, no.7, pp. 966-977, Jul. 2014.
2. S. Singh, H. S. Subramania, and C. Pinion, "Data-driven framework for detecting anomalies in field failure Data", in Proc. IEEE Aerosp. Conf., pp. 1–14, 2011.
3. R. Chougule and S. Chakrabarty, "Application of ontology guided search for improved equipment diagnosis in a vehicle assembly plant", in Proc.IEEE CASE, pp. 90–95, 2009.
4. 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.
5. D. Wang, W. H. Tang, and Q. H. Wu, "Ontology-based fault diagnosis for power transformers", in Proc. IEEE Power Energy Soc. Gen.Meeting, pp. 1–8, 2010.
6. S. Singh, S. W. Holland, and P. Bandyopadhyay, "Trends in the development of system-level fault dependency matrices", in Proc. IEEEAerosp. Conf., pp. 1–9, 2010.
7. T. J. Felke and J. F. Stone, "Method and Apparatus for Developing Fault Codes for Complex Systems Based on Historical Data", US Patent003318 A1, Jan. 2004.
8. S.P. Eagleton and T. Felke, "Method and Apparatus using Historical data to Associate Deferral Procedures and D-matrix", US Patent 6,725,137 B2, Apr. 2004.
9. T. Felke, "Application of model-based diagnostic technology on the Boeing 777 airplane", in Proc. 13th AIAA/IEEE DASC, pp. 1–5, 1994.
10. G. Ramohalli, "The Honeywell on-board diagnostic and maintenance system for the Boeing 777", in Proc. IEEE/AIAA DASC, pp. 485–490, 1992.
11. E. Miguelanez, K. E. Brown, R. Lewis, C. Roberts, and D. M. Lane,"Fault diagnosis of a train door system based on semantic knowledge representation railway condition monitoring", in Proc. 4th IET Int.Conf., pp. 1–6, 2008.
12. 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.
13. M. Schuh, J. Sheppard, S. Strasser, R. Angryk, and C. Izurieta, "Ontology-guided knowledge discovery of event sequences in maintenance data", inProc. IEEE AUTOTESTCON Conf., pp. 279–285, 2011.
14. S. Deb, S. K. Pattipati, V. Raghavan, M. Shakeri, and R. Shrestha, "Multi-signal flow graphs: A novel approach for system testabilityalysis and fault diagnosis", IEEE Aerosp. Electron. Syst.., vol. 10,no. 5, pp. 14–25, May 1995.
15. S. Singh, A. Kodali, K. Choi, K. R. Pattipati, S. M. Namburu, S.C. Sean,D. V. Prokhorov, and L. Qiao, "Dynamic multiple fault diagnosis: Mathematical formulations and solution techniques", IEEE Trans. Syst.,
16. Man Cybern. A, Syst. Humans, vol. 39, no. 1, pp. 160–176,Jan. 2009.M. Schuh, J. W. Sheppard, S. Strasser, R. Angryk, and C. Izurieta, "A Visualization tool for knowledge discovery in maintenance event sequences", IEEE Aerosp. Electron. Syst. Mag., vol. 28, no. 7, pp. 30– 39, Jul. 2013.
17. P. M. Frank and J. W¨ unnenberg, "Robust fault diagnosis using unknown input observer schemes", in Proc. Fault Diagnosis Dynamical Syst.: Theory Appl., pp. 47–98, 1989.
18. N. Viswanadham and R. Srichander, "Fault detection using unknown input observers", Control-Theory Ad. Tech., vol. 3, pp. 91–101, 1987
19. P. M. Frank, "Fault detection in dynamic systems using analytical and knowledge-based redundancy—a survey and some new results",Automatica, vol. 26, no. 3, pp. 459–474, 1990.
20. A. S. Willsky, "A survey of design methods for fault detection in dynamic systems", Automatica, vol. 12, no. 6, pp. 601–611, 1976.
21. V. Venkatasubramanian and S. H. Rich, "An object-oriented two-tier architecture for integrating compiled and deep-level knowledge for process diagnosis", Comput. Chem. Eng., vol. 12, no. 9–10, pp. 903– 921, 1988.
22. C. Charniak and D. McDermott, Introduction to Artificial Intelligence. Reading, MA, USA: Addison Wesley, 1985.
23. V. R. Benjamins, "Problem-solving methods for diagnosis and their role in knowledge acquisition," Int. J. Expert Syst.: Res. Appl., vol. 8, no. 2, pp. 93–120, 1995.
24. M. Iri, K. Aoki, E. O'Shima, and H. Matsuyama, "An algorithm for diagnosis of systems failures in the chemical process", Comput. Chem.Eng., vol. 3, nos. 1–4, pp. 489–493, 1979.

BIOGRAPHY

Sneha P. Mendhe is PG Student in Dept. of Computer Science & Engg., Priyadarshini Bhagwati College of Engineering, Rashtrasant Tukdoji Maharaj Nagpur University.

K.N.Hande is Assistant Professor in Dept. of Computer Science & Engg., Priyadarshini Bhagwati College of Engineering, Rashtrasant Tukdoji Maharaj Nagpur University.