ISSN:0975-9646

An Ontology-Based Enhance Term Prediction Algorithm and Union of Divers D-Matrix

Seema Jadhav[1], Prof.Santosh Chobe[2],

[1, 2]*Department of Computer Engineering,*
DYPIET Pimpri, Savitribai Phule Pune University, India

***Abstract*— To squeeze the reviewed framework level fault diagnostic data comprising of conditions between finding symptoms and lack of success of modes associated with a framework Fault Dependency (D)-matrix is an organized diagnostic model.There is need to gather data in regards to servable symptoms and failure modes to modify the fault dependency matrix which can be useful to build precise and proficient fault diagnosis. An ontology based D-matrix describes an ontology based text mining procedure for consequently constructing and updating a D-matrix innumerable repair verbatim assembled the analysis episodes. To represent unstructured knowledge, ontology based data mining innovation can be useful which manages gathering of unstructured information considering likenesses and contrasts between them. Ontology is fabricated which gives the normally observed relationship in fault diagnostic domains. By utilizing diverse text mining algorithm, important concepts like symptoms, failure mode and their relation with unstructured information can be found. In our strategy we first build the D-matrix for various datasets and use C4.5 classifier for classification. Then generating graph model for each produced D-matrix and utilized the graph comparison calculations to develop new graph and using that graph create single comprehensive D-matrix.**

***Keyword*—Data mining, Text mining Data mining, Text processing, Fault diagnosis ontology, D-matrix**.

I.INTRODUCTION

Typically, created system lives up to expectations in its pre recognized working conditions and done with its given assignment as expressed. There is nothing to stress over the framework until it is working precisely and gives worthy results. In the event that the consequence of the system is not as per the expectation, this as introduction of fault in system. Identification of faults and its rectification is a subfield of control designing which relate itself with dealing with a framework, perceiving when an issue has happened, what are the reasons behind it and discover the sort of fault and its location. It is essential to find the fundamental driver of a fault on the grounds subsequent to there might be possibility that other interconnected subsystems may similarly give fault indications that might possibly cover the hidden driver. Systems like On-Board Diagnostics (OBD) referring to a vehicle self-diagnostics and reporting capability. This framework gives mechanics or proprietor of the vehicle access to the status of the different vehicle subsystems [6]. After resolving the different issues it is essential to note down its causes, effect of the cause on the framework in organized way so that use this data latter while developing to the framework to make it perfect.

Fault dependency D-Matrix is a systematic analytic model to get the different framework level fault diagnostic data comprising of dependencies between recognizable symptoms and failure modes associated with a system. But, there is need to gather information regarding servable symptoms and failure modes to modify the fault dependency matrix which can be helpful to build accurate and efficient fault diagnosis. So, an ontology based D-matrix depicts an ontology based text mining strategy for consequently mining so as to build and upgrading a D-matrix by mining countless repair verbatim (typically written in unstructured text) accumulated during the diagnosis scenes.

The purpose of Text Mining [9] is to prepare unstructured information, bring out significant indices from the textual data utilizing mining algorithms. Ontology is a mechanism that describes the concepts in a domain furthermore the relationships that hold between those concepts. So as to develop an ontology framework for the Fault Diagnosis (FD) of vehicles frameworks, it is important to analyze numerous concepts and relationships exhibited. An ontology model is produced for exact and efficient fault diagnosis for vehicles frameworks. The D-matrix is built as it is one of the standard diagnostic models specified in IEEE Standard 1232 to analyze the faults in automobile frameworks [10]. The methodology of FD starts by removing the mistake codes from an objective structure and, focused around the watched lapse codes, the experts take after specific finding procedure close by their experience to analyze the deficiencies. At the time of fault determination, a few data sorts are assembled, for instance, error codes, analyzed benefits of working parameters associated with broken part/structure, repair verbatim, and some more. The assembled data is then treated to the OEM database and particularly the repair verbatim data gathered over a time can be mined to make the D-matrix analytic models. Such models can be used by the field experts and different partners to perform exact FDD.

II.LITERATURE REVIEW

Dnyanesh G. Rajpathak et al.[1]focused on the fault information is captured and formalized in the fault diagnosis ontology, which is expanded in light of the new information. The resulting ontology based data mining algorithms that uses this data model encourages in-time FD. To make a D-Matrix systematic model, principled

approach is proposed by analyzing the unstructured repair verbatim data connected with the different structures in parallel through the development of philosophy based substance mining computations. It overcomes the confinement faced in the real business of expecting to construct the D-Matrix investigative models physically or using first principles. Further in proposed system have capacity to capture the cross-structure conditions which helped to fundamentally enhance the execution of FDD. The relations from the fault analysis Ontology are used to discover the conditions between the manifestations and the failure modes contrasting with various structures. It upgraded the execution of proposed system when compared with the Latent-Dirichlet Allocation (LDA) procedure [11].

M. Schuh et al.[2]developed a tool that facilitates knowledge discovery from aircraft maintenance data through sequences of maintenance events. A few interconnecting ontologies were produced in Ontology Web Language(OWL), in view of formally characterized IEEE principles, and use these ontologies to manage the information change, information mining, and intelligent representation forms. The device gives a simple to-utilize interface that produces pertinent groupings of information in a significant setting in a small amount of the time it would take space specialists to recover and show comparable data. To present an innovative review of the present device and expansions, which now incorporates a few perception alternatives, ties together a few philosophy based information sources, and empowers a technique for getting indicative development suggestions.

S.Strasser et al.[3] focused on model-based diagnostic algorithm assumed that the model is correct. In the event that the model is incorrect, the analytic calculation might analyze the wrong fault, which can be exorbitant and tedious. Utilizing past upkeep occasions, one to have the capacity to make rectifications to the model all together for indicative calculation to accurately analysis faults. In this system, a development methodology is proposed which utilizes the graph theoretic representations of Timed Failure Propagation Graph (TFPG) models and indicative sessions in light of as of late institutionalized symptomatic ontologies to decide factual errors between that which is normal by the models and that which has been encountered in practice. These inconsistencies are then examined to produce proposals for developing the demonstrative models. Development proposals incorporate recognizing new conditions and wrong or shaky conditions.

T. Felke [4] focused on the application of model based diagnostic technology to the Central Maintenance Computer for the Boeing 777 Airplane. The paper introduces an outline of model based indicative innovation as connected to the 777 with an emphasis on the model era process. It recognizes the distinction of building up a model with the suitable level of determination. A model that is excessively broad won't give the required exactness while a model that is excessively definite will be troublesome, making it impossible to create and keep up. The way that information passage exercises for parts of the model expected to begin before the model was totally indicated was the wellspring of extra troubles that the procedure needed to succeed. The paper displays how these issues were illuminated on the 777 and how this experience can be advantageous in future projects.

J.Sheppard et al.[5] focused on, there is no doubt that system complexity is increasing. There are various ramifications of this expansion in multifaceted nature, other than higher execution. On one hand conventional method for testing are being overpowered by the many-sided quality. Framework information is divided by time and teaches. Early outline information is not accessible amid the operational stage. Outline information is frequently isolated from test information. Indeed, even inside of a specific control, e.g. diagnostics, information is divided. One way to deal with managing the multifaceted nature issues is to incorporate these wellsprings of data into a solitary photo of the condition of the framework. This is the methodology taken by the SCC20 Diagnostic and Maintenance Control (DMC) subcommittee. The DMC is building up a group of gauges that are item data trade guidelines for test, analysis, and upkeep. This paper depicts the present endeavors by the DMC to coordinate information from an extensive variety of sources into a "photo" of the symptomatic and support condition of a framework.

O. Benedittini et al. [6] worked onIntegrated vehicle health management (IVHM) is an accumulation of information pertinent to the present and future execution of a vehicle framework and its change into data can be utilized to support operational choices. This outline and operation idea grasps an incorporation of sensors, correspondence advancements, and computerized reasoning to give vehicle wide capacities to analyze issues and prescribe arrangements. This article intends to report the presenting so as to cut edge of IVHM examination an efficient audit of the writing. The writing from various sources is ordered and broke down, and the major rising subjects are introduced. On this premise, the article depicts the IVHM idea and its development, examines setups and existing applications alongside primary drivers, potential advantages and boundaries to appropriation, outlines plan rules and accessible systems, and distinguishes future exploration challenges.

T.Jadhav et al.[7] focused in data annotation methods and An ontological approach is proposed for extracting and structuring data from unstructured documents posted on the web. The data extraction method is based on conceptual modeling, this approach focuses specifically on unstructured documents which are rich in data, narrow in ontological breadth, and contain multiple records of information for the ontology. So to automatically extract data in multi-record documents and label them, it employs ontologies together with several heuristics. However, it is necessary to construct ontologies manually for different domains.

H.Liu et al.[8]focused on the use of Natural Language Processing (NLP) methods for information extraction and thought indexing in the biomedical region, a methodology that rapidly and proficiently consigns the right sentiment an uncertain biomedical term in a given setting is required all the while. The present status of Word Sense Disambiguation (WSD) in the biomedical area is that high quality principles are used in perspective of significant material. The insults of this procedure are (i) producing WSD runs physically is a period expending and dull assignment, (ii) maintenance of principle sets turns out to be progressively troublesome after some time, and (iii) high quality guidelines are regularly fragmented and perform ineffectively in new areas involved particular vocabularies and distinctive classifications of content. This work shows a two-stage unsupervised system to make a WSD classifier for a dubious biomedical term W. The chief stage normally bodes well marked corpus for W, and the second stage surmises a classifier for W using the decided sense-labeled corpus as a planning set. A formative examination was performed, which fiend began that classifiers arranged on the decided sense named corpora accomplished a general precision of around 97 percent, with more conspicuous than 90 percent exactness for each individual uncertain term.

III.IMPLEMENTATION DETAILS

In this section mentioned the system overview in detail, proposed algorithm, of the proposed system.

A. System Overview

The objective of this paper is to give An Ontology-Based Comprehensive D-Matrix Using Graph Comparison Algorithm, which is consisting of developments of D-matrix from repair verbatim data. We create a graph for each D-matrix, as soon as the generation of D-matrices from different datasets. Then we combine the graph in such a way that only one common pattern is merged from generated various D-matrices to build a single, generic D-matrix. To build the D-matrix [1], following steps have to be created:

- The fault diagnosis ontology by using dataset.
- Ontology-based text mining.

Fault diagnosis ontology is developed using vehicle dataset that shows the concept and sub concept and relation between that and their instances. Ontology-based text mining including steps Document Annotation, Term Extraction and Phrase Merging. The proposed system creates two D-matrix for two dataset respectively. Then, the undirected graph is generated depending on the D-matrix. At first, the fault diagnosis ontology is created .Then in ontology based text mining, the following steps are performed.

1) Document Annotation

The document annotation is used to sort the required data which is used for analysis and it removes the irrelevant data.

The repair verbatim data focuses are composed by recovering them initial the OEMs database, which are recorded through field FD. In the initial step, the conditions, for instance, part, symptom, and failure mode, applicable for the D-matrix are annotated from every one repair verbatim by building up the document annotation calculation. A repair verbatim includes a couple parts, symptoms, failure modes and actions and the right associations must be built up between the essential terms in view of their vicinity with each other. Here, a repair verbatim is first part in various sentences by using the sentence limit recognition standards and the tern appearing in the same sentence are co-related with each other. The stop words are deleted that are non-descriptive terms. Finally, the terms from the handled verbatim are coordinated utilizing the examples as a part of the fault diagnosis ontology.

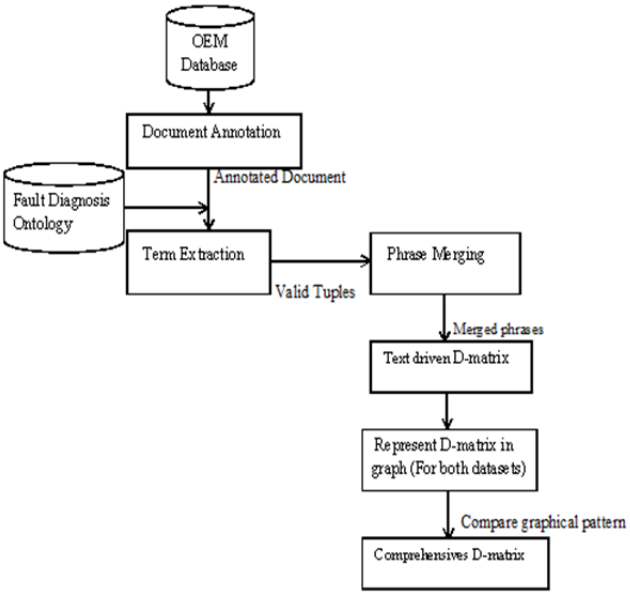

Fig.1. System Architecture

2) Term Extraction

After the annotation algorithm, there is a need to have the terms which are used to develop the D-matrix. For this purpose, the term extraction algorithm is used. From each one annotated repair verbatim the tuples, for example parts - $P_a \in \{P_1,P_2,\ldots.P_i\}$, symptoms- $S_b \in \{S_1, S_2, \ldots.S_j\}$, failure modes - $f_c \in \{f_1, f_2,\ldots.f_k\}$, and $(S_b\ P_a\text{-}f_c) \in \{S_1, P_1\text{-}f_2, S_1, P_1\text{-}f_2,\ldots., Sj,Pi\text{-}fi, SjPj\text{-}fj\}$ are developed by using the term extraction estimation to populate a D-matrix. There are more tuples generated in that similar tuples also present so tuple validation is necessary. In validation process tuple weight is calculated. If that weight is greater than threshold that tuple is consider as valid tuple.

3) Phrase Merging

In this step extract the failure mode phrases from term extraction phase and to avoid the ambiguous references from those failure modes, use phrase merging. Parts, symptoms, failure mode, and action this parameters co-occurring with the expressions, which is used to measure the conditional probabilities and the expression with their probability score over the particular threshold are combined. At last whatever phrases merged using that we develop the D-matrix.

B. Algorithm

In our proposed system we use fault diagnosis ontology and vehicle unstructured information for building D-matrix. D-matrix shows the dependency between symptom and failure mode. In D-matrix symptoms take as column and part failure mode as row. D-matrix is in binary format. i.e.0 and 1. If 1 is output of D-matrix for particular column and row, then the fault is detected between that part failure mode and symptoms. D-matrix of two datasets represent as graph. Symptoms and failure mode take as vertex in graph and according to their relation edge is formed. In this way the graph is formed from the created D-matrix. Next, our framework analyzes two graphs and combined the common patterns appears in both graphs. Development of a graph from D-matrix gives better visualization and analysis.

Steps of proposed system-

1) D-matrix1 from datset1
2) D-matrix2 from dataset2
3) Graph1 from D-matrix1
4) Graph2 from D-matrix2
5) List of co-ordnance from graph1 and graph2. As columns from D-matrix1 related to columns from D-matrix2. Rows from D-matrix1 related to rows from D-matrix2
6) Similarity depending upon the edges from vertices.
7) Similarity graph generation i.e. graph3
8) Heterogeneous D-matrix from graph3.

To handle the abbreviation disambiguation problem we use the C4.5 classifier that required less processing time compared to Naive-Bayes model in existing system.

IV. Result and Discussion

A. Experimental Setup

This system is developed on Java framework (version JDK 1.8.0) and NetBeans (version 8.0.2) used as a development tool with windows platform. The system doesn't require any specific hardware requirement to execute as well as it execute on any common machine.

B. Dataset Discussion

In this system we used repair verbatim vehicle dataset in which content vehicles parts, symptoms, failure modes of a dataset. Also we create ontology on vehicle dataset and give to the system as an input.

C. Result

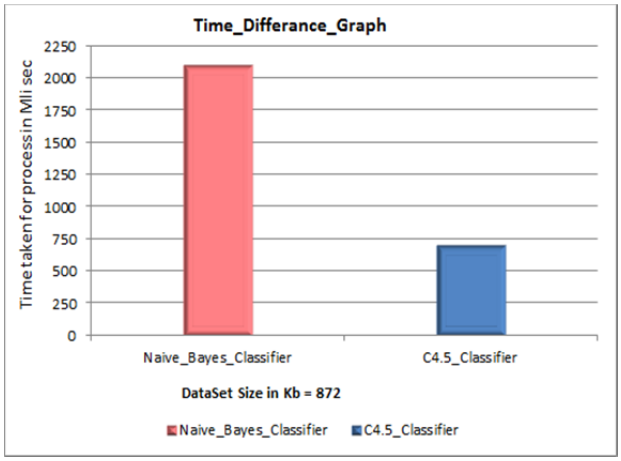

Fig.2. Time Graph comparison between Existing and Proposed System

The graph shows the time required for the existing system with naive bayes classifier is more than the time required for the proposed system with C4.5 classifier.

Following Fig.3 shows the final D-matrix generated after graphical representation and union process of both dataset D-matrix. This final D-matrix improves the fault detection.

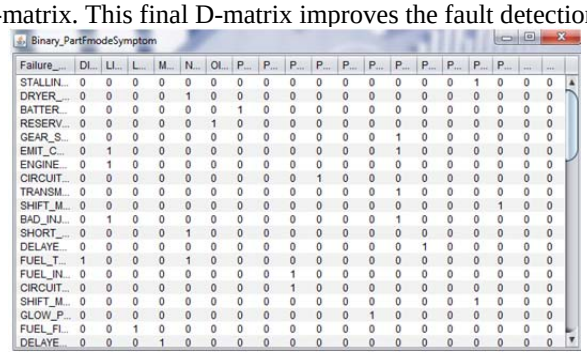

Fig. 3. Final D-matrix

V. Conclusion

By considering the applicability of making common D-matrix from different D-matrices, create the D-matrix from each unstructured repair verbatim information by utilizing text mining algorithm. Proposed framework shows the undirected graphs that are produced for every D-matrix which is created from the unstructured repair verbatim information. The graph comparison calculation is utilized to create D-matrix such that the normal examples rising up out of the heterogeneous D- matrices which can be utilized to develop single, comprehensive D-matrix. This D-matrix improves the fault detection process and save the space. Using C4.5 classifier we save the system processing time.

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] 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. 3039, Jul. 2013.

[3] 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, 2011, pp. 112.

[4] Felke, T., "Application of model-based diagnostic technology on the Boeing 777 Airplane".i n *Procdure. 13th AIAA/IEEE DASC,* 1994,pp. 1–5.

[5] J. Sheppard, M. Kaufman, and T.Wilmering, "Model based standard for diagnostic and maintenance information integration," in Proc. IEEE AUTOTESTCON Conf., 2012, pp. 304-310.

[6] O. Benedittini, T. S. Baines, H. W. Lightfoot, and R. M. Greenough, "State-of-the-art in integrated vehicle health management," J. Aer.Eng.,vol. 223, no. 2, pp. 157-170, 2009.

[7] Tushar.Jadhav, Santosh Chobe, "Data Extraction and Annotation Methods Using Tag Value Structure" in Proc.IJSR,Volume 4 Issue 7, July 2015.

[8] Hong fang Liu,1 Yves A. Lussier,,2 and Carol Friedman, "Disambiguating Ambiguous Biomedical Terms in Biomedical Narrative Text: An Unsupervised Method", *J. Bio. Inf* ., vol. 34, no. 4, pp. 249–261, 2001.

[9] T. Hearst, "Untangling text data mining," in *Proc. 37th Annu. Meeting Assoc. Comput. Linguist.*, 1999, pp. 3–10.

[10] *IEEE standard for artificial intelligence exchange and service tie to all test environments (AI-ESTATE)*, IEEE Std. 1232–2002, 2002.

[11] S. K. Lukins, N. A. Kraft, and L. H. Etzkorn, "Bug localization using latent dirichlet allocation," *Information. Software. Technology*. Volume. 52, number. 9,pp. 972–990, 2010