 ISSN 2349-6193
Impact Factor (PIF): 2.243

IJESMR
International Journal OF Engineering Sciences & Management Research

THE METHOD FOR UNION GRAPH GENERATION OF HETEROGENEOUS D-MATRIX

***Miss. Poonam Madhukar Jagdale**
*Department of Computer Engineering, JSPM's Imperial College of Engineering & Research, Wagholi, Pune, India.

Keywords: D-matrix, Data mining, fault detection and diagnosis, ontology , graph comparison algorithm, text mining**.**

ABSTRACT
To obtain the structure level issue analytic data between the detectable indications and failure modes, the dependency matrix (D-Matrix) for deficiencies is delivered. The fundamental first rule to make d-matrix by using the domain information is especially delayed undertaking. Later, in the wake of finding the new symptoms and disappointment modes for first time, mining it into the delivered d-matrix is a troublesome task. Along these lines, the text mining system centered around the ontology is created by mining number of unstructured dataset i.e. repair verbatim data assembled while issue of fault diagnosis system to create and update the d-matrix. In this methodology, ontology for fault conclusion is assembled first which identifies with the ideas and connections display in the determination domain. The content mining calculation is used as a part of next technique. In our strategy, we used the system to make d-matrix for two datasets independently. By then, the chart is delivered for every one made d-matrix. The compare between 2 D-matrixes is procured by connection the each graph for generating the ultimate graph.

INTRODUCTION
Finding the problem when a complex automotive structure happen is easy, if you've got the crucial data. We can divide this data in to two areas: comprehension of the framework inside of which the issue exists; and accordingly the capacity to apply a intelligent demonstrative schedule. The framework group up with it's as one with to execute an arrangement of assignments by keeping up its execution inside of a commendable extent of resistances. Any deviation of a structure from its exemplary execution is controlled as a fault. The fault detection and diagnosis (FDD) is performed to understand the deficiencies and diagnose the essential drivers to reduce the time period of a structure. Fault detection and diagnosis could be a key piece of varied operations organization mechanization systems. Fault detection sees that a problem went on, paying very little mind to the approach that you simply do not nevertheless recognize the essential main cause. Faults could likewise be seen by a mixed pack of quantitative or subjective implies that. Fault diagnosis is pinpointing one or also shrouded base of issues, to the point wherever remedial move are frequently made. This can be moreover recommended as "fault separation", particularly when focusing on the refinement from issue location. In like manner, accommodating utilization, "fault diagnosis" each currently and once more fuses deficiency distinguishing proof, thus "fault disconnection" underscores the refinement. The unusualness of automotive structures has created and in this way the related demonstrative limits ought to a great many. Because of constantly intending to be mechanical movement that is installed inside of the vehicle frameworks, for case pushed programming embedded systems [3], symptomatic sensors, and web, so on the rationality of FDD finds the opportunity to be a testing movement inside of the occasion of part or structure glitch. Obviously, once each conclusion scene the teachings learnt are preserved in an exceedingly number of databases to search out and diagnose the faults. This database contains bunches of data required for the analysis methodology like facet symptoms, parts, faults, failure mode, error codes, so forth.
The explanation behind Text Mining is to method unstructured (text based) information, separate vast numeric records from the content, and, on these lines, build the data contained within the substance open to the various data mining (quantifiable and machine learning) figuring's. Data may be differentiated to induce rundowns for the words contained within the reports or to figure diagrams for the records focused around the words contained in them. In [1], projected a text mining technique to guide the characteristic data removed from the unstructured repair verbatim in a very D-matrix [4]. In our projected framework, we tend to used information set file that contain unstructured repair verbatim data as an input dataset. The D-matrix is one amongst the standard demonstrative models decided in IEEE standard 1232 [7].After making d-matrix from [1], in our framework, we modify over this matrix into the graph i.e. an undirected graph. Complex system's broad thought has been for

ISSN 2349-6193
Impact Factor (PIF): 2.243

IJESMR
International Journal OF Engineering Sciences & Management Research

creating analytic systems in light of a selected showing perfect model-reliance displaying. Varied instruments diagram models into "D-matrix" (dependency system) and find symptomatic systems from the matrix [2].
In this paper we study about the related work done, in section II, the implementation details in section III where we see the system architecture, modules description, mathematical models, algorithms and experimental setup. In section IV we discuss about the results and at last we provide a conclusion in section V.

RELATED WORK

In the current arrangement of fault modeling [5], [6], [7], [8], the restricted attempts are seen to congregate a D-framework by dismembering unstructured repair verbatim data. Merely starting late [10] the device is recommended that finds the data by conveying material blueprints from the on-board finding and facilitates information by utilizing the logic based mostly data mining. notwithstanding, the crucial models proposed in are thought to be done and static; but in authentic due to outline and delineating changes and new vehicle basic masterminding dispatches the new appearances and failure modes are seen creating such models old style. In existing framework improvement of D-matrix is finished by physically or utilizing 1st customary. Generally, the D-frameworks are created by utilizing the history data, building data, and determinable data [5], [6], [7], [8], [9] for example, however a practically no comprehension is given regarding the revelation of novel responses and failures modes saw shockingly and their consolidation within the D-matrix models.

To defeat this issue, in [1] the fault information is gotten and formalized within the fault identification ontology, that is extended in light of the new information. The subsequent ontology based mostly substance mining figuring's that uses this data model supports in-time FD.
To make a D-lattice instructive model, principled logic is proposed by dissecting the unstructured repair verbatim data identified with the different structures in parallel through the progress of ontology based substance mining figuring's. It beats the suppression went up against inside of the veritable business of hoping to collect the D-matrix asking models physically or using beginning standards. Further, in our theory we've the point of confinement get the cross-system conditions that had any sort of effect to in an extremely broad sense enhance the execution of FDD. The relations from the fault diagnosing cosmology are used to find the conditions between the manifestations furthermore also the failure modes diverse with contrasting frameworks. It updated the execution of our framework once cut up with the Latent-Dirichlet Allocation (LDA) strategy.
The D-matrix are made by utilizing the history data, building information, and considerable data [7], [8], [9], [10], for instance, however a practically no comprehension is given regarding the revelation of latest symptoms and failure modes saw shockingly and their thought within the D-matrix models.

In the current [1], the d-grid is produced using the main dataset. Inside of this methodology, it can't be useful once there's a prerequisite of developing d-matrix from different datasets. This gets the opportunity to be incredibly dull assignment once the new dataset given for delivering d-matrix is perceived with the past made d-matrix.

IMPLEMENTATION DETAILS

A. System Overview

This paper proposes Graph Comparison Algorithm for An Ontology-Based Comprehensive D-Matrix. This system comprises the developments of D-matrix from the repair verbatim data. After the creation of the D-matrices from the two datasets, we generate the graph for each d-matrix. Then, the graphs are combined such that common patterns along with unique patterns are merged from the generated heterogeneous D-matrices to construct a single, generic D-matrix. To construct the D-matrix [1], following steps has to be created:

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

This step describes three steps, such as 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

IJESMR
International Journal OF Engineering Sciences & Management Research

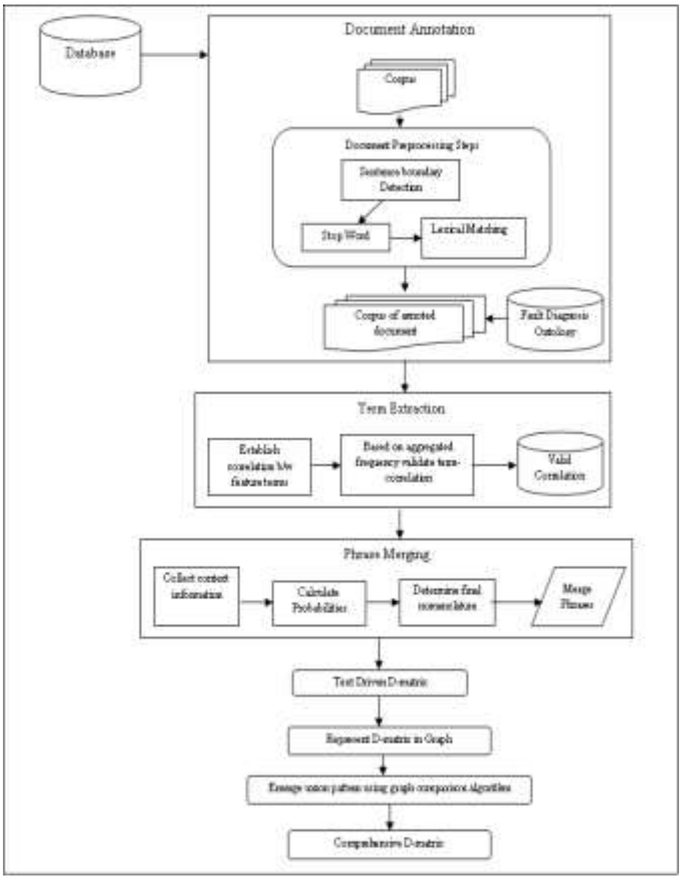

Fig.1 System Architecture

1. *Fault Diagnosis Ontology*

The fault detection ontology in particularly is a lightweight ontology, which is formalized by using the ontology development technology. It gets the terms and the relationship viewed in the zone of vehicle deficiency finding. At a coarse grained level, the fault diagnosis ontology has the shape,

$$\mathrm{FaultDignosis}_{\mathrm{onto}} = (C_i, C_{i\mathrm{subconcept}}, I_{ci}, R_{ci} \rightarrow C_{ji})$$

Where,

C_i =usually determined ideas within the FD domain, like System, Sub- system, Part, Fault, Symptom, failureCondition, Action, Cause, failureMode, problem Description, and parameter value.

$C_{i\mathrm{subconcept}}$ = More specific concepts are formalized by using the concept sub- concept hierarchy. For example, the concepts fault Codes and textual Symptom specializes the top-level concept Symptom, where faultCodes formalizes the class of symptoms identified by the faulty system parameter value and textualSymptom captures the textual problem description.

I_{ci} = captures the info values determined in planet applications by instantiating the ideas.

$R_{ci} \rightarrow C_{ji}$ = the parallel relations between any two ideas, say C_i and C_j determine how two ideas are identified with one another in our space.

2. *Ontology-Based Text Mining*

- Document Annotation: The sentence boundary detection (SBD), are required to split a repair verbatim into separate sentences, the stop words are erased to remove the non-descriptive terms, and the lexical matching identifies the correct meaning of abbreviations. Likewise the terms of the processed verbatim are compare using the fact in the fault diagnosis ontology.

a) Sentence Boundary Detection
Here, a repair verbatim is first split in various sentences with the help of the sentence boundary detection rules and the terms shows in the same sentence are co-related with each other.

 ISSN 2349-6193
Impact Factor (PIF): 2.243

IJESMR
International Journal OF Engineering Sciences & Management Research

b) Lexical Analysis
The probabilistic learning technique, lexical matching is proposed based on the encompassing contextual data. Beneath, we demonstrate how the abbreviations used to shows the part terms are disambiguated and the same procedure is used to disambiguate the symptom and failure mod abbreviations [1].

- Term Extraction: From each clarified repair verbatim the tuples, for instance, parts, manifestations, disappointments modes are worked by using the term extraction calculation to assess a D-grid. The weights are dispensed to each tuple using the going with mathematical statement and the tuples with their weights over the specific limit are considered as the countable individuals

$$(T_w)_{i,j} = T_{i,j} * idf_{Ti}$$

$$T_{i,j} = \frac{\eta_{i,j}}{\sum_k n_{kj}}$$

Where $\eta_{i,j}$ is the number of co-occurrences of a given tuple, Ti, that appears in a repair verbatim $T_j$and the denominator is the sum of number of co-occurrence of all tuples in T_j

$$idf_{T_i} = \frac{log|V|}{|\{v: T_i \in v\}|}$$

where,
$|V|$ is the total number of repair verbatim in a corpus,
$\{v: T_i \in v\}$ is the number of repair verbatim with T_i;

- Phrase Merging: The expression consolidating is utilized to discard far fetched references of the disappointment mode phrases, where the disappointments mode grants that are made by using a clashing vocabulary. The savvy information co-happening with the announcements, i.e., parts, side effects, disillusionment mode, and activities is utilized to gage the surprising probabilities and the illustrations with their probability score over the specific edge are joined.

3. Text driven D-Matrix

The text-driven D-matrix approach achieved higher fault detection, higher fault isolation, and lower ambiguity group size due to textual symptoms (in addition to DTC symptoms) and the corresponding failure modes included in the text-driven D-matrix.

4. Graph Comparison Algorithm

In our system, we used the graph merging algorithm which takes the generated d-matrix from [1] as input. The same procedure [1] for constructing the D-matrix is done in our proposed system i.e. Graph Comparison Algorithm for An Ontology-Based Comprehensive D-Matrix., for two times for two repair verbatim data. Now, the graph is deployed by using the d-matrix. The columns and rows from the d-matrix are treated as the vertex for the graphs. As the d-matrix shows the dependencies in the binary format, the edges for the graphs is decided by using this binary information i.e. if 1 is the output in d-matrix for specific column and row, then, the edge is formed between that column and row. In this manner, the graph is formed from the developed d-matrix. Next, our system compares two graphs and merged the common details along with the unique content appearing in both graph. For this, the system used the graph merging algorithm.

B. Algorithm

Pseudo code of proposed system is
Step 1: D-matrix1 from datset1
Step 2: D-matrix2 from dataset2
Step 3: Graph1 from d-matrix1
Step 4: Graph2 from d-matrix2
Step 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
Step 6: Union graph generation i.e. graph3

 ISSN 2349-6193
Impact Factor (PIF): 2.243

IJESMR
International Journal OF Engineering Sciences & Management Research

Step 7: Heterogeneous d-matrix from graph3

C. **Mathematical Model**

System S is represented as, S = {D, M, G, C, H}

- Database
  D= {d1, d2,……, dn}
  Where, D is the set of datasets which contains the repair verbatim data and d1, d2,……, dn are the number of dataset.
- D-Matrices
  M= {m1,m2,….mn}
  Where, M is the set of d-matrices generated from the datasets D and m1,m2,….mn represents the number of d-matrix.
- Graph Generation
  G= {V, E}
  Where, G is the set of graphs generated from the d-matrices.
  V= {v1, v2,…….Vn}
  Where, V represents the list of corresponding vertices from the d-matrices and v1, v2,……. Vn is the number of vertices. Where the columns and row from the d-matrix are managed as the vertex for the graphs.

  E= {e1, e2,….., en}
  Where, E shows the set of edges depending on the data in the d-matrix and e1, e2,….., en represents the number of edges. The edges for the graph is picked by using this binary data i.e. if 1 is the yield in d-framework for specific row and column, then, the edge is organized between that row and column.
- Graph with union patterns
  C = {X, Y}
  Where, C represents the graph of union patterns by merging the graphs.
  X= { x1, x2,……, xn}
  Where, X represents the set of union vertices from the generated graphs and x1, x2,……, xn represents the number of vertices.
  Y= { y1, y2,……., yn}
  Where, Y is the set of union edges from the generated graphs and y1, y2,……., yn represents the number of edges.
- Final D-matrix
  H={C}
  Where, H represents the final d-matrix generated from merging the graphs.

D. **Experimental Setup**

The system is built using Java framework(version jdk 6)on Windows platform. The Netbeans (version 6.9) is used as a development tool. The system doesn't require any specific hardware to run; any standard machine is capable of running the application.

RESULTS AND DISCUSSION

A. Dataset

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

B. Result

- *Fault Detection*

IJESMR
International Journal OF Engineering Sciences & Management Research

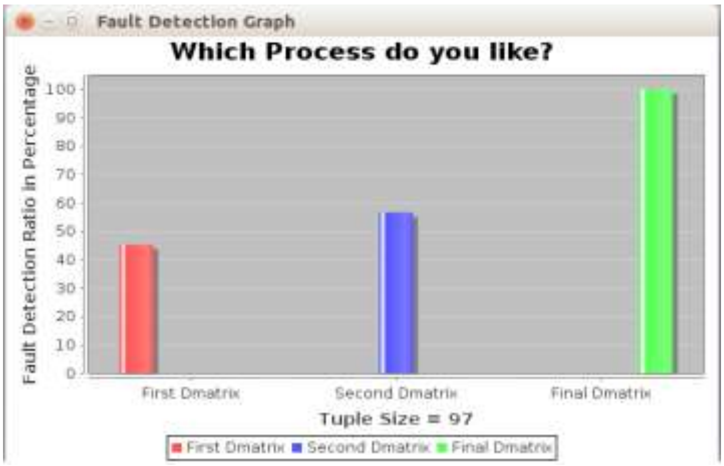

Fig 2. Comparison of Fault Detection between D-matrix1, D-matrix2 vs. Final D-matrix

The fig 2 shows the Comparison of Fault detection between D-matrix1, D-matrix2 vs. Final D-matrix. The proposed system shows that final D-matrix have high fault detection ratio than the D-matrix1 and D-matrix2.

- *Fault Isolation*

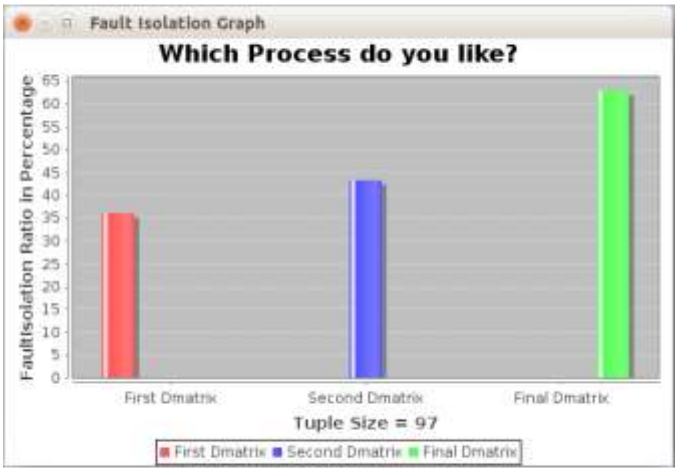

Fig 3. Comparison of Fault Isolation between D-matrix1, D-matrix2 vs. final D-matrix

The fig 3 shows the Comparison of Fault isolation between D-matrix1, D-matrix2 vs. final D-matrix. The proposed system shows that final D-matrix have high fault isolation ratio than the D-matrix1 and D-matrix2.

- *Disambiguity Graph*

IJESMR
International Journal OF Engineering Sciences & Management Research

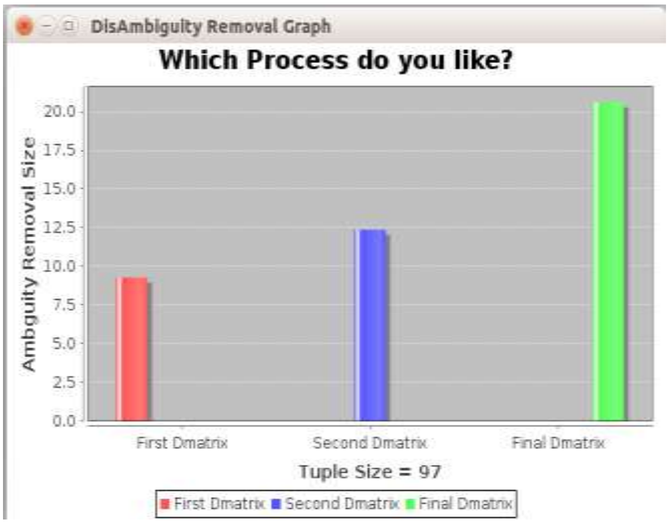

Fig 4. Comparison of disambiguity between D-matrix1, D-matrix2 vs. Final Dmatrix

The fig 4 shows the Comparison of disambiguity between D-matrix1, D-matrix2 vs. Final D-matrix. The proposed system shows that Final D-matrix have high disambiguity ratio than the D-matrix1 and D-matrix2.

CONCLUSION AND FUTURE SCOPE

This framework considers the pertinence of making union d-matrix from different d-matrices. Framework makes the d-matrix from two unstructured repair verbatim data like way in [1] by using content mining algorithms. System structure, the undirected graphs are produced for two d-matrix which is generated from the unstructured repair verbatim data. The graph examination estimation is utilized to make d-matrix like ordinary cases ascending out of the heterogeneous d-matrices which can be utilized to develop single, extensive D-matrix. The recently developed d-matrix that is inclusive D-matrix is additionally compared with D-matrix1 and D-matrix2 with parameter of fault detection. Fault isolation and Disambiguate. The proposed Inclusive D-matrix has better execution in every one of the three parameter than the - matrix1 and D-matrix2.This application is used in automobile industry and if ontology accessible for different fields then is likewise utilized for different fields for investigating the one of a kind patterns in text data instead of referring the content.

ACKNOWLEDGMENT

The authors would like to thank the researchers as well as publishers for making their resources available and teachers for their guidance. We also thank the college authority for providing the required infrastructure and support. Finally, we would like to extend a heartfelt gratitude to friends and family members.

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. W. SHEPPARD AND S. G. W. BUTCHER, "A Formal Analysis of Fault Diagnosis with D-matrices", The Johns Hopkins University, Baltimore, MD 21218, USA.
[3] 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.
[4] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S. Kavuri, "A review of process fault detection and diagnosis Part III: Process history based methods," *Comput. Chem. Eng*., vol. 27, no. 3, pp. 327–346, 2003.
[5] 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* ., 2008, pp. 1–6.
[6] J. Sheppard, M. Kaufman, and T. Wilmering, "Model based standard for diagnostic and maintenance information integration," in *Proc. IEEE.*
[7] M. Schuh, J. Sheppard, S. Strasser, R. Angryk, and C. Izurieta, "Ontology-guided knowledge discovery of event sequences in maintenance data," in *Proc. IEEE AUTOTESTCON Conf* ., 2011, pp. 279–285.

IJESMR
International Journal OF Engineering Sciences & Management Research

[8] 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. 1–12.
[9] 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.
[10] R. Studer, V. R. Benjamins, and D. Fensel, "Knowledge engineering: Principles and methods," *J. Data Knowl. Eng.*., vol. 25, nos. 1–2, pp. 161–197, 1998.
[11] 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.
[12] S. Singh, H. S. Subramania, and C. Pinion, "Data-driven framework for detecting anomalies in field failure Data," in *Proc. IEEE Aerosp. Conf.*, 2011, pp. 1–14.