Graphical Analysis on Text Mining Unstructured Data Using D-Matrix

Sneha P. Mendhe [1]
Dept. of CSE PBCOE, Nagpur.
Nagpur, India
Snehapmendhe28@rediffmail.com

Kapil N. Hande [2]
Asst. Prof Dept. of CSE PBCOE, Nagpur.
Nagpur, India
kapilhande@gmail.com

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.

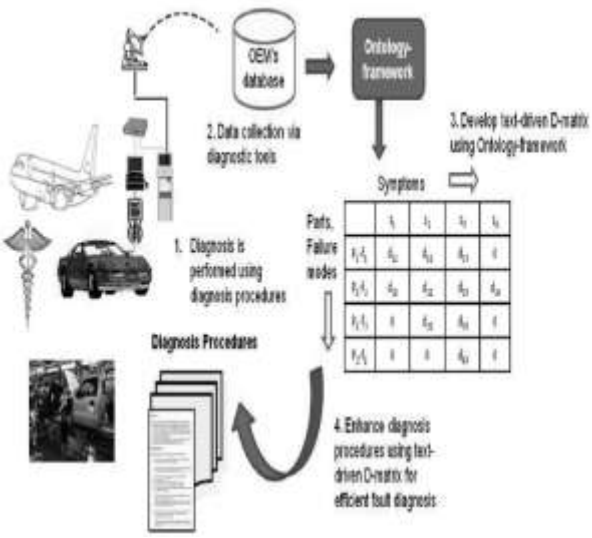


Figure 1: Scope of the ontology-based text mining framework.

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 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 is 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.

II. OBJECTIVES

The fault detection and diagnosis (FDD) is performing to detect the faults and diagnose the root-causes to minimize the downtime of a system. However, the huge size of the repair verbatim data restricts an ability of its effective utilization in the process of FDD.

After applying the data mining techniques, one can able to identify patterns in the facts or numbers. This may lead to new discoveries or research area about the types of fault that can or cannot occur, or relationship between types of fault and particular fault, and so on.Time Factor is play a key role while searching data from database, so to retrieve fast data access (CAR, TRUCK, BUS, and BIKE) only selected vehicle can be searched rest is not considered.

III. METHODOLOGY

Firstly, the repair unstructured data points are collected by retrieving them from the raw database, which are recorded during field Fault Diagnosis. In the first module, the terms, such as part, symptom, and failure mode, relevant for the D-matrix are annotated from each repair verbatim by developing the document annotation algorithm.

A repair verbatim data consists of several parts, symptoms, failure modes and actions and the correct relation must be established between the relevant terms based on their proximity with each other. Here, a repair verbatim is first split in different sentences by using the sentence boundary detection rules and the terms appearing in the same sentence are co-related with each other. Frequency of fault occurrences calculation that gives the numeric value, on the basis of probability D-matrix by Bayes theorem. Initially, the causal relationship between the relevant symptom-failure mode pairs is identified to make sure that only the correct pairs are extracted.

The annotated terms are extracted in the following combinations (Pi FMj) and [Sj (Pi - FMk)],where by starting from the left side and moving toward the end of a repair verbatim the position of a first part term, say Pi appearing in a repair verbatim is identified and the word window of four terms is applied on the either side of Pi.

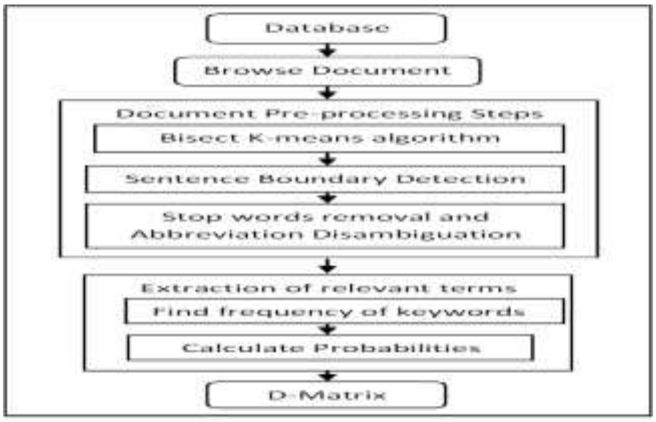


Fig.2: Flow diagram of proposed work

Due to the human error while capturing the repair verbatim data an inconsistency has been observed in the data in terms of the nomenclature used to record the failure modes, e.g., Tank Sensor−Inop or Tank Pressure Sensor− Internal Short. Hence, we check whether two different failure modes are the variations of essentially the same failure mode such that they can be merged before populating a D-matrix. The phrase merging algorithm takes as input a set of failure modes extracted by the term extractor algorithm.

D1 = {D1,D2, D3, ...Dn }

Where D is represent as a set of merged phrases and D1, D2, D3, ...Dn is a number of merged phrases. Above Fig 2 shows the methodology contains two main module namely document pre-processing steps and extraction of relevant terms with probability calculations involved in ontological text mining method.

IV. EXPERIMENTAL RESULT

The D-Matrix framework is created using proposed methodology. The real-life vehicle data is collected from the automobile service industry. A text driven D-Matrix is created using the symptoms shown in column and failure mode in rows. After execution of the first module following result shows the particular repair verbatim splits and the separated using stop word removal module from the sentences that will help out the find the term extraction numeric values data

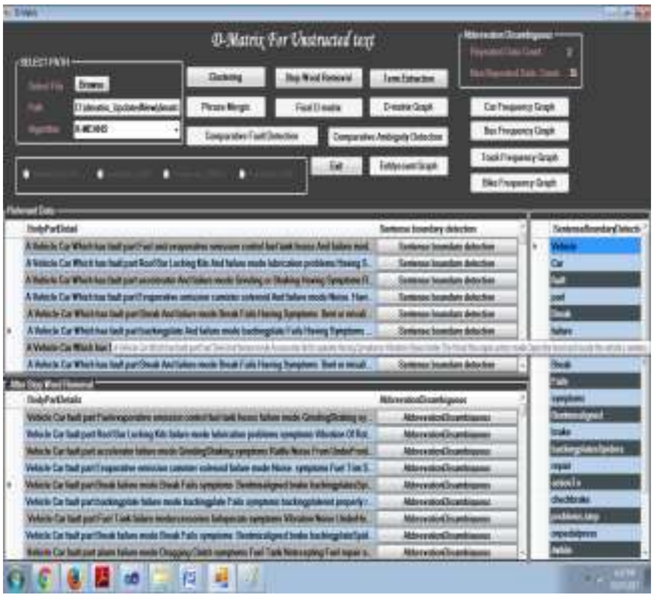


Fig. 3: After clustering and Stop Word Removal

For getting numeric values the term extractor is used that gives particular fault, symptoms and number of repair action count gives more accurate data for analysis. Figure 4 is the frequency count for the CAR having fault part and their frequency of failure mode.

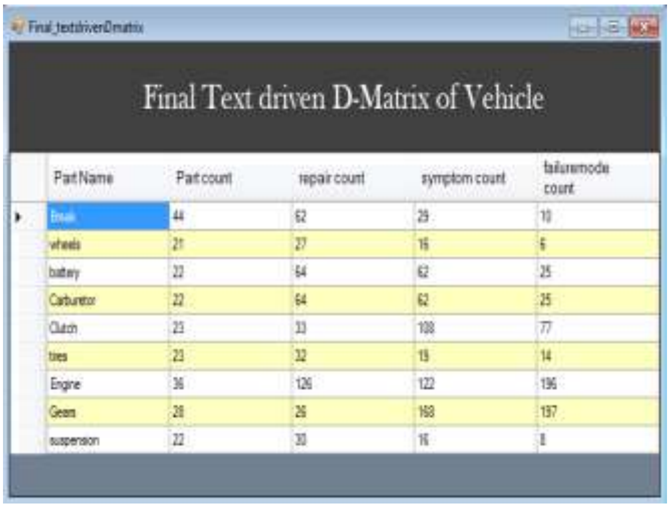
 ISSN: 2321-8169
1412 – 1416

	Fault Part	Count of frequency of failure mode
▸	backing plate	4
	Anti-lock braking	9
	Bumper1	8
	Alternator bracket	4
	Headlight moto	2
	control system	3

Fig. 4: Frequency count of failure mode

Figure -5 final text driven D-matrix gives the selected attribute part name and their respective part count, repair count, symptoms and failure mode count for exact interpretation.

Figure 6 shows the representation of D-matrix into Graph which consist of car components. This graph represents the fault parts and their frequency of occurrences. Due to limited size, it is not possible to show every field in D-matrix therefore all the fields are shown in graph.



	Part Name	Part count	repair count	symptom count	failuremode count
▸	Break	44	62	29	10
	wheels	21	27	16	6
	battery	22	64	62	25
	Carburetor	22	64	62	25
	Clutch	23	33	108	77
	tires	23	32	19	14
	Engine	36	126	122	196
	Gears	28	26	168	197
	suspension	22	30	16	8

Fig. 5: Text Driven D-Matrix

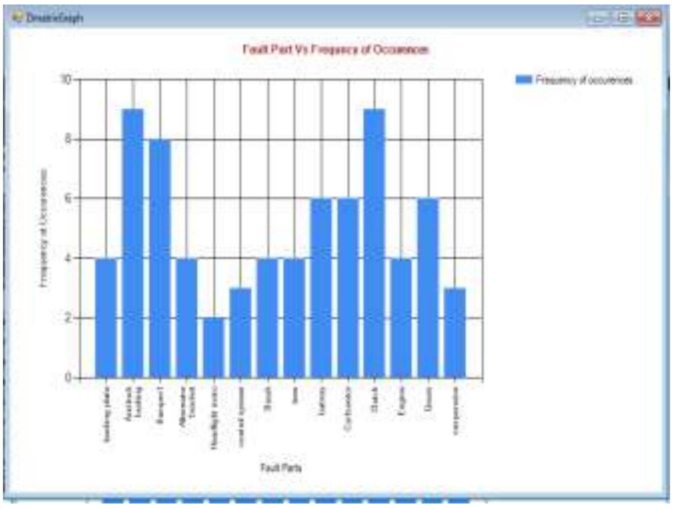


Fig. 6: D-Matrix Graph

Figure 7shows the comparison of the previous technique with the proposed technique in terms of fault detection. The fault detection (FD) is given as the number of faults detected by the symptoms by observing the failure modes associated with a system. It was used to evaluate the fault coverage. Blue colour shows the historical data and yellow colour recent retrieved data. Also figure 8 shows the comparative ambiguity size of D-Matrix with previous and recent approach.

Finally the all vehicles fault part frequency entity count graph shows by figure 9.

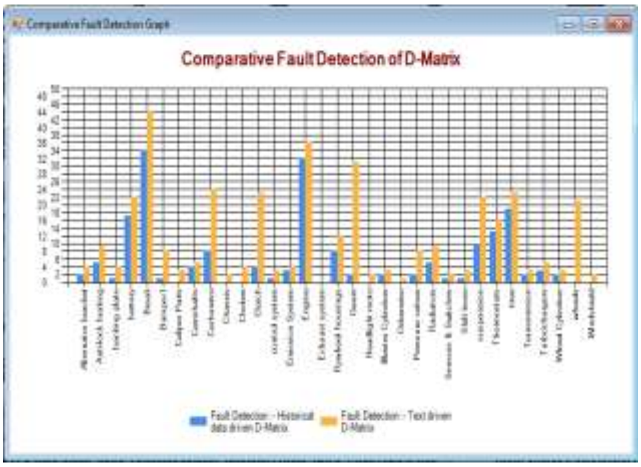


Fig. 7: Comparative Fault detection of D-Matrix Graph

 ISSN: 2321-8169
1412 – 1416

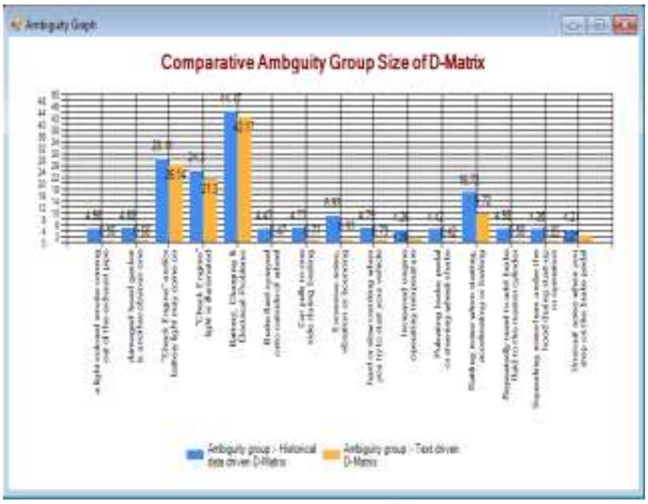


Fig. 8: Comparative Ambiguity Size Of Dmatrix

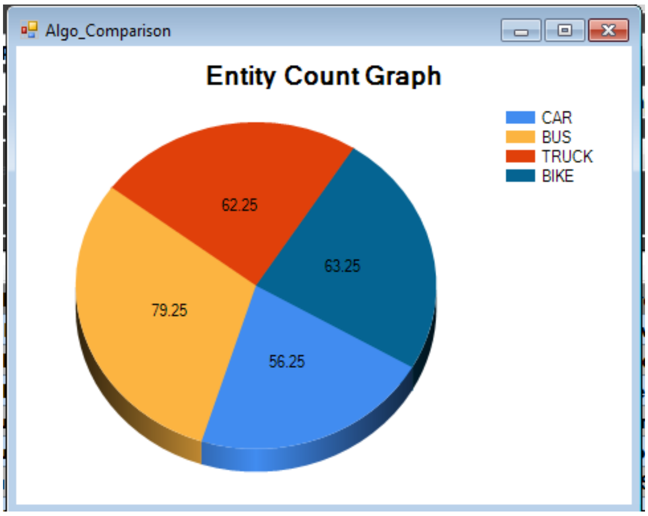


Fig. 9: Vehicle Failure Part Count Entity

V. 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 of 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.

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., 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.

[16] 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.

[17] N. Viswanadham and R. Srichander, “Fault detection using unknown input observers”, Control-Theory Ad. Tech., vol. 3, pp. 91–101, 1987

[18] 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.

[19] A. S. Willsky, “A survey of design methods for fault detection in dynamic systems”, Automatica, vol. 12, no. 6, pp. 601–611, 1976.

[20] 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.

[21] C. Charniak and D. McDermott, Introduction to Artificial Intelligence. Reading, MA, USA: Addison Wesley, 1985.

[22] 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.

[23] 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.