

A Decision Support System based on Ontology and Data mining to Improve Design using Warranty Data

Abstract

Analysis of warranty based big data has gained considerable attention due to its potential for improving the quality of products whilst minimizing warranty costs. Similarly, customer feedback information and warranty claims, which are commonly stored in warranty databases might be analyzed to improve quality and reliability and reduce costs in areas, including product development processes, advanced product design, and manufacturing. However, three challenges exist, firstly to accurately identify manufacturing faults from these multiple sources of heterogeneous textual data. Secondly, accurately mapping the identified manufacturing faults with the appropriate design information and thirdly, using these mappings to simultaneously optimize costs, design parameters and tolerances. This paper proposes a Decision Support System (DSS) based on novel integrated stepwise methodologies including ontology-based text mining, self-organizing maps, reliability and cost optimization for identifying manufacturing faults, mapping them to design information and finally optimizing design parameters for maximum reliability and minimum cost respectively. The DSS analyses warranty databases, which collect the warranty failure information from the customers in a textual format. To extract the hidden knowledge from this, an ontology-based text mining based approach is adopted. A data mining based approach using Self Organizing Maps (SOM) has been proposed to draw information from the warranty database and to relate it to the manufacturing data. The clusters obtained using SOM are analyzed to identify the critical regions, i.e., sections of the map where maximum defects occur. Finally, to facilitate the correct implementation of design parameter changes, the frequency and type of defects analyzed from warranty data are used to identify areas where improvements have resulted in the greatest reliability for the lowest cost.

Keywords: Ontology, Self-Organizing Maps, Warranty data, text mining, decision support

1. Introduction

Customer-oriented approaches provide opportunities for manufacturing to address the challenges of global competition and produce high-quality, highly reliable products with low lead time and cost. Automotive companies tend to offer services on explicit vehicle sub-systems and provide prolonged warranty coverage for additional costs (Yang et al. 2004). It has been observed that the auto manufacturers spend a substantial amount of their sales returns in repairing the defects that occur during the warranty period. Automobile manufacturing companies spend a substantial percentage of their sales revenue (2.5% to 3.0%) in fixing vehicles under the warranty period (Sureka et al. 2008). The auto industry

spends around 45 billion dollars to 50 billion dollars a year on recalls and warranty claims (Nelson 2013). General Motors and Ford together spent around 8.6 billion dollars on warranty claims for capturing the bulk of the market share (Sureka et al. 2008). Moreover, they have invested billions of dollars in process engineering and product design changes to reduce this warranty burden. The automotive industry continually seeks new and innovative ways to reduce their costs and improve their profit margin (Vernick et al., 2003). It has been shown that providing better service, gains customer confidence in products, leading to increased revenue for the company (Khare et al. 2012). It is therefore essential that fault diagnosis is conducted appropriately by collecting the necessary information from the customers and insights need to be derived from collected data to improve the design of the products.

Customers' satisfaction and feedback play a crucial role in marketing the product and prevent existing problems from recurring. They also assist in identifying the No Fault Found (NFF), phenomenon in failure analysis of the product failure data which describes the event when the defect cannot be identified and repaired at the service station by the technician. Hence, warranty data analysis is important to provide useful information for design activities, evaluate design changes, compare experiences of different groups of products and study the variations in manufacturing and usage environments. It is essential for automobile manufacturers to manage customer satisfaction to stay in the global market and continuously enhance the product reliability depending upon the customer's perception of product quality. For customer satisfaction, the product design needs to be improved by focusing on the weaknesses of the product, which can be identified by analysing the data acquired from the warranty database and customer's feedback about the product performance. Customers visit service centres for product repairs and the product defects will be examined. In this way, plenty of data is collected which can be a useful source for failure analysis of the products (Kang et al., 2017). Managing the volume and variety of customer feedback or warranty information is difficult, but valuable in identifying the range and varieties of defects. Therefore, improving the ability of the knowledge-based systems to derive information from customer feedback systems will assist in the development of innovative products and will also help in improving the existing products (Kang et al., 2017). Understand the connections between warranty data and design and manufacturing are essential as the warranty data can identify opportunities for quality and reliability improvement. Knowledge of failing automobile parts, the occurrence of the failures, the cost associated with the failures and the

impact of the product failure on customer satisfaction is important. Thus, manufacturers need to address the challenges of effectively utilising heterogeneous customer feedback and warranty data to resolve recurring issues associated with the product defects at design and manufacturing phases.

The volume and variety of customer feedback information are increasing. Furthermore, warranty information plays a vital role in determining the reliability of the product as it provides a contractual obligation between the manufacturer and the consumer and aims to protect the consumers against any failure of the product in the future. The organisation is liable to pay for all the charges incurred in repairing the defect within the warranty period, so cost is a major incentive for reducing defects. Warranty data is a strong indicator of in-field product performance and reliability, and it also helps in determining the post-production cost incurred by the manufacturer, since warranty cost accounts for the major portion of these it therefore has significant impact on the total profit in the manufacturing business.

The following three challenges are associated with analysing customer feedback information available in the warranty database.

- To accurately identify manufacturing faults in the increased volume and variety of textual data sources coming from geographically heterogeneous service centres and customer feedback information, while taking into account the ontological differences.
- To capture the nonlinear relationship of high dimensional data and visualise the defect feature clusters on a low dimensional display. Furthermore, manufacturing faults identified should be accurately mapped with the design errors like defects in parameters' tolerance values for the components.
- To simultaneously optimise cost and reliability for design parameters' tolerances for components taking into account the abovementioned information.

A Decision Support System (DSS), is proposed which collects the warranty failure information from the customers' feedback in textual forms such as customer emails, customer feedback surveys, customer reviews of the product on company websites etc. DSS primarily supports service centres and helps in identifying the critical component(s) leading to the majority of the faults and provides this information to the manufacturing facility. An ontology-based text mining approach is adopted to extract hidden knowledge existing in the customer warranty and feedback databases. Furthermore, a data mining approach using Self Organizing Maps (SOM) has been proposed to draw information from the warranty database

and link it to the manufacturing data. The defects obtained can be linked to two causes – (a) Faulty parameter setting (b) Tolerance error. Statistical analysis is conducted to re-evaluate the parameter values or the tolerances depending on the nature of the results obtained. To implement these changes in practice, a cost analysis is performed to determine the cost incurred for making the design changes.

The rest of the paper is organised as follows: Section 2 presents an extensive literature review in the relevant areas to uncover the current limitations and existing research gaps; Section 3 presents a framework for the decision support system and explains its various components. Furthermore, it also discusses the ontology development and text mining approach used in this research; the SOM and how it is applied in the present context is discussed in Section 4; Section 5 presents a numerical illustration along with a cost optimisation model; Section 6 concludes this research with a note about future research.

2. Literature Review

Warranty or field performance data provide crucial information related to the quality and robustness of the product as perceived by the customers. Analyzing the warranty data offers valuable insights regarding possible design changes to improve product reliability. Warranty data analysis is considered as an important step in Product Lifecycle Management, and as a result, there is a plethora of research on warranty analysis (Luo et al., (2018), Liao (2016), He et al., (2016)). Current literature on warranty data investigation aims to predict the warranty failure and reliability analysis (Chien et al., 2007), examining the warranty policies and cost associated with it (Luo et al. 2018, Liao 2016) and the early prediction of warranty failures by considering the analysis of failure rates. The analysis of present research indicates a need to develop an adequate methodology for efficiently analyzing the warranty data and preventing any product defect in the future. **The first subsection is dedicated to the relevant research in the domain of fault diagnosis and methodologies employed. The research work associated with the analysis of warranty data and the need to examine product failure data is described in the second sub-section.**

2.1. Fault Diagnosis and Methodologies

Hui et al. (2000) proposed a customer service support system by considering a data mining approach, although they did not consider an ontological approach. Ajayi et al., (2008) designed a framework for improvement of the design of the product using the information available at the service centers. Further research is being carried out in developing text mining methods like text tagging and query processing. Surekha et al., (2008) developed an upper manufacturing ontology to draft a common semantic net in the manufacturing domain. The majority of the published research deals with the development related to analysing the textual data present in the customer feedback form or the warranty database. There is very limited research, which primarily focuses on identifying the fault region and the reevaluation of the tolerances using a warranty database. Mannar et al., (2006) proposed a Fault Region Localization (FRL) methodology to relate warranty failure to defect/process parameters using rough set-based analysis. This analysis was only applicable to categorical variables and overlooked the necessity of incorporating data mining approaches for a more accurate prediction. Söderholm (2007) studied the No Fault Found phenomenon and identified the steps for maintainability and design improvement to prevent the product failures. Several research works have been reported in the literature using text mining techniques for analysing the data obtained from customers in regard to product defects (Buddhakulsomsiri et al., (2006), Rajpathak (2013)). Buddhakulsomsiri et al., (2006) developed an association rule generation methodology to mine automotive warranty data.

Some early researchers considered the cost issues related to warranty failure and subsequent quality improvements of the product by considering recommendations obtained from failure diagnosis. The majority of these earlier works overlook the need to identify the fault region of the product by analysing the tolerance levels obtained from the warranty databases and providing appropriate recommendations for product design changes.

2.2. Analysis of Warranty Data

Several researchers have focused on addressing superior warranty policies while dealing with the product failure with the aim of enhancing customer satisfaction. Thomas et al., (1999) analysed warranty costs and articulated the warranty policy for the customer considering the coverage period and remuneration to the customer. Yeh et al., (2007) investigated the effects of a free repair warranty on the periodic replacement policy for a repairable product. Pham et al., (1999) developed a cost model with warranty and risk costs for software systems. Blischke et al., (1992) identified a variety of costs related to the failure

of a product during the warranty period. The majority of the publications such as Bouguerra et al., (2012), Pham et al., (1999), Blischke et al., (1992) have discussed ways to adopt newer warranty policies for attracting more customers without focusing much on the failure diagnosis for improving the product design aspect. Some of the earlier papers addressed the emphasis of warranty policies in mitigating the cost components of the manufacturers and overlooked the need to extract valuable insights while analysing the product failure and employing it in the design purpose. Bouguerra et al., (2012) developed a mathematical model to study the opportunity provided by extending the warranty to the consumer from a manufacturer's perspective. **Although a limitation of the work is that it does not link analysis of large-scale consumer data related to product defects with devising warranty policies.** Some research work, such as Park et al., (2017) addresses the need to reduce the maintenance cost of the defective products during the warranty period and to devise new warranty policies based on the failure time and repair time of the product. **Most early research does not examine large and varied forms of warranty failure information in order to improve the product design in the manufacturing process. Earlier works dealt with improving the warranty policies while examining the product failure data.** For the automotive domain, Rajpathak (2013) developed an ontology-based text mining system for retrieving knowledge from diagnosis data. The warranty data are collected to identify the best practices while repairing the product related to the automobile domain. Employing advanced data mining techniques could have been a possible scope of extension for the work. **It is therefore important to incorporate data mining techniques to reveal insights from the warranty data to improve the further design of the product in the manufacturing process.**

2.3. Research Gaps and Current Limitations

A review of the needs and importance of warranty databases shows that companies are conscious of the potential benefits although they are not taking full advantage of the opportunities offered to learn from them. Several researchers determined that the majority of organisations do not fully utilise the information obtained from the warranty database due to a lack of awareness about the availability of critical knowledge which might be used to subsequently improve the design process (Park et al., 2017), (Bouguerra et al., 2012). The warranty database is an enormous potential source of knowledge that could assist decision making in product design and manufacturing processes, but it is not routinely examined to identify reasons associated with recurring problems or behavioural patterns in product defects which may relate to manufacturing or design. Warranty and customer feedback generate

substantial amounts of data which are routinely stored in databases and investigation of these could provide learning and ensure valuable enhancements in product design processes of the organisation.

This paper aims to fill the research gap by employing ontology-based text mining and SOM based data mining to retrieve valuable information from the warranty database and relate it with the manufacturing data. Moreover, SOM is robust as it employs competitive learning rather than error-correction learning and it preserves the topological properties of the input space by using a neighbourhood function.

The next section of the paper will discuss the overall framework of the system and the application of knowledge discovery in text and text mining in analysing the warranty database for extracting valuable information.

3. A framework for the Decision Support System

A conceptual framework for the development of the proposed decision support system is shown in figure 1. The service centers of automobile manufacturers collect data from customers pertaining to product defects and different issues of the product faced by the customer. The data is stored in the customer feedback database and this provides input to the decision support system. An ontology has been developed to facilitate the process of knowledge search and knowledge discovery within the domain of the automobile manufacturing. To satisfy the requirements of exploratory Text Mining experiments, it is essential to identify common terms indicating the type of knowledge likely to exist in the customer database. An ontology has therefore been built, comprised of the key manufacturing components and automobile parts, in order to link the defects to the manufacturing design and derive knowledge from the existing customer feedback database. In this case the MASON (Manufacturing Semantic Ontology) ontology is used to capture manufacturing and design information and the ontology was developed using Protégé software. The ontology will help in locating the region of defect which needs to be further examined to identify the specific defective component. A brief overview of knowledge discovery in text and text mining process is presented in the section 3.1, following a similar approach to previous published research in Choudhary *et al.*, (2009a), Carrillo *et al.*, (2011) and Choudhary *et al.* (2009b), Shukla *et al.* (2016) and Choudhary *et al.*, (2018).

The application of Self Organizing Maps (SOM) is proposed for detecting the specific defective component, which may help in identifying the new tolerance values for preventing the occurrence of those defects in future. Applications of SOM to map failure modes and effects analysis are well documented in the works of Chang et al., (2017). Self-Organizing Maps (SOM) is a popular method used in recent research. Several authors have employed this technique in the domain of efficient condition monitoring and fault diagnosis of manufacturing machines (Li et al., (2018). Yu et al., (2015) performed a risk-based fault detection using Self Organizing Maps (SOM) as SOM is able to handle non-linear variations of the system. The application of SOM in monitoring facilities and determining the risk of faults has become very popular with several researchers (Malek et al., (2018). SOM is a data mining technique for knowledge discovery in databases (KDD), generally applied for discovering relevant knowledge, such as patterns, associations, changes, anomalies and significant structures from large amounts of data stored in databases, data warehouses, or other information repositories. SOM has been selected for its robustness, and has been employed here to recognize the defective component and provide new tolerance values, thus preventing further occurrences of such defects. A detailed discussion of SOM is presented in section 4. Furthermore, cost analysis is conducted to determine the feasibility of recommendations of tolerance revaluation by considering the cost and reliability factors. The manufacturing setup change cost and variable costs have been considered. Thus, cost feasibility analysis along with the identification of the region of defect provides a comprehensive framework as shown in figure 1, which may be adopted by automobile industry to prevent recurring design and manufacturing faults.

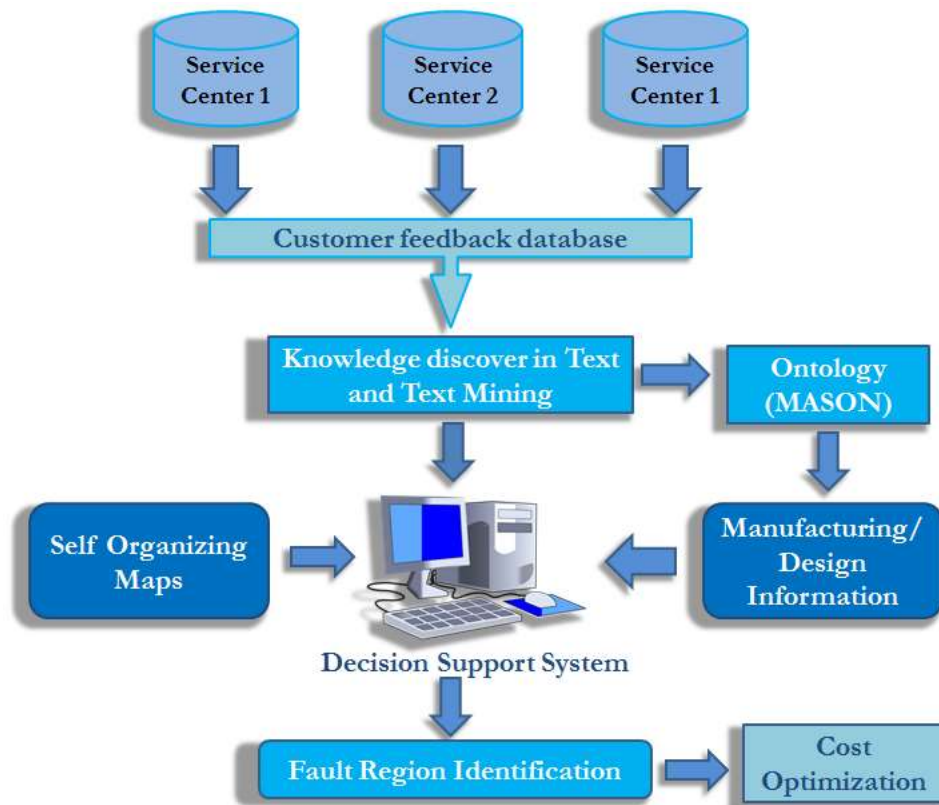


Figure 1: Components of the Decision Support System

The following discusses the details of individual component of the conceptual framework for decision support system and their applicability in the present context.

3.1. Knowledge Discovery in Text and Text Mining

Knowledge Discovery in Text can be defined as “the nontrivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in unstructured data” (Fayyad et al., (1996). It involves various information extractions, clustering categorization, data mining techniques and knowledge management. Knowledge discovery in text refers to the process of conversion of unstructured or semi-structured text data in to knowledge or high-level information. On the other hand, Text mining can be used to extract the relevant information from the warranty database by the combination of natural language processing and the data mining techniques. This may help in linking the functional failures identified in the warranty database to the design parameters which are responsible for the defects. Text Mining uses the semi-structured or unstructured data for extracting information and further investigating it for discovering implicit or hidden meaning within the text. Knowledge Discovery in Text is comprised of three steps which are discussed as follows,

1. Document Collection: Includes identification and collection of documents for analyzing purposes such as customer emails, customer feedback on product website, customer reviews, twitter feeds about the product etc.
2. Retrieving and Pre-processing documents: Retrieved documents need to be transformed into a form appropriate for employing text mining methods. Pre-processing style and text mining technique to be applied will vary for different kinds of documents. Initially, the unwanted texts are removed to reduce the size of the text and then the document is converted into XML or other format. The processed document is then used for retrieving basic linguistic information pertaining to its content.
3. Text Mining: Various algorithms and techniques are used in this step for extracting metadata or high-level information for obtaining behavioral patterns between the extracted information. Knowledge obtained using the text mining techniques can be considered by the decision makers to make improvements in the design processes.

Following the collection of documents in the form of customer feedback, customer emails etc., the textual data is transformed into a suitable format. Pre-processing of the textual data is carried out after loading the data into the system and pre-processing is generally implemented to reduce the excessive information and generate metadata. In text mining, the non-informative words and unwanted texts are removed to reduce the redundancy in computation and avoid information overload. After pre-processing, text mining is performed to extract patterns and valuable knowledge. In order to execute Text Mining successfully, it is imperative to build an ontology of common terms observed in the warranty database and ontology development is discussed in section 3.2. Figure 2 provides the linkage between domain expertise and text mining expertise and the process of knowledge discovery in text. Information obtained by text mining is used to make minor changes in the current product design process to improve customer satisfaction. Insights obtained from the text mining process may be used in the future decision-making processes.

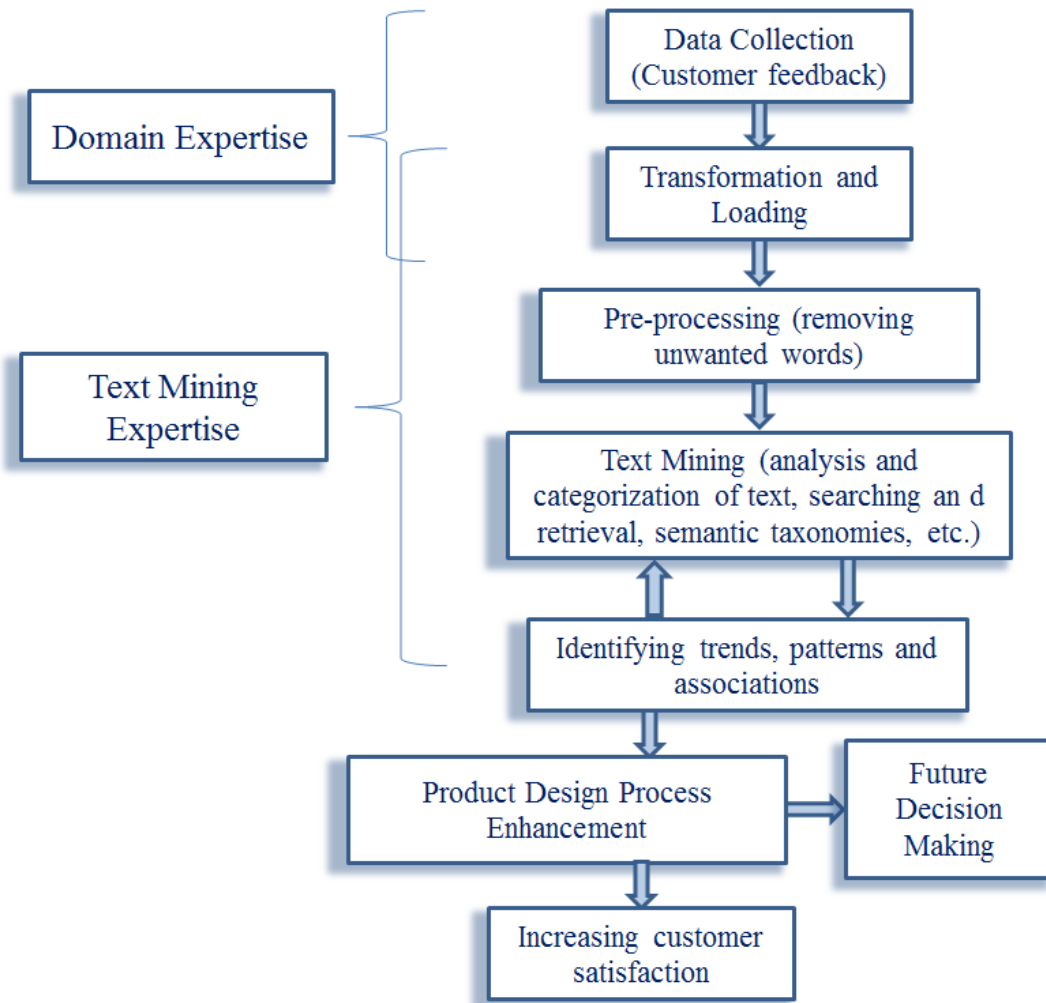


Figure 2: Connection between domain expertise and Text mining expertise and the KDT process

3.2. Ontology Development

Ontologies are semantic tools that help in representing the concepts behind the text in a formalized manner and further identifying the taxonomic and semantic relationships that exist between the concepts. Ontology was defined by Gruber in 1995 as both conceptualization of the domain and specification of this conceptualization. The MASON (Manufacturing Semantic Ontology) ontology has been used to represent knowledge for the manufacturing of automotive components. The Web Ontology Language (OWL) has been used to represent the ontology as it offers a complete framework for describing ontologies. OWL supports the following elements Classes, Taxonomic relations, Datatype properties, Objects properties, Individuals and Restrictions where classes show the concepts of the domain and relationships are defined between the classes and further properties of the classes (Bechhofer 2009). The protégé environment developed at Stanford University was used to

develop the MASON ontology and Graphviz software was used to represent the relationships between the classes in the graphical form. The screenshots in figures 3, 4 and 5 show the various classes and subclasses designed.

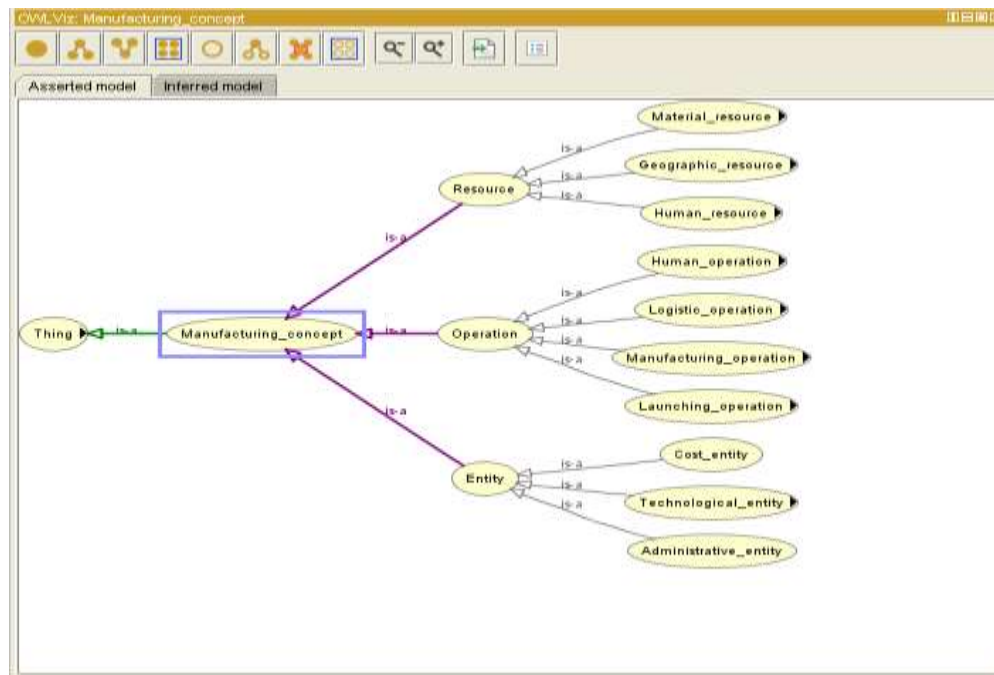


Figure 3: development of ontology using Protégé

The ontology is developed to facilitate the process of knowledge search and knowledge discovery by analyzing the three manufacturing concepts, Resource, Operation and Entity (see figure 3). The warranty database can be used to identify the problems related to the manufacturing concept as each component has its specific function that can be related to the type of defect. Figures 4 and 5 give detailed knowledge about the subclasses of MASON's three major classes – Resource, Operation and Entity. The relevant keywords showcasing the region of defect are obtained by application of KDT on warranty failure/customer feedback data gathered in the past. The Knowledge discovery process can be further enhanced by applying domain expertise to link the design data to the warranty data. A set of hierarchies have been developed to capture the various kinds of faults related to the warranty failure. Each hierarchy has a single root which shows the main topic which will be dealt with as shown in figure 6.

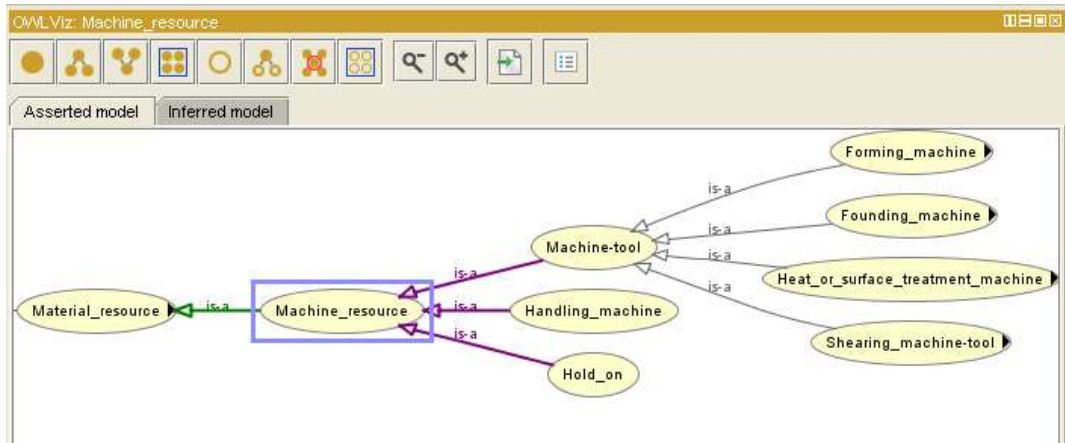


Figure 4: Developed ontology for resource using Protégé

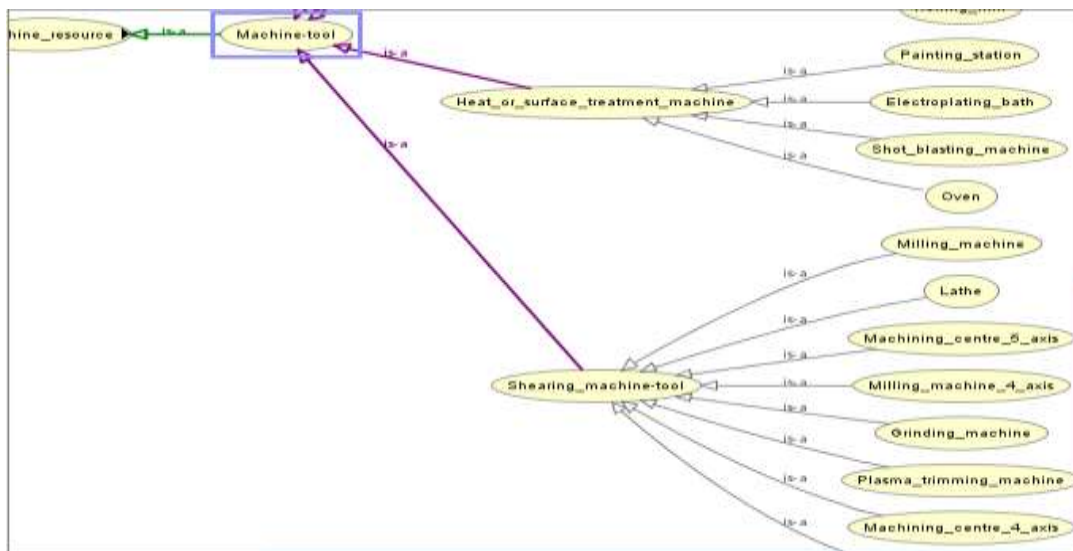


Figure 5: Developed ontology for machine-tool using Protégé

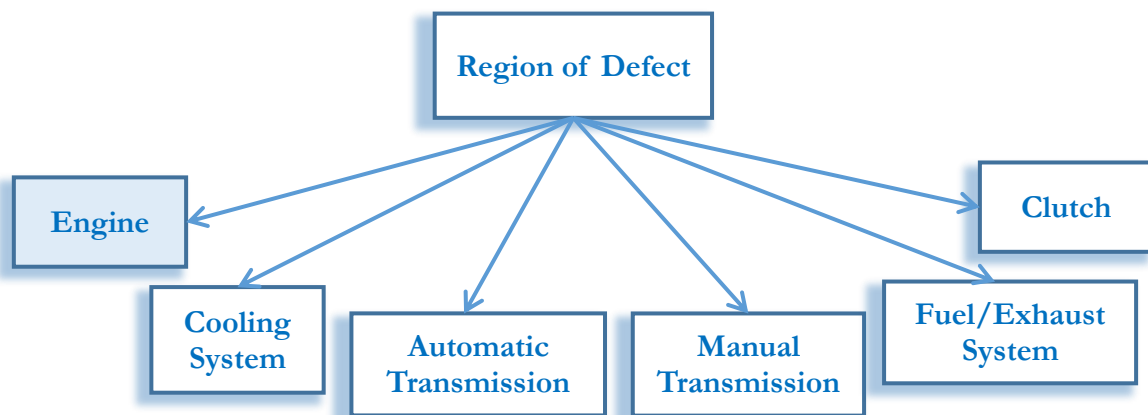


Figure 6: Top level Hierarchy related to warranty database

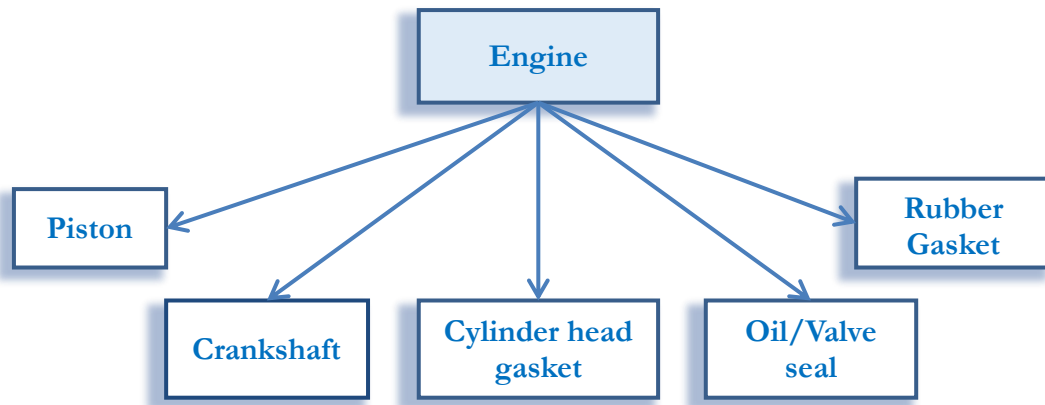


Figure 7: Hierarchy relating engine from top level hierarchy

The top-level hierarchy (figure 6) shows that the region of defect can be linked to a number of parts like engine, automatic transmission, cooling system and fuel and exhaust system. A further set of hierarchies help to identify the defective component or the design feature. Each area is considered in further detail in separate hierarchies using specific design features associated with them. Figure 7 shows a sample hierarchy for the case of the design features or components related to the class engine. In figure 7, the middle-level hierarchy shows that the defect in engine might be linked to the defects associated with piston, crankshaft, cylinder head gasket, oil/valve seal and rubber gasket. A sample domain expertise has also been constructed to help to identify the main root cause of the functional failure. Table 1 shows the key points and information which may help in identifying the defects and adapting remedial measures accordingly. Inclusion and modification of particular terms are considered based on context. A combination of the dictionary feature of the text mining software and an ontology based approach have been adopted to address the semantic issues of the work. The upper ontology developed will help in identifying the region of defect which can then be further analyzed using SOM to identify the defective tolerance value and extent of defect. The Self Organizing Maps are discussed in detail in the next section.

Table 1: An example showing a sample Domain Expertise

Engine	Engine fails to rotate when attempting to start? Engine rotated, but does not start?
--------	---

	<p>Engine difficult to start when cold?</p> <p>Engine difficult to start when hot?</p> <p>Starter motor noisy or excessively rough in engagement?</p> <p>Engine starts, but stops immediately?</p> <p>Engine idles erratically?</p> <p>Engine misfires at idle speed?</p> <p>Engine misfires throughout the driving speed range?</p> <p>Engine hesitates on acceleration?</p> <p>Engine stalls?</p> <p>Engine lacks power?</p> <p>Engine backfires?</p> <p>Oil pressure warning light illuminated with engine running?</p> <p>Engine runs-on after switching off?</p> <p>Engine noises?</p>
Cooling Systems	Overheating? Overcooling? External coolant leakage? Internal coolant leakage? Corrosion
Fuel and Exhaust Systems	Excessive fuel consumption? Fuel leakage and/or fuel odor? Excessive noise or fumes from exhaust system?
Clutch	<p>Pedal travels to floor? Clutch fails to disengage (unable to select gears)?</p> <p>Clutch slips (engine speed increases with no increase in vehicle speed)?</p> <p>Judder as clutch is engaged? Noise when depressing or releasing clutch pedal?</p>
Manual Transmission	<p>Noisy in neutral with engine running? Noisy in one particular gear?</p> <p>Difficulty engaging gears? Jumps out of gear? Vibration? Lubricant leaks?</p>
Automatic Transmission	<p>Fluid leakage? Transmission fluid brown, or has burned smell? General gear selection problems? Transmission will not downshift (kick down) with accelerator fully depressed? Engine will not start in any gear, or start in gears other than Park or Neutral? Transmission slips, shifts roughly, is noisy, or has no drive in forward or reverse gears?</p>

4. Self-Organizing Maps

SOM is a machine learning algorithm for tasks requiring clustering, dimensionality reduction or the visualization of high dimensional data. It also provides greater insights into the underlying learning and has an advantage over other methods as it can perform modelling of data even with missing values (Shukla et al. 2018). In this paper, SOM is used as a data mining tool to orderly map a high dimensional distribution to a regular low dimensional grid. Thereby, it is able to convert complex, nonlinear statistical relationships between high-

dimensional data items into simple geometric relationships on a low-dimensional display (Kohonen, 1995). SOM has therefore been employed to identify the defective component and help in recognizing the tolerance values to prevent further defects.

SOM is an unsupervised neural network used for clustering purposes and it has the ability to deal with a non-linear, smooth mapping of high-dimensional input space onto a low-dimensional array of neurons. The input vectors resembling each other are mapped together with nearby neurons in the output space to preserve the topology. A detailed description of SOM is presented in Kohonen (2006). However, for the sake of completeness, the following brief overview is provided. Each n neurons associated with the SOM is related to a d -dimensional weight vector and a neighbourhood relation is developed for connecting the topology of the map to the adjacent neurons. SOM consists of two stages – competitive learning and cooperative learning. In competitive learning, one sample vector is chosen from the input data and the Euclidean distance with each of the weight vector of SOM is estimated. The neuron, which is closest to the sample vector, is considered as the best matching unit (BMU). In cooperative learning, the topological neighbours and the weight vectors of the BMU are updated in order for them to move closer to the input vectors. The learning equation can be represented as $W_l(t+1) = W_l(t) + h_{BMU,l}(t)[x(t) - W_l(t)]$. Here, $h_{BMU,l}(t)$ is the decreasing neighbourhood function having relationship with training time and distance between BMU. Due to the learning process, the weight vectors get attracted towards highly crowded areas within the input region. SOM provides valuable information about the system depending upon the clusters formed when the weight vectors gets attracted towards the high density regions. Due to such robustness, SOM is highly desired to perform complex tasks such as machine perception, process analysis etc. (Li et al. 2018).

A two-dimensional topological map having $x \times y$ neurons and weight vector, $W_l = [W_{l,s}, W_{l,o}, W_{l,d}]$ is taken into consideration. Each of the neurons are denoted as N_l , where $l = 1, 2, 3, \dots, n$ and $n = x \times y$. The steps associated with Self Organizing Maps are as follows,

Step 1: Two-dimensional topological map having n neurons and weight vectors assigned to each neuron are initialized randomly. Also, setting $k = 1$

Step 2: A sample vector $x(t)$ is chosen randomly from the input data and determining the Euclidean distance between all the neurons and the sample vector using the following

$$\text{equation, } \sqrt{(W_{l,s} - s_k)^2 + (W_{l,o} - o_k)^2 + (W_{l,d} - d_k)^2} \quad (1)$$

The neuron with smallest Euclidean distance is considered as the Best Matching Unit (BMU) and the corresponding neuron is denoted as N_{BMU} and the weigh vector associated with the neuron is represented as $W_{BMU} = [W_{BMU,s}, W_{BMU,o}, W_{BMU,d}]$.

Step 3: Using the Kohonen learning rule, W_l is updated, $W_l(t+1) = W_l(t) + h_{BMU,l}(t)[x(t) - W_l(t)]$. $h_{BMU,l}(t)$ is the neighbourhood function computed using the following

$$\text{equation, } h_{BMU,l}(t) = a(t) \exp\left(-\frac{\sqrt{(W_{BMU,s}-W_{l,s})^2 + (W_{BMU,o}-W_{l,o})^2 + (W_{BMU,d}-W_{l,d})^2}}{2\beta^2}\right).$$

Here, $a(t)$ is the Learning rate and β is the distance on the topological map between N_l and N_{BMU} .

Step 4: Setting $k = k + 1$, if $k = \text{Max. iterations}$, then end the process or else go to Step 2.

The next section discusses the cost optimisation in order to make trade offs between warranty costs for repair or replacement and warranty improvement costs which include modification in production process or the design changes to improve the reliability of the product.

5. Cost optimization

Quality control decisions and design changes define product characteristics such as durability, reliability and reparability and these characteristics can be achieved through more efforts to achieve manufacturing improvements. Decisions to make improvements in manufacturing or product design need to be considered in the context of several aspects including customer satisfaction, frequency of failure, rate of improvability for specific subsystems and repair costs incurred by the company. Not all improvements are cost effective as there is a trade-off between different potential benefits. Hence, it is imperative that a cost analysis is performed to determine the repair or replacement cost before investing too many resources for a slight improvement in the product quality.

When the product is shipped to the customer, the manufacturer incurs additional cost related to repair and replacement of the components during the warranty period. This cost should be minimized in order to increase the profitability of the manufacturer. In this subsection, the warranty cost incurred is analyzed and a comparison is made with the warranty improvement cost which includes modifications in production processes or design changes which would improve the reliability of the product. The modification in the design

specifications suggested in table 3 may not be feasible to implement as they might not prove to be cost effective. Therefore, an analysis has to be done before these changes are put in practice. For this purpose, the model proposed by Yang et al. 2004 is employed with a modification made in the improvement cost of the original model and a fixed component regarding setup changes also considered.

Let the cost be divided into two components vulnerability cost and the improvement cost. The vulnerability cost includes the repair and replacement cost of losing its relevant function. Different functions may have different levels of relevance depending on which costs are calculated. The improvement cost includes the variable cost and the fixed cost which the manufacturing organizations are liable to pay when implementing design changes. All these activities should be carried out within a limited budget, so an optimized solution is obtained to maximize the profitability.

Nomenclature

C_{VK} - Expected cost of vulnerability due to component failure on time t and this cost is proportional to the component replacement cost and cost of losing the relevant functional requirement.

$$C_{VK}(t) = p_k(t)C_{Rk} + p_k(t) \sum c_{ik}C_{FRi} \quad (6)$$

$p_k(t)$ is the probability of k^{th} component failing before time t . Each component lifetime follows an exponential distribution, then $p_k(t)$ can be expressed as,

$$p_k(t) = 1 - e^{-t/\alpha k}, \text{ here } \alpha k \text{ is the mean life of } k^{th} \text{ component.}$$

C_{Rk} - Repair cost of k^{th} component

C_{FRi} - Cost of losing the relevant i^{th} function

c_{ik} - k^{th} component effect on i^{th} function

i - Vehicles sold

t - Month

C_{Ik} - Improvement cost for implementing the changes on the k^{th} component

C_{Vik} - Variable component of the improvement cost for implementing the changes on the k^{th} component

C_{FIk} – Fixed component of the improvement cost for implementing the changes on the k^{th} component

The mathematical model has been developed without considering a time dimension as the total cost is investigated for a specific useful life of the product and the time dimension has therefore been omitted in the interests of maintaining simplicity. The objective is to minimize the overall cost of the various subsystems or components with limited resources. The total cost can be reduced by lowering the probability of failure and assuming that the vulnerability and improvement cost remains constant.

The total cost function is given as:

$$\text{Minimize } TC = \sum TC_k = \sum p'_k C_{Vk} + \sum (p_k - p'_k) C_{VIk} + \sum C_{FIk}(p) \quad (7)$$

Equation (7) aims to minimize the total cost of the component comprising of the vulnerability cost and the improvement cost. Here, TC_k is the total cost of component k , p_k is the current failure probability of k^{th} component and p'_k is the reduced failure probability of k^{th} component.

Constraints associated with the mathematical model are given as follows,

$$\sum (p_k - p'_k) C_{Ik} \leq B \quad (8)$$

$$p_k \leq p'_k, \text{ for all } k \quad (9)$$

$$p_k > 0 \text{ and } p'_k > 0, \text{ for all } k \quad (10)$$

Here B is the total manufacturing budget available. Now, the improvement cost is related to the variable component and fixed component for implementing the changes on the k^{th} component in the following way,

$$C_{Ik} = C_{VIk} + C_{FIk} \quad p_k > p'_k > p_k/2 \quad (11)$$

$$C_{Ik} = C_{VIk} + C_{FIk} \quad p'_k < p_k/2 \quad (12)$$

Let $d_k = p_k - p'_k$ to simplify the equations. So, the objection function presented in equation (7) can be expressed in the following way,

$$\text{Minimize } TC = \sum p_k C_{Vk} - \sum (C_{Vk} - C_{VIk}) d_k + \sum C_{FIk}(p) \quad (13)$$

Subject to,

$$d_k \leq p_k, \text{ for all } k \quad (14)$$

$$\sum(C_{Ik} d_k + C_{FIk}) \leq B \quad (15)$$

$$d_k > 0, \text{ for all } k \quad (16)$$

Only the constraint (15) includes all the variables and the remaining constraints consider one variable. For each of the corresponding variable, the difference between the improvement cost and vulnerability cost helps in reducing the total cost. The next section discusses a numerical illustration of the proposed framework.

6. A Numerical Illustration

In order to improve the quality of the product it is necessary to identify the defective components and make relevant changes in the parameter or tolerance settings. Different components failing in a system have different consequences and sometimes failure of a very trivial part can generate a substantial impact on system's performance and thereby on customer satisfaction. Failure or breakdown of a component may result in the failure of certain major product functions and it is therefore essential to understand the design structure and association between components and functions. The design process needs to be carried out by mapping from the functional requirements (FRs) in functional space to the design parameters (DPs) in the design space to the components in the component space (CSs). The effects that each component has on the system vulnerability is dependent on the design hierarchy. A multilevel hierarchical model can be used to relate the functional requirements to the components responsible for the execution of that particular function. This approach may help in identifying the critical components which may be responsible for major functional faults. Figure 9 shows a hierarchical diagram relating functional factors to the design parameters which are then linked to one or more components. The ontology developed helps in identifying the link between the functional requirement (FR) and the design parameter (DP) i.e., the design defect leading to the particular functional error.

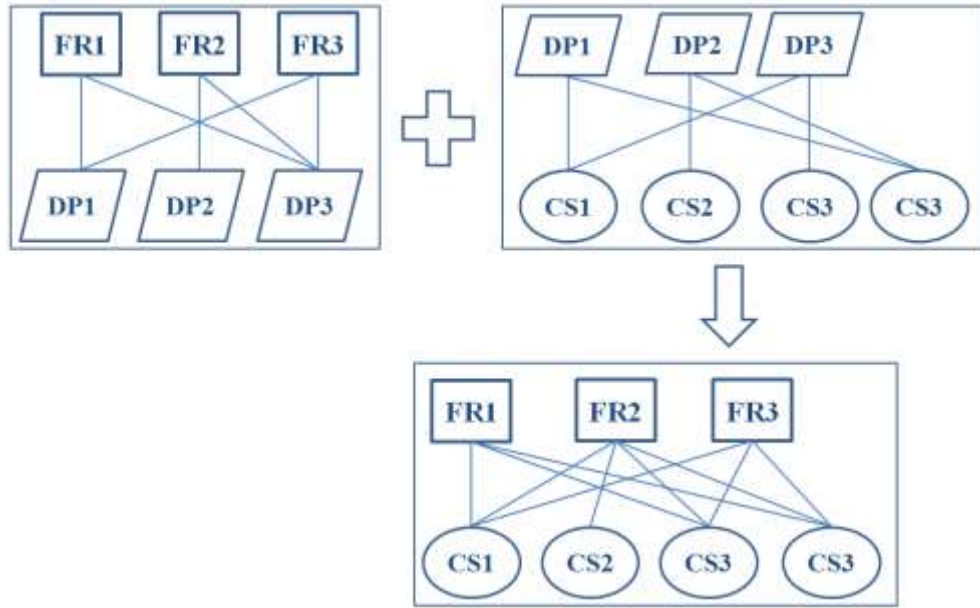


Figure 9: A multi-level hierarchical diagram relating functional requirement, design parameter and components

Consider a design parameter DP1 where defects have been identified. Let (C1, C2, C3) be the components which constitute the design parameter DP1. A set of 10,000 manufacturing data and 1000 warranty failure data related to DP1 were analyzed to identify the main defective component. These data sets were grouped into two dimensional grids comprising of (5 X 10) neurons arranged in a hexagonal grid pattern. An input vector is assigned to the neuron with the nearest weight vector and subsequently the weight vector of neuron and the corresponding neighborhood neurons are adjusted according to the input vector and the neighborhood function as shown in equation (2). Table 2 shows the warranty failure data and the manufacturing data which are clustered into 50 different neurons. The majority of equations associated with Self Organizing Maps (SOM) can be found in Chang et al., (2017) and Yu et al., (2015).

$$W_l(t+1) = W_l(t) + h_{BMU,l}(t)[x(t) - W_l(t)] \quad (2)$$

Where $h_{BMU,l}(t)$ represents the neighbourhood as shown by equation (3).

$$h_{BMU,l}(t) = a(t) \exp \left(- \frac{\sqrt{(w_{BMU,s} - w_{l,s})^2 + (w_{BMU,o} - w_{l,o})^2 + (w_{BMU,d} - w_{l,d})^2}}{2\beta^2} \right) \quad (3)$$

$a(t)$ = Learning factor which is considered to be 0.8 in this case.

Table 2: Mean value and count of manufacturing and warranty data arranged in two dimensional grid of 50 neurons.

S.No	Manufacturing Data			count	Warranty Failure Data			Count
	C1	C2	C3		C1	C2	C3	
1	10.55101	18.96843	20.94361	415	11.07879	19.08431	21.03802	10
2	8.483217	17.75551	19.64217	47	0			
3	9.721756	21.45246	19.49893	473	10.66825	21.80442	19.25218	1
4	11.13036	18.81956	19.52244	539	11.8331	18.95486	19.81426	89
5	7.648737	22.46332	20.47007	1	0			
6	8.758142	20.61662	19.96867	565	0			
7	9.010184	22.19387	21.49219	17	0			
8	9.541421	17.97081	22.06666	5	0			
9	9.845446	20.051	21.81692	249	10.61797	20.37224	21.94375	1
10	11.09983	21.99073	19.86663	124	11.63909	22.32786	20.12021	26
11	8.281033	20.22919	18.79477	156	0			
12	9.527174	19.1344	18.41678	286	0			
13	8.635951	18.62159	19.26973	98	0			
14	10.47553	20.96334	20.79101	586	10.91661	20.85614	20.80677	17
15	11.71363	21.12462	19.7726	142	12.23524	21.34059	19.94179	135
16	8.428841	19.33559	20.60936	236	0			
17	8.572805	17.56218	21.44275	7	0			
18	9.337446	18.78864	20.68822	282	0			
19	11.54612	21.13548	21.4464	58	12.0262	21.06607	21.45621	97
20	11.22664	18.06416	20.99601	57	11.90039	18.38954	21.29727	31
21	7.507553	20.76465	22.06825	5	0			
22	9.583892	20.33299	17.97287	137	0			
23	11.07326	22.87021	21.7321	8	11.34851	22.46843	21.67827	4
24	7.528346	19.72343	20.13243	85	0			
25	9.643401	19.17304	19.53007	1185	10.03604	19.27746	19.43061	2
26	9.681582	19.05289	21.19553	206	0			
27	10.7503	20.71125	16.79192	5	11.79187	21.82889	17.64193	1
28	10.85045	20.57757	18.54035	471	11.48178	20.26043	18.68474	28
29	7.770355	21.27176	18.12595	1	0			
30	9.329488	21.11714	18.54112	154	0			
31	10.96038	18.53649	22.2114	24	11.5935	18.98324	22.35297	6
32	10.58733	18.28541	18.19471	60	10.61174	16.50716	18.90054	1
33	11.66498	18.38052	18.16571	7	12.33106	18.58684	18.16043	5
34	12.08978	17.38278	20.1278	5	12.56894	17.82302	20.4576	5
35	11.87907	22.39758	18.44441	5	12.27014	22.07668	18.23274	6
36	8.210988	18.53457	22.23573	16	0			
37	10.09794	20.14504	20.11071	2210	10.95323	20.11904	19.67599	10
38	11.59279	20.01887	20.54397	566	12.00613	20.01954	20.61809	433
39	8.541952	17.61661	18.02493	6	0			
40	10.23342	22.07902	21.30138	48	10.69339	21.97654	20.72718	3
41	12.60105	21.7822	21.42898	4	12.488	21.97321	21.57434	27
42	8.048715	21.71943	21.93443	10	0			
43	12.45446	20.38884	18.6359	40	12.52111	20.47897	18.8921	41
44	8.845264	20.62083	21.42792	262	0			
45	10.90969	18.88433	17.29366	5	0			
46	7.607948	21.4266	19.95639	30	0			
47	9.050813	23.04001	19.4712	20	0			
48	8.93014	22.25977	20.52998	22	0			
49	10.59547	20.64536	22.60601	40	11.59355	20.5104	22.88462	9
50	12.35972	18.70606	18.80551	20	12.45652	18.63674	19.07386	12

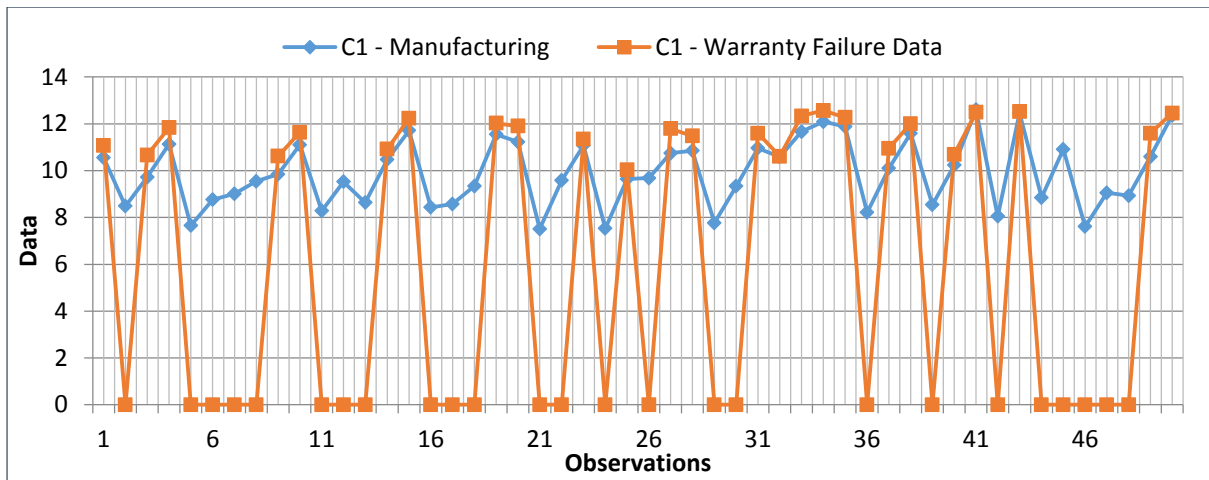


Figure 10(a): Graph between mean value of the manufacturing data and warranty data with respect to the neuron allocated for parameter C1

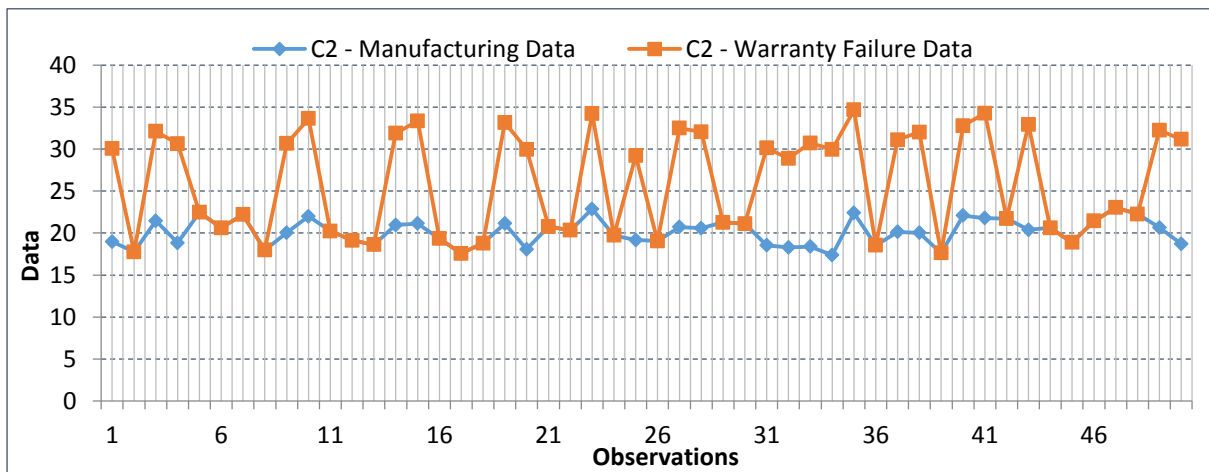


Figure 10(b): Graph between mean value of the manufacturing data and warranty data with respect to the neuron allocated for parameter C2

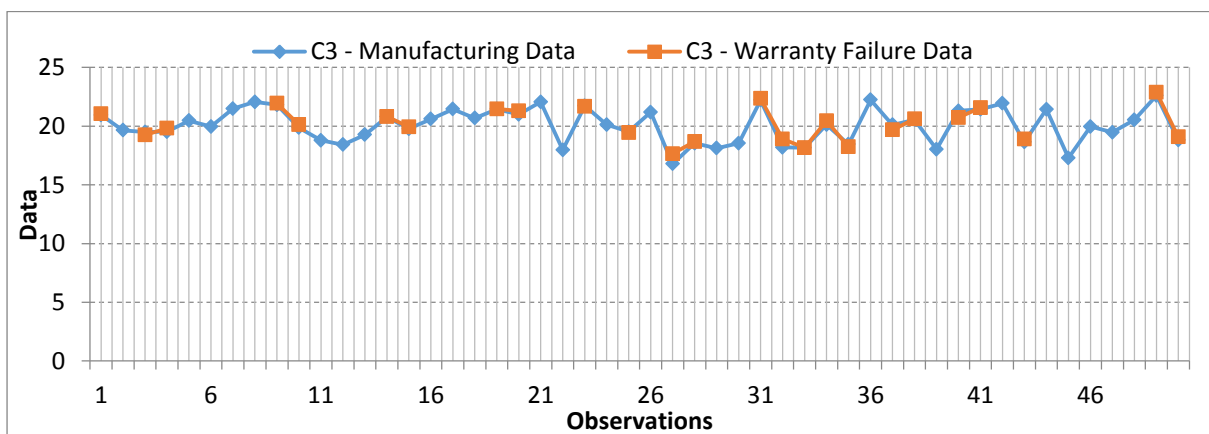


Figure 10(c): Graph between mean value of the manufacturing data and warranty data with respect to the neuron allocated for parameter C3

From the figure 10 and by comparing the means of the components value obtained from manufacturing data and warranty data it can be observed that the C1 parameter is the main cause of the defects observed in DP1. Thus, the specification of the component C1 should be modified in order to reduce the probability of occurrence of any defects in future. If we analyze the neurons 38 and 15, it is seen that it comprises of maximum number of defects 433 and 135 respectively. From the figure (10a), it can be seen that the majority of problem are encountered when the value of C1 reaches the upper control limit. Thus the new control limits are obtained by analyzing the cells where the maximum defects have been observed. In this case, cells 38 and 15 were analyzed and new USL were obtained using the equation (4).

$$USL_{new} = (w1 * C1_m + w2 * C2_w) / (w1 + w2) \quad (4)$$

$$LSL_{new} = LSL_{old} \quad (5)$$

In the above calculation, the manufacturing data is also being considered along with the warranty data as it is observed that the particular neuron corresponds to failure and the manufacturing data available may also comprise of items which are defective in nature. Table 3 shows the new values of control limit obtained when cells (38, 15) and (15) are considered independently.

Table 3: Modified control limits for the defective parameters

C1 (cells considered)	USL	LSL	Mean
38,15	11.88	7	9.44
15	11.9744	7	9.4872

The tolerance data obtained by examining the control limits of the defective components is used for determining the variable component of the improvement cost for implementing the changes in the manufacturing process for the defective component. The variable component associated with each of the defective component is computed in the similar way. The results obtained after solving the objective function (13) and constraints (14), (15) and (16) are presented in table 4. The values obtained for the probability of a specific component failing and the total cost is estimated which comprises of the improvement cost and vulnerability cost. The model developed is solved using LINGO and the results obtained are shown in table 4. Thus, from the cost structure it may be noted that it is more beneficial and cost effective to

change the USL and LSL value of component C4 to 19 and 13.909. Subsequently, design changes should be brought in parameter C1 depending on the manufacturing budget. Thus, observing the cost structure along with the changes in maximizing the return for an organization.

Table 4: Cost incurred in improving/repairing the defective components

S.No	Parameter	Initial Specification			Final Specification			pk	$p'k$	C_{VK}	C_{Ik}		$C_{VK} - C_{VIk}$	TC
		LSL	Mean	USL	LSL	Mean	USL				C_{VIk}	C_{FIk}		
1	C1	7	10.5	12	7	10.5	11.9744	0.1	.0487	3000	200	1000	2800	1143.64
2	C1	7	10.5	12	7	10.5	11.89	0.1	.0436	3000	200	2500	2800	2657.92
3	C4	13	16	19	14.1257		19	0.1	.0417	2000	300	1300	1700	1399.11
4	C4	13	16	19	13.909		19	0.1	.0558	2000	300	1000	1700	1075.14

Customer satisfaction can be increased by using information obtained from the performance of the product to determine which product design improvements should be made. Hence, the challenge lies in utilizing the warranty data to prioritize the essential improvement areas of the product whilst simultaneously allocating limited resources so that benefits are maximized for the manufacturers. A relationship between warranty data and design vulnerability has been presented to determine the cost of a failure of an automobile component which includes the repair or replacement cost and cost associated with losing relevant functions. Design vulnerabilities are expressed in terms of the cost related to the automobile subsystem or component failure and the mathematical model presented helps to estimate the vulnerability cost by providing the relevant functions affected by the failure of the component. Warranty data is used to estimate the measure of design vulnerability by considering an automobile industry example and determining the design weaknesses and obtaining feedback for possible design changes. Design improvements can be incorporated taking into account of the limited engineering resources in terms of budget requirement. The cost optimization method provides valuable insights regarding the allocation of maximum resources when the difference between vulnerability cost and improvement cost is the highest for a component or a subsystem. The vulnerability cost indicated the weaknesses associated with the product design and accordingly provides feedback for design improvements. After investigating the control limits associated with the defective components, the tolerance data is obtained which helps in estimating the variable cost related to incorporating the changes in the design process of the defective component.

7. Conclusion and Future Work

A Decision Support System for analyzing the warranty database is proposed in this research. A text mining based ontological approach has been used to identify the type of defect using the warranty database. The domain knowledge of automobile manufacturing unit is used to identify the manufacturing processes related to the defective parameters. In the problem considered, the MASON ontology is used to create an upper ontology of the manufacturing process. Further, a Protégé platform is used with the support of Graphviz to represent the ontology in hierarchical form showing the classes, subclasses and their relationship. With the detailed knowledge of the manufacturing process, more complex ontologies can be developed to capture appropriately detailed manufacturing knowledge. In this paper, Self-Organizing Maps are used to cluster the defective data and the normal data. The warranty data is compared with the manufacturing data and clusters of data are formed using the Self Organizing Maps. The cost factors associated with the design and parameter changes are then studied to determine the feasibility of the design change and appropriate measures are proposed. The Cost Analysis is done to identify the feasibility of implementing the changes and tighter tolerances that have been recommended by using SOM. Both improvement and vulnerability costs are identified and depending on the manufacturing budget recommendations are made.

A more complex ontology can be developed to cover all the aspects of the system. For this, the entire product (e.g. an automobile) can be divided into parts to study it in more detail and to include the complex mechanisms which have led to particular types of defect, so that this knowledge can be used for future defect identification. The data from the automobile industries can be collected and the applicability of the system can be studied in real scenarios. Further the cost data at the repair centre can be identified in order to study the applicability of the proposed changes.

Acknowledgement: The authors are grateful to the (Raytheon Chair for Systems Engineering) for funding.

References

Ajayi, M., & Smart, P. (2008). Innovation and learning: exploring feedback from service to design. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 222(9), 1195-1199.

Braasch, A. How to Reduce Your Warranty Costs. <https://www.babtec.de/en/news/professional-articles/best-practice-for-warranty-management.html>

Bechhofer, S. (2009). OWL: Web ontology language. In *Encyclopedia of Database Systems* (pp. 2008-2009). Springer US.

Blischke, W. R., & Murthy, D. N. P. (1992). Product warranty management—I: A taxonomy for warranty policies. *European Journal of Operational Research*, 62(2), 127-148.

Bouguerra, S., Chelbi, A., & Rezg, N. (2012). A decision model for adopting an extended warranty under different maintenance policies. *International Journal of Production Economics*, 135(2), 840-849.

Buddhakulsomsiri, J., Siradeghyan, Y., Zakarian, A., & Li, X. (2006). Association rule-generation algorithm for mining automotive warranty data. *International Journal of Production Research*, 44(14), 2749-2770.

Carrillo, P., Harding, J., & Choudhary, A. (2011). Knowledge discovery from post-project reviews. *Construction Management and Economics*, 29(7), 713-723.

Chang, W. L., Pang, L. M., & Tay, K. M. (2017). Application of self-organizing map to failure modes and effects analysis methodology. *Neurocomputing*, 249, 314-320.

Chien, C. F., Wang, W. C., & Cheng, J. C. (2007). Data mining for yield enhancement in semiconductor manufacturing and an empirical study. *Expert Systems with Applications*, 33(1), 192-198.

Choudhary, A. K., Harding, J. A., & Tiwari, M. K. (2009a). Data mining in manufacturing: a review based on the kind of knowledge. *Journal of Intelligent Manufacturing*, 20(5), 501.

Choudhary, A. K., Oluikpe, P. I., Harding, J. A., & Carrillo, P. M. (2009b). The needs and benefits of Text Mining applications on Post-Project Reviews. *Computers in Industry*, 60(9), 728-740.

Choudhary, A. K., Harding, J. A., Tiwari, M. K., & Shankar, R. (2018). Knowledge management based collaboration moderator services to support SMEs in virtual organisations (Article in press)

- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, 17(3), 37.
- Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing?. *International journal of human-computer studies*, 43(5-6), 907-928.
- He, Y. H., Wang, L. B., He, Z. Z., & Xie, M. (2016). A fuzzy TOPSIS and rough set based approach for mechanism analysis of product infant failure. *Engineering Applications of Artificial Intelligence*, 47, 25-37.
- Hui, S. C., & Jha, G. (2000). Data mining for customer service support. *Information & Management*, 38(1), 1-13.
- Kang, S., Kim, E., Shim, J., Cho, S., Chang, W., & Kim, J. (2017). Mining the relationship between production and customer service data for failure analysis of industrial products. *Computers & Industrial Engineering*, 106, 137-146.
- Khare, V. R., & Chougule, R. (2012). Decision support for improved service effectiveness using domain aware text mining. *Knowledge-Based Systems*, 33, 29-40.
- Kohonen, T. (2006). Self-organizing neural projections. *Neural networks*, 19(6), 723-733.
- Kohonen, T., & Maps, S. O. (1995). Springer series in information sciences. *Self-organizing maps*, 30.
- Li, Z., Fang, H., Huang, M., Wei, Y., & Zhang, L. (2018). Data-driven bearing fault identification using improved hidden Markov model and self-organizing map. *Computers & Industrial Engineering*, 116, 37-46.
- Liao, G. L. (2016). Optimal economic production quantity policy for a parallel system with repair, rework, free-repair warranty and maintenance. *International Journal of Production Research*, 54(20), 6265-6280.
- Luo, M., & Wu, S. (2018). A mean-variance optimisation approach to collectively pricing warranty policies. *International Journal of Production Economics*, 196, 101-112.
- Malek, S., Gunalan, R., Kedija, S. Y., Lau, C. F., Mosleh, M. A., Milow, P., & Saw, A. (2018). Random forest and Self Organizing Maps application for analysis of pediatric fracture healing time of the lower limb. *Neurocomputing*, 272, 55-62.

- Mannar, K., Ceglarek, D., Niu, F., & Abifaraj, B. (2006). Fault region localization: product and process improvement based on field performance and manufacturing measurements. *IEEE Transactions on Automation Science and Engineering*, 3(4), 423-439.
- Nelson, G. (2013, October 28). How data mining helped GM limit a recall to 4 cars. <http://www.autonews.com/article/20131028/OEM11/310289980/how-data-mining-helped-gm-limit-a-recall-to-4-cars>
- Park, M., Jung, K. M., & Park, D. H. (2017). Optimal maintenance strategy under renewable warranty with repair time threshold. *Applied Mathematical Modelling*, 43, 498-508.
- Pham, H., & Zhang, X. (1999). A software cost model with warranty and risk costs. *IEEE Transactions on Computers*, 48(1), 71-75.
- Rajpathak, D. G. (2013). An ontology based text mining system for knowledge discovery from the diagnosis data in the automotive domain. *Computers in Industry*, 64(5), 565-580.
- Shukla, N., Hagenbuchner, M., Win, K. T., & Yang, J. (2018). Breast cancer data analysis for survivability studies and prediction. *Computer Methods and Programs in Biomedicine*, 155, 199-208.
- Shukla, N., & Kiridena, S. (2016). A fuzzy rough sets-based multi-agent analytics framework for dynamic supply chain configuration. *International Journal of Production Research*, 54(23), 6984-6996.
- Söderholm, P. (2007). A system view of the No Fault Found (NFF) phenomenon. *Reliability Engineering & System Safety*, 92(1), 1-14.
- Sureka, A., De, S., & Varma, K. (2008). Mining automotive warranty claims data for effective root cause analysis. In *Database Systems for Advanced Applications* (pp. 621-626). Springer Berlin/Heidelberg.
- Thomas, M. U., & Rao, S. S. (1999). Warranty economic decision models: A summary and some suggested directions for future research. *Operations Research*, 47(6), 807-820.
- Vernick, J. S., Mair, J. S., Teret, S. P., & Sapsin, J. W. (2003). Role of litigation in preventing product-related injuries. *Epidemiologic reviews*, 25(1), 90-98.
- Yang, K., & Cekecek, E. (2004). Design vulnerability analysis and design improvement by using warranty data. *Quality and reliability engineering international*, 20(2), 121-133.

Yeh, R. H., Chen, M. Y., & Lin, C. Y. (2007). Optimal periodic replacement policy for repairable products under free-repair warranty. *European Journal of Operational Research*, 176(3), 1678-1686.

Yu, H., Khan, F., & Garaniya, V. (2015). Risk-based fault detection using Self-Organizing Map. *Reliability Engineering & System Safety*, 139, 82-96.