Automotive Warranty Management: Leveraging IT for Efficient Root Cause Analysis

ASHISH SUREKA, KISHORE VARMA & SUDRIPTO DE
APRIL 2007

Executive Summary

The top 50 U.S.-based warranty providers together reported USD 23 billion in warranty claims during 2006, up 5.1 percent from 2005. It is interesting to note that auto manufacturing companies in USA such as General Motors Corporation and Ford Motors Corporation are amongst the top 50 U.S. based warranty providers of product warranties in terms of the total dollar amounts they reported in warranty claims during calendar 2006. GM and Ford together spent USD 8.6 billion on claims during 2006 and auto manufacturers grab the bulk of the pie in the overall extended warranty market. Auto manufacturing companies are spending around 2.5-3.0 percent of their revenue in fixing vehicles under warranty. This puts a tremendous pressure on auto companies to come up with innovative ways to reduce the overall cost of warranty claims. Bridgestone and Firestone’s recall of 6.5 million tires used primarily on Ford Explorer vehicles, and the deaths of more than 100 people in accidents blamed on the failure of those tires triggered legislators to introduce a new U.S. law called as Transportation Recall, Enhancement, Accountability and Documentation (TREAD) Act which makes it mandatory for automakers to compile quarterly reports on consumer complaints and warranty claims. "How can a warranty management leverage unstructured textual for an efficient root cause analysis and defect discovery".
**Claims Analysis is based on Unstructured Data**

Automotive warranty data is generally gathered by filling a claim form by a customer and a technician. The form is either filled on a paper which is later scanned and imported into a database or the information is directly entered online. A form can contain many fields to be entered by a customer or a technician. Some of the fields require information such as the product code, model number, date and timestamp and customer id. This information falls into the category of structured data in the sense that the information has a well defined format and requires close-ended answers i.e. there are finite choices from which a selection can be made. Usually the form also contains a comments section where a customer or a technician can provide detailed information about the problem. This is the section where information is entered in the form of a natural language text or a free-form text. The data entered in comment sections falls in the category of structured data and is a key element in diagnosing and understanding the problem. Following are some of the real world examples of the customer complaint, technician comments and action taken field data.

**Customer complaint**
- Air conditioning not working
- Poor performance from a/c system
- Water ingress into passenger foot well
- Room lamp flickering when switched ON

**Technician Comments**
- Found expansion valve defective
- Thermostat by-pass valve not working
- No engine cranking noise. Solenoid check
- AC Knob found broken

**Action Taken**
- Removed and replaced the seals on the elbow joints
- Cable set properly & refitted
- AC cable replaced resulting smooth movement
- Connected the coupler to compressor

If the auto manufacturer suspects a recurring problem they sift through the claims data and manually go through the customer and technician comments to see if they can find any kind of patterns or clue that can help them finding the cause of the problem. A high level analysis of a defect can be done from data stored in structured data fields. However, a drill down analysis or an in depth analysis requires a warranty claim analyst to read the free-form textual data fields also. The main challenge is that the manual process of reading each and every comment is impractical and time consuming. Hence there is a strong business case in automating the process of analyzing the natural language text data stored in claim forms for an efficient data analytics. There is a significant value in a system that not only extracts information out of structured data but also leverages unstructured data which is currently not exploited to its full extent. In the next section, we give a high level overview of a solution based on a natural language processing technique called as text tagging and annotation that can make the process of information extraction from warranty claim forms more effective.

**Text tagging and annotation: Adding structure to unstructured textual data**

Text Tagging and Annotation also called as Named Entity extraction forms an important component of many language processing tasks such as text mining, information extraction and information retrieval. Named Entity extraction consists of identifying the names of entities in free-form or unstructured text. Some of the common types of entities are proper nouns such as person names, products, organization, location, email addresses, vehicle, computer parts and currency, temporal entities such as dates, time, day, year, month and week, numerical entities such as measurements, percentages and monetary values. There can be numerous domain specific entities also. For example, named entities that can be extracted from the customer complaint, technician comments and action taken field of an automotive warranty claim forms can be a vehicle part, location of a defect, symptom, defect type, technician action, effect, reason for a defect etc.

The process of text tagging and annotation helps in adding structure to unstructured textual data. The idea behind applying named entity extraction to the textual data fields in an automotive warranty claim form is to bring the natural language text data recorded as part of the customer complaint, technician comments and action taken field into a format which can
then be easily queried using SQL (Structured Query Language). The process of gaining useful insights from warranty claim forms consists of two phases. In phase one, the unstructured data present in the claim form is first converted to structured data using a text tagging engine. As illustrated in Figure 1, named entities such as part, symptom, and condition have been extracted from a customer complaint: “Noise from AC system when switched ON”. Text tagging and annotation brings the data into a format which can be easily queried using SQL (Structured Query language). In phase two, the processed data from phase one is analyzed using a query engine which takes in high level queries from a warranty analyst and displays reports, tables and charts. Similarly, important names entities from technician comment and action taken have also been extracted. As illustrated in Figure 2, a warranty analyst can plot a pie chart showing a distribution across failed components or parts in a vehicle. The pie chart and the table shown in Figure 2 are created using data obtained from phase one process.

AutoSeek: A text tagging and query processing engine.

AutoSeek is a tool that brings in automation to the process of analyzing warranty claim forms in automotive industry. AutoSeek is made up of two parts. One part is called as “Text Tagging and Annotation Engine” and the other part is called as “Query Engine”. Figure 1 and 2 are screenshots from AutoSeek. They illustrate how text tagging and annotation coupled with query processing can provide a solution to the problem of analyzing unstructured data stored in automotive warranty claim forms. In the prevailing circumstances of more than 100 million claims being filed per annum, it becomes practically impossible for any warranty analyst to go through the text of the claims manually. As a result most of the info reported in the text goes unnoticed and undiscovered. Auto-Seek Text analytics will help integrate and automate this process of deciphering info from text. This will reduce detection time by an average of 50%.

Figure 1: AutoSeek screen showing the results of text tagging and annotation
Source: Infosys Research

Figure 2: AutoSeek screen of the query engine user interface showing distribution chart across vehicle parts
Source: Infosys Research
**Key Words:** Text Analytics, Text tagging and annotation, Root cause analysis, automotive warranty claims.

**References:**

1. Warranty Week, The Newsletter for Warranty Management Professionals  
3. Chrysler Jeep Liberty Recall for Defective Steering Part  
5. Automotive Warranty Management: Paying the Bill and Solving the Problem by Kevin Prouty,  
   A report on Manufacturing, AMR Research, November 01, 2000
6. Walking the automotive industry tightrope: Keeping customer and your brand safe by Rick Fournier, Thomas Shovelton and Larry Stolle  
7. Text mining tools take on unstructured data, June 21, 2004 (Computerworld)  
   [http://www.computerworld.com/databasetopics/businessintelligence/story/0,10801,93968,00.html](http://www.computerworld.com/databasetopics/businessintelligence/story/0,10801,93968,00.html)
   [http://www.dmreview.com/article_sub.cfm?articleId=1009161](http://www.dmreview.com/article_sub.cfm?articleId=1009161)
9. How to deal with structured and unstructured data, an article on Techworld.  

**About the Authors**

**Ashish Sureka,** PhD (NCSU, 2005), is a Research Associate at SETLabs. His research interests are in data mining and text mining. He has authored several papers in various national and international conferences and journals. He can be reached at Ashish_Sureka@infosys.com.

**Kishore Varma,** MTech (IIT Mumbai, 2006), is a Junior Research Associate at SETLabs. His research interests are in Data mining, Text mining and Algorithms. He can be reached at kishore_varma@infosys.com.

**Sudripto De** is a senior consultant of the manufacturing and supply chain group, Domain Competency Group, Infosys Technologies. Sudripto has over 14 years of automotive industry and consulting experience, having partnered several automotive majors in supply chain management and new product development. He is an Engineer from Delhi College of Engineering with an MBA from Faculty of management Studies, Delhi. Sudripto can be reached at Sudripto_de@infosys.com.