Modelling an Optimized Warranty Analysis methodology for fleet industry using data mining clustering methodologies with Fraud detection mechanism using pattern recognition on hybrid analytic approach

R. Srinivasan*, Dr. S. Manivannanb, Dr. N. Ethirajc, Dr. S. Prasanna Devi, S. Vinu Kirane

a Asst. Professor, Mechanical Engineering Department, Dr. M.G.R Educational and Research Institute, Chennai 600107, India
b Deputy Dean, Dr. MGR Educational & Research Institute University, Chennai 600107, India
c Professor, Mechanical Engineering Department, Dr.MGR Educational & Research Institute University, Chennai 600107, India
d, e Department of Computer Science & Engineering, Apollo Engineering College, Chennai 602105, India

Abstract

In this paper we have analyzed the huge volume of warranty data for segregating the fraudulent warranty claims using pattern recognition and clustering methodology. Recent survey of automotive industry shows up to 10% of warranty costs are related to warranty claims fraud, costing manufacturers several billions of dollars. Most of the automotive companies are suspecting and aware of warranty fraud. But they are not sure of the extent and ways to eliminate it. The existing methods to detect warranty fraud are very complex and expensive as they are dealing with inaccurate and vague data, causing manufacturers to bear the excessive costs. We are proposing model to find anomalies on warranty data along with component failure data and patterns based on historic warranty claims data under particular region and for specific component as the data are of high volume.

We are managing to isolate all the impacting the factors that indicate a claim, that has a high probability of fraudulence such as failure date and claim date, mode of failure etc., In addition to this we discover suspecting claims that have the greatest adjustment potential for further review by claim process. We altogether integrating data with with claims processing, reports and business rules along with reported mode of failure as we are minimizing changes to existing systems, since the analysis is carried out by identifying patterns. Since we are working with factual data, it gives more room to identify the actual cost involved on warranty claim.

Keywords: Data Analysis; Optimization; Warranty; Manufacturing; Fleet Industry; Predictive Analysis; Clustering; Pattern recognition; Fraud detection; Rosner; Grubb; Algorithm

1. Introduction

Automobile manufacturing industries are highly aware that fraudulent and questionable claims can significantly raise warranty costs and affect customer satisfaction. With the help of recent technology on data analytics, manufacturers can accept the data driven fact identification results on each warranty claims. Using a hybrid approach with clustering and pattern identification, industries can integrate traditional business rules for
warranty process with anomaly detection, advanced analytics and social network analysis to automatically score each claim, customer and the service provider of a particular claim. Our approach can quickly detect patterns on the available data sets for required period of time, that individual auditors might otherwise miss when checking manually. Industries that put this type of solution in place can start catching bad claims right away based on actual fact data sets. We have predicted for sample datasets using Minitab.

2. Warranty Fraud – An Overview

The word fraud means deceit. However, in many cases, honest mistakes may lead to invalid warranty claims. Most of such claims are flagged due to misunderstanding of policies or insufficient training or innocent clerical errors. Many of fraudulent claims are not cases of outright fraud but they could be questionable or suspicious claims. When a manufacturer identifies that true fraud is involved, they find that it comes in all shapes and sizes. Fraud ranges from consumers knowingly or mistakenly sending in counterfeit defects on automobiles to be repaired to repair technicians using repairing parts more than needed, corrupt auto-repair service centres inventing serial numbers and trying to repair non-damaged parts. The nature of fraud implications are extending beyond the immediate costs. However, all the cause and questionable claims affect warranty costs and customer satisfaction.

3. Detection of Fraudulent claims on Warranty data

There could be chances of replaced parts being sold on the black market. The insurance costs and additional security needs make the problem even more expensive. Automobile industry have put many systems and processes in place to detect and reduce fraudulent claims, but current practices are not enough to locate all such claims. An additional benefit of fraud detection is that once those fraudulent claims are isolated from the datasets, the issue detection and claim definition become much more accurate. By isolating these data early warning systems can more precisely detect significant changes in claims rates based on available datasets. Problem definition is more effective, since the claims being analyzed are actual customer issues, by allowing product engineers to better understand failure modes and the factors. True fraud is most likely to occur when a third-party service organization is utilized.

The following are the traditional approaches on identifying such fraudulent claims;

3.1 Manual adjudication:
The process of manually reviewing each individual claims thoroughly with the use of expert resources. Sometimes the identified data variation may high and patterns are often missed because different people are looking at different batches of claims. Some data may be missed out due to human error.

3.2 Automated (rules-based) adjudication:
Automated rules-based systems are now commonly used among organizations. Almost all claims transaction engines are equipped with this type of adjudication checking mechanisms. This approach is currently working for finding types of fraud that the organization is already aware of. However, these systems will not have dynamic working mechanism as new claims are processed, since they are hard-coded with business rules. Through trial & error, by making telephone calls to auditors, fraudsters quickly learn the rules and work around them for making benefit.

3.3 Return validation:
Requiring service-providers to return failed parts are reducing fraudulent behavior to some extent(at least for claims with parts). To do this, logistics and resources necessary on a large scale process.

3.4 Field audits:
Manually auditing each individual claims and service providers is a very useful way to find patches of fraud in the automobile service. Most of the times, it is very difficult for auditors to know where to focus while inspection. It is hard to determine the highest volume service providers or the claims with the highest cost or the claims with no parts etc.,
3.5 Analytics:
Based on ranking or scoring of claims and service providers based on the likelihood of fraud, data analytic techniques can be used to focus where the most value exists. Analytics can dynamically done and can learn from automobile service provider and inspection auditor behavior over time to find increasingly more fraud. Analytics can also identify new business rules that can be applied through the rules engine.

4. Actual issue detection on warranty data
Identification of an issue in the automobile that is claimed for warranty is the first step toward fixing it. Manufacturers are keen on reducing the detection-to-correction cycle. The early detection of emerging issues allows the entire warranty claim process to begin sooner than actual. If the issue has been identified, it is very critical to use all available data to perform analysis and to understand which attributes are driving the issue. It is very significant to identify the dependent and independent variables. It is necessary to understand that what is driving the issue focuses engineers on which processes to examine or which parts to tear down and whether a recall might be necessary. Questions may include:

- Does the issue often reported on specific combination of supplier, assembly location and usage?
- Does the problem only occur in certain geographic regions?

Several hours may be saved out of the issue detection and problem definition process that reduces a large number of potentially defective auto parts or products.

4.1 Data integration:
Warranty and service claims should be integrated with sales data in order to predict and understand failure rates. Product information such as manufacturing location, auto model, auto part suppliers, customer selected options and important dates greatly enhances the identification of issue detection and problem definition. Additional data sources such as customer complaints data, customer surveys, and field service data can give a more complete picture of automobile and service field performance. In many cases, these data are primary indicators of warranty issues. Bringing all the automobile service data together and applying warranty-specific business rules are the first steps to an effective identification and definition system.

4.2 Automated analytics:
Many automobile concerns are trying to identify new issue patterns through brute force technique. Analysts are now deep diving into thousands of warranty claim records, Pareto analysis and trend charts will eventually picking up new issues. Most of the valuable time is much better spent fixing problems instead of searching for them. Automating the data driven analytics can be put in place to detection of significant changes in failure rates, costs etc., Strong focus is needed to look for changes across multiple attributes. This includes production period, time in service, claim period and usage each yield a different insights on field failures. Once these issues are identified, analysts and data engineers will be able to isolate failure modes pretty much quickly and can understand what factors are driving the claims. A standard set of analytics is needed that identifies the patterns within user's comments and quantitative fields, problem definition can be greatly accelerated.

4.3 Communication:
Identification of issues are useful only if someone is doing something about it. Communication is essential on this case. When the analytic engine detects a new issue, it should be capable of notifying the appropriate engineer immediately. The analytic team members should be able to classify a specific sets of potential issues (for example, electrical issues on a specific model and brake issues across all product lines or water pump issues across all make and model of engines) allowing this to focus on specific alerts that are relevant to the potential ones. Problem definition on updating issue status and sharing analysis among team members, reduces time wasted on manual intervention and accelerates the problem solving process. Once the issue has been defined effective communication allows all of the stakeholders to understand the current status and the expected outcome of the fix.

5. Finding out anomaly, fraudulent data
A. Rosner Test:
Here we are referring anomaly as fraudulent data. Identification steps are as follows;
1. The sample mean and standard deviation are calculated based on the n sample values. k equals the number of suspected anomalies.
2. The sample value with the largest deviation from the mean is used to calculate the test statistic Ri as follows:

\[
R_i = \frac{|X(i) - \bar{X}(i)|}{S(i)}
\]

X(i) is the value with the largest deviation from the mean but can be either the largest or smallest value in the sample.
3. The sample value X(1) is then removed from the sample, and the mean X(2), S(2), and R(2) are calculated from the n-1 values.
4. The previous steps are repeated until all k suspected anomalies have yielded corresponding R(k) test statistics.
5. Each R(i) is compared in sequentially reverse order to a table of critical values for Rosner’s test. If the computed statistic R(i) is greater than the table value, then there are I number of anomalies.

|   | Mean | Std Deviation | Y    | R    |
|---|------|--------------|------|------|
| R1| 323,350 | 204,601        | 1,031,831 | 3.46 |
| R2| 303,108 | 167,056        | 878,047  | 3.44 |
| R3| 286,198 | 135,801        | 701,373  | 3.06 |
| R4| 273,617 | 116,054        | 540,379  | 2.30 |

Table 1: Rosner’s Test results for sample dataset

B. Grubb’s Test

Grubbs’ test is a statistical test developed by Frank E. Grubbs to detect anomalies in a univariate data set [Grubbs 1969]. Grubbs’ test is also known as the maximum normed residual test [Stefansky 1972]. Grubbs’ test is defined for the statistical hypothesis:

\[H_0: \text{The data set does not contain any anomalies (Frauds).}\]
\[H_a: \text{There is at least one anomaly (Fraud) in the data set.}\]

The test statistic is the largest absolute deviation from the data set mean in units of the data set standard deviation and is defined as:

\[G = \frac{\max |X_i - \bar{X}|}{s}\]

where

\(\bar{X}\) is the sample mean of the data set
\(s\) is the standard deviation of the data set

The hypothesis, H0, is rejected at the significance level, \(\alpha\), if

\[G > \frac{n - 1}{\sqrt{n}} \left[ t_{a/2, n-2}^2 \right]^{1/2}\]

Where \(t_{a/2, n-2}^2\) denotes the upper critical value of the \(t\) distribution with (n-2) degrees of freedom and a significance level of \(a/2\n\). Grubbs’ test detects one anomaly at a time. Multiple iterations are executed until no anomalies are discovered.
6. Key outcomes on using pattern identification of anomalies(frauds) and clustering of analysis of large-scale Warranty data for fraudulent claim detection

Automotive companies are strongly focusing to minimize the number of failures that occur during the early stages of a new product release. Factual data on real-time claims information and regular updates would enable companies to catch and stop mistakes earlier. Using this machine learning and predictive analytics on clustered data, automotive companies can decrease costs and increase revenues. This approach can accurately forecast further unethical behavior based on identified patterns as shown in Figures 2 and 3 respectively. It can also help dealers and automotive industries identify and improve training issues, increasing customer satisfaction improving product quality and reliability to minimize the warranty claims.

7. Conclusion

In the current situation, retaining and maintaining customer satisfaction and reducing costs are highly focused objectives of any automotive industry. The warranty claims are significant area for potential cost reduction. Applying clustering methods and pattern identification on warranty data analytics across the claim process would significantly reduce warranty costs. Nowadays, automobile manufacturers are applying best practices on analysis to deny the claims that are fraudulent and reduce costs associated with legitimate claims. These savings can be improved by using the problem solvers and automating much of the analysis work on solving problems instead of finding them. These fraud detection can accurately identify and helps to avoid bad claims that used to be noise in the data as the results of issue detection can be more accurate and problem definition can be more precise.

References

[1] http://en.wikipedia.org
[2] Karim MR, Suzuki K. Analysis of warranty claim data: A literature review. International Journal of Quality and Reliability Management 2005;22(7):667-686.
[3] Wu S. Warranty data analysis: A review. Quality and Reliability Engineering International 2012;28(8):795-805.
[4] Kalbfleisch JD, Lawless JF. Statistical analysis of warranty claims data. In: Blischke WR, Murthy DNP, editors. Product Warranty Handbook, New York: Marcel Dekker; 1996, p. 231-259.
[5] Suzuki K, Karim MR, Wang L. Statistical analysis of reliability warranty data. In: Rao CR, Balakrishnan N, editors. Handbook of Statistics: Advances in Reliability, Amsterdam: Elsevier; 2001, p. 585-609
[6] Wang L, Suzuki K, Yamamoto W. Age-based warranty data analysis without datespecific sales information. Applied Stochastic Models in Business and Industry 2002;18(3):323-337
[7] Lawless JF. Adjustments for reporting delays and the prediction of occurred but not reported events. Canadian Journal of Statistics 1994;22(1):15-31.
[8] Kalbfleisch, J, Lawless, J. Truncated data arising in warranty and field performance studies, and some useful statistical methods. 1991; RR-91-02

[9] Rai B, Singh N. Customer-rush near warranty expiration limit, and nonparametric hazard rate estimation from known mileage accumulation rates. IEEE Transactions on Reliability 2006;55(3):480-489

[10] Ion RA, Petkova VT, Peeters BHJ, Sander PC. Field reliability prediction in consumer electronics using warranty data. Quality and Reliability Engineering International 2007;23(4):401-414

[11] Mohan K, Cline B, Akers J. A practical method for failure analysis using incomplete warranty data. Proceedings of Annual Reliability and Maintainability Symposium. 2008:193-199.

[12] Wu S. Warranty claim analysis considering human factors. Reliability Engineering and System Safety 2011;96(1):131-138.

[13] Karim MR, Suzuki K. Analysis of warranty data with covariates. Journal of Risk and Reliability 2007;221:249-255

[14] Rai B, Singh N. Hazard rate estimation from incomplete and unclean warranty data. Reliability Engineering and System Safety 2003;81(1):79-92

[15] R.Srinivasan, S.Manivannan and S.Prasanna Devi, A CRITICAL REVIEW PAPER ON WARRANTY ANALYSIS FOR FLEET INDUSTRY USING DATA MINING TECHNIQUES, International Journal of Applied Engineering Research (IJAER), Volume 10, Number 55 (2015) Special Issues, pp. 1132-1134