WARRANTY DATA ANALYTICS

Francisco Figueroa, Bobbi McFadden, Gerald Oden, Richard Ording, and Andy Pawlowski
Continental Automotive Systems

P. Vikram Chowdhary and Keith Thompson
Ubiquiti Inc.

ABSTRACT
The appropriate use of analytics help identify root causes, resolve “no trouble found” issues, and provide means to get early warning of problems. Based on case study data and our experience, we exemplify several important techniques using warranty information. Our findings go beyond the boundaries of our own organizations, and include interactions with our business partners. Our approaches generalize to address similar issues more widely, and we explain how to do this. Also our techniques have helped in improved quality as well as significant direct cost savings.

1. OVERVIEW
Analyzing repairs & maintenance data in the automotive sector is a difficult, widespread problem that has significant cost ramifications. The analysis is not only to identify individual components that may fail in normal use, but also, to rapidly identify the causes of system-level failures due to unexpected interactions among components that form a system (rather than just the component-level issues). Approaches to address the analysis problems may need to “mine” repair datasets generated by technicians who perform repairs. Data may need to be obtained from different data sources (e.g., the same repair may have separate records for the reimbursement claims, the parts returned & tested for failures, and with text narratives stored in different repositories; also, repairs over the history of usage of a vehicle will result in separate records that are collected for each individual repair instance). Repairs performed under warranty account for ~$25B annually, and post-warranty costs are a few factors more in costs; in many cases, these costs could be controlled if the appropriate data analytics were to be performed. In addition to explicit costs, there are implicit costs, such as Customer loyalty, which may even exceed the explicit costs.

Data analysis may be the most effective and efficient solution to identify problems because it provides a cheaper, centralized means – without a need to transport physical components or having analysts travel to repair locations. The reasons why data analytics has eluded prior efforts include the inherent complexity of the tasks and “messy” text data; the latter is addressed by Ubiquiti for such datasets. The market opportunity represented by these problems is large; and the importance in identifying safety issues, and the benefits to manufacturing & transportation sectors, would probably be even greater.

Repair and maintenance of manufactured goods, especially in the automotive sector, represents a sizable portion of the world economy. An under-estimate of the annual repair & maintenance costs per vehicle is $300, and there are over 200M vehicles in the US alone. There has been a concerted effort in the automotive sector to reduce these costs by focusing on reducing repairs, warranty, and safety issues; there is also growing interest in “Early Warning Systems” (e.g., see www.aiag.org/scriptcontent/event_presentations/files/E6EWSC01SP/EWS_final.pdf ). Despite these efforts, repair & maintenance costs are difficult to control, since once sold, vehicles are geographically dispersed, and indirect means by way of data collected at repair points must be used to analyze and assess the issues. We focus attention on the automotive sector, although the ideas reported on here have wider applicability.
Repairs are done at a dealership when a vehicle is under warranty, or in repair-shops otherwise. Often, obvious misbehaving components get replaced even though it may not fix the root-cause. This leads to repeat failures, high costs, and “No Trouble Found” for replaced components when tested under a “Parts Return” program by the manufacturer. Industry experts characterize the associated costs as being 40% or higher considering all the vehicular repairs and maintenance costs, which is large. Certainly, significant effort is expended in analyzing warranty, repairs & maintenance data at automotive OEMs (e.g., Ford and GM) and their Suppliers (e.g., Lear Corp., ArvinMeritor, and Continental). Supplier companies tend to have 3-15 people who manually analyze data, and the OEMs have 50-500 individuals; and yet, personnel costs are small as compared to the costs due to delays in identifying problems (see [1]).

2. TECHNICAL BACKGROUND

After describing the basic datasets and the relevant analytic software, we discuss our experience and initial findings using some actual example cases—which also helps explain our approaches.

Datasets

The datasets tend to be large (e.g., 100K records, each 1-2KB), which implies that significant summarization is needed for analytics. Automotive warranty records typically have 20-50 fields. Most fields contain structured data such as dates, costs, or categories. A few structured fields provide the problem symptom, the failed component, and the failure type. The complementary text extraction improves the accuracy of these data fields using the unstructured text narrative information within the records. For system-level issues, the fields may not provide data of much direct consequence (e.g., NTF other than for the failed component). Note: In our description, metrics are assumed to be on dataset sizes normalized by the vehicle production volumes.

Figure 1: Screenshot with a subset of the concepts assigned to a warranty repair record.
Ubiquiti “extracts” domain-specific “descriptors” from available categorical and numeric fields, and also from text narratives in each record. These descriptors are assigned to each record from a reference domain-specific ontology which Ubiquiti provides; detailed ontologies are available due to client engagements. Software, rather than human analysts, extracts information from text due to high accuracy, sufficient for analytical tasks of interest. Figure 1 shows a warranty record to the left with some descriptors in an ontology red-outlined to the right. Structured metadata for each record, shown to the right, uses a domain ontology. Ubiquiti Codes are drawn from text narratives, whereas other codes and parts information is structured data sent in by the repair locations. Analytics are applied to all the structured and the unstructured information – as depicted in Figure 2.

Figure 2: With domain-specific ontology, both structured & unstructured data is analyzed.

3. Case Study 1: Root-Cause Analysis

We discuss the incorporation of specific techniques useful in detecting root-causes in many cases of interest at Continental. In doing so, we differentiate between two approaches to analytics:

1. In a classic approach, the records are organized into categories by statistical means, and each record fits into multiple categories based on its contents; thereafter, the analyst looks for the top issues by examining the overall distribution.

2. In our newer approach, the records are each categorized into a single, most-likely root-cause category, based on set of known issues for a given product; and the reason is that the results are more accurate, given complex nature of our products (and replicates manual data review).
Figure 3: Screenshot of using the classic approach to analytics.

3.1 Overview of the Product and Analysis
Our focus was on a complex product: Engine Controllers. A relevant dataset was reviewed manually to find the most-likely Root Cause category. Using the reviewed dataset as a baseline, a list of Known Issues was created. Working with Ubiquiti, logic was incorporated into software that determined the most-likely Root Cause categories. This approach was improved iteratively, and verified against manual review results.

3.2 Product and Data Details
- **Product**: Engine Controller for Automotive and Light Truck
  - **High Volume & Complex**: Controlling Engine and Transmission functions
  - **System-Critical**: Safety and Emissions-related, interfaces with the entire under-hood Electrical system
- **Data Set**: From three separate databases
  - **Claims Data**: 244 Claims related to given part
  - **Narrative Data**: 1367 Narratives related to Claim VINs (each with three Narrative fields)
• Narrative, “stream-of-consciousness” format; with misspellings, grammatical errors; seldom includes P-Codes (standard codes describing failure)

• **Analysis Data:** Sixty-seven analysis results related to Claim VINs
  • Narrative form, describing Root Cause analysis on returned parts

---

**Figure 4:** Continental’s Engine Controller product.

---

3.3 **Classic Analytics Detail**

• Ubiquiti software merged Claims & Narratives (linked data from different repositories and/or times; usually assigned codes to each individual record).

• Ubiquiti software binned each Claim into every appropriate Root Cause category (i.e., each of which was then a Candidate A-Code in the newer A-Code approach); and provided the ability to review the overall distribution of all possible Root Causes to determine the top issues. However, the classic approach did not assign codes to the linked group as a whole – which helps find the most-likely Root Cause (i.e., the A-Code in the newer approach, described below).

3.4 **Newer A-Code Analytics Detail**

Ubiquiti determines the most-likely Root Cause category (A-Code). This was done by:

• Implementing VIN History logic: assigning Candidate A-Codes to groups of Claims rather than individual Claims, choosing the groups to represent all the Claims related to a given issue; these groups were identified by combining all Claims for each given VIN, within a given range of Mileage; this mileage range was tested to see what range would provide the best correlation to the reference data set; thus, all the Claims within the group are assigned all the same Candidate A-Codes.
• Determined priority for each Candidate A-Code, based on Known Issues list (includes code name, description of issue, description of suspect population - e.g., vehicles, engines, transmissions, build dates etc.): a Candidate A-Code was assigned to each issue in the list; then, each Candidate A-Code was given a priority, from one to four, based on how likely that category represented a root cause - the higher the number, the more likely a root cause; this number was assigned by qualitative analysis of each Candidate A-Code, considering factors such as how big the known issue was, and how specifically it was defined (e.g., vehicle X, built between Y and Z, with symptoms A, B, and C).

• Implemented A-Code Logic to select A-Code based on a priority on Candidate A-Codes: if a single Candidate A-Code was found, it became the assigned A-Code; else the Candidate A-Code of highest priority was selected (and a priority conflict was set if multiple Candidate A-Codes for given Claim had the same priority); the A-Code was not assigned (i.e., set to “A-Code Not Assigned”) if no Candidate A-Code matched.

Figure 5: Screenshot of using the newer (A-Code based) approach to analytics.
3.5 Results Detail

- **Categorization Percentage**: Automatic categorization of ~67%-75% of Claims; the rest of the Claims needed manual review. Among those for manual review, there were those with “A-Code Not Assigned” (which meant Ubiquiti did not find any matching category), and “Priority Conflict” (which meant Ubiquiti found multiple possible Root Causes with the same likelihood or priority).

- **Reduced analysis cycle time**: Taken for MIS (Months-in-Service dataset) from the original monthly manual review time of two weeks to two days.

- **Categorization Accuracy**: with *Classic* approach (i.e., with all Candidate A-Codes), one of the Candidate A-Codes matched manual analysis 75% of the time; no information was available 7% of the time; none of the possibilities matched manual analysis 17% of the time; and the *Newer* A-Code logic exactly matched the most likely Root Cause for 68% of Claims.

This accuracy has shown to be enough to identify top issues and emerging issues. And analysis now takes around two days (vs. the two weeks it used to take), making it practical to analyze all Claim data available; rather than just the Return data.

The accuracy is continuously being improved, as new data is entered, and manual review takes place. The Known Issues list is a living document, and resulting Ubiquiti logic is constantly being reviewed, updated, and improved.

Next steps for this project include a continual updates of the Known Issues list and resulting Ubiquiti logic for this product; and expanding this process to products across the Division.

4. **CASE STUDY 2: RESOLVING “NO TROUBLE FOUND” ISSUE**

We discuss the case of a part, which when replaced, was returned and passed production tests – thereby suggesting a system-level issue, an adjacent component, or a diagnostic issue. The case involved *Fuel Injectors*, which when returned, tested as NTF.

- **Background of the Issue**: The *symptom* was the MIL light on, and "Misfire" indications; the *vehicle analysis* indicated that recommendations by the OEM Warranty Call Center was to *Repair & Replace* complete engine injectors sets (then returned for analysis); the *parts analysis* indicated several engine sets were NTF; *other information* included that the adjoining parts (i.e., spark plugs, ignition coils, engine controller) were also indicated.

- **Ubiquiti Analysis**: Using merged data (from two different databases), binned several thousand warranty claims identified as “warranty center” repairs; linked NTF returns with warranty center related claims; binned components repaired in the merged claim records; and exhibited the time-line of multiple repairs for each VIN.

- **Analysis Result**: Narratives revealed a “shotgun” approach to solving “Misfire” concern; data analysis revealed that injector was not the last sequential repaired component that solved the customer concern.

- **Corrective Action**: Suggest to customer warranty specialist that the “Warranty Center” stop recommending engine sets of injectors replaced for "Misfire" detection; data indicates that this recommendation was effective in reducing engine sets of injector returns analyzed to be NTF.

- **Cost impact**: The Supplier-related Cost Avoidance is estimated to be $50k / year; additionally, there is significant savings to the OEM based on reduction in repair, labor, and shipping costs.

- **Next Steps**: Create *dynamic record set* (explained in the sequel) in Ubiquiti software that automatically identifies and bins warranty records referencing “Warranty Center” text in OEM supplied verbatims; then trend and review records on regular basis.
5. CASE STUDY 3: IMPROVING QUALITY

This study involved the interaction of hardware and software over a period of time (suggested by problems that arose in an escalating warranty situation with expensive repairs). The study shows how otherwise complex issues may be handled with relative ease with appropriate data analytics, and thereby, helps resolve issues rapidly and accurately.

5.1 Overview of the Product and Analysis

- **Background of the Issue**: Symptoms indicated were significant increases in Claims for Engine Controllers beginning Sept’07; the parts analysis showed a large percentage of returned modules had shorted/damaged components related to a single node in the circuit.

- **Ubiquiti Analysis**: Merged Warranty Claim data with Return Analysis data; plot of the counts of Claims over time by Model Year helped compare distribution of Failure Codes for times before and after the onset of the issue; the Narratives and Co-occurrence mining patterns were also investigated.

- **Analysis Result**: Plot by Model Year over time showed increase in Sept’07 independent of model year; distribution analysis showed significantly more High-Side than Low-Side failure codes; pattern and co-occurrences indicated correlation between issue and Software change.

- **Corrective Action**: Suggested software change to Customer to reduce stress on module; implemented hardware changes to make module less susceptible to stress.

- **Cost Impact**: The Cost Avoidance related to this issue is estimated to be $4.8M / year.

5.2 Analysis Details

The Figures 6a and 6b show the significance of the problem pictorially; the pivot table counts and the identification of the areas of interest are as shown.

![Figure 6a: Pivot Table count of VIN_Cd (not normalized).](image-url)
That the Fault Distributions correlate with “Fault Code shift” is easily seen in Figure 7.

Figure 7: Fault Distributions correlate with “Fault Code shift”
After the initial examination and identification of the areas of interest, a second-round of studies was conducted. These studies helped to pinpoint the issues, and lead to their resolution.

Analyze in Ubiquiti as follows:

1. Identify records in vehicle history where module was replaced
2. Identify records in vehicle history where module was Re-Flushed
3. Identify vehicles where record in Step 1 was after record in Step 2
4. Separate vehicles from Step 3 into two cases:
   - Most recent Re-Flush prior to March 2007
   - Most recent Re-Flush during March 2007 or later

Figure 8a: Second round of study – data examined for the different periods.

Figure 8b: Second round of study – data downloaded and examined in software.
6. Case Studies 4: Cost Avoidance

A significant aspect of interest, particularly for automotive Suppliers, is to analyze data in order to identify the correct party responsible for a given failure issue. This is needed to ensure that the appropriate parties pay the associated costs, and this activity falls under Cost Avoidance. Now, we consider three quick cases of cost avoidance as follows.

6.1 Based on Faster Analysis

- **Background of the Issue**: Components in question was an Engine Controller, and the issue was that the monthly manual analysis of Warranty Claim data was very time-consuming.
- **Corrective Action**: Configured Ubiquiti Ontology to analyze data with less manual review.
- **Result**: Continental reduced two week analysis cycle down to two days.
- **Cost Impact**: The estimated annual cost avoidance has been $233k.

6.2 Based on NTF Resolution

- **Background of the Issue**: The product was an Engine Controller, and Continental was being charged for known system-level issues (documented in Technical Service Bulletins, TSBs).
- **Ubiquiti Analysis**: Using advanced searches of MY07 Claims data analyzed with Ubiquiti software, the detailed analysis took about 40 hours of analysis time.
- **Analysis Results**: Identified Claim population related to TSBs.
- **Cost Impact**: Lowered Continental’s Warranty Responsibility percentage by removing Claims related to TSBs from Recovery population; this resulted in an estimated annual net savings of $222k.
- **Next Steps**: Expanding Warranty Recovery data to other products.

**Figure 9**: Combined Warranty & Parts Return report – showing the Parts Return information.
6.3 Based on Adjoining Component

- **Background of the Issue**: Symptoms were “long crank times” or “hard starting” complaints; the issue was that the injectors were replaced for every vehicle with this complaint; *vehicle analysis* showed known injector failure modes (for a specific engine family), and the customer SQE had formally notified dealers to replace engine set of injectors for this complaint; and *other information* included the adjoining component Fuel Tank was indicated.

- **Ubiquiti Analysis**: Conducted rapidly (in 4 man-hours several hundred claim and narratives were imported into Ubiquiti and analyzed, with a formal report submitted to the customer).

- **Analysis Results**: Several vehicles were identified with fuel tank contamination; abnormal (heterogeneous) distribution of repairs in one particular cylinder location was found; and the fuel tank contamination associated to engine sets of returns analyzed to be NTF.

- **Corrective Action**: Based on the data and analysis presented to customer, it was agreed that the position held by Continental was valid; a reduction in claims indicated that the dealer notification for injector Repair & Replace was revoked; and removing the dealer notification eliminated unnecessary warranty repairs.

- **Cost Impact**: Estimate Cost Avoidance of up to $321k for the model year.

7. **APPROACHES FOR EARLY WARNING**

The fastest, least costly means for early warnings is by data analysis; and simple considerations help significantly. First, there should be breadth of data sources that should be utilized; second, automated means of getting alerted should be available; and third, there should be appropriate means to project costs going forward. We discuss and illustrate the first two considerations, and the third is typically available in common analytics and statistical tools already.

7.1 Experience at Continental

At Continental, before Ubiquiti software was implemented, all Claims were manually reviewed, which was tedious and often inconsistent among different analysts. The best information was derived from modules returned, and these were a small percentage of the overall Claims (i.e., about 10-20%). Since use of the Ubiquiti software system, all Claims can be reviewed quickly; charts for Claims are obtained, insight is available into 80% of Claims where modules are not returned. By manual review of Claims that are not recognized by Ubiquiti, new or emerging issues get detected. Having a means to study large data quantities is particularly useful during product launches, where the opportunity to make the most impact on Warranty is available.

7.2 Software Techniques for Early Warning

We briefly describe three basic approaches that help in early warning: first, some *Data Mining* techniques to find unexpected and emerging patterns within the data; second, *Alerts* which raise flags for emerging issues in real-time as new data is loaded; and third, *Dynamic Record Sets* which contain Claims that meet preset criteria, and need to be examined as new data is loaded.

7.2.1 Various *Data Mining* Techniques

Ubiquiti provides data mining of specific use in particular domains, and for automotive warranty, the following are commonly used. *Co-occurrences* within records (similar to “Itemset Counting” in market-basket data analysis, but with ontology-structured data) finds frequent co-occurrences considering each repair (e.g., vehicle type, Model Year, Failure Mode etc.); *VIN History Patterns* use repair “histories” for individual vehicles, and analyzes time-sequence and order of repairs; and *Distribution Differences for Datasets* compare “baseline” against “comparator” datasets based on various distributions (e.g., Geography, Mileage, Production and/or Repair Dates etc.), and indicates cases where the distribution differences is significant based on various metrics.
7.2.2 Software Alerts

Alerts are triggered when pre-set conditions are met, and are checked automatically (usually as new data arrives). Evaluation can also be initiated manually (e.g., to test Alerts). Various mining algorithms, described above, have been incorporated into the alerting techniques as well. Note – the Alerts may be set to be quite simple (e.g., “trigger if count > 100”), or fairly complex (e.g., “trigger if any combination of vehicle type and model year has more than 50% repeat repairs for a particular failure mode where the relatively frequency of failure mode is more than 3 times as high as the relative frequency of this failure mode for the same vehicle across all model years”).

Figure 10: Screenshot of Sequence Mining on VIN Histories.

Figure 11: Example Alert – Mining Thresholds used for Trigger Metric.
7.2.3 “Dynamic Record Sets”
Similar to database Views, Ubiquiti provides Dynamic Record Sets wherein a “standing query” is executed that is applied to each incoming record to see if it meets pre-determined, user-defined criteria (i.e., any that can be applied in a Search). These recordsets can be set up (to search) for data for common anticipated issues, and checked as and when such issues are detected. Examples of such issues are may be problems appearing after a “clean-date”, specific issues in particular Model, MY, Geography etc., or any combination of constraints applied in a Search valuation.

Figure 12: Any Search criteria can be set as a Dynamic record Set.

8. SUMMARY
Data analytics on warranty data provides a very significant means for reducing costs, identifying problematic issues, and can have a huge impact on the Quality, Engineering and Safety areas. Our experiences, as sketched here, provide an example of utilizing state-of-the-art techniques in applied analytics in this increasingly important area – in particular, as focus increases on vehicle repairs and maintenance as opposed to new vehicle sales. We welcome the opportunity to work with others in these areas, and we hope to learn and share more in this important endeavor.
9. Next Steps

Continental and Ubiquiti will continue our close working relationship as we enhance and expand our application of Data Analytics using Ubiquiti software. We plan to expand expertise and use of Ubiquiti tools across the Division, including but not limited to the use of Early Warning Techniques (e.g., setting up Dynamic Record Set searches and review data on a regular basis); the Root Cause Analysis process (for applicable products); and the TSB Claim Review process. As we do this, we will continue to catalog examples, and sum up the cost impact resulting from our activities.

References


