An Early Warning System for Vehicle Related Quality Data

Matthias Grabert[1], Markus Prechtel[1], Tomas Hrycej[1], and Winfried Günther[2]

[1] DaimlerChrysler AG, Research and Technology
[2] DaimlerChrysler AG, Global Services and Parts
Matthias.Grabert@DaimlerChrysler.com

Abstract. Vehicle production audit tests, warranty claims and car control unit data are stored in a central data warehouse for data mining analysis. Neural network based part failure rate estimations, adjusted for mileage and seasonality, are used for monitoring warranty claims. Association and sequence analysis connect production audit data, car control unit data and warranty claims for an early detection of quality changes both in production state and car field usage. Calculations are performed via grid computing.

1 Data Driven Quality Management

A major aspect of quality management is avoidance of quality defects. Nevertheless, 100% freedom from error in motor vehicle production and during their lifecycles cannot be achieved with these highly complex products. Therefore, a continuous quality monitoring at production time and during the lifecycle is mandatory to early address upcoming quality problems.

Vehicle service centers document repair cases with standardized codes, and all reports are later on transferred to the central claim database at least during the warranty period. Beside repair cases, protocols of car control units are read out during service center visits. Newest generation cars collect load spectrum data continuously and reports are later transferred at WLAN hotspots or via GSM to the carrier.

Data sources are of different kinds: While failure codes during warranty periods are completely documented and therefore available for all cars, only approximately 1% of manufactured motor vehicles are checked carefully directly before delivery. These production audits happen definitely earlier and allow thus a rapid identification and elimination of quality problems. Therefore, an early as well as detailed quality monitoring system optimally consists of different data sources in different lifecycle stages.

An ultimate goal of such a quality monitoring system is - beside a rapid detection of already existing problems – to give a reliable forecast of upcoming events or increasing failure numbers as early as possible.

“As early as possible” has to be seen in the context of structural conditions of each car component. Upcoming failures are mainly dependant on car age, mileage and – e.g. in commercial vehicles – usage profile.

P. Perner (Ed.): ICDM 2004, LNAI 3275, pp. 88–95, 2004.

The following data sources have been available at DaimlerChrysler for our early warning system:

- repair codes during warranty period (called in this paper “DB1”)
- messages from control unit codes (called in this paper “DB2”)
- results of production audits before delivery (called in this paper “DB3”)

An overview of the software system as well as deployed data mining methods will be given in the rest of the article.

2 Models and Methods Overview

The early warning system is working on a two-steps base (Fig.1): First, the system is trained with data from the past (usually 24 months): complex mathematical distribution models are learned for warranty claims (DB1) taking into account seasonal changes of failure rate (e.g. air-condition fails usually during summer due to heavy usage). Models are adjusted for mileage and age of the vehicles. For vehicles without warranty entries, mileage has to be estimated since no information is then available in DB1. Furthermore, associations between warranty claims and control unit messages (DB2) as well as production audit results (DB3) are calculated in step one. In a second step, current DB1 data (usually last 12 months) are tested on a monthly base for significant changes against the calculated models of step one (Fig. 2). Grid computing is used for DB1 training and testing (~ 10,000 models are processed during testing only). Furthermore, the probability of a warranty case under the condition of a pre-occurring DB3 (audit result) or DB2 (control unit message)

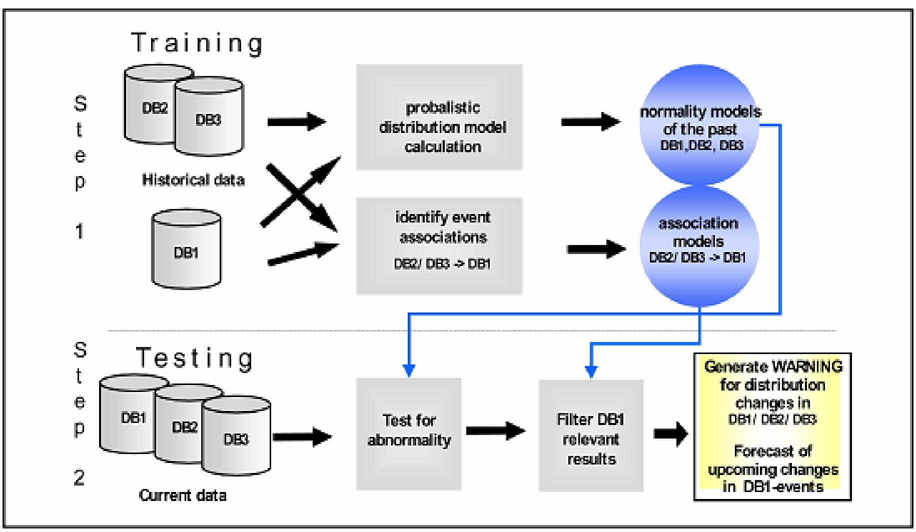

Fig. 1. Model related system architecture

event are calculated to quantify the association strengths. Significant associations are determined by a Chi-Square-Test. Warnings for the system user are then generated and displayed. Every warning is a candidate for a possible production weakness and has to be carefully evaluated for necessary consequences by a human expert.

2.1 Software Architecture and Implementation Details

Figure 2 displays the EWS software architecture. User and system administrator navigate through the password protected system via standard web browser within the DaimlerChrysler intranet. The system administrator configures the system and starts training or testing steps, when needed. Data is loaded via PERL::DBI from different data sources into a MySQL [1] based data warehouse. Motor vehicles with unreliable date of initial registration or repair date (e.g. before production date) are excluded from later analysis *(Data Cleansing*) [2]).

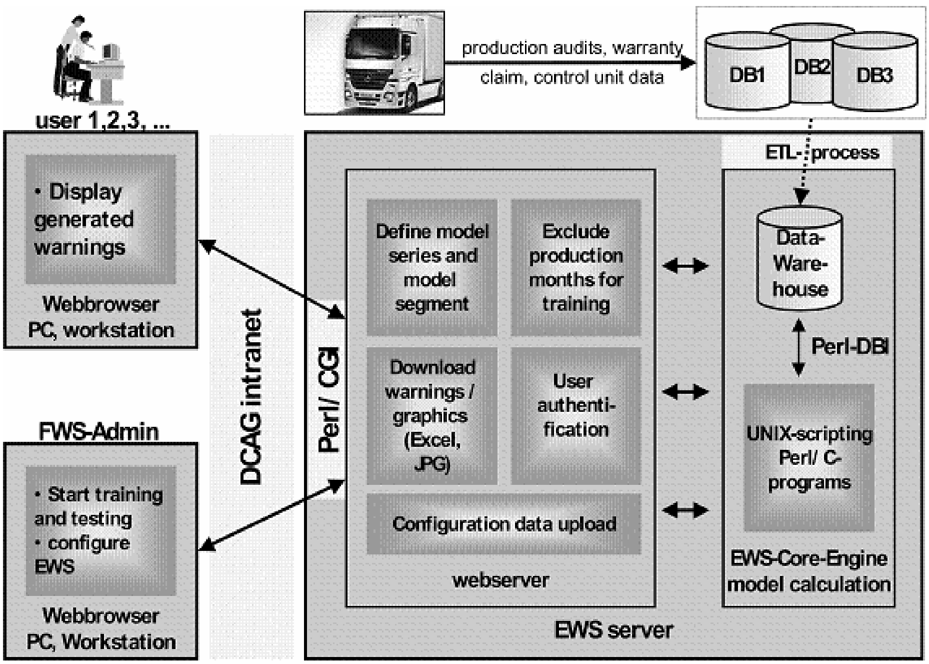

Fig. 2. Overview on system software architecture

Software components (Perl, Unix-tools, MySQL database, gnuplot to generate graphics as well as C programs for mathematical modeling) have been chosen regarding the ease of portability. The system is currently running on a two-processor HP A500 with 440MHz, 2 GB RAM under HP-UX 11.00. Time for porting to a different UNIX system (such as Solaris, Linux) is below one day.

2.2 Car Warranty Claims

Mercedes-Benz's warranty database (DB1) contains the complete repair history of every Mercedes vehicle during warranty time (at least first 12 months after initial registration, another 12 months depending on mileage). Currently data of 17 Mio. cars are collected in the database. Part failures are standardized documented during car repair using over 15,000 so called "damage codes".

The following part of the early warning system solely depends on DB1 data analysis using the above mentioned two-step architecture:

1. Calculate a mathematical model for the damage frequency of each damage code appearing in a specified model series for a given training period (typically 24 months of production).
2. Test the damage data of newly produced cars (e.g. last year of production) for significant distribution changes against the trained models (test period has to be different from training period!).

Tests are performed on a car group level. Vehicles in the group should be as similar as possible to guarantee a homogenous random sample for the stochastic modeling. Grouping criteria are:

- model series (e.g. E-Class, Sprinter, Actros)
- segments within the model series (e.g. Sprinter 2,8t or Sprinter with right hand drive)
- engine (power rating)
- model type (determined by the first 6 digits of the unique vehicle identification number)

Car groups can be defined in database tables by the system administrator, thus enabling the system to be easily adapted to new model series.

A distribution model is only built for a car group if enough failure cases are in the specified group for a specified damage code. The level of specialization is ascending for the describing criteria: After a calculation on model series level (lowest level of specialization), a calculation on segment level will be tried and so on. Specialization is aborted if no stochastic model can be calculated for a specific level (too few data).

Each stochastic model is finally evaluated by testing the model with the training data: If the trained model is not good enough to pass the test with data used for the learning phase, it will be discarded. Atypical production months (e.g. after production start of a new series) can be excluded from the training period on damage code level.

For the Sprinter model series, about 550 stochastic models are typically generated for about 90,000 cars and different damage codes during training. Training phase including data transfer from the central warranty database to the data warehouse (~100MB) takes about 6 days on a single machine. For a final model calculation, up to 600MB of prepared data have to be processed for each car group (estimation of missing mileage or date of initial registration, part failure rate for each month after registration). The parameters of the generated distribution model include no more than 500 bytes.

These long training periods could be reduced by 50% with the help of grid computing since calculation of each model is independent from another.

During test phase, typically the last 12 months of production will be tested against the trained models of the past. For each production month all available months after registration will be tested separately for a specific car group.

A warning is produced when the number of failures for a car group exceeds the model based significance level (figure 3) in a specific damage code. A 12-months test for the above mentioned Sprinter group takes about 1.5 days of computation time.

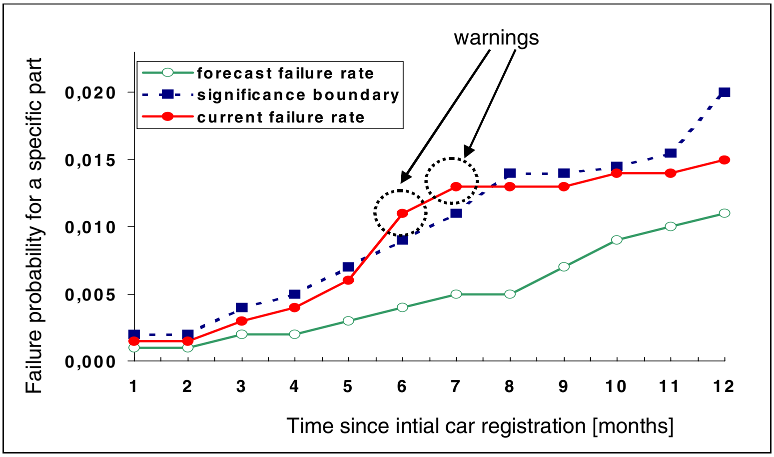

Fig. 3. The x-axis corresponds to months 1 ... 12 after initial registration. Car group, production month and damage code are fixed. The y-axis shows the cumulated failure probabilities (forecasted failure rate, current failure rate and calculated significance boundary). The current failure rate exceeds in two months the significance boundary (month 6 and 7). Thus, two warnings will be generated by the system for the specific car group and damage code

2.3 Production Quality Audit

Beside all quality monitoring performed during car manufacturing, approximately one percent of all cars are carefully audited directly before leaving the factory. Found quality issues are encoded according to the warranty damage codes and then entered into a database (DB3).

The early warning system for DB3 data should recognize significant changes in production quality and reveal a possible correlation between production failures and later on warranty cases (DB1).

An approach similar to the DB1-system is used to monitor production quality: Audit data are separated into training and testing set. Since neither time nor mileage is relevant for any failures directly after manufacturing, the failure probability can simply be estimated using the relative frequency of failures within the training period (usually several months of production). Warnings during the test phase are generated, if the observed number of failures exceeds the expected number of failures. To identify only significant changes in the failure rate, a hypothesis test based on the binomial distribution is applied.

The probability for a warranty failure under the condition of a preceding production failure is calculated to identify a possible correlation between both failure types. Significance of correlation is assured by a chi-square test [3].

2.4 Car Control Unit Messages

Modern motor vehicles like the Mercedes S-Class contain up to 60 control units to control important car functions such as the Anti-lock Brake System (ABS), Electronic Stability Program (ESP) and fuel economy. Beside the control task, these units monitor all connected devices (e.g. sensors) for correct functionality. Errors are internally stored and offer valuable diagnostic data at each garage visit. Since 1999, control unit messages are collected in a special central database (DB2) at DaimlerChrysler to early detect upcoming failures in specific morel series.

The early warning system for DB2 is built according to the second part of the production audit early warning system: The probability for a warranty failure under the condition of a preceding control unit error message is calculated and significant correlations are identified by a chi-square test. These correlations are valuable forecasts for potentially upcoming part failures and thus a further garage visit can be avoided for the client.

Unfortunately, not all Mercedes garages are connected to the DB2 databases. Therefore, not all control unit messages can be collected in the database which makes it impossible to identify changes in the error message distribution.

3 Details of Stochastic Models

Failure processes are stochastic processes and thus their appropriate representation is done with help of stochastic models.

The focus of interest is whether the frequency of newly occurred failures has increased in comparison with a reference time interval. Theoretically, comparisons with fixed reference values can be done, too. In practice, this fails because of the number of different failure types (about 15000), for which such reference values would have to be specified for each model series. Furthermore, the variability of the vehicle mission (described for example by the vehicle age, monthly mileage rate, or season of the year) would have to be taken into account. So a certain failure rate can be normal with a monthly rate of 20,000 kilometers while it is too high with a monthly rate of 2,000 kilometers. This makes the comparison with a reference time interval the only viable alternative.

In statistics, concepts for quantification of statements about the differences between two data sets have been elaborated. This quantification takes place in terms of probabilities. A typical instrument, frequently used in quality control, is the hypothesis test. In our case, it is hypothesized that the quality has not changed. In the statistical language, this corresponds to the assumption that the frequency of the newly occurred failures can be explained by the failure probability distribution observed in the reference interval.

To be able to do this, the reference probability distribution has to be represented by a model received directly from the data. For this, various distribution classes have been proposed (Weibull distribution, Gamma distribution, Fatigue Life distribution

[4]). Unfortunately, none of these standard distributions is flexible enough to capture the abundance of the really observed failure distributions, two examples of which are presented in Fig. 4.

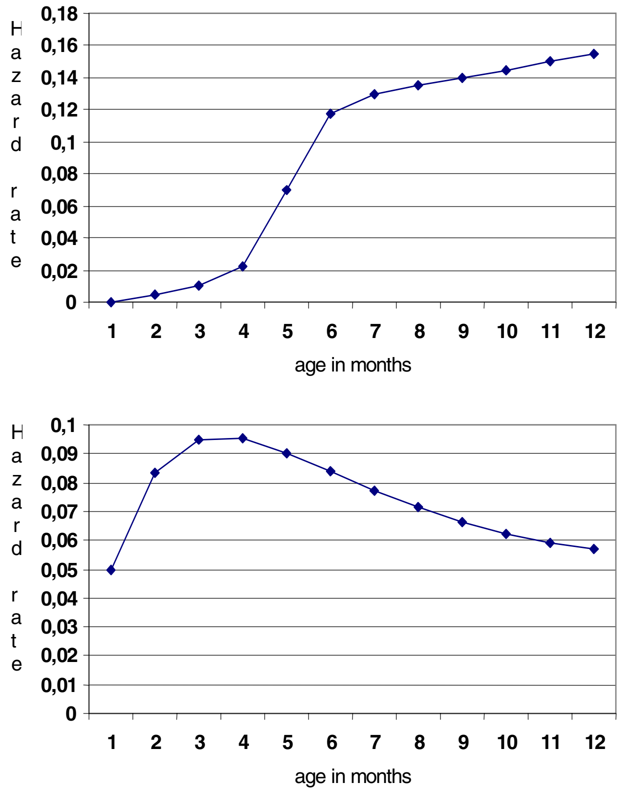

Fig. 4. Observed hazard rate, i.e., the failure probability at month T of vehicle age, conditional on this failure not having occurred until month T-1, for two hypothetical, but realistic failures types

The charts depict the curve of the observed hazard rate, that is, of the conditional probability that a failure occurs at month of age T if it had not occurred in the months 1 to T-1. The Weibull distribution can express only hazard rates proportional to gT^{g-1}, that is, depending from a power of T. The hazard rates of Gamma distribution are proportional to $T^{g-1}\ e^{-T}$. Furthermore, there is a scaling parameter. Also the Fatigue Life distribution has merely two free parameters. None of these distribution classes can reasonably approximate the two cases of Fig. 4.

It is important to point out that that the approximation of the failure distribution in the early warning system has to be considerably accurate. Otherwise, the test criterion might be violated merely through the approximation error even if applied to statistically "normal" data, launching a false alarm.

These high requirements are satisfied by no classical distribution family. This is why we adopted a numerical approach. The dependence of the hazard rate is directly approximated by the multi layer perceptron (a widespread type of neural net). In this way, we gained the generality not to be limited to a single dependence from time (as would be the case with standard distributions). Rather, multiple characteristics such as mileage rate and season of the year, whose effect on the failure rate can be expected, could be included. So a particularly accurate statistical representation of the normal failure process has been attained. The parameters of a multilayer perceptron are numerically estimated from the failure data minimizing the cross entropy of the observed and the theoretical distributions [5].

Beside this, the approach can be applied to monitor really occurring failures that had lead to warranty claims. But this framework is equally applicable to indirect quality indicators. For example, quality problems observed in the manufacturing outlet control can be tested for increased frequency. To quantify the relevance of these problems for warranty costs, a model of relationship between these indirect indicators and the warranty claims is needed. In the simplest case, such a model can consist of empirical conditional probabilities

$$P(\mathit{failure} \mid \mathit{outlet\ problem})$$

To receive significant results, statistical tests such as the Chi square test for independence on the contingency table can be used.

If an outlet problem occurs with an abnormal frequency **and** this problem is statistically associated with a subsequent warranty claim, **then** a corrective action is required.

4 Conclusion

Through a joint analysis of the databases DB1, DB2 and DB3, the chain from manufacturing over online diagnostics to really occurring failures is closed. The product quality can be monitored simultaneously on several interconnected levels. The system has been in test operation since the end of 2003.

References

1. DuBois, P.: MySQL & Perl Developer's Guide: Webanwendungen programmieren mit Perl DBI, mod_perl und CGI, Markt&Technik Verlag, München, 2002.
2. Kimball, R. et al.: The Data Warehouse Lifecycle Toolkit. Wiley, 1998.
3. Hartung, J.: Statistik, R. Oldenbourg, München, 1987.
4. Stuart, A.; Ord, K.: Kendall's Advanced Theory of Statistics, Volume 1: Distribution Theory, Arnold, London, 1994.
5. Stützle, E.; Hrycej, T.: Forecasting of conditional distributions - an application to the spare parts demand forecast, Proc. 2001 IASTED Int. Conf. on Artificial Intelligence and Soft Computing, Cancun, 2001.