DOI : 10.3745/JIPS.2008.4.3.113

Decision of Abnormal Quality Unit Lists from Claim Database

Sang-Hyun Lee*, Sang-joon Lee, Kyung-li Moon*** and Byung-Ki Kim***

Abstract: Most enterprises have controlled claim data related to marketing, production, trade and delivery. They can extract the engineering information needed to the reliability of unit from the claim data, and also detect critical and latent reliability problems. Existing method which could detect abnormal quality unit lists in early stage from claim database has three problems: the exclusion of fallacy probability in claim, the false occurrence of claim fallacy alarm caused by not reflecting inventory information and too many excessive considerations of claim change factors. In this paper, we propose a process and methods extracting abnormal quality unit lists to solve three problems of existing method. Proposed one includes data extraction process for reliability measurement, the calculation method of claim fallacy alarm probability, the method for reflecting inventory time in calculating claim reliability and the method for identification of abnormal quality unit lists. This paper also shows that proposed mechanism could be effectively used after analyzing improved effects taken from automotive company's claim data adaptation for two years.

Keywords: *Claim, Production, Abnormal Quality, Decision Support*

1. Introduction

With increasing demand of client about high quality service and augmentation of related regulations such as PL(Product Liability), the importance of claim management is more emerged in current environment[9]. Claim means a kind of a complaint from clients when products did not satisfy client's needs with their function, quality and the delivery date [10].

The systematic and efficient claim management system based on accurate data analysis could lead to computerized claim management in analyzing the causes, cost, complaints of claim and alternatives. And finally, this claim management system leads to business automation and reduction of claims [6].

Most enterprises have controlled claim data related to marketing, production, trade and delivery. Claim management intending to prevent recurrence of claim could be divided into two categories: one is to take measure against each claim and the other one is to use claim information system effectively. In manufacturing industry, by using claim information system each company can establish more robust production line, increase satisfaction and assurance of clients. In addition, by using claim analysis, companies can extract engineering information such as unit reliability from claim data and detect critical reliability related problems in early stage [10].

Most manufacturing companies have faced and will continue to face serious reliability problems, most often caused by one or some combination of the following: an unanticipated failure mode, harsher than expected operating environment, an unknown change in raw material properties or supplier quality, an improperly verified design change, etc[12].

To solve above problems, companies should build up claim information system and claim database by introducing the concept of reliability by design prior to goods manufacturing. They could avoid serious claims and reliability matters in the manufacture of goods through conducts pre-verification by identifying random failure units and preparing claim change of units.

This paper is organized as follows. Section 2 describes related works and section 3 proposes abnormal quality claim list extracting method. Section 4 delineates extracting VIN (Vehicle Identification Number) lists comparable to unit identification list in the automobile industry. Section 5 includes conclusion.

2. Related Work

In terms of claim matters, building up an early claim alarm system refers to a sort of hue Warranty Planning System that needs introduction of AI (Artificial Intelligence)

Manuscript received August 23, 2008; accepted September 21, 2008.
Corresponding Author: SangHyun Lee
* Department of Computer Engineering, Chonnam National University, Yongbong-dong, Buk-gu, Gwangju, 500-757, Korea (leesang64@gmail.com, bgkim@chonnam.ac.kr)
** Department of Business Administration, Chonnam National University, Korea (s-lee@chonnam.ac.kr)
*** Department of Computer Engineering, Honam University, Korea (kimoon@honam.ac.kr)

technology and the like for matters of system awareness based on reliability problems(percent defective tolerance, error variation, control limit deviance, etc.) and inference matters like judgment of risk index. There are many studies conducted in terms of this development: Data mining modeling in the automotive industry[4], claim process modeling of the automobile[3], software cost model for quantifying the gain and loss associated with claims[11], software-based reliability modeling[5].

The study of extended warranty repair strategies is performed[1]. It is associated closely with the concept of extended warranty repair in terms of claim matters. It shows that applying the quasi-renewal process to find out critical points of claims and deciding the extended duration of warranty repair based on these points leads to better reliability of goods and actually has more cost-saving effects on dealing with claims.

In regard to building up an early claim alarm system, a variety of topics relevant to reliability of goods is introduced[2]. For matters of reliability in close relations with claim data, possible forecast methods is proposed[6]. How to detect change points using adjacent distribution of warranty data in terms of identifying sub-units of product serviceable during the warranty period is proposed[7,8].

Existing method which could detect abnormal quality unit lists in early stage from claim database has three problems. The first one is the exclusion of fallacy probability in claim and the second one is the false occurrence of claim fallacy alarm caused by not reflecting inventory information. The final one is the too many excessive considerations of claim change factors.

3. Abnormal Quality Claim List Extracting Process

The process of claim analysis is a one kind of data mining for reliability analysis, which is equivalent to reliability knowledge extracting process. The important thing among reliability knowledge is decision-making using information included in each cell of claim information table. The main problem is that quality guaranty cost has been gradually increased since the decision-making fallacy resulted from the uncertainty of information included in each cell has been overlooked. In reliability analysis, phenomena against defect could be regarded as random variable, and we can make probability distribution. From such viewpoints, we could regard the calculation of claim fallacy alarm probability by statistical hypothesis verification as the first step. The second step is a claim table reconstruction process through an appropriate assumption of inventory period. In reality, claim table construction based on production and marketing amount with the ignorance of a point of inventory time leads to much difference in calculating probability distribution for it makes the difference between production amount and marketing amount remarkable. Final step is to extract abnormal quality unit identification lists and the most important thing is to identify that lists among bad quality units.

3.1 Data Extracting Process for Claim Reliability Measurement

The reliability of claim data could be measured by operation group and production group. Data flow for measuring reliability in each operation group is shown in Fig. 1. The applied unit contains unit code, previous models, applicability of previous models and point of application. New units interlock the number of units and that of claims with each other in the previous model to extract data. Unit sales master includes unit number, classification code, model code, model index and point of production and sales. OP(OPeration)-specific claim

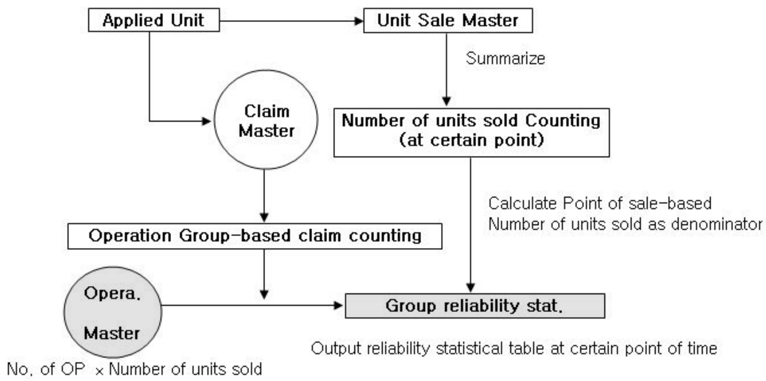


Fig. 1. Flow chart of reliability measurement data in each OP group

counting typically calculates with duration of use and the number of claims, and total sales amount. The purpose of the OP master is to multiply the number of individual OP's that comprise the OP group by the number of units sold for the sake of correcting defective percentages. Group reliability statistics include model codes, OP codes, points of production, and the shape and scale parameter of Weibull distribution as the reliability function.

The data flow chart for measuring reliability in each part group is illustrated as shown in Fig. 2. To estimate both shape and scale parameters for OP group, it is required to extract data. Part master includes unit classification, part group, part code and part name. PG(Program Group)-based claim counting includes unit classification, model code, model index, group code, duration of use at point of sales, and number of claims.

3.2 Claim Fallacy alarm Probability Calculation

To estimate the probability of claim fallacy alarm, suppose that n_i is the number of units produced at time point i; n_{ij} is the number of units produced at time point i and sold at time point $i+(j-1)$; and R_{ijk} is the number of warranty reports serviced at k th point of time with regard to certain claim codes under consideration. The decision-making process of an early claim alarm system works based on the frequency of claims in the past. It is necessary to allocate a power function that allows us to detect questions in regard to service life of different units and it is also possible to focus on suspected claim codes, if applicable.

Suppose that R_{ijk} is subject to Poisson distribution with independent parameter, n_{ij}, λ_k. Here, λ_k means the amount of units reported during k th serviceable period for certain claim code under consideration. λ_R^0 equivalent to the reference value of λ_k may be obtained on the basis of reports in past. The process of detecting lower reliability refers to testing a hypothesis $\lambda_1 \le \lambda_1^0, \lambda_2 \le \lambda_2^0, \cdots, \lambda_M \le \lambda_M^0$. Here, M refers to future period of reporting units noted and is also predetermined value. In terms of overall error alarm rate, the increase of M must turn to decrease in the aspect of power function. Since R_{ijk} is independent of $R_{ijl}(l \ne k)$, the test for hypothesis $\lambda_1 \le \lambda_1^0, \cdots, \lambda_M \le \lambda_M^0$ may be conducted in form of individual test. The test for a hypothesis $\lambda_1 \le \lambda_1^0$ (corresponding to report during the first serviceable period) can come to conclusion of $\lambda_1 > \lambda_1^0$ at time point $i+1$, if $R_{i11} \ge C_{i11}$ for appropriate critical values. During follow-up period, additional information can be still be accumulated on the basis of this information. In general, if $S_{ij1} \ge C_{ij1}$ at time point $i+j$, we can come to the conclusion of $\lambda_1 > \lambda_1^0$. Here, S_{ij1} means the sum of $R_{il1}(l=1,\dots,j)$ values. That is, S_{il1} is equivalent to the cumulative frequency of claims reported during the first serviceable period in terms of units manufactured at time point i. The probability of TYPE Ⅰ error alarms can be expressed in the following equation:

$$\alpha_1^* = 1 - p[s_{i11} < c_{i11,} \cdots, s_{iM1} < C_{iM1}] \qquad (1)$$

This is less than or equal to the probability of actual Class 1 errors. At time point $i+(j-1)+k$ (i.e., j th time point after information is available at service point k), a hypothesis $\lambda_k \le \lambda_k^0$ can be accepted, if $S_{ijk} \ge C_{ijk}$. Thus,

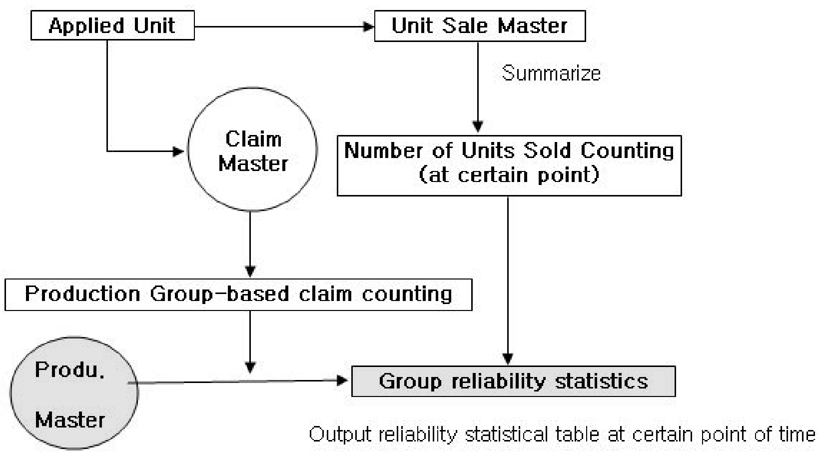


Fig. 2. Flow chart of reliability measurement data in each part group

TYPE Ⅰ error can be expressed in the following equation:

$$\alpha_k^* = 1 - P[S_{j1k} < C_{i1k}, \cdots, S_{i,M-K+1,K} < C_{i,M-k+1,k}] \le \alpha_k \qquad (2)$$

The probability of error alarms according to test of hypothesis based on independence assumption can be calculated as follows:

$$\alpha^* = 1 - \prod_{K=1}^{M}(1-\alpha_K^*) \le 1 - \prod_{K=1}^{M}(1-\alpha_K) = \alpha \qquad (3)$$

Before determining α_k and critical values, the monitoring of specific unit claim codes with these concepts applied can be outlined as shown in Fig. 3. This figure shows alarms arising during the 1st serviceable period from units manufactured at the 2nd production point. As the point of production passes by, monitoring goes on, but something terminates. As shown in this figure, "+" indicates S_{ijk} as the number of cumulated reports, and "–" indicates C_{ijk} as the critical value corresponding to S_{ijk}. In terms of significance level α and M (M =4 as shown in figure), $[i(i-1)]/2$ charts can be drawn at point of production, $i(i \le M+1)$. And up to $[M(M+1)]/2$ monitoring charts can be considered in regard to the point of production totaling $M+1$.

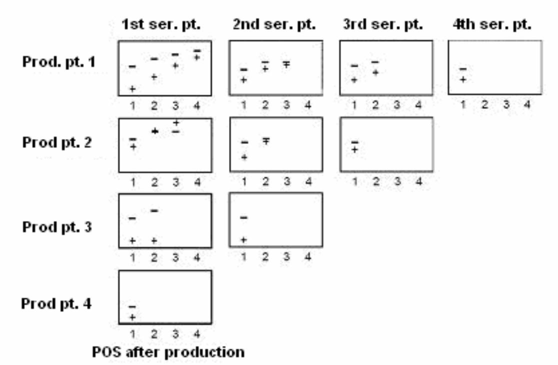


Fig. 3. Monitoring of unit claim codes

3.3 The Method for Reflecting Inventory Time in Calculating Claim Reliability

Until now, the construction of early claim alarm system has focused on the calculation of the probability of pure claim fallacy alarm by using probabilistic approach. In that method, claim fallacy alarm could be wrong when inventory is not reflected to the calculation. That problem could be improved by reflecting inventory amount as well as production amount during claim table reconstruction.

Most of reliability-based claim analysis algorithms are based on inventory periods subdivided in certain point of time. But the matter is that if we run the algorithms in such a way, there is actually less number of units produced and sold according to the point of inventory even with only one claim received, so that the value of $1-\Pi\alpha^*$ is converged on 0. To prevent this convergence, it is necessary to determine the optimal inventory period and thereby classify whole production-to-sales intervals into appropriate inventory points. In particular, the purpose of the reliability-based information system is to extract information on correct point of production and use of a good in regard to any abnormality of units, and thereby provide certain indicative information for the system that supports product services. So it is necessary to build up an information system for extracting an identification list of bad quality units.

The optimal inventory period can be determined in the following procedures: first, input the interval of inventory period and the least unit of each interval, and increase the least unit by an increment of 1. At each step, calculate skewness and determine the interval of inventory period at point of least skewness value to be the optimal inventory period. Fig. 4 shows a process of determining the optimal inventory period, if the interval of inventory period is 3 and the least unit of each interval is 7.

3.4 Identification of Abnormal Quality Unit Lists

Claim change is affected by many factors, and we need

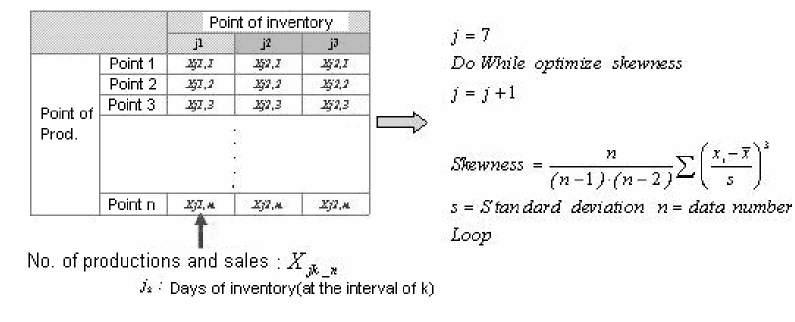


Fig. 4. Process of determining the optimal inventory period

to extract reasonable abnormal quality units by calculating pure claim change. Claim change can be divided into variation caused by positive effects and that caused by accident, while the change of positive effect can be classified into change caused by point of use and that caused by point of production.

The method for identifying abnormal quality unit lists based on reliability includes management of fraction defective, information system implementation related to a point of production time, calculating power function based on claim report frequency, and prediction of fraction defective calculated by the poisson distribution. That is, if differences in parameter λ based on previous and current claim data are calculated by the product of significance level α, and the final significance level exceeds the limit of certain significance levels, there is an identification list of units available in regard to the point appropriate as shown in Table 5(e.g. point of production/use).

Particularly, it is essential to provide an identification list of bad quality units for such units. The list of identified bad quality units based on positive effect analysis can be obtained through the following decision-making system:

$$\text{IF } \sum C(R_{ijk}) > \alpha^*$$

As an intermediate step for extracting more accurate abnormal quality unit identification lists, following improvement effect analysis could be conducted.

IF (AV_Ratio > threshold value) and (average of pre-improvement > average of post-improvement) then positive effects=Yes, else positive effects=No. Here, AV_Ratio refers to net variation of positive effects in terms of claim variations; it is the ratio of reproducibility variation to claim variations. And threshold value refers to a value offered by users. For example, it is usually set around 30% in automobile makers. Once it is judged that there is any positive effect of improvement, we can convert all alarm items not longer than 6 months after the alarm arose in those of appropriate class, and simply output an identification list of units before improvement. But if it is judged that there is no positive effect, keep all the alarm items as they are and just output an identification list of units.

4. Case Study: Application of Automotive Company

This section introduces the practical cases of applying the extraction of identification list of bad quality units as proposed herein to the extraction of bad quality VIN lists from claim information systems in the automobile industry. To extract bad quality VIN lists, this study used claim data arranged for report to Motor Company in the Korea. In particular, claim data related to pad kit – front disk as parts of front brake system of car; dating from September 2004 to August 2005. Table 4 shows claim data loaded to suit the point of production and sales and the serviceable point with regard to pad kit – front disk brake. By estimating parameters of the Weibull distribution, it is found that m = 2.0056, η = 44.6, and the intercept value of least square regression line, i.e. t_0 = 2031.98. By using these parametric values, we can get average life cycle = 11 months, the viability in 3 years (t =36) = 47.8 % and time of multiple failures = 2.6 years

Poisson probability detection is disadvantageous in a sense that when there are a few number of units sold even without any claim, $1-\prod\alpha^*$ becomes zero, as discussed in above Section 3. To complement this demerit, this study applied the Z-test method for failure rate in parallel with the Poisson test so as to calculate the optimal interval of production and sales based on inventory point in regard to 1- to 2-month period of use, see Fig. 6.

Fig. 7 shows output of standardized failure rate data and the results of positive effect analysis for a portion marked with dotted lines as well. Data available at point of improvement (a portion as shown with arrow mark) in Fig. 7 is considered contaminated and at once is excluded from the analysis. In Fig. 8, the results of output show that the variation of reproducibility as net positive effect amounts to 52.92%, which indicate improvement in more than half the whole production process of units.

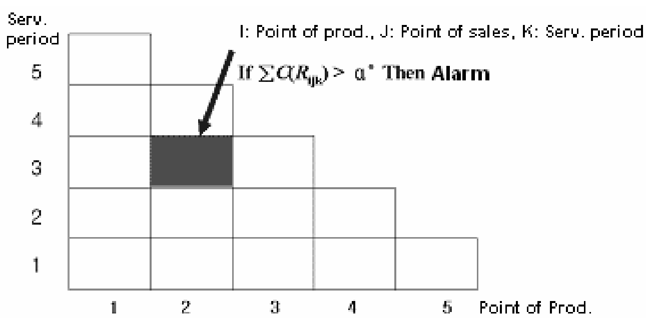


Fig. 5. Unit identification list

Ref. failure rate : λ_0	period of use	Statistios	sep. 04			oct. 04			nov. 04		
0.000793	2nonths	No. of claims	8			2					
		Z-value	2.45	-1	-1.1	-1.4	-0.7	-0.9	-1.9	-1.1	-1.6
		Z_Δ	0.01	0.83	0.87	0.92	0.76	0.8	0.97	0.87	0.95
		$Z_{S\Delta}$	0.00518			0.56677			0.80145		
		Poisson %	0.99	0.39	0.29	0.10	0.60	0.48	0.03	0.27	0.07
		P_Δ	0.01	0.61	0.71	0.90	0.40	0.52	0.97	0.73	0.93
		$P_{S\Delta}$	0.003915			0.187245			0.654592		
0.000793	1month	No. of claims	1						2		
		Z-value	-0.6	-0.7	-0.8	-1.6	-0.5	-0.6	0.2	-0.8	-1.2
		Z_Δ	0.71	0.76	0.78	0.95	0.7	0.73	0.42	0.79	0.88
		$Z_{S\Delta}$	0.42354			0.48097			0.29185		
		Poisson %	0.48	0.62	0.54	0.07	0.77	0.69	0.75	0.52	0.26
		P_Δ	0.52	0.38	0.46	0.93	0.23	0.31	0.25	0.48	0.74
		$P_{S\Delta}$	0.091682			0.066283			0.090036		
Point of inventory			7Days<	14days<	14days>	7Days<	14days<	14days>	7Days<	14days<	14days>
No. of productions and sales			4353	1186	1547	6656	647	926	4317	1643	3338
			7086			8229			9298		

Fig. 6. Calculating the optimal interval of production and sales

	2004-09	2004-10	2004-11	2004-12	2005-01	2005-02	2005-03	2005-04	2005-05	2005-06	2005-07	2005-08
Use 12 Month	0.015%											
Use 11 Month	0.015%	0.014%										
Use 10 Month	0.021%	0.044%	0.014%									
Use 09 Month	0.014%	0.032%	0.021%	0.015%								
Use 08 Month	0.000%	0.028%	0.023%	0.020%	0.011%							
Use 07 Month	0.002%	0.017%	0.017%	0.030%	0.036%	0.000%						
Use 06 Month	0.011%	0.015%	0.011%	0.022%	0.022%	0.027%	0.013%					
Use 05 Month	0.000%	0.014%	0.018%	0.011%	0.010%	0.041%	0.039%	0.000%				
Use 04 Month	0.004%	0.000%	0.006%	0.009%	0.004%	0.017%	0.038%	0.037%	0.025%			
Use 03 Month	0.000%	0.000%	0.004%	0.000%	0.005%	0.007%	0.025%	0.012%	0.054%	0.022%		
Use 02 Month	0.038%	0.007%	0.023%	0.000%	0.015%	0.011%	0.024%	0.017%	0.110%	0.094%	0.056%	
Use 01 Month	0.000%	0.036%	0.011%	0.000%	0.014%	0.020%	0.022%	0.032%	0.029%	0.000%	6.900%	0.054%

Fig. 7. Positive effect analysis and relevant statistic value

	Before improvement				After improvement				Part Avr.
	Month of use (#1)	Month of use (#2)	$\overline{x_a}$	R_a	Month of use (#1)	Month of use (#2)	$\overline{x_b}$	R_b	
Month of prod (unit#1)	0.032	0.017	0.0245	0.0150	0.069	0.056	0.0625	0.0130	0.0435
Month of prod (unit#2)	0.029	0.110	0.0695	0.0810	0.054	0.044	0.0489	0.0102	0.0592
Basic Statistics	$\overline{\overline{x}}_a = 0.0470$ $\overline{R_a} = 0.0480$				$\overline{\overline{x}}_b = 0.0557$ $\overline{R_b} = 0.0116$				
Analysis Statistics	▸ $EV_{repeatibility} = 5.15\sigma^2{}_{repeatibility} = 5.15 \cdot (\overline{\overline{R}} / d_2) = 5.15 \cdot \frac{(0.0298)}{1.128} = 0.1360$ ▸ $AV_{reproductivity} = 5.15\sigma^2{}_{reproductivity} = \sqrt{(5.15 \cdot \overline{X}_{diff} / d_2^*)^2 + reproductivity^2 / nr}$ ▸ $R\,\&\,R_{repeatibility+reproductivity} = \sqrt{EV^2 + AV^2} = \sqrt{0.1360^2 + 0.092^2} = 0.164$ ▸ $PV_{product} = 5.15\sigma_{product} = Rp / d_2^* = \frac{(0.592_{Max} \sim 0.435_{Min})}{1.411} = 0.057$ ▸ $TV_{total} = 5.15\sigma_{total} = \sqrt{PV^2 + R\,\&\,R^2} = 0.174$ ▸ $\boxed{\%AV = \frac{AV_{reprod}}{TV_{total}} \times 100(\%) = 52.92\%}$								

Fig. 8. Analysis statistics

The output results of bad quality VIN show that all the units produced appeared in ‘alarm’ status at certain point just after 7-month use, which indicates considerable faults in reliability. The unit with this problem detected resulted in serious loss of cost including warranty repair based on claims. Follow-up ongoing process improvement and more has contributed to a remarkably better quality of units produced since October 2004.

5. Conclusion

In this paper, we conduct research about claim information system structure related to reliability in claim management system development and improvement method of research related to early detection of abnormal change.

The research for improvement method is conducted by applying existing research resulted from claim data of automotive company during last two years and by complementing exposed problems. Proposed abnormal quality unit lists extracting method shows more reasonable result when we applied it claim data of automobile components.

In the future research, we plan to systematize it as an early detection alarm system for abnormal quality.

Reference

[1] Bai, J., & Pham, H., “Repair-limit Risk-free Warranty Policies with Imperfect Repair”. IEEE Transactions on Systems, Man, and Cybernetics, Part A, Vol. 35, No. 6, 765-772, 2005

[2] Blischke, W. R., & Murthy, D. N. P., “Warranty Cost Analysis”. New York: Marcel Dekker, 1994.

[3] Hipp J., & Lindner, G., “Analysing Warranty Claims of Automobiles an Application Description Following the CRISP-DM Data Mining Process”. Proceedings of 5th International Computer Science Conference, Hong Kong, 31-40, 1999

[4] Hotz, E., Grimmer, U., Heuser, W., Nakhaeizadeh, G., & Wieczorek, M., REVI-MINER: “A KDD-environment for Deviation Detection and Analysis of Warranty and Goodwill Cost Statements in Automotive Industry”. Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, San Francisco, 432-437, 1999

[5] Jeske, D. R. & Zhang, X., “Some Successful Approaches to Software Reliability Modeling in Industry”. Journal of Systems and Software, Vol. 74, No. 1, 85-99, 2005

[6] Kalbfleisch, J. D., Lawless, J. F., & Robinson, J. A., “Methods for the Analysis and Prediction of Warranty Claims. Technometrics”, Vol. 33, No. 3, 273-285, 1991

[7] Karim, M. R., Yamamoto, W., & Suzuki, K., “Change-point Detection from Marginal Count Failure Data”, Journal of the Japanese Society for Quality Control. Vol. 31, 318-338, 2001

[8] Karim, M. R., and Suzuki, K., “Analysis of Warranty Claim Data: A Literature Review”, International Journal of Quality & Reliability Management, Vol. 22, No. 667-686, 7 2005

[9] Kastetter, T. E., “Quality Concepts and Products Litigation”, TQM Magazine, Vol. 11, No. 4 (1999) 264-274.

[10] Nagoya QS Research Group, “Claim Management”, Japanese Standards Association, Tokyo, 2001.

[11] Teng, X. & Pham, H., “A Software Cost Model for Quantifying the Gain with Considerations of Random Field Environments”. IEEE Trans. Computers, Vol. 53, No. 3, 380-384, 2004.

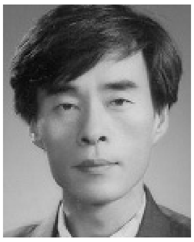
[12] Wu, H., & Meeker, W. Q., “Early Detection of Reliability Problems Using Information from Warranty Databases”, Technometrics, Vol. 44, No. 2, 120-133, 2001.

Sang-Hyun Lee

He received the B.S. degree in Computer Science from Honam University, Gwangju, Korea in 2002 and the M.S. degree in Computer Science from Honam University, Gwangju, Korea in 2004. He is a Ph.D. candidate majoring in Computer Science Engineering. His research interests include artificial intelligence, Software Engineering, Early Warning System, claim analysis.

Sang-Joon Lee

He received the BS, MS and Ph.D. degrees in Computer Science and Statistics from Chonnam National Univ. in 1991, 1993 and 1999, respectively. From 1995 to 2006, he was in Seonam Univ. and Shingyeong Univ. as assistant professor. Since 2007, He has been with Chonnam National Univ. as an assistant professor in the School of Business Administration. His research interests include Management Information Systems, Software Engineering, IT Service and Ubiquitous Business.



Kyung-li, Moon
Ph.D, is a professor at the Department of Computer Engineering, Honam University in Gwang-Ju, Korea. His theoretical work began at Seoul University as a statistical computing scientist, and then expanded into complexity science, chaos theory, and cognitive science – “generative” sciences.

Byung-Ki Kim
He received the BS and MS in Mathematics from Chonnam National Univ. in 1978 and 1980, respectively. He received Ph.D. degrees in Computer Science from Chonbuk National Univ. in 2000. He has been a professor at Chonnam National Univ. since 1981. His research interests are Software Engineering and IT Service.