Proceedings of the World Congress on Engineering and Computer Science 2018 Vol I
WCECS 2018, October 23-25, 2018, San Francisco, USA

Close the Last Gap of R&D in Automotive Design

Yuan Li, Zhang Wang, Lei Wang, Jiaying Lv

Abstract -- **Although there are many design processes and design methodology progresses in automotive design in the last few decades by using computer technology such as CAD, CAE, PDM, PLM and etc., one the real gaps between customers and the vehicle designers still exists. That is, the vehicle OEMs still don't know what exactly customers want and feel about their products in real time and how to meet those demands. We used tools like VOC, QFD, CRM, marketing research, dealer visiting, warranty database, but all these tools and data lack of real time, data infidelity, small sampling pool and high cost. In this paper, we present a technology that help on closing the gap by using the Internet+ technology or the big data analysis technology. Specially, we utilized and advanced the big data analysis technology to help the engineers on vehicle system design rather than a high level public sentiment assessment. The process and the computer system we developed are also described.**

Key words: Automotive, Data acquisition, Data analysis, Text mining, Information extraction

I. Introduction

Although there are many design processes and design methodology progresses in automotive design in the last few decades by using computer technology such as CAD, CAE, PDM, BOM, PLM and etc., one the real gaps between customers and the vehicle designers still exists. That is, the vehicle OEMs still don't know what exactly customers want and feel about their products in real time and how to meet those demands. We used tools like VOC, QFD, CRM, marketing research, dealer visiting, warranty database, but all these tools and data lack of real time, data infidelity, small sampling pool and high cost. They require manual intervention and hardly an automated process.

Nowadays, however, with the advancement of the internet and mobile technology, it's the habit of vehicle buyers and consumers to express their comments on various vehicles, show off their vehicles and most of the time publish their complains about their vehicles in the internet sites such as Autohome (40% market share of the vehicle online shopping sites in China), Yiche, Renrenche, Weibo (Microblog), Weichat, Youtube and etc. in the format of text and photo in chat rooms, user group discussion rooms and etc. With close integration to these websites, we can further dig down to more specific information such as buying intents and budget. Today, we simply call them cloud information [1], [2]. This information is valuable to OEM if obtained. It contains indefinite information about the car quality feedback and complains of existing buyers, the appearance and styling that delighted the shoppers, option preference and etc. The information can be very specific to certain model and reflects the true feeling. When used well, it can not only alert the OEMs about the immediate quality crisis and malfunctions but also be used to identify future vehicle trends. It's valuable information for vehicle planning people and engineers for their future design and the vehicle annual upgrades. It can be used for most of vehicle development phases such as strategic intention, strategic concept and all the way to design frozen as show in Fig. 1. The cloud information has the characteristics of large sample pool, fast, cost free and genuine. Traditionally, we need spend time and money to gather these information from vehicle yearbooks, industry reports, company internal analysis reports, warranty database and etc. however, most of the time, these information are small in sample quantity and slow in response time. Therefore, how to use well the cloud data may actually define how close companies to their customers and the future of these companies. In fact, based on our experience, it also can point the direction that is hard to provide in traditional process and sometimes presented as a surprise to the company.

In this paper, we present our efforts and methods on utilizing the data. Specially, we utilized and advanced the big data analysis technology to help the engineers on vehicle system design rather than only a high level public sentiment assessment. The process is almost an automatic process and can be integrated with other computer systems such as ECM (engineering change management, BOM (Bill of Material), PDM (Product Data Management) and etc. [3]

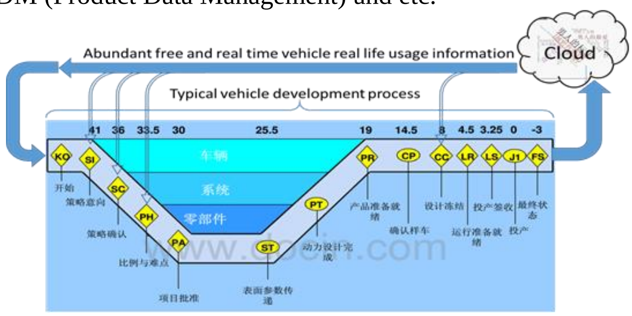

Fig 1: Abundant free and real time vehicle real life usage information can be utilized

By the integration of these systems, it not only streamlined the design change process but also provided the opportunities to utilize the Linked Data [7] property in data in these systems which strengthened the Key Words searching scope. The

Manuscript received April 21, 2018; revised May 21, 2018.

Yuan Li, member of IAENG, SAE, and CAA, is with the Beijing Motor Company, No. 99 Shuanghe Street, Renhe Town, Shunyi District, Beijing, China (mobile phone: +86-18500293291, office phone: +86-10-56637112; e-mail: 18500293291@163.com).

Zhang Wang, Lei Wang, Jiaying Lv are also with the Beijing Motor Company No. 99 Shuanghe Street, Renhe Town, Shunyi District, Beijing, China.

existing systems in general exchange data in an encrypted data stream [8], this requires the data be decrypted process when systems were integrated.

In our effort on this topic started 4 years ago, we found then that how to effectively do the text mining and data cleansing in auto industry is still in its infancy and there are no mature software system, process and successful business examples or in other words, the accuracy of the results are not able to reach to a satisfaction level. Based on this finding, we decided to develop our own system and algorithm. This system focused on engineer's daily work and enabled the engineers to design the vehicle and its sub-systems to meet custom needs. We called this function as "deep guidance" and we also believed this is a paradigm change both in a company's development process as well as the information system architecture. Rather than getting the striped down information through marketing and planning department of the company in the traditional or over-the-wall way, the engineers can access these data first hand which help them understand the company's vehicle planning and option decision at least as shown in Fig. 2.

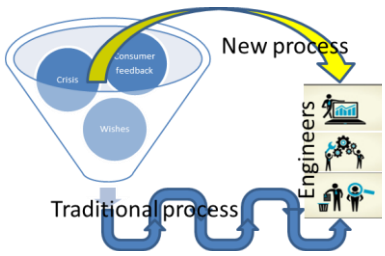

Fig. 2: New Process vs. Traditional Process

II. SYSTEM ARCHITECTURE DESCRIPTION:

The system developed has three main layers: data collection, data analysis, and result presenter as shown in Fig. 3. The data collection layer has six main modules. They are, data planning, data acquisition, information extraction, data warehousing, text mining, and content analysis and management. According our company vehicle style and future direction, we plan and define the data we want to crawl in the data planning module and then, the distributed multi-thread data acquisition module automatically crawls the data to our database. In this way, we optimized the data collected to be the most valuable asset rather than all the data from internet. We have configured the system to crawl data from 200+ automotive related sites where posts containing text and photos about vehicle information and feedback running multi-million lines daily. Data from company owned sites and contracted sites were also merged with crawled data in the data warehouse module and these data has more data dimensions then the crawled data. In the data acquisition, we crawled data such as the industrial trends, benchmark vehicle data, government regulation updates, customization trends, and public sentiments. It supports data formats of Words, PDF, PPT, Excel, CAJ and etc. And also, in the data acquisition module, the method to penetrate the "data block mechanism" has been implemented for each individual site so we can maximize the data volume and it supports RSS sites and agents. The geological data of the publisher has been identified by the IP address so we have a geological distribution of the data feedback. In the database management, we used cluster technology to distribute the data load thus insure the data process speed.

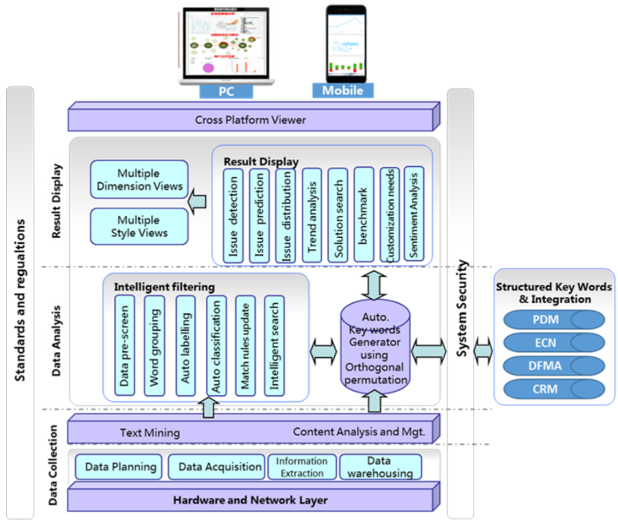

Fig. 3: System architecture

The cross platform viewer use the login information to identify user and personalize his display and match his specialty and interests so that the system only notify the related events and warnings and it supports both PC and smart phone usage.

Once the data is our local database, the information extraction module then use our NLP (nature language processor) analyzer and concept relation dictionary to extract necessary information from the text about the vehicle. The extracted data are stored alongside with internal vehicle quality data and warranty data in the data warehouse. Then, the text mining module matches the text find the quality, benchmark and etc. related information in the data management module. The data warehouse was not only served as a company data asset reservoir but also as the bases for fake、spam and duplicated data identification together with intelligent filtering module where modified factor graph model were implemented. [12]

The data analysis layer has two main modules: Intelligent filtering and key word generation. The intelligent filter has the functions of data and user ID pre-screen, word grouping and segmentation, automatic labeling and automatic classification, matching rule upgrades, and intelligent search. The fake and fabricated data and users were detected and excluded in these modules. The searching speed can reach 2100GB/s. The similarity detection can reach one million article/20ms. The maximum data table can contain 4 billion entries and the maximum data capacity is 8TB. The word grouping and segmentation module can match words 5MB/s and full split speed 2MB/s and the average accuracy is 97.3%. The key word generator utilized the key words defined in the traditional systems such as PDM, ECN, BOM, FEMA, and CRM. By doing the orthogonal permutation, the key words were generated. They form the foundation of the key words

ISBN: 978-988-14048-1-7
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

due to their well-structured classification and fields. The well-structured words are natural fit to the concept relation dictionary which can greatly increase the matching accuracy. One the advantages of this method is the traditional systems and their key words were maintained and updated regularly by a rigorous process which reflect the new technology and industrial trends. These key words were then wrap with their synonyms and antonyms. The preliminary ontology of these words were also built and we have seen an increase of the searching accuracy in terms of the relevance to the subject. The interface that allows the user to add new key words were also built.

The automatic classification module implements algorithm such as KNN, VSM and Bayes. These key words then cross matches the text mined from the internet in the intelligent search module.

The result presenter has two main modules: Viewer and result presenter. The Viewer can display result in multiple dimension and styles. The result generator has the functions of issue detection， issue prediction, issue distribution, trends analysis, solution search, benchmark, customization needs, and sentiment analysis. In the solution search module, the system searches the company internal knowledge base, paid article and patent sites as well as public sites such as CNKI, Baidu and etc. for a solution based on the geological information and recent events in that area. Once identified then the solution recommendation module sent the most useful solution articles to the engineer. In the benchmark module, the competitor vehicle performance data such as the fuel consumption, EV range and all the way to their sub-systems performance and properties were benchmarked in multiple dimensions such as the cost, reliability and etc. for engineers to view.

System usage:

The system can be used by high rank decision makers and engineers in the field of marketing, quality and planning.

1. Used by the decision makers

The data is organized by adopting the Kano model where

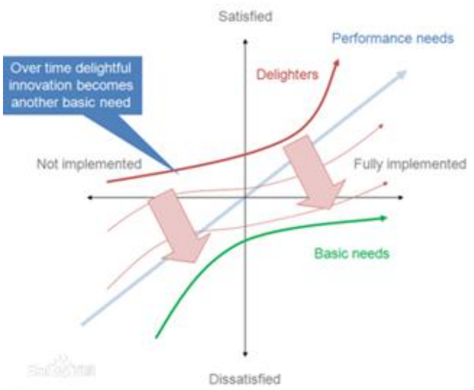

Fig. 4. The Kano Model

the consumer's needs are classified in three aspects, basic needs, performance needs, and delighters as shown in Fig. 4. In general, the standard options provided with by most successful vehicles in the market of the same class produced by Chinese OEMs were considered as the basic needs, the corresponding options in production by the joint venture companies are considered as the performance needs, and the concepts published and highly recognized by public in events and shows were considered delighters. In this way, the feedback are aligned with the internal marketing research and planning process. The data and analysis results used by the planning people and decision makers to decide what's options and features should be equipped on the next vehicle. The engineers used these classification to decide the urgency to change for certain quality issues.

2. Used by the engineers

The engineers who registered in this systems can instantly know the vehicle consumer feedback. The issues are organized in a fish bone diagram and the nodes indicated the root, the vehicle level, system level and sub-system level and component level as shown in the Fig 5. The competitors' quality level on the sub-systems and components were also displayed to engineers so they can decide what level of improvements are needed with cost consideration.

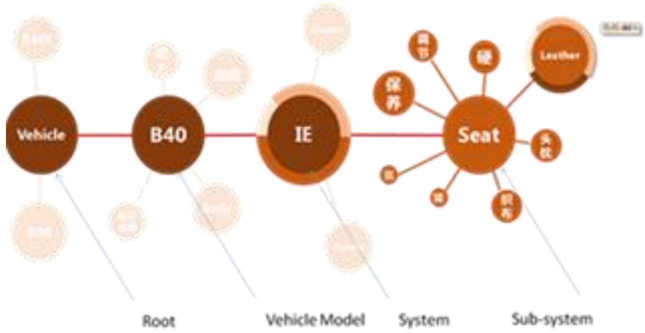

Fig. 5: The issue fish bone network.

The nodes and levels are predefined according to company's Vehicle Partition and Part Structure. It consists of

Fig. 6. The issue description published by the consumer where shown the leather worn

8 levels and 4897 nodes and the mechanism and process to update both levels and nodes are developed. By clicking on nodes, the engineers then can use the system to dig into the problems. The size of the nodes relatively reflects the number of issues. By click on the node shown "Seat" which is under "IE" which stands for interior and exterior of the vehicle in Fig. 5, for instance, we can find the leather is one of the main issues needs fixed. Once clicked on the leather node, the consumer concerns and photo can be read as shown in Fig 6. So far the engineer users spread cross engineering fields including design and release, quality, marketing, customization and aftermarket. The computer screen capture

of the node network and bar chart where the number of issues occurred and their distribution in last month are shown in Fig. 7.

The system also provided the users with subscription functionality where specific fields, subsystems and roles can be defined. Once subscribed, the system will notify the engineer of related field on the issue of the car as soon as the consumer complaints about the vehicles in the internet. This minimize the engineer effort to traverse through the issue network. For a pre-defined catalog of urgent problems, the system will automatically initiate an engineering change in the ECM system integrated thus speed up the problem solving process.

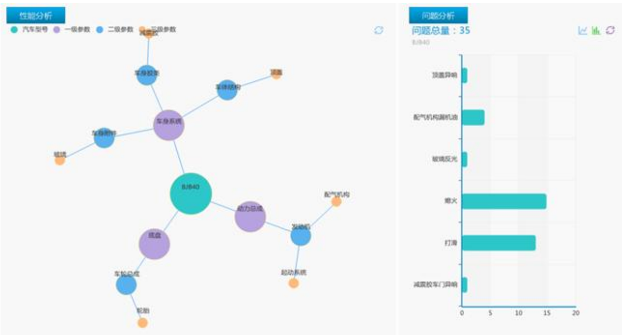

Fig. 7: The number of issues occurred in last month

III. Next Step

The system still has many improvements pending such as increase the accuracy of the nature language analysis and increase the key word database. The Linked Data property in both internal systems and cloud information will be examined further. Also, we would like to improve the solution recommendation module by sending the most useful solution segments instead of the whole article to the engineer.

IV. Conclusion

From Oct. 2014 to Dec. 2015, we have successfully developed the big data based intelligent analysis platform for vehicle development. Since then, it went through several upgrades. So far the system has identified 696 issues and guided the vehicle design and planning of 12 vehicle model-years. The user base has reached 3000 inside Beijing Motor Company. It created direct links between vehicle consumers and vehicle designers. It has not only reduced the issue solving cycle time but also prevented critical issues that lead vehicle safety issues and potentially accidents. It also changed the company's IT architecture and became one of the fundamental system in the company.

References

[1] C.C. Jia, M.Y.Wang, D. Zhen, X. X. Zhang "A Review on the Research of Data Mining in Microblog Based on Cloud Computing", Journal of Anhui Agri. Sci. 2014, 42(31); 11193-11195.

[2] F. Qian "Review of Domestic Reaserch on Data Mining", Jounral of Information, Volume 10, 2008. P11-13.

[3] F. Q. Xiong "Review on Web Data Mining", Electronics World, 2016. Page 98-99.

[4] H.B. Zhou "Cloud Computation: Technology, Application, Standard and Business Model" Electrical Industry Publisher, 2011-6.

[5] W.Y. Lei "Cloud Computation: Technology, Platform, and Application Examples", Tsing University Publisher, 2011-5.

[6] Q.J. Guo, M.L. Jia, Y. Li "Research on Components' Relationship and Configuration Management for Multi Model Structure", Journal of New Industrialization, DOI: 10.3969/j.issn.2095-6649.2015.05.09, 2015, 5(5); 62-66.

[7] M. Fischetti "The Web Turns 20: Linked Data Gives People Power", Scientific American 2010 October 23.

[8] Y. Li "Increase the PDM data security level by Data Encryption", Journal of CAD/CAM and Industrial Manufacturing. Volume 11, 2009, Page 37-39.

[9] Y. Li, J. Chen, W. Huang, F. Yu, J.Y. Lv, Y.P. Pan "Automobile R&D Oriented Big Data Analysis System", Journal of Technology and Economics, June 2018, p25-28.

[10] E.R. Cadotte; N. Turgeon (1988). "Dissatisfiers and satisfiers: suggestions from consumer complaints and compliments", Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior. 1: 74–79.

[11] A. Sylla, D. Guillon, E. Vareilles, M. Aldanondo, T. Coudert, L. Geneste, "Configuration knowledge modeling: How to extend configuration from assemble/make to order towards engineer to order for the bidding process" Computers in Industry, Volume 99, August 2018, Pages 29-41.

[12] Y. Lu, L. Zhang, L. Xiao, Y. Li"Simultaneously detecting fake reviews and review spammers using factor graph model．" New York: ACM Press：In: Davis HC, Halpin H, Pentland A, eds. Proc. of the 5th Annual ACM Web Science Conf. (WebSci 2013)，2013.