Quality Big Data Analysis and Management Based on Product Satisfaction Index

Y F Sun, A P Lu, L Zhuo, G Li, J Jia, W Liu, C J Hu

Senior Manager, PCSD Quality, Global Supply Chain, Lenovo Ltd, 100193, P.R. China

E-mail: sunyfa@lenovo.com

Abstract. The core idea of quality work is to create enduring quality by providing users with reliable, easy-to-use, and surprising products. In the Internet+ era, user demand for quality has changed with the times. Customers are no longer satisfied with the requirements of basic quality functions, and more in pursuit of ease of use, pleasure and even identity. Even if the product has “satisfied with the specifications” in quality, users are still dissatisfied and express their dissatisfaction through the Internet. If we can use the Internet, big data, artificial intelligence and other technologies to collect and analyze user expectations or dissatisfaction with product quality and find improvement directions, then quality management will see great progress. Based on this, we can create a new type of quality big data analysis and management system, which we can call the Product Satisfaction Index (PSI) management system. Through the analysis and transformation of PSI, the customer's textual expectations are turned into numerical quantifiable indicators. Through the management of PSI indicators, quality big data mining and analysis are performed to improve the product to meet customer expectations. In advancing the product satisfaction index (PSI) indicator management system, we should fully implement the definition of PSI indicators, strategic position, data analysis methods, closed-loop management models, review and continuous improvement, and data platforms. Visualize customer needs, analyze the matching between product characteristics and customer needs, and promote continuous improvement of PSI indicators through closed-loop management. The new PSI index management method opens up new ideas and methods for quality performance management, and helps build high-quality development and create customer satisfaction.

1. Introduction

In the Internet+ era, emerging technologies are empowering consumers to change the business landscape. With the advancement of technology, users’ expectations for products are getting higher and more transparent [1]. The company uses data-driven decision-making, a strategy of transition to smart quality, combined with the background of traditional quality industry leaders, users are still dissatisfied, and it is imperative to transform to a customer centric transformation. The work of quality management can rely on the Kano model [2] to innovatively propose a quality performance management strategy shown in Figure1. In terms of protect, we need to continuously reduce the failure rate of products, meet the realization of basic functions of product quality, and defend the cornerstone of the enterprise [3]. In terms of attack, we can improve user satisfaction and even surprise. By mining big data on the Internet, combining machine learning and semantic analysis technology [4], we can transform user expectations into quantifiable product satisfaction index (PSI)
and continue to promote products [5]. The satisfaction index improved to leading customers to a better experience.

![Quality Big Data Analysis Report](image1)

**Figure 1. Protect + Attack Quality Strategy**

2. Definition of PSI indicator
The PSI (Product Satisfaction Index) refers to the statistical user satisfaction converted from fragmented Internet user reviews, which intuitively reflects the degree to which user needs are met.

Data source: User comments on product.
Calculation formula:

\[
PSI = \frac{10(\text{Positive Reviews} - \text{Negative Reviews})}{\text{Positive Reviews} + \text{Negative Reviews}}
\]

Score range: negative 10 points to positive 10 points. The higher the PSI score, the better. The worst is a negative 10 points, preferably the positive 10 points.

Example: a user have a comment on the product: My A product keyboard feels good, the display clarity is not high, and the fan noise is a bit loud when playing key games.

- Fragment comment: 3 fragments in total. The number of positive comments is 1 segment: the keyboard feels good; the number of negative comments is 2 segments: the display clarity is not high, and the fan noise is still a bit loud.

- The PSI indicator score is -3.3, and the formula is \( PSI = 10 \times \frac{1-2}{1+2} = -3.3 \).

3. PSI indicator strategic position
In the "customer-centric" quality management, the PSI index fits the user's definition of product quality satisfaction, and improves the quality management from guaranteeing survival based on failure rate to leading success based on satisfaction. Enterprises should implement PSI indicators into organizational assessment and employee assessment. Leaders at all levels of quality and all employees, research & development, product managers, project managers, marketing and other organizations should implement PSI indicators in performance evaluation, and the important position of PSI indicators should map to "customer-centric" strategic practice.

4. PSI indicator data analysis method - data visualization
The PSI indicator data platform automatically collects public data related to product quality feedback from users on the Internet and customer service through data collection methods such as crawlers. The analysis is performed through technologies such as big data processing, machine learning, and natural language processing. Combining with the daily analysis logic of the quality department, finally through visualization technology to build a system tool designed to promote efficient and excellent quality work, continuous improvement of product quality and user satisfaction.

The PSI indicator data platform solution adopts industry-recognized technologies such as machine learning and natural language processing, which has mastery, replication and extensive expansibility.

Two sets of highly practical data visualization systems are the basis of PSI indicator management, including two major modules: (1). Continuous monitoring and analysis of PSI around indicator management, (2). Observing customers' evaluation of products from a big data perspective statistics and focus. These two modules are easy to operate. You can drill down through the data with one
mouse button to directly obtain key products, user concerns and improvement directions directly from the system page. In addition, the advantages and disadvantages that the user is most concerned about can be displayed intuitively and sorted. The system users are easy to operate, the decision is intuitive, and it is efficient and practical.

**Figure 2.** Function module (1): Continuous monitoring and analysis of PSI around indicator management, flexible selection and multi-dimensional display

**Figure 2.** Function module (2): Observe the statistics and focus of product evaluation by customers from the perspective of big data, which can be controlled globally and interpreted in detail
5. PSI indicator data analysis methods - three types of analysis structure

According to the principle of business improvement, the quality department can summarize the inductive analysis methods, unify the analysis steps and purposes, so that the overall PSI analysis quality is high, the effect is good, and the improvement is significant. The following three are the summary of the company's many years of experience:

- The life cycle of the product, continuously ensuring that the product meets customer needs
- Iterative product upgrades, continuously improving the product to meet customer demand changes
- Product market competition, the joy of creating superior products to lead customer demand

5.1. The life cycle of the product, continuously ensuring that the product meets customer needs

Enterprises should be committed to continuous improvement of a product after it is launched. The quality department of the enterprise shall regularly monitor the PSI indicators of the products in each market segment. Once abnormalities are found, further analysis and improvement will be made immediately. In the product life cycle, the product's PSI index is compared with the score, target trend analysis, benchmarking analysis, and influence factor analysis, etc., to identify gaps and continuously ensure that the product meets customer needs.

Example: For a product that was launched in January 2018, the first month after listing, PSI analysis found that the product scored less than 5 points. With reference to the target and similar product scores, realize that the score is very low and perform a key analysis. Through PSI analysis and comparison of the refined attributes of the product, the team quickly found that users had a large number of negative reviews and low scores in the three aspects of the product: Display, Noise, and Network Card. To users' normal expectations on similar products. The company quickly and targeted the analysis and improvement. In the following March and April, the product's PSI score increased month by month.

![Figure 3. PSI analysis, product life cycle trend analysis](image)

![Figure 4. PSI analysis, attribute analysis of product life cycle](image)
5.2. Iterative product upgrades, continuously improving the product to meet customer demand changes

Commitment to the new generation of products to continuously meet the expectations of ever-changing customers is the key to continuous improvement of product quality during product iteration. With the development of the industry and changes in usage scenarios, some product attributes that once met customer expectations will gradually become unsatisfactory. The enterprise quality department will regularly analyze the new and old products' PSI, identify the new requirements and new features of the new generation of products, and output the customer demand and satisfaction change analysis report to the product manager to plan the layout of the new product.

Example: It is a case of performance configuration improvement for a new generation of gaming notebook products. By analyzing the PSI data of the previous generation gaming notebooks, the quality department realized that a considerable number of users have become a trend for playing games with notebooks, and the subsequent change is that users have increasingly higher requirements for notebook performance. PSI data shows that when the configuration of a gaming notebook is too low, the user's performance score will be very low. Combining the changing trends of the industry and the analysis reports delivered by the quality department, the product manager decides that performance needs to be improved on the new generation of products. Through product configuration adjustment, the solid-state hard disk is adjusted from the previous optional configuration to a mandatory configuration, and large-capacity solid-state hard disks have also become the first choice for high-end gaming books. In addition, the PSI score of the new generation of gaming notebooks has been comprehensively improved. It has been praised by users in the first month of listing, and it continues to maintain high growth and high profits.

In fact, the improvement of the new generation of gaming notebook is far more than just performance. Through the analysis of PSI data, companies have made changes in performance, overall design, display, keyboard, and hard disk partitions according to customer needs. Adapt to market improvement and achieve better user satisfaction. For changes in customer needs of different natures, when determining improvement, it is combined with qualitative prioritization to match organizational resources to achieve the best improvement effect.

![Figure 5. PSI analysis, product iteration upgrade analysis](image)

5.3. Product market competition, the joy of creating superior products to lead customer demand

In a smart device industry, on the one hand, companies need to grasp market trends in real time to maintain their leading edge, and on the other hand, they are one of the main driving forces for industry progress. The enterprise quality department compares the PSI data of different merchant products in the segmented market to help the enterprise grasp the trend of customer demand and the advantages and disadvantages of each manufacturer's products, and provide analysis and reference for users to create happy products.

In the past year, the enterprise quality department analyzed PSI data and found that users are generally concerned about the product cooling experience. The customer experience is poor on many
products in the enterprise and in the industry. However, the internal testing and evaluation results of research and development show that the product design and the finished product are Meet industry quality standards. This allows companies to locate the overall market disadvantage through PSI analysis and launch new thermal design and innovation modules and software applications. Users can set whatever they want to meet the different needs of users in different use scenarios. The enterprise’s PSI thermal score is the highest in the industry.

Figure 6. PSI analysis, product market competition

Summarizing the above three commonly used PSI analysis models, it is not difficult to see that in many cases the analysis mostly uses a comparative strategy, but the benchmarks to be selected are different for different purposes. The PSI index could be widely used in various departments of the company, penetrating all aspects of the product from design, development, and production to sales, and it is an effective baton for the organization to reach consensus and continue to improve product quality.

The last but not the least, the purpose of the PSI analysis is to explore the opportunities for quality improvement through effective PSI data analysis, and ultimately increase user satisfaction on e-commerce reviews, promote quality satisfaction and increase sales. Through data research, the PSI index has a strong positive correlation with the user's rating on the e-commerce platform. The improvement of the PSI index is directly reflected in the improvement of the user's rating, then furtherly drive sales.

Figure 7. Strong positive correlation results (1)
6. Closed-loop management mode of PSI indicators
The closed-loop management mode of PSI indicators, including the general principles of PSI indicator annual target setting (including the principles and procedures of target setting, stipulating the concept that senior leaders are responsible for full participation, and the division of responsibilities of each department of the PSI indicator), and improving the quality decision-making closed-loop system. Combined with customer-oriented cultural propaganda, the entire staff explained the closed-loop improvement culture of PSI indicators, and further promoted the "customer-centric" quality management strategy.

7. Review and continuous improvement of PSI indicators
"Fu Pan" is a term derived from Chinese Go. Fu Pan is a deep reflection and summary of experience after the action. Through continuous learning, summary, reflection and refinement to continuously improve organizational wisdom, ensure that the same mistakes will not be made again, and inherited experience and improvement of ability ultimately turn failure into wealth and success into ability. In practice, the PSI indicator is exerting higher value through continuous review. Among them, an important review of PSI is to transform PSI's result-based analysis based on a single data source (public review data) into a multidimensional data set (public review data, research and development data, production data, product parameters, customer portraits, customer behavior records, etc.) to improve the PSI analysis method and improve the PSI indicator system. This round of review and research will lead the quality performance system to pioneering exploration of digitalization and intelligent reform, and will provide a model basis for customer satisfaction quality analysis based on big data, leading the industry's quality analysis reform.

8. Screening and guarantee of data quality
Good data quality is the first prerequisite for effective data analysis. As mentioned earlier, our data mainly comes from open e-commerce and forum platforms. Users, stores or platforms will release some for invalid purposes. Of comments, fake comments, or a lot of duplicate comment records. In order to be able to more effectively identify the true evaluation of customers, we need a certain mechanism and means to process these data. After continual exploration and practice, we summarized and implemented the processing mechanism and method for typical noise data. Summarized as follows:

8.1. By analyzing the structure of the web page, the relevant data of the target product is locked, and the irrelevant data is removed.
  • On the page of a specific product, in addition to being able to browse to the related comment records of that specific product, the comment records of similar products are usually also
listed. By analyzing the structure of the web page, we can establish a reasonable data acquisition path to lock the review records of the products we care about

- For some e-commerce platforms, different products may be sold on the same website in different periods. When these data enter the database, we establish the correct data association method to associate the correct product description with its corresponding review record

8.2. By observing the content of the comments / accounts, records with certain specified characteristics are removed.

- Check the comment length (too short <5 characters), illegal characters, empty words, etc. directly removed.
- Check the relevance of the content of the review and the product on the page, and remove the records that talk about objects that are obviously irrelevant to the target product.
- Check the similarity of the comments and duplicate the comments with a very high similarity.
- Check the account posted by reviews. Accounts that frequently post multiple reviews in different products within a short period of time are considered to be pseudo-customers (water army), and the reviews posted by them are removed.
- Check the account posted by reviews. Accounts that have posted reviews multiple times under the same product are considered to be fake customers (water army), and the review records posted by them are removed.

8.3. Through human observations, conduct investigations on data fluctuations that clearly violate the laws of common sense and handle them if necessary

- For a promotional season or a new product release of a certain product, we will have an expectation that the number of reviews will increase rapidly in a short period of time. When a specific product appears in a specific store in a short period of time, the number of reviews rapidly increases. These are indications that this product needs to be checked and confirmed when it is being promoted.
- Cross-validation, periodically spot testing key products to observe the amount of data on different platforms, and the degree of consistency of users when talking about content. Obviously, they need to be checked and confirmed.
- For a specific product, if the time of review and record is obviously beyond the sales range of the product, it is necessary to carry out investigation and confirmation

The data processing mechanisms and methods established above are based on the most basic laws of nature, and their effectiveness is obvious. At the same time, we are also aware that the complexity of data on the Internet is dynamically changing. Over time, it is expected that there will be many unexpected so-called fake data. We need to continuously observe and summarize to accumulate the laws that identify the authenticity of data. And dynamically improve the data guarantee system.

9. Data platform for PSI indicators
Through unique crawler technology, mixed model sentiment analysis, product attributes based on actual business, and rule-based classification, a unique PSI indicator data platform was developed.
9.1 Data acquisition module

The source of the data is mainly related to user feedback from major e-commerce websites, social media, forums, and customer service. The crawler technology is applied to the PSI index data collection. In order to facilitate efficient data crawling, Lenovo has created a unique Lenovo data acquisition platform. This platform supports different types of crawling tasks, including width-first, depth-first, requesting open APIs to obtain data, and simulating browser crawling tasks. In addition to actively crawling data, it also supports uploading user-defined data, such as customer service center data and user research data. Lenovo data acquisition platform adopts user configuration to support users to flexibly create, monitor, update, and delete crawl tasks. And adopt multi-threaded distributed processing to solve the problem of data crawling efficiency and stability. It also supports proxy switching and abnormal alarm functions.

![Figure 10. Data Crawl Module Program (1)](image1)

![Figure 11. Data Crawl Module Program (2)](image2)
9.2 Data preprocessing module

9.2.1 Data extraction and cleaning.
The data collected by the data acquisition module is mostly unstructured. Before data analysis, data extraction needs to be performed to convert it into structured data. Data extraction Extract the necessary fields for product quality analysis from the original data, such as review publisher, review content, product information, time, etc., and discard information that is not related to quality analysis.

The results of data extraction include web page formatting tag data, space newlines, and other special character data or data content format that does not meet the analysis requirements. These need to be cleaned before further data analysis.

9.2.2 Data alignment.
Data alignment is an essential module of a product review analysis system. Because the data sources are diverse, the name description of the same product will be different on different data sources.

9.2.3 Data storage.
Data storage can be divided into storage of structured data after data preprocessing and storage of raw data files such as web pages obtained from data collection. Structured data is stored in a relational database for further analysis and statistics, while the original data files can be archived and stored in a distributed file system.

9.3 Data analysis module

9.3.1 Word segmentation and part-of-speech tagging.
Understanding and analyzing the pre-processed data requires the use of natural language processing technology. Of which the segmentation and part-of-speech tagging are the basis of all natural language processing. The accuracy of the segmentation and part-of-speech tagging determines the accuracy and credibility of all subsequent analyses.

The so-called word segmentation is the process of splitting an originally entered paragraph or sentence into meaningful phrases according to the context. English word segmentation is relatively simple, and can be completed according to spaces and punctuation. However, there is no clear space symbol between different words in Chinese word segmentation, and special segmentation techniques need to be trained to handle it.

9.3.1 Sentiment Analysis.
Sentiment analysis technology is the core of the product review analysis system. All subsequent functional modules are built on the basis of sentiment analysis technology. The sentiment analysis model needs to identify the evaluation object and the emotional tendency (positive, negative, neutral) from the original text.

The construction of sentiment analysis model first needs to manually label a large amount of training corpus, and then use natural language processing and machine learning technology to perform model training on the labeled corpus to obtain the sentiment analysis model. The sentiment analysis model obtained from the training is used to construct a product review analysis system, and the sentiment analysis errors found during use are fed back to the system for continuous optimization and improvement of the sentiment analysis model.

PSI indicator data analysis uses natural language processing technology. In addition to word segmentation and part-of-speech tagging, sentiment analysis technology is the core of the entire PSI indicator data platform. The sentiment analysis model needs to identify the evaluation object and the emotional tendency (positive, negative, neutral) from the original text. The construction of sentiment analysis model first needs to manually label a large amount of training corpus, and then use natural language processing and machine learning technology to perform model training on the labeled corpus to obtain the sentiment analysis model. Use the sentiment analysis model obtained from the training to build a product public comment analysis system, and feedback the sentiment analysis
errors found in the process to the system for continuous optimization and improvement of the sentiment analysis model. In the construction of sentiment analysis models, a hybrid model based on CRF sequence labeling + rules is used, with an accuracy of more than 90%.

![Figure 12. Sentiment Analysis Module](image)

10. Implementation effect of PSI indicators
From the perspective of the quality closed-loop management system of PSI indicators, enterprises have fully implemented the definition of PSI indicators, strategic status, data analysis methods, closed-loop management models, review and continuous improvement, and data platforms, which will help enterprises to achieve a comprehensive quality management level. Promotion. Within the organization of the company, a consensus has been formed on the quality culture. From the results of the operation indicators, we have made great progress. The company's management level, PSI index and new quality performance management mode provide a network-wide quality analysis model, which improves the level of product quality analysis and enhances the competitiveness of the company.

In summary, in the field of closed-loop quality management, based on in-depth insights into industry and quality changes, through the combination of advanced technology and quality management, companies continue to invest in the field of intelligence, boldly explore, constantly summarize and practice these results, and further consolidate Lenovo's level and influence in the field of quality management also help corporate brand building.

11. Conclusion
In this paper, by creating a new type of quality big data analysis and management system based on product satisfaction index, the customer’s textual expectations will be converted into digital quantifiable indicators, so as to realize high-quality big data mining and analysis, to meet customer expectations and improve product quality. The new PSI index management method opens up new ideas and methods for quality performance management and helps build high-quality development and create customer satisfaction. With theoretical support, it has a good promotion effect.

References
[1] Wu F, Wang Z Q, Zhou X B and Zhou G D 2020 Joint model for sentiment analysis and review quality detection with user and product representations Journal of Software 31 2492-507
[2] Gao X D and Wang A 2020 Customer Satisfaction Analysis and Management Method Based on Enterprise Network Public Opinion Operations Research and Management Science 29 232-9
[3] Jin J H, Wu L T, Zhang T T and Yan X B 2020 Research of scoring for takeout merchants based on sentiment analysis Journal of Management 33 66-75
[4] Meng Y, Wang H W and Wang W 2020 The effect of electronic word-of-mouth on sales through fine-gained sentiment analysis Management Review 29 144-54
[5] Li K M and X Y 2020 Analysis on product (service) quality index Quality Exploration 1 64-9