Amax Photo/Hoxton/Martin Barraud via Getty Images

Take a Big BYTE

Know the effects of big data on quality improvement to better solve problems and address customer concerns | by Chao-Ton Su

Quality improvement is a key principle of total quality management and a never-ending process for every organization. A systematic approach is crucial for improving process and product quality to enhance organizational competitiveness.

This article discusses quality improvement and big data, emphasizing the effect of big data on quality improvement. In addition, some related techniques and examples of big data analysis are covered.

Quality improvement

Many organizations have attempted to develop a systematic approach that uses specific techniques to improve quality and reduce waste in processes and products. In various industries, quality improvement is usually accomplished through teamwork. Three elements are crucial for successfully implementing quality improvement (see Figure 1, p. 25).

1. **Quality concepts:** Realizing various quality concepts could help address quality problems efficiently. Through appropriate on-the-job training, many enterprises guide their employees to learn the philosophies of the quality gurus to have a better attitude toward quality improvement. Examples of quality concepts include: quality is defined by the customer; quality means conformance to requirements; quality comes from prevention; less variation in performance results in higher quality products; and attractive quality elements should be created to satisfy the latent needs of customers. These basic quality tenets continue to have a considerable effect on the improvement of quality performance.
2. **Management models:** We must use management models to solve quality problems. Frequently used management models in practice include the plan-do-check-act cycle, quality control story, Ford's 8 disciplines (8D), and the define, measure, analyze, improve and control (DMAIC) approach. Among these, 8D is widely used in high-tech companies because it simultaneously emphasizes the values of containing, correcting and preventing problems.
3. **Quality improvement techniques:** Frequently used quality improvement techniques can be divided into three categories.

Just the Facts

Big data has affected the task of quality improvement in three significant ways: understanding the voice of the customer, collecting and analyzing the data, and developing prediction models.

In addition, Quality 4.0 has introduced more progressive techniques and tools to analyze big data including artificial intelligence, machine learning, data mining, data preprocessing and feature selection.

It is up to the quality professional to ensure the big data are correct and not missing values, to know methods to quickly process and extract actionable information, and to determine whether the data can lead to solving problems or customer concerns.

1. Statistical methods, including basic statistics, hypothesis testing, regression, and design of experiments (DoE) or the Taguchi methods.
2. Quality tools, including quality function deployment, the basic seven, the new seven, statistical process control, process capability analysis, measurement system analysis, and failure mode and effects analysis.
3. Toyota Production System and lean thinking, which have been commonly used in industries for waste elimination, cost and cycle-time reduction, and quality improvement.

Quality improvement requires that problems be resolved. Figure 2 (p. 27) presents the data-driven logic for solving a quality problem. For a given problem, you collect data, conduct an analysis by using suitable tools and determine ideal solutions. Ideal solutions are modified to obtain practical solutions.

In this process, identifying the appropriate improvement opportunity, deconstructing a problem and interpreting analytical results are crucial. Furthermore, a quality improvement project must be linked with the business strategic goal, the voice of the customer (VOC), process and engineering.

Big data

Big data has attracted the attention of scholars and practitioners. Generating an increasing amount of data is an inevitable trend in the development of modern technology. For example, you easily can install sensors and smart chips in machines and products to obtain relevant information, such as product characteristics and operational conditions. Big data is now imperative for many organizations, from providing services to manufacturing.

Big data has no universal definition. Author Doug Laney[1] defined big data in terms of three Vs:

1. **Volume**—the size of the data set.
2. **Velocity**—the speed of data in and out.
3. **Variety**—the diverse range of data types and sources.

The three Vs have appeared as a popular framework for describing big data. In addition to these three, two other common dimensions of big data include veracity, which is the quality or trustworthiness of the data, and value, which is the worth of the data being extracted.

Big data does not concern data itself. Rather, it concerns the identification of more effective problem-solving strategies. In this article, big data is considered a holistic approach to exploring the five Vs (volume, velocity, variety, veracity and value) to enable the acquisition of actionable knowledge for enhancing the competitiveness of enterprises. Technology will continue to evolve over time, and we will develop different viewpoints on big data in the future.

Effect of big data on quality improvement

Three key effects of big data on quality improvement are:

1. **VOC:** Understanding the VOC, including internal and external customers, is crucial because this information is valuable to an organization when making decisions regarding quality improvement directions. Conventionally, organizations use surveys, interviews, focus groups, warranty data, field data and complaints to ascertain customer desires. Big data, however, boasts a superior capability for precisely assessing VOC. Large data sets and sophisticated tools enable you to identify actual customer purchases and motivations. Most vital customer requirements can be determined based on the VOC, thus providing ample opportunities for quality improvement.
2. **Data collection and analysis:** Data collection plays a crucial role in quality improvement. Traditionally, various data collection methods are used depending on the situation.[2] There are three types of data:

+ **Experimental data—**Data are collected from a designed experiment. DoE or Taguchi methods frequently are used to address this type of data.
+ **Observational data—**Data are sampled through a planned observational study. Regression analysis or cause and effect analysis generally are used to analyze this type of data.
+ **Historical data—**Data have already been collected. Computational intelligence and a data mining approach are imperative for solving problems with historical data.

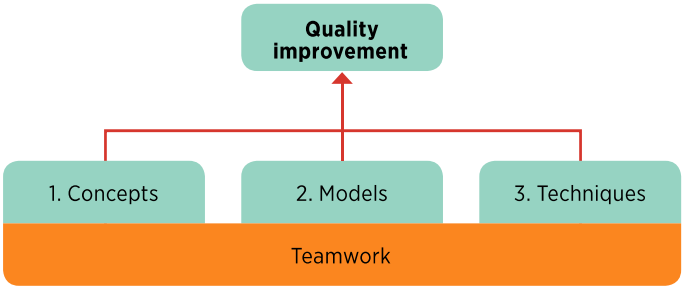
When DMAIC is used for problem solving, for example, usually you can apply DoE or Taguchi methods to optimize the process. The major reason is that Six Sigma supports engineering good sense, and you can fully understand the context of the problem. At this moment, a well-planned experiment can be conducted to collect required data for problem solving.

On the other hand, when dealing with a problem that is not as well-defined (or understood), you may have only historical data. As a result, some intelligent approaches could be effective in investigating data.

In the age of big data, data can be collected from sources such as social media, transactions,

FIGURE 1

Method for quality improvement



public data and machine-to-machine data. In manufacturing, a considerable amount of real-time field data—in addition to historical data—construct the big data. For these big data, often you don't know what to analyze. You may perform trial and error many times. Each attempt helps you to further understand the context of a problem.

Big data can challenge and encourage you to apply more advanced techniques to analyze massive sets of structured and unstructured data that have been collected. Using the correct approaches may reveal the hidden meaning within big data.

For example, a pizza restaurant monitored social media and analyzed the text and pictures that were posted to determine the root cause of customer dissatisfaction. The restaurant designed a system to resolve the primary problem, which was caused by the delivery driver.[3] Another example is applying an association rules algorithm to determine the influence of different machine combinations on yield in a foundry.

Using the correct approaches may reveal the hidden meaning within big data.

3. **Prediction:** Big data may allow you to more accurately predict the future. Through big data prediction, you can enhance process and product performance while enabling superior risk management. You can develop a prediction model to identify major quality problems in the operating machines before they are prone to break down. For example, a casting company employed plant process parameters to predict tensile strength using a neural network.

Preventive maintenance has been widely employed in the manufacturing industry. Many organizations, with the help of big data, prefer to implement predictive maintenance, which is designed to help determine the condition of equipment in service to predict when maintenance is required.

For example, a semiconductor manufacturing company in Taiwan applied an equipment degradation model to predict the useful lifetime of equipment. A laptop manufacturing company attempted to use the reason for customers requesting repairs to predict the parts needed for repairing laptops.

Prediction is a practical result of big data analysis. Constructing an efficient prediction model is not easy, however, because big data

usually are highly complex. The entire picture often is difficult to understand, and unknowns always remain. Something not occurring in the past does not guarantee it won't happen in the future. Furthermore, because big data is generated in real time, how to accurately and quickly predict change points in advance is critical.

Based on W. Edwards Deming's teaching,[4] if a process is in statistical control (that is, in a stable state), the variation to expect in the future is predictable. If a process is unstable, the performance is unpredictable. Predicting big data requires a careful assessment of whether the process is in a stable condition. Data obtained from an unstable process may provide unreliable predictions of the future.

Related techniques

Big data projects usually are associated with large and complex problems. Accordingly, the problem typically is vague and unclear. Therefore, you must first decide what problems to confront and who should join the team. The big data implementation method is like the framework presented in Figure 2. However, some more progressive techniques and tools to analyze big data are required.

Big data analytics: Analytics is applying math and statistics to discover meaningful patterns in data. Using conventional data analysis methods to handle big data can be hard. Therefore, in addition to mathematics and statistics, artificial intelligence (AI), machine learning (ML) and data mining approaches often are suggested for processing these large amounts of data. The discipline of analyzing big data to unearth the potential value of big data and obtain helpful insights for making better business decisions is referred to as big data analytics (Figure 3, p. 28).

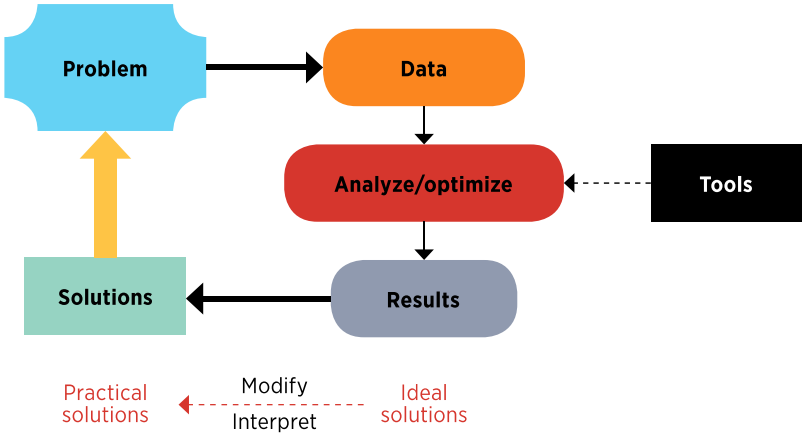
AI: AI involves using computers to solve problems that involve perception or intelligence. Through data processing and algorithm operation, AI attempts to produce meaningful information and render the machine smarter than people.

Generating an increasing amount of data is an inevitable trend in the development of modern technology.

Amax Photo via Getty Images

FIGURE 2

Data-driven logic for solving quality problems



AI techniques include natural language processing, logic-based reasoning, computer vision, search algorithm and ML.

ML: ML is a branch of AI. For a given problem, ML collects training data, selects features from data and constructs a model. This model can be regarded as the result of learning and is used to make predictions or solve highly specific problems. ML methods include regression, neural networks, support vector machines, decision trees and Naïve Bayes.

Data mining: Data mining is the process of discovering patterns and establishing relationships in a large data set to obtain a better understanding of a studied system. The typical process involves problem definition, data collection and preparation, modeling, validation and application. Common data mining tasks include classification, clustering, association and prediction. Text mining, also known as text data mining, is the process of extracting useful information from unstructured text.

Data preprocessing: Big data is usually unstructured. To improve the quality of raw data and the implementation results, more time for performing data preprocessing is required. Three basic methods are used for data preprocessing.

1. **Data cleaning:** Data can be incomplete, noisy and inconsistent. Therefore, you should provide missing values, identify outliers, smooth out noisy data and correct inconsistent data by using domain knowledge.
2. **Preprocessing categorical data:** Categorical data must be converted into suitable numeric values. One-hot encoding is a common technique used for working with categorical data.
3. **Data transformation:** Data must be converted to an appropriate scale for processing. Normalization and standardization are well-known techniques for data transformation.

Feature selection (FS): FS is the process of selecting a number with valuable features or attributes that help predict or identify the output (Y) from inputs (Xs). FS can be used to simplify, improve accuracy and strengthen the understanding and explanation of the model. Frequently used FS methods include correlation analysis, the relative importance of input variables in a neural network, meta-heuristic algorithms and decision trees.

Examples

Performing data analysis on the internet may require a quality information platform that can assist in:

1. **Monitoring:** Understanding the current situation.
2. **Analyzing:** Determining the cause of the problem.
3. **Predicting:** Predicting possible outcomes and complications to be prepared for or prevent the recurrence of problems.
4. **Optimizing:** Optimizing goals.

Example one: A casting company implemented a project to collect field-based data for process improvement. Based on engineering knowledge, the project team identified 17 possible factors that may affect the process output—tensile strength (y).

First, the team sought to determine the significant process factors that affect y. Five feature selection techniques, including neural network, random forest, support vector machine, rough set theory and regression analysis, were implemented. Based on majority rule, the team selected nine key process factors to study further.

Next, the neural network was used to construct a nonlinear relationship between nine control factors and response (y). The trained network was used as the fitness function in the genetic algorithm (GA). The control factor values were transformed into a vector (chromosome) to represent the possible solution, and the GA was used to optimize the solution.

In this study, the GA was executed in 20 runs. The implementation results revealed that the standard deviation of the 20 runs is small, demonstrating the robustness of the obtained solution. The optimal solution (with the highest tensile strength) was selected from these 20 possible solutions. Using this optimal combination would increase the tensile strength by about 13.5%.

The central idea of process optimization is shown in Figure 4.[5]

Example two: One company operated an electronic toll collection (ETC) system for freeways. The system uses a sensor to emit radio waves and detect radio frequency identification (RFID) tags attached to a car.

The average number of daily transactions is about 15 million. The relatively low rate of vehicle detection accuracy causes substantial financial loss to the company. The company formed a big data project to analyze ETC data for vehicle detection to identify crucial features affecting RFID tag detection and to improve the RFID tag detection rate.

Of the five vehicle types, a large truck was used as an example. The accuracy of the vehicle detection rate for a large truck is about 83.4%. The 170,500 large trucks record data were sampled from the database.

In the original data, 190 variables may have affected the vehicle detection rate. After data preprocessing, 170,000 records remained: 141,700 were detected and 28,300 undetected. Data were separated into a training set (120,000) and a test set (50,000). In the training set, 100,000 samples were detected and 20,000 were undetected, demonstrating an unbalance in the data.

Therefore, an oversampling technique was performed by adding more examples from the smaller set. That is, the amount of extracted, undetected vehicle data matched that of detected vehicle data. Therefore, the final training set included 100,000 detected and 100,000 undetected samples. This training set was used for further analysis. The test data set, which was used for validation, did not perform oversampling.

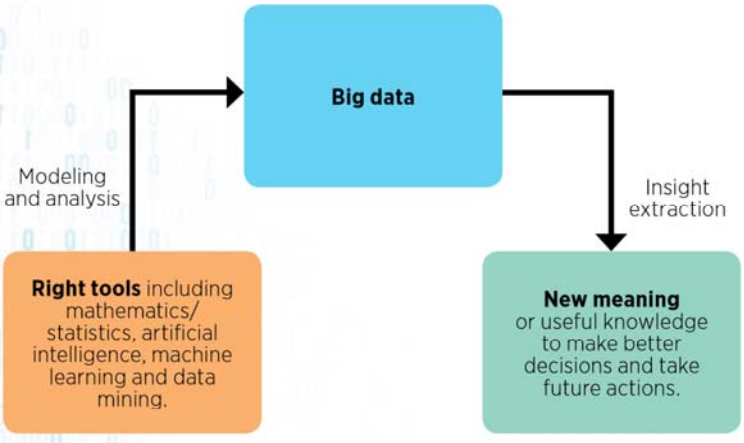
Five feature selection algorithms were employed to select the most critical features. The implementation results identified 29 crucial variables. Some useful rules were generated from the decision tree/C4.5 algorithm.

Additionally, the project team selected some controllable variables from 29 variables and applied a neural network and GA to determine the optimal settings for the controllable variables.

Big data can challenge you to apply more advanced techniques to analyze massive sets of structured and unstructured data that have been collected.

FIGURE 3

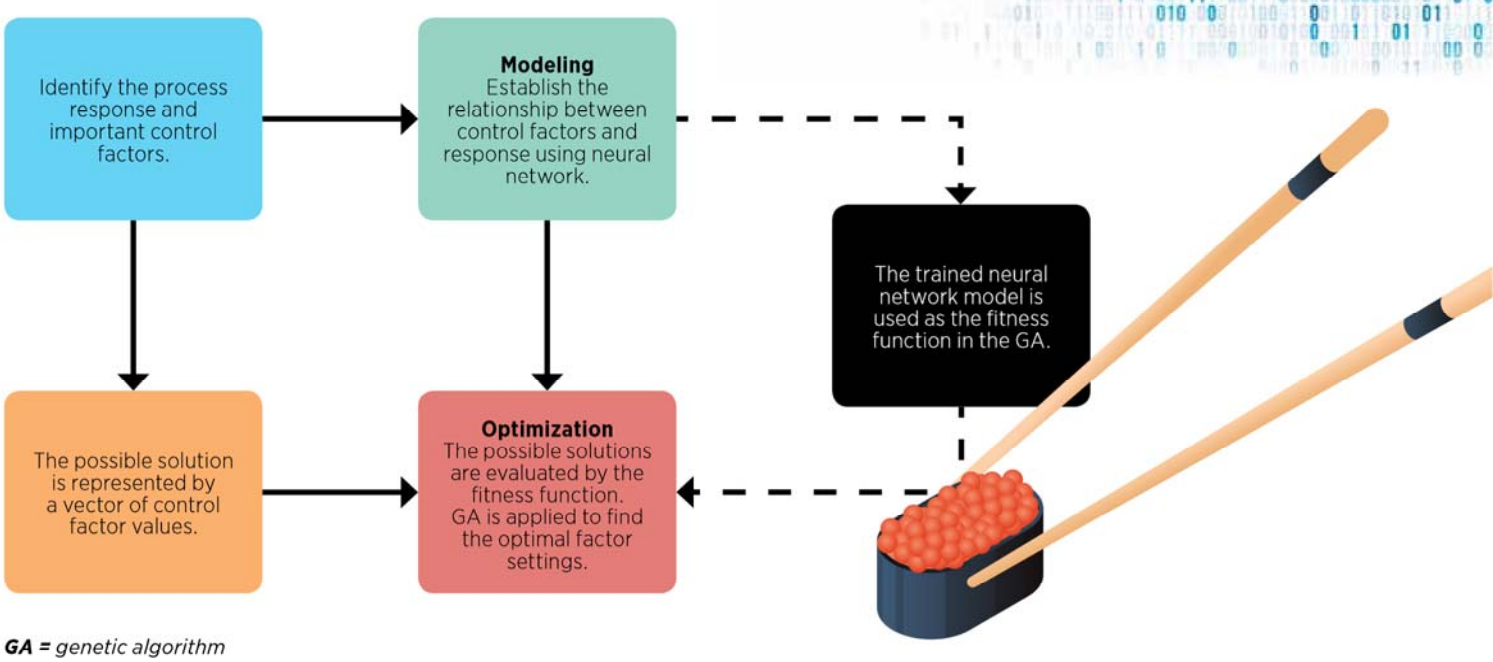
Big data analytics



ONYXprj via Getty Images

FIGURE 4

The central idea of process optimization (Example 1)



***GA** = genetic algorithm*

Based on these analyses, several valuable insights—such as vehicle speed control, RFID tag placement position and usage time, and determination of traffic volume—can be obtained to enhance the vehicle detection rate.

Key principles to remember

Big data have affected the task of quality improvement considerably. When using big data for quality improvement, understanding the engineering problem itself is crucial. However, three key principles also are useful for successful big data analysis:

1. **Data quality:** whether the data pose problems, such as those related to measurement, incorrect records or missing values.
2. **Methods of data analysis:** how to effectively and quickly process massive amounts of data and extract actionable information through appropriate tools.
3. **Customer perspective:** whether the problem being solved is a customer concern.

This situation is like making delicious Japanese sushi: To succeed, you must possess high-quality ingredients and be skilled in the craft to be able to create the right taste that meets customer needs. QP

REFERENCES AND NOTES

1. Doug Laney, "3D Data Management: Controlling Data Volume, Velocity and Variety," *Application Delivery Strategies*, Meta Group, File 949, Feb. 6, 2001.
2. Geoff Vining, "Technical Advice: Scientific Method and Approaches for Collecting Data," *Quality Engineering*, Vol. 25, No. 2, 2013, pp. 194-201.
3. Jim Duarte, "Data Disruption," *Quality Progress*, September 2017, pp. 20-24.
4. W. Edwards Deming, *The New Economics*, Massachusetts Institute of Technology Press, 1993.
5. For similar examples, see Chao-Ton Su's, *Quality Engineering: Off-Line Methods and Applications*, CRC Press/Taylor & Francis Group, 2013.

Chao-Ton Su is the chair professor of the department of industrial engineering and engineering management at National Tsing Hua University in Hsinchu, Taiwan. He received his doctorate in industrial engineering from the University of Missouri in Columbia. He is an academician of the International Academy for Quality, an ASQ fellow, a Chinese Society for Quality fellow and a Chinese Institute of Industrial Engineers fellow. Su authored *Quality Engineering: Off-Line Methods and Applications* (CRC Press/Taylor & Francis Group, 2013).