Selecting Key Product Characteristics to Improve the QMS in Automotive Sector

Laila Benzaza, Najlae Alfatehi and Abdelouahid Lyhyaoui

April 21, 2022

Selecting key product characteristics to improve the QMS in automotive sector

Laila Benzaza[1], Najlae Alfatehi [2], Abdelouahid Lyhyaoui[3]

[1]Laboratory of innovative technologies (LTI), National school of applied sciences (ENSA), Abdelmalek Essâadi University, Tangier 90000, Morocco
[2]Pluridisciplinary laboratory of research and innovation (LPRI), EMSI Casablanca, Morocco
[3]Laboratory of innovative technologies (LTI), National school of applied sciences (ENSA), Abdelmalek Essâadi University, Tangier 90000, Morocco

Abstract. In the automotive industry, one of the major factors contributing to customer satisfaction is the product’s quality.
To improve this latter, companies are implementing quality management systems: QMS, which are requiring usage of quality tools especially statistical ones, to control the manufacturing processes, by monitoring the product’s characteristics. But, must we track all the characteristics to achieve the product compliance and system efficiency? This paper aims to give a solution model that will allow the organizations to select the key product characteristics that should be monitored by quality tools as they have the most influence on the product’s quality.
The goal is to use mathematics in industrial field to improve quality management systems and manufacturing processes performances.
The paper also gives architecture of the discussed problematic and a perspective on the next work which will be an application of the proposed solution in the automotive sector.

Keywords: Automotive Industry · Data Mining · Linear regression · Machine Learning · Product characteristics identification · Quality Improvement

1 Introduction

Quality in automotive industry plays a crucial role in increasing business competitiveness, that’s why the organizations started a long time ago to adopt the quality management systems, studies about this topic have been performed in companies 70 years ago by Dr. Edward Deming and Dr. Joseph [1], then, international standards were created to give guidelines for QMS: Quality Management system implementation.
The most widely accepted quality standard is ISO9001, which mention as principles, customer focus and continuous improvement [2], the certification in accordance with this kind of standards provides the chance for companies to work with new customers, who seem difficult to reach without certification [3].
In the automotive sector, “IATF 16949” is the quality standard which is a complementary of ISO 9001 [4], it gives more specific requirements for automotive

sector [5], and its certification has increasingly became a must and can be a requirement from the customer, who may conduct audits to check the compliance of the QMS [6], this proves that the customer satisfaction is strongly linked to the supplier's Quality Management System due to its impact on product's quality.

One of the requirements of IATF is the mastery of common quality methods to ensure a compliant system; these methods are called the "5 core tools" and attempt to gain advantages in term of quality, cost and delay [7-8-9-10-13, 14], among these tools, there's "SPC" and "MSA" which are statistical tools acting on real data. The first one "SPC", helps the organizations to improve their processes by monitoring the product characteristics , and establishing suitable control charts whose interpretation determine whether a process is in or out of control, and afterwards acting on the main root causes behind the process instability and incapability[11-12, 13], generally these causes are schematized in the 7M diagram, among these 7M causes, there's "measure", which can directly influence SPC results since the output of the measurement system is the input of SPC which is the measured or tested product characteristics. So if the measure is not correct/real, automatically the results of SPC won't be properly interpreted [13], here comes the importance of "MSA" which allows the assessment of the whole measurement system including the appraisers (people who make the measurements) and measurements equipment/tools [14].

There are many other quality methodologies that can be used by companies and which are also depending on the data and the more data we have, the more in depth analysis we can get, but the problematic is what product characteristics should be controlled to maintain the quality of our processes? And if there are many product characteristics, wouldn't it be a waste of time to track all of them while not all of them directly influence the product quality?

To solve this problematic, we are going to use a data mining method to facilitate the identification of the key product characteristics that impacts the conformity of the product.

For this, we will proceed as follows: Section 2 of this paper presents the related work which is an overview on data mining methodology and a description of the problematic, section 3 contains an overview on the linear regression models and the proposed solution, the Section 4 provides a description of the case study that will be discussed and presented in the next work and finally section 5 gives some concluding thoughts and perspectives.

2 Related work

2.1 Data mining methodology

The data mining is a step of KDD (knowledge Discovery from Databases) used to produce an enumeration of patterns, called models, through the collected and analyzed data [15]. This data is then validated after different steps like data cleaning through the elimination of invalid, redundant or incomplete data, data correction and data assessment etc. The management of data before usage makes it more reliable and consistent [16-17], after that, millions of patterns can be generated, but only a few of them are interesting. When we say interesting, we mean that the pattern answer to the need, or it can prove a hypothesis already defined [17].

Data mining can be applied in many different fields and domains, it helps to avoid the subjectivity usually related to human being in making decision [18], this can be done by going from the known information which is the collected data to the unknown information, by analyzing and understanding the data tendency.

The data mining allows either the description or the prediction taking into consideration the inputs (data), this can be made using different models generated by data mining software like (SPSS, R, IBM, SAS) [18], many methodologies are

available and integrated in datamining like statistics, patterns recognition, and machine learning [17].
As we mentioned before, the data mining can be used for the prediction, thing that interests us the most in this paper since the work will be basically the prediction of the compliance of the product based on information about its characteristics, the aim is to easily identify the characteristics that have the bigger probability to influence this prediction.
Some of the methodologies for data mining indicated above and that can be used for prediction are: statistics and machine learning, or we can call it statistical learning, it is the application of statistical methods to make predictions or estimations using data, this can be done through the extraction of interesting patterns from data and the understanding of their meaning; this is what we call: [learning from data] [19].
This methodology is used in different fields like chemistry, medicine, cyber security, finance and industry, where the output of statistical learning can be schematized by a variable "**y**" and the input by a variable "**x**", these variables can be qualitative or quantitative [19].

2.2 Problematic

In our case, the input variables are the measured (quantitative) or tested (qualitative) product's characteristics, so if we consider "**x**" our input, then **"x"** is a vector of components x_i, which are the characteristics of the product.
Now, if we consider the product's conformity which is represented by the output "**y**", there's a relationship between **y** and the vector **x** [x_1, x_2... x_n].
In fact, if the product's characteristics x_i are not in line with the requirements, then, the product would not be compliant neither, when we say requirement, generally it's a target to achieve, for quantitative variables, this target is an estimated value with a tolerance margin, and for qualitative variables, the target is a compliant situation (OK variable/ pass variable), and this target is the reference with which a comparison is made to judge the conformity of the variables, we will call this target, the estimated variable.
The machine learning approach will attempt to predict the output "**y**" by learning from the vector "**x",** that means that the model is predicting the conformity of the product relying on the conformity of its characteristics [20]. However, the output "**y**" doesn't consider all the product characteristics (x_i) in an equal way, there are some characteristics that influence the conformity of the products more than others.
Let's take as an example of product in the automotive industry, a small electrical wire used in the manufacturing of other automotive products like engine wire harness; there are many characteristics that could be identified in this product, hence, the vector "**x**" may be [*color*, *length*, *section*, *insulation type strands type*], if we chose the two following characteristics : "*color"* and "*section"* of the wire and compare them, we will notice that the section is more important than the color to judge the conformity of the wire, because the section directly impacts the functionality of this product which is the wire connectivity, that means that the conformity of an "electrical wire" depends mostly on its section rather than its color.
To better understand this statement, let's consider as a threshold "a green wire with *4mm²* as a section", a blue wire with *4 mm²* will have the same performance as the threshold, because the electrical current that will pass through these two wires is the same, even if they don't have the same color.
But the electrical current passing through another wire with green color (same as the estimated variable) and a section equal to *2mm²* (less than the estimated variable), is small because the resistance here is greater due to the small section (according to the generic formula of electricity: U=R*I, where R: resistance, I: current and U: voltage)

and this may influence the wire functionality.

In the first example, the product quality is close to the estimated one even if the product is not the same color as the threshold, while in the second example; the quality of the product is poor even if the color of the product is the same as the threshold.

Therefore, the conformity of the product "**y**" in this example mainly depends on its section rather than its color.

Certainly in this example, we cannot exclude the wire color from the characteristics list to ensure -because it may be a customer requirement- and a blue wire with *4mm²* as section is considered as NG product because it has a NG characteristic (color) compared to the threshold (Green wire with *4mm²* as section), however, the characteristic to be closely monitored by statistical quality tools in this example, is the section of the wire, because it influence the most and directly its conformity (connectivity) and it can lead to a customer claim in case of discrepancy.

This paper tries to answer to a central question: how can we identify the product characteristic that the product's conformity mostly depends on, using Data mining and statistical learning?

What we are searching for is a data mining method that can help companies in automotive sector to choose the product characteristic "x_i" which is one of the variables considered as inputs, but with the most influence on the product's conformity "**y**" (the output), and consequently should be controlled and monitored the most by statistical quality tools or other quality methodologies instead of controlling all products characteristics arbitrarily.

If we consider again "**y**" the conformity of the product, and **x** [x_1, x_2 ...x_n] the characteristics of that product, we have:

$$\boldsymbol{y} = \mathrm{f}(\boldsymbol{x}) \tag{1}$$

Where "x_i" is quantitative or qualitative inputs (measured or tested characteristics) and "**y**" is a qualitative output (ok product /NG product).

First, let's define the function "**f**".

We notice that there's a direct relationship between the conformity of the product and the conformity of its characteristics means that if the value of "x_i" increases, the value of "**y**" increases too, and vice versa, therefore, there's a linear relationship between "**y**" and "**x**".

3 Solution: LR model and LS estimation

The linear regression LR is a data mining model that tries to establish a linear relationship between an output variable "**y**" called explained variable and one or more input variables "x_i" called explanatory variables,

The generic form of the linear regression model is

$$\begin{aligned} y &= f(x_{1,}x_{2,\dots,}x_n) \\ &= x_1\beta_1 + x_2\beta_2 + \dots + x_n\beta_n + \varepsilon \\ &= \textstyle\sum_{i=1}^{n}(x_i\beta_i) \quad + \varepsilon \end{aligned} \tag{2}$$

Where "ε" is the random error component which is the difference between **y':** the estimated output variable**,** and the actual **y**.

As the conformity of the product "**y**" depends on all its characteristics "x_i", then, we will use the linear regression model to estimate "**y**".

The reason why we chose the linear regression is that this method allows an easy interpretation of the coefficients as probabilities using the LS estimator (Least Square) and an effective examination of the input effects on the output [21, 22].

For a product "y_j" where **j** ∈ {*1... k*} and *k* is the number of analyzed samples of a population (products), we have:

$$y_j=\mathrm{f}(x_{j1}, x_{j2}...,x_{jn})=\sum_{i=1}^{n}(x_{ji}\beta_{ji})+\varepsilon_j \quad (3)$$

To estimate the parameters "β_{ji}" where **i** ∈ {*1... n*} and "n" is the number of product characteristics, the maximum likelihood (ML) is one of the approaches used by statisticians for estimating a mathematical model parameter.
This estimation in multiple linear regressions is equal to the least squares (LS) which is another approach for parameters estimation but gives the same results for linear regression analyses when the dependent variable is normally distributed [23].
The principle of LS is to make the sum of squares of "**ε**" as small as possible.
So, if we have ***k*** samples, and we consider a normal distribution of "x_i",
The sum of squares of errors "**s**" would be [24]:

$$s=\sum_{j=1}^{k}(\varepsilon_j)^2 \quad (4)$$
$$=\sum_{j=1}^{k}(y_j-y_j')^2$$

Where (y_j) is the output and (y_j') is the estimated output of the sample **j**.

And $$y_j=\sum_{i=1}^{n}(x_{ji}\beta_{ji})+\varepsilon_j \quad (5)$$

And $y_j'=\sum_{i=1}^{n}(x_{ji}\beta_{ji}')$ because there's no error in the estimated yj'

So $$\varepsilon_j=(y_j-y_j'))$$
$$=y_j-\sum_{i=1}^{n}(x_{ji}\beta_{ji}') \quad (6)$$

From (4) and (6), we get:

$$s=\sum_{j=1}^{k}(y_j-\sum_{i=1}^{n}(x_{ji}\beta_{ji}'))^2 \quad (7)$$

To find the unknown parameters, β_{ji}, we minimize s,

$$\min(s)=min\sum_{j=1}^{k}(y_j-\sum_{i=1}^{n}(x_{ji}\beta_{ji}'))^2$$

Now differentiate "s" with respect to β_{ji},

$$\frac{ds}{d\beta_{ji}}=-2\sum_{j=1}^{k}x_j(y_j-\sum_{i=1}^{n}(x_{ji}\beta_{ji}'))=0$$

After verification, we obtain: $$\sum_{j=1}^{k}x_j(y_j-\sum_{i=1}^{n}(x_{ji}\beta_{ji}'))=0$$

$$\sum_{j=1}^{k}y_j-\sum_{i=1}^{n}(x_{ji}\beta_{ji}')=0 \quad (8)$$

After the identification of the parameters "β_{ji}", which will represent the percentage of influence of the relevant characteristic "x_{ji}" on the global conformity of the product "y_j", we will hereafter select the bigger "β_i" and then associate the key characteristics "x_i" that have the most influence on the product conformity "y".

Conclusion: We will be able to easily identify the key characteristics of a product based on an analytical and statistical study.

4 Case study

In this section, we will describe the case study that will be the subject of the next work; it consists on applying the proposed solution in an affiliate of an international company for vehicles wiring harnesses production.

The picture below gives a simplified description of the manufacturing process that begins by "cutting" where the wires are cut, then stripped (insulation removed from the end of the cut wire), then "crimping" (mechanical operation which consists in assembling two parts by deformation, in our case the two parts are wire and terminal), After that comes the pre-assembly process P2 where some of the wires –depending on the wire harness design-are welded (joined) and others are twisted (these two sub-processes are done in a parallel way), and the last process P3 consists to insert terminals into the connectors, make the lay outing and taping operations and assembly of all the semi-finished products coming from P2 process to constitute a complete wire harness, ready to be shipped to the customer after its electrical and visual inspection.

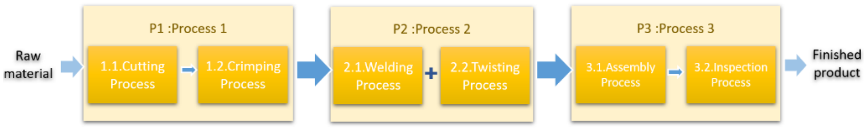

Figure 1: Manufacturing processes of Wire harness

In each of these processes, there is a set of characteristics to ensure, for example the output of the twisting process has as characteristics the length and the twisting pitch.

According to quality data (PPM: part per million) during the 5 last years, the processes with high ratio of defect are "crimping" and "welding" processes, that's why, the company controls them by SPC tool, this tool is considering all the characteristics in the relevant processes.

The objectives of the upcoming work are:

1-to prove the reliability of the proposed solution in this paper,

2-to make sure that the company is controlling the right processes,

3-to decrease the number of characteristics by process to be considered in SPC methodology,

4-to make the process control method easier and faster.

The steps of the next work will be basically the collection of data, data analysis and application of the solution presented in this paper.

5 Conclusion and perspectives

To effectively improve the QMS of a company in the automotive industry, statistical tools can be used to control the manufacturing processes using as input the product's characteristics; the guidelines related to such tools give only the steps and requirements about their implementation but the identification of the product characteristics is left to the company's decision, this latter may consider all the characteristics in her method, which can lead to a waste of time and a complicated operation.

This paper proposed a solution to facilitate the identification of the key product characteristics which have the most influence on the product quality, and described the architecture of the discussed problematic that will be the subject of our next work.

References

1. Janis Priede, Implementation of Quality Management System ISO 9001 in the World and its Strategic Necessity, 8th International Strategic Management Conference, University of Latvia, Riga, LV-1050, Latvia (2012)
2. ISO, ISO 9001 Quality Management Systems Requirements, 5th edition, International Quality standard book (2015)
3. Joaquı´n Texeira Quiro´s University of Extremadura, Badajoz, Spain, and Maria do Rosa´rio Fernandes Justino ISCAL-Polytechnic Institute of Lisbon, Lisbon, Portugal, A comparative analysis between certified and non-certified companies through the quality management system, Journal, pp 958-969 (2013)
4. IATF: International Automotive task force, Quality Management System requirements for automotive production and relevant service parts organizations, 1st edition, Automotive Quality Management Standard book (2016)
5. Iker Laskurain-Iturbe, Germán Arana-Landín, Iñaki Heras-Saizarbitoria & Olivier Boiral, How does IATF 16949 add value to ISO 9001? An empirical study, Journal pp 1478-3371 , DOI: 10.1080/14783363.2020.1717332, (2020)
6. Hector Hernandez, Quality audit as a driver for compliance to ISO 9001:2008 standards, , Journal pp 454-466, Westport Innovations, Vancouver, Canada (2008)
7. Jigar A. Doshi1, Darshak Desai, Overview of Automotive Core Tools, Applications and Benefits, Review paper, Gujarat, India, DOI 10.1007/s40032-016-0288-z (2016)
8. AIAG, Ford Motor Company, Chrysler Corporation, and General Motors Corporation, Advanced Product Quality Planning and control plan, Reference manual book, second edition (2008)
9. AIAG, Ford Motor Company, Chrysler Corporation, and General Motors Corporation: Production Part Approval Process, Reference Manual book, Fourth edition, (2006)
10. AIAG, Ford Motor Company, Chrysler Corporation, and General Motors Corporation: Potential Failure Mode and Effects Analysis (FMEA) Fourth Edition book, (2008)
11. Radu Godina, João C.O. Matias, Susana G. Azevedo, Quality Improvement With Statistical Process Control in the Automotive Industry, International Journal pp 1-8, University of Beira Interior, Covilhã, Portugal (2016)
12. Jiju Antony and Tolga Taner, A conceptual framework for the effective implementation of statistical process control, Journal, pp 473-489, DOI 10.1108/14637150310484526, Warwick Manufacturing Group, University of Warwick, Coventry, UK and Institute of Biomedical Engineering, Bogazici University,Istanbul, Turkey (2003)
13. AIAG, Ford Motor Company, Daimler Chrysler Corporation and General Motors Corporation: Statistical Process Control, reference Manual book, Second edition (2005)
14. AIAG, Ford Motor Company, Daimler Chrysler Corporation and General Motors Corporation: Measurement System Analysis, reference manual book, fourth edition (2010)
15. Ralf Mikut and Markus Reischl, Data mining tools, advanced review pp 431-443, DOI: 10.1002 (2011)
16. Suranga C. H. Geekiyanage, Andrzej Tunkiel, Dan Sui, Drilling data quality improvement and information extraction with case studies, Journal of Petroleum Exploration and Production Technology, University of Stavanger, Stavanger, Norway (2020)
17. Jiawei Han, Micheline Kamber, Jian Pei, Data Mining Concepts and techniques, third edition, published book by Elsevier, Simon Fraser University (2012).
18. Stéffane Tufféry, Data Mining et statistique décisionnelle L'intélligence des données, book 4th edition, University of Rennes, France (2012) (In French)
19. Trevor Hastie, Robert Tibshirani, Jerome Friedman, the elements of statistical learning, Data mining, Inference, Prediction, Springer series in Statistics, second edition (2013).
20. Rodrigo Fernandes de Mello & Moacir Antonelli Ponti, Machine learning A practical approach on the statistical learning theory, ebook, Library of Congress Control Number: 2018947414 (2018)
21. Robin Gomila, Logistic or Linear? Estimating Causal Effects of Experimental Treatments on Binary Outcomes Using Regression Analysis, Journal, Princeton University (2020)
22. Ottar Hellevik, Linear versus logistic regression when the dependent variable is a dichotomy, Original paper published by Springer Science+Business Media (2007)
23. David G. Kleinbaum · Mitchel Klein, Logistic Regression, A Self-Learning Text, Third edition, springer, USA (2010)
24. T. Daniya, M. Geetha, B. Santhosh Kumar, R. Cristin, Least Square Estimation of Parameters for Linear Regression, International Journal, India (2020)