Optimizing the Request for Quotation process with machine learning-based connector clustering

Alexander Lemmens

Master of Electronics and ICT Engineering Technology

Introduction

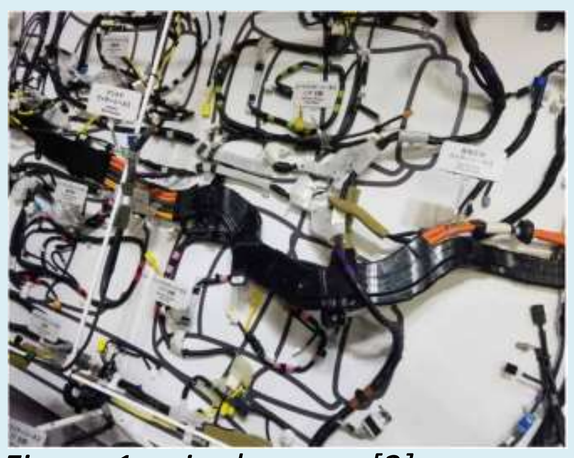

Figure 1: wire harness [2]

Yazaki is a global automotive supplier to all of the major Original Equipment Manufacturers (OEMs).

An OEM can require particular parts, like **wire harnesses**. A wire harness in the automotive industry assembles different components such as cables, terminals, **connectors**, etc. [1], [2]. A wire harness is shown in Figure 1.

Connectors are essential parts of a wire harness because they are **weak points** in systems, can be made with **high-profit margins**, and are **complex parts**, including 123 characteristics [3]. A connector is illustrated in Figure 2.

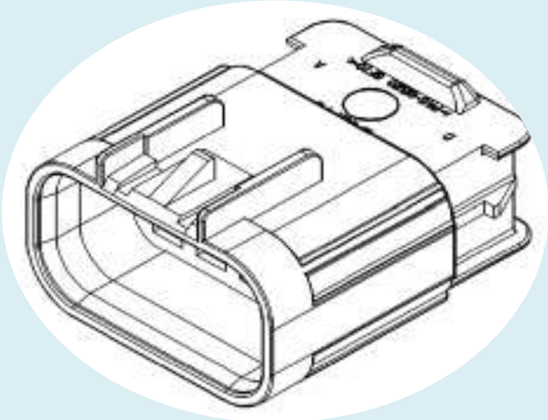

Figure 2: connector [11]

The requirements are communicated using the **Request for Quotation** (RfQ) process. The supplier then reacts to it with a response [3]. This response consists of a technical and commercial offer. The technical part consists of the product requirements, and the commercial part consists of the price, annual volumes, etc. Because of this, clear specifications are necessary **for accurate pricing and component selection** [4].

Problem and objective

Recent studies show that **time restrictions** are becoming more important during the RfQ response process. The methods now used process the data **manually**, meaning that interchangeable connectors must be found by manually searching the available databases. This results in long lead times and a disruption of the day-to-day schedule, which causes delays in current business [4].

The main objective of this thesis is **to find similar connectors** based on their characteristics to **speed up the RfQ process** using machine learning-based clustering methods.

However, before using clustering techniques, this thesis aims to **create preprocessed datasets** to optimize the grouping results and performances. Different clustering techniques will be applied to the preprocessed dataset and **compared** to investigate which ones are best for grouping the connectors based on their characteristics without knowing the type of connectors.

Another objective is to suggest a clustering approach using the K-means method specifically to group wire harness connectors, using **additional information** about the connector types to validate and compare various approaches.

Finally, this study aims to find the **advantages and limitations** of the suggested approaches and discusses **feature work** to improve these suggestions.

Methods

The methodology to achieve the objectives is split into three parts: **data preprocessing**, **machine learning clustering**, and **validation and testing**.

1. Data preprocessing involves **cleaning**, **reducing**, and **transforming** the data to enhance quality. Figure 3 shows a visual overview of the preprocessing steps used. Data cleaning consists of **handling outliers** and **missing values**. Data reduction aims to reduce the form of the dataset while maintaining its essential features, including **feature selection** and **dimensionality reduction**. Data transformation will change the dataset's structure, so it becomes appropriate for clustering methods using **normalization**, **standardization**, and **encoding categorical values**.

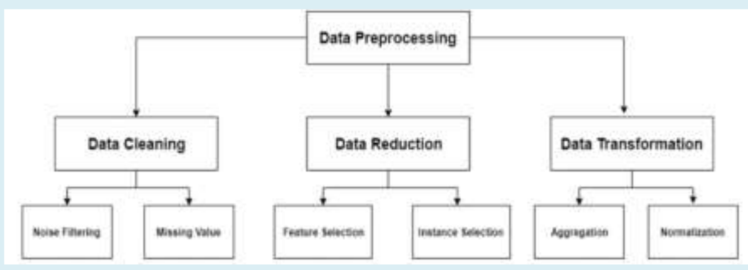

Figure 3: Classification of data preprocessing techniques [7]

3. Validation and testing are crucial to **evaluate** the different clustering methods. This ensures that the results meet the objectives and provide meaningful insights. For this, both **confusion matrices** and **silhouette scores** were performed.

2. The machine learning clustering part of the methodology involved **unsupervised** learning containing **K-means clustering**, **balanced K-means** clustering, the **Gaussian Mixture Model** (GMM), and **Density-Based Spatial Clustering of Applications with Noise** (DBSCAN). These methods were performed on a dataset of approximately 6,500 connectors without true labels. Additionally, **supervised clustering** was performed on a dataset of 1020 connectors with true labels to **optimize** the K-means clustering method to group the connectors into 17 groups. Finally, to identify the most similar connectors based on their characteristics and find interchangeable connectors, the K-means and DBSCAN methods were **combined**.

Results

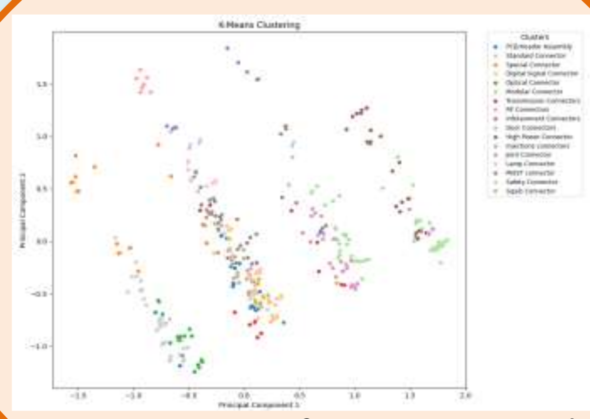

Figure 9: Best performing approach

Figure 4: K-means clustering

Figure 4 validates that the results of the K-means algorithm produce clusters of **the same shape and size** due to how the algorithm works. It also doesn't handle noise, forcing data points to clusters it may not belong to.

Figure 5 demonstrates that DBSCAN has identified nine clusters and a group called noise. Contrary to the similar size and shape in the K-means results, it identifies clusters of **different densities and shapes**. This gives more insights into datasets with more **complex structures** and **high dimensions**.

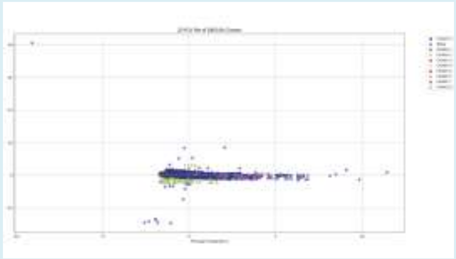

Figure 5: DBSCAN clustering

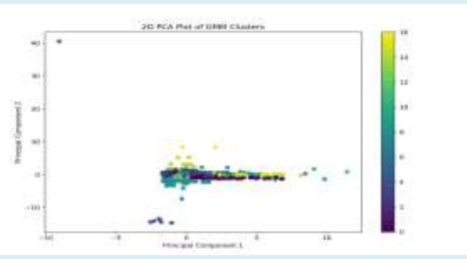

Figure 6: GMM clustering

Figure 6 shows that most of the data points lie in the center of the plot after grouping the connectors using GMM. The clusters are **more compact** compared to the results of K-means and DBSCAN.

Figure 7 illustrates the confusion matrix of the **best-performing approach**, namely K-means, without standardizing the data, without adding weights, using the average values of the connectors per type as the initial center, and handling the outliers. The accuracy of this method is **62%**, and Figure 9 shows the plot of this approach.

Figure 7: Validation of best-performing approach

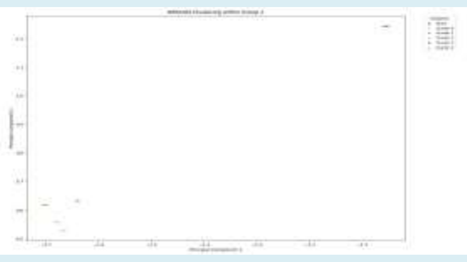

Figure 8: DBSCAN results for MOST connectors

Figure 8 illustrates the **third round of clustering** by **combining** the K-means and DBSCAN results. This approach is **successful** because it is tested with information from Yazaki.

Conclusion

For the dataset of 6,500 connectors without true labels, it was shown that DBSCAN identifies clusters of different shapes and densities better than K-means. The plots of the results after using GMM demonstrated that the clusters are more compact compared to K-means and DBSCAN. The second part of the thesis used the dataset of 1020 connectors, which included true labels. Here, the best-found approach was the K-means method, without standardizing the data, without adding weights, using the average values of the connectors per type as the initial center, and handling the outliers gives the best performance of K-means in terms of confusion matrices and the silhouette scores. Finally, the most similar connectors of the MOST connector group were successfully found.

Supervisors / Co-supervisors / Advisors: Rosa Rocha and Prof. Nikolaos Tsiogkas

[1] Y. Corporation, "YAZAKI Corporation," *YAZAKI Corporation*. https://www.yazaki-group.com/en/ (accessed Mar. 23, 2024).
[2] L. Chen, "Automotive Wire Harness Connectors: An ultimate guide on wiring connectors for automotive applications," *Wire Harness and Cable Assemblies Manufacturer - Cloom*, Sep. 27, 2022. https://www.wiringo.com/automotive-wire-harness-connectors.html (accessed Mar. 23, 2024).
[3] "Degradation of road tested automotive connectors," IEEE Journals & Magazine | IEEE Xplore, Mar. 01, 2000. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=833055 (accessed Mar. 26, 2024).
[4] Car - centre for automotive research, "The strategic value of information in the RFQ respons process," Center for Automotive Research, Aug. 2007.

YAZAKI

De opleiding industrieel ingenieur is een gezamenlijke opleiding van UHasselt en KU Leuven

UHASSELT KU LEUVEN