METHODICAL APPROACH TO CLUSTER CONFIGURATIONS OF PRODUCT VARIANTS OF COMPLEX PRODUCT PORTFOLIOS

Mehlstäubl, Jan (1);
Pfeiffer, Christoph (1);
Kraul, Ralf (2);
Braun, Felix (2);
Paetzold-Byhain, Kristin (1)

1: Technische Universität Dresden;
2: MAN Truck & Bus SE

ABSTRACT

Companies are increasingly struggling to manage their complex product portfolios. Since they do not fully understand the complexity, intelligent solutions are required. Emerging technologies and tools offer new ways to deal with existing problems. With the help of clustering, similarities between product variants can be identified automatically, and complexity can be systematically reduced. This article aims to develop a methodological approach to identify correlations between product variants in complex product portfolios automatically by using clustering algorithms. The approach includes the systematic cleaning and transformation of product portfolio data. In addition, a guide for algorithm selection and evaluation of clustering results is presented. As the last step, the results are systematically analysed and visualised. To validate the methodical approach, it is applied to a real-world data set of a commercial vehicle manufacturer and the usefulness of the results is confirmed in an expert workshop.

Keywords: Complexity, Machine learning, Portfolio management

Contact:
Mehlstäubl, Jan
University of the Bundeswehr Munich
Germany
jan.mehlstaeubl@unibw.de

Cite this article: Mehlstäubl, J., Pfeiffer, C., Kraul, R., Braun, F., Paetzold-Byhain, K. (2023) 'Methodical Approach to Cluster Configurations of Product Variants of Complex Product Portfolios', in *Proceedings of the International Conference on Engineering Design (ICED23),* Bordeaux, France, 24-28 July 2023. DOI:10.1017/pds.2023.265

1 INTRODUCTION

The number of product variants offered by companies has increased significantly in recent years due to changes in customer demands and increased global competition in many industries (Krause and Gebhardt, 2018). The growing external variety leads to an increase in internal variety and thus to higher complexity and costs (Schuh *et al.*, 2018). To keep internal variety and complexity as low as possible, companies pursue different strategies for product structuring (e.g. platforms or modular systems). Nevertheless, the large number of product variants is increasingly challenging for companies (Schmieder and Thomas, 2005). For example, a BMW 7 can have up to 10^{17} possible product variants (Hu *et al.*, 2008). The number is much higher for commercial vehicles (Kusiak *et al.*, 2007). Although the complexity is no longer manageable manually, the activities in product portfolio and variety management are driven by the experiential knowledge of the developers (Mehlstäubl, Braun, *et al.*, 2022). Machine learning techniques offer great potential in these disciplines and make it possible to gain insights from large amounts of data and thus provide a meaningful basis of information for decision-making processes in product portfolio and variety management (Mehlstäubl *et al.*, 2023). Especially clustering algorithms offer great potential for identifying similarities among product variants (Hochdörffer *et al.*, 2018). Compared to manual approaches, there are advantages in terms of time and costs as well as the objectivity of the results. The described motivation and problem statement results in the following research question: How can correlations between product variants in complex product portfolios be detected automatically using clustering algorithms?

In the following terminological and methodological background of complex product portfolios and clustering is described in section 2. Subsequently, in section 3, state of the art on the clustering of complex product portfolios is presented and the need for research is derived. Section 4 presents the research methodology used to conduct this research. Section 5 describes the methodological approach for clustering complex product portfolios. Subsequently, in section 6, the validation of the results through a case study at an industrial partner from the commercial vehicle sector is conducted. The validation includes the application of the approach to a real-world data set as well as the assessment by experts in a workshop.

2 BACKGROUND

2.1 Complex product portfolios

The product portfolio refers to all products and/or services that a company offers on the market (Jonas, 2013). It consists of several product families, which represent a selection of similar product variants developed on a common product platform (Kissel, 2014). Product variants are products of similar form or function with a generally high proportion of identical groups or parts (DIN 199-1, 2002). In systems theory, complexity is understood to be the number, variety and relationships of the elements as well as their states and variability (Krause and Gebhardt, 2018). Complex product portfolios result from the number, variety and temporal variability of features and components, their relationships, and the resulting product configurations. Since a lot of product variants are not profitable due to complexity costs, many companies are focusing on projects to rationalise product variety. Examples of complex product portfolios are those of passenger or commercial vehicle manufacturers with hundreds or thousands of features that can be selected by the customer (Greisel *et al.*, 2013).

2.2 Clustering as a part of machine learning

Machine learning is the science that gives computers the ability to learn without being explicitly programmed (Samuel, 1959). A computer program learns from experience E concerning a task T and a performance measure P if its performance on T, measured against P, improves with experience E (Mitchell, 1997). Murphy (2012) extends the definition to include the use of patterns for prediction and decision support. He defines machine learning as a set of techniques that can automatically detect patterns in data and use them to predict future data or make other types of decisions under uncertainty. Clustering is a technique of unsupervised machine learning. In unsupervised learning, patterns are extracted from unlabelled data. The aim is to identify similar instances in a data set and to divide them into homogeneous groups (Géron, 2017). This can be expressed as the partitioning of n data points of a data set D = {x1, x2, ..., xn} into k disjunctive subsets C1, C2, ..., Ck. Each data point is described as

https://doi.org/10.1017/pds.2023.265 Published online by Cambridge University Press

a vector of feature values (Kubat, 2021). This results in different approaches for clustering procedures. A distinction can be made between density-based, distance-based, probabilistic, and hierarchical clustering algorithms. The clustering itself is only a single phase in a data analysis process. For the industrial application of clustering, a data analysis process must be run through entirely. In this article, the CRoss Industry Standard Process for Data Mining (CRISP-DM) by Wirth and Hipp (2000) is used to develop the methodological approach to cluster complex product portfolios. This process focuses the industrial application of data analysis and is characterised by the phases of business and data understanding at the beginning. These phases are particularly important in industry and product design due to the complexity of the domains and the data challenges (Mehlstäubl, Gadzo, *et al.*, 2022).

3 STATE OF THE ART

The state of the art is based on the literature review conducted by Mehlstäubl *et al.* (2021) on the use of data mining in product portfolio and variety management. For this article, the relevant approaches are those that use clustering for variety reduction and control as well as for market analysis. Zhang et al. (2007) investigate ways to identify product families based on market segmentation. They define product families directly from the preferences of the respective customer group, which result from the feature combinations sold. Fuzzy clustering is used for this purpose, which considers the fuzziness of customer preferences. Tucker *et al.* (2010) propose a methodology for the top-down development of product families. This makes it possible to derive the optimal number of product families based on customer preference data and to minimise product portfolio cost. They apply the ReliefF algorithm for weighting product features and the X-Means algorithm for clustering. In the context of individualised mass production, Kusiak *et al.* (2007) investigate possibilities for the standardisation of product configurations to reduce variety-induced complexity. They cluster historical sales data of trucks using the k-means algorithm to identify central customer configurations. Chan *et al.* (2012) investigate ways to identify ideal points within market segments. They consider that evaluating product features from the consumer's point of view is not fully separable when making a purchase decision. They use also fuzzy clustering for market segmentation based on customer survey data.
Romanowski and Nagi (2004) are developing an approach for the automated generation of generic BOMs. First, purchased parts are unified into part groups using a hierarchical clustering algorithm. Subsequently, all existing sub-assemblies are grouped using the k-medoid algorithm. In the k-medoid algorithm, the most centrally located data point of a cluster is defined as the cluster centre. Finally, using the same algorithm, groups of similar products, i.e. product families, are identified from all existing parts lists. Neis (2015) applies a two-stage clustering to BOMs. In the first step, the product portfolio is structured with a hierarchical clustering and the number of clusters or product families is determined. In the second step, a k-medoid procedure is used to determine the reference product structure. The medoids serve as the starting point for the reference product structure. Ma and Kim (2016) present a method in which product architecture candidates are generated using the k-means clustering algorithm for different values of k. The method is based on a time series analysis. Using a time series analysis, the expected profits are then modelled to determine the optimal position and number of product architectures among the product architecture candidates. The current state of the art shows a need for further research regarding three main fields associated with the defined research question. First, none of the existing approaches is clustering data of complex product portfolios with a variety of features and sold configurations. Second, there is a need for research regarding the optimisation of clustering by using and comparing different algorithms. Third, the existing approaches do not evaluate the quality of clustering by applying different validity indices.

4 RESEARCH APPROACH

The research approach of this article is based on the Design Research Methodology (DRM) type 5 according to Blessing and Chakrabarti (2009). In the first phase, the research question is defined based on a literature review and the challenges of an industrial partner. Subsequently, the descriptive study I provides a deeper understanding about theoretical as well as methodological background. In the prescriptive study, the methodological approach to cluster complex product portfolios is introduced. It is derived from the CRISP-DM process model and describes the individual phases for the analysis of complex product portfolios with clustering. The approach consists of five phases and is validated in the final descriptive study II with a case study in the

https://doi.org/10.1017/pds.2023.265 Published online by Cambridge University Press

commercial vehicle industry. Due to the global competitive situation and the multitude of transport tasks and application scenarios, commercial vehicle manufacturers have a particularly broad and deep product portfolio (Kreimeyer *et al.*, 2013) and are therefore particularly well suited for a case study. The descriptive study II consists of a validation of the applicability as well as the success of the introduced approach. In the application validation, a real sales data set with configurations of product variants was analysed. In the success validation, the results are evaluated with eleven experts from the industrial partner in a workshop.

5 METHODICAL APPROACH TO CLUSTER COMPLEX PRODUCT PORTFOLIOS

The developed methodical approach enables the clustering of data that represent a complex product portfolio. An overview of the approach is given in Figure 1. The first step is to describe and select the target data, i.e. the data set on the basis of which the clustering of a product portfolio can be carried out (section 5.1). The preparation of the data includes the handling of missing values, encoding and dimension reduction (section 5.2). In clustering, the algorithms are selected and implemented systematically (section 5.3). The next step is the evaluation of the clusters identified with the help of clustering validation indices (section 5.4). Finally, the clusters are analysed and the optimal clustering is visualised (section 5.5).

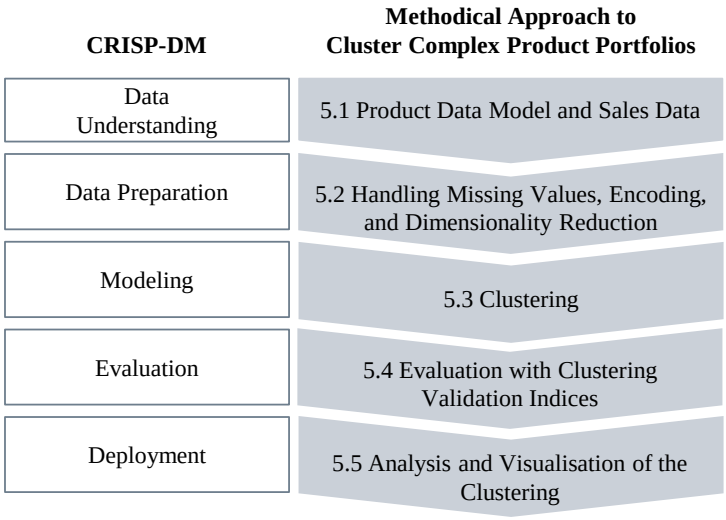


Figure 1: Overview of the methodical approach

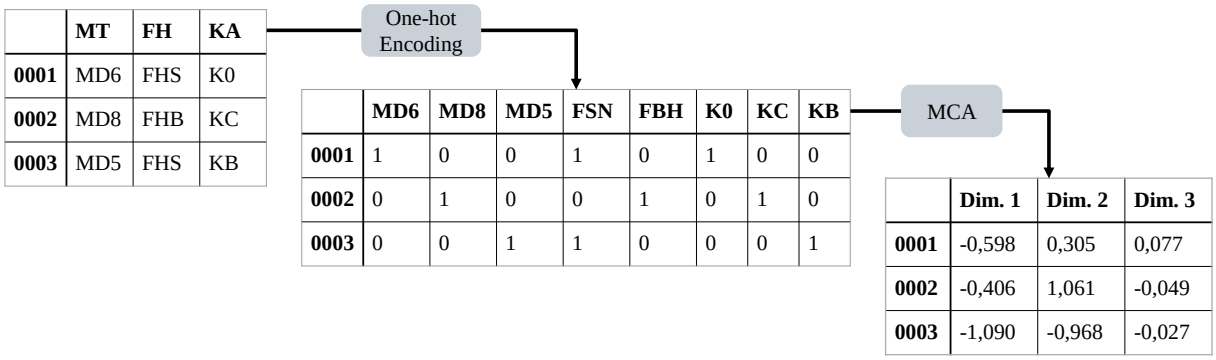
5.1 Product data model and sales data

In complex product portfolios, the product variants are derived from a generic product data model that is valid for all product variants (Braun *et al.*, 2017). The main elements of a product data model are the feature structure, the product structure and the rule sets (Kreimeyer *et al.*, 2016). The feature structure contains all feature categories (e.g. suspension) and the associated features (e.g. air or leaf suspension) that can be selected by the customer and allows a complete product description of all possible product variants from a sales perspective. The dependencies and constraints between the individual features are mostly defined by the Boolean combinatoric rules (Mortensen *et al.*, 2000). The product structure represents the corresponding engineering view and structures all technical solution modules. It contains the components (e.g. steering wheel) and the associated component variants (e.g. steering wheel standard or plastic). The component variants represent the smallest elements, the technical modules, which can be combined into a complete product. In the configuration process of a product variant, one feature from each feature category is selected by the customer based on the specifications and restrictions of the product data model. The resulting component variants are picked based on the features with Boolean part selection rules. Clustering can be carried out at both the feature level and the component level.

https://doi.org/10.1017/pds.2023.265 Published online by Cambridge University Press

5.2 Handling missing values, encoding, and dimensionality reduction

The product data model is in constant motion as feature categories and components are added or removed in line with product development cycles. For this reason, the sales data contains missing values, which must be cleaned up. Feature categories or product variants can be deleted if the majority of the entries contain missing values. If only individual values are missing, they are replaced by a fixed value.
An attempt is made to replace the missing value with the value with the highest probability (Han *et al.*, 2012). In cases with categorical input data with no ordinal relationship, the most probable value is the most frequent expression. Clustering algorithms can only process numerical data, which is why encoding is required. Since there is no ordinal relationship between most of the features, one-hot encoding is used. In one-hot encoding, a discrete nominal variable x, which can assume the values x1, x2, ..., xn is converted to a binary vector v (Hancock and Khoshgoftaar, 2020). If a certain expression xi of x is to be coded, every element of v receives the value 0 except the i-th element, which takes the value 1. Figure 2 shows how the original form of the target data is changed by one-hot encoding.
The computational effort for most clustering algorithms increases rapidly with the number of dimensions. With a dimensionality reduction, the original data can be projected into a lower dimensionality. For a one-hot encoded dataset, Multiple Correspondence Analysis (MCA) can be applied (Abdi and Valentin, 2007). It is an extension of Correspondence Analysis, which makes it possible to examine the relations of several categorical attributes. The MCA provides two important results. First, the singular and eigenvalues as well as the percentages of the explained variety are derived. Secondly, the projection of the original data points onto the coordinates of a low-dimensional space is provided (Figure 2).



	MT	FH	KA
0001	MD6	FHS	K0
0002	MD8	FHB	KC
0003	MD5	FHS	KB

	MD6	MD8	MD5	FSN	FBH	K0	KC	KB
0001	1	0	0	1	0	1	0	0
0002	0	1	0	0	1	0	1	0
0003	0	0	1	1	0	0	0	1

	Dim. 1	Dim. 2	Dim. 3
0001	-0,598	0,305	0,077
0002	-0,406	1,061	-0,049
0003	-1,090	-0,968	-0,027

Figure 2. Encoding and dimensionality reduction

5.3 Clustering

In this phase, different clustering algorithms are implemented. If the same amount of data is clustered using different algorithms, the clusters and their quality determined in each case differ. Table 1 shows an overview of the characteristics that positively or negatively influence the performance of clustering algorithms. Product portfolio data has a high variance and high dimensionality. Besides the choice of algorithms, the second important influence factor for the clustering quality is the number of clusters. Therefore, the data set must be clustered several times, varying the number of clusters searched for in each case.

5.4 Evaluation with clustering validation indices

As explained in the previous section, the quality of an identified clustering depends on both the selection of the appropriate algorithm and the number of clusters. Therefore, the clustering results must be analysed considering these two factors. For this purpose, so-called clustering validation indices (CVI) are used. Each of these indices follows a different calculation formula and thus focuses on a specific aspect of the structure of a clustering. For the evaluation of the identified clusters, three different CVIs are used, which can be applied independently of the algorithm. The first CVI is the Davies-Bouldin-Index, which maps the average similarity of each cluster to the most similar cluster (Davies and Bouldin, 1979). Clusters that are further apart and less dispersed have smaller values of the Davies-Bouldin-Index and indicate higher quality of the clustering. The second CVI is the silhouette coefficient, which indicates how accurate the assignment of instances to a cluster is by

https://doi.org/10.1017/pds.2023.265 Published online by Cambridge University Press

relating their similarity to instances in the same cluster to their similarity to instances in the nearest cluster (Rousseeuw, 1987). The clusters of clustering are denser and better separated, the higher the silhouette coefficient. The last CVI is called inertia and is the sum of the squared distances of the instances to their respective nearest centroid. The smaller the inertia of clustering, the closer the instances are to the corresponding centroid and the more compact the individual clusters.

Table 1. Overview of algorithms and their characteristics

Algorithm	Cluster Definition	Characteristics	Application
k-Means, *distance-based*	Clusters are groups of instances represented by their centroid	• Simple calculation • Depending on initialisation	+ Spherical clusters - Cluster density & size strongly varies
Mini-batch k-Means, *distance-based*		• High speed for large amounts of data • Lower accuracy	+ Spherical clusters - Cluster density & size strongly varies
EM, *probabilistic*	Clustering consists of several probability distributions	• Shape and density depending on the distributions • Different covariance matrices (e.g. round, spherical)	+ Elliptical or spherical clusters - High dimensionality - Few instances per cluster
DBSCAN, *density-based*	Clusters are areas of high density	• Identification of outliers possible • Density parameters difficult to determine	+ Clusters of any shape - Cluster density varies greatly
Linkage, *hierarchical*	Clustering consists of a multi-level hierarchy of nested clusters	• k does not have to be given in advance • Different linkage functions applicable	+ Overlapping clusters + Spherical clusters - Depending on the linkage function: varying cluster size & density, clusters very close to each other

5.5 Analysis and visualisation of the clustering

The result of the clustering consists of the assignment of the identification numbers to the clusters. To enable the interpretation of the results, the identification numbers must be merged with the feature categories and features of the product variants. Subsequently, the clusters can be analysed regarding the three aspects characteristic features, distances, and market-specific attributes. Features are characteristic for a cluster if they occur in all or almost all product variants of the cluster. This enables the analysis of unique and common features of the individual clusters. The distances between the clusters can be considered by visualising them in two-dimensional or three-dimensional space or in a matrix. Through visualisation, conspicuous clusters can be identified easily and through comparison in a matrix, the exact values can be examined. For the analysis of the product portfolio, market-specific attributes of the clusters can be considered. The unit numbers can be taken directly from the input data. In addition, data on costs and revenues are relevant, for example to estimate whether and how profitable individual clusters in the product portfolio are.

6 CASE STUDY IN THE COMMERCIAL VEHICLE INDUSTRIE

6.1 Application validation

For the implementation, a sales data set with the configurations of the sold product variants of the industrial partner from the commercial vehicle industry was analysed, with a time range from April 2020 to March 2022. The data set contains 189 802 configurations as well as 986 feature categories and a total of 12 511 features. The Python libraries pandas, scikit-learn, and prince were used for the implementation. In the first step, the 986 feature categories were reduced with experts to 246 feature categories that have high relevance for the characteristics of the product variant. In the data preparation, the duplicates were removed as they do not have any added value in terms of the variety

https://doi.org/10.1017/pds.2023.265 Published online by Cambridge University Press

of the product portfolio and to increase the efficiency of the calculation. The remaining 65 456 unique vehicle configurations and 246 feature categories were examined for completeness in the next step. One feature category and four vehicle configurations were removed due to too many missing values. The remaining missing values were replaced with the feature that occurs most frequently in the respective feature category. The resulting dataset contains 65 452 vehicle configurations and 245 feature categories with a total of 1 828 features. Subsequently, a one-hot encoding and a dimension reduction with an MCA were conducted. This reduces the one-hot encoded data from 1 828 to 118 dimensions which represent 99% of the variety among the product configurations. Figure 3 shows the CVI of the algorithms k-Means, Mini-batch k-Means, Ward-Linkage, and EM (Gaussian Mixture Model spherical). The interval for the number of clusters k was set between 10 and 100 together with the industrial partner. A higher number of clusters would make the interpretation and subsequent analysis of the clusters too time-consuming.

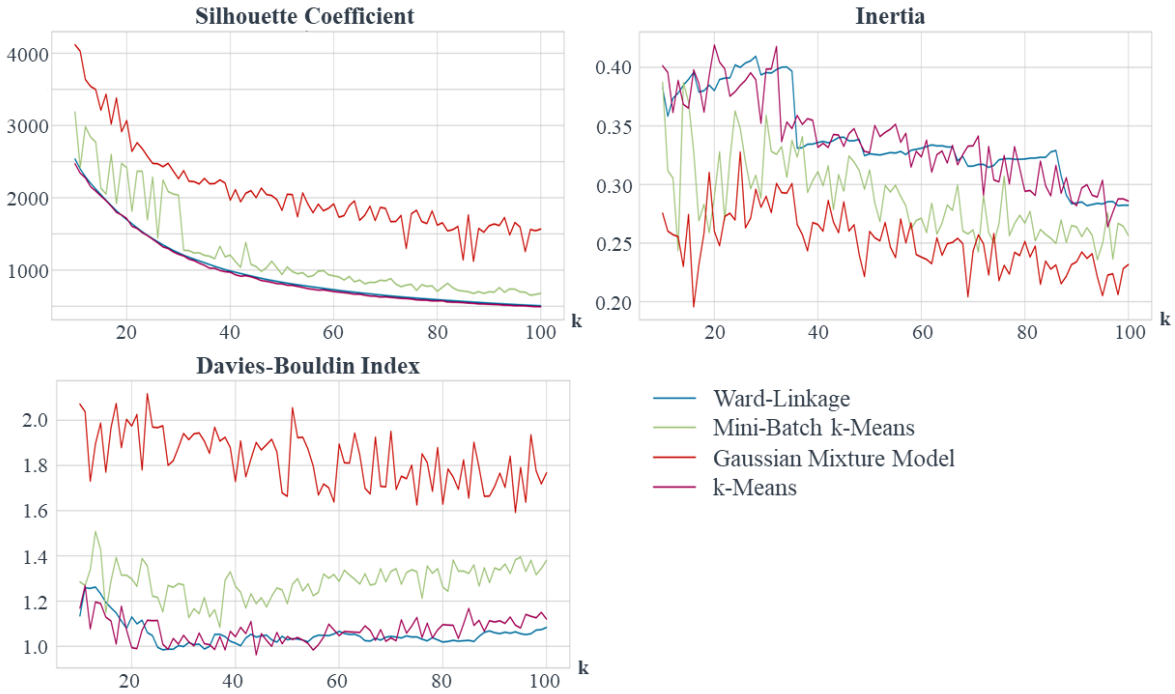


Figure 3. Evaluation of the implemented algorithms

The DBSCAN algorithm identified two clusters and a lot of noise. Since two clusters are not enough for the analysis of the product portfolio of the industrial partner and almost all of the data points were in one of the two clusters, the results of the DBSCAN are not considered further. The graph for inertia shows that k-Means and Ward-Linkage have almost identical values, which are lower and thus better than those of the other two algorithms. For the silhouette coefficient and the Davies-Bouldin index, these two algorithms also provide the best values. In a detailed analysis, the Ward-Linkage algorithm performs slightly better than the k-means algorithm. The best CVI values were achieved with the clustering of 31 clusters (see Figure 4 below). The identified clusters were analysed in terms of their distances from each other and characteristic features. In Figure 5 below, the identified clusters were visualised in three dimensions. It can be seen, for example, that clusters 19 and 20 are close together and far away from all other clusters. Based on this, the analysis of the characteristic features can be used to conclude what differentiates them from the other clusters. For example, there are features such as the X and Y types, the 8X4/4 wheel formula, the ceram clutch, or the A engine family that are found exclusively in these two clusters. In the same way, commonalities of clusters that are close to each other, such as clusters 15, 16, 17, and 18 can be investigated. They are all chassis, have the engine families B or C, the cab compact or comfort, and the construction type normal high.

6.2 Success validation

For the success validation the results were presented to eleven experts of the industrial partner from different departments and with different backgrounds. The methodological approach was considered

https://doi.org/10.1017/pds.2023.265 Published online by Cambridge University Press

as useful by the experts for identifying correlations in complex product portfolios. The form and content of the analysis were deemed suitable for more in-depth analyses. The possibility of subdividing a large number of product variants into a small number of groups by means of the method developed represents a high added value for analysing and controlling the width and depth product portfolio of the industrial partner. However, the added value of the methodological approach is estimated to be higher than the value of the clustering carried out. The experts considered the use of the method to be promising and potentially profitable if implemented as a flexible method within a software tool. The input feature categories must be selectable and thus individual and targeted analyses become possible. Regarding the selection of the 246 feature categories, it was noted that these represent a variety of different perspectives on product variants and therefore the statement is not specific enough for the potential user group. Due to its adaptability, the approach can be used by different users in different contexts to identify an optimal grouping of product configurations and can also be easily transferred to components or even BOMs.

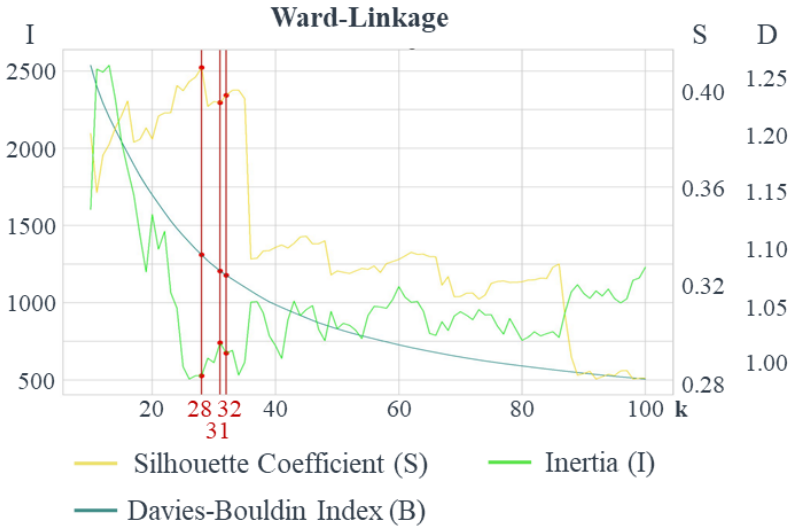


Figure 4. Identification of the optimal cluster number for the Ward-Linkage algorithm

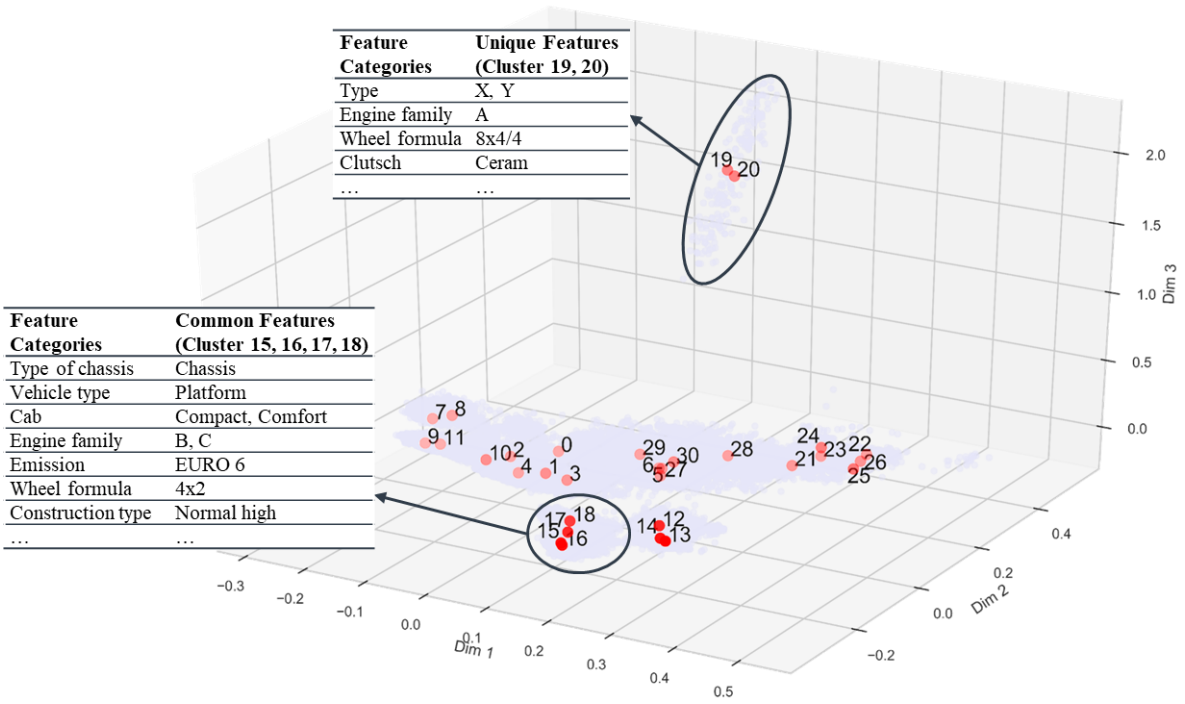


Figure 5. Analysis of the identified clusters

https://doi.org/10.1017/pds.2023.265 Published online by Cambridge University Press

7 CONCLUSION AND OUTLOOK

The developed methodical approach makes it possible to cluster complex product portfolios automatically and thus identify the correlations between the product variants. This is confirmed by an application at a commercial vehicle manufacturer with a particularly complex product portfolio and thereby, the research question is answered. In the approach, different clustering algorithms are applied and objectively compared to identify the optimal clustering. In addition, the benefit of the methodical approach is confirmed by eleven experts from the industrial partner. Clustering enables the objective identification of similar variants, special configurations and the characteristic features that occur together within the clusters. However, the approach does not address how the clustering results can be used to derive concrete measures for adapting and optimising the product portfolio. This is to be explored in the following research activities together with the experts of the industrial partner. Furthermore, the approach only considers the positioning of the variants sold. It is not examined where these are located in the total possible variant space of the product portfolio. This requires the inclusion of all information of the product data model. Moreover, it also becomes apparent that a one-time clustering is not enough, but that a flexible tool is necessary to carry out a clustering depending on the different perspectives of the portfolio managers. This requires the implementation in a software tool. The tool must allow the user to individually select the features and/or components for clustering and automate the analysis. This enables an individual and goal-oriented use of the approach. Moreover, the application of the approach has only been applied and validated in one company. The approach should therefore be applied to further companies from other industries to confirm its general applicability.

ACKNOWLEDGMENTS

This research work is part of “FORCuDE@BEV - Bavarian research association for customized digital engineering for bavarian SME's“ and funded by the "Bayerische Forschungsstiftung (BFS)”. We direct special thanks to the Bayerische Forschungsstiftung (BFS) for financial support of the whole research project.

REFERENCES

Abdi, H. and Valentin, D. (2007), “Multiple correspondence analysis”, Encyclopedia of Measurement and Statistics, Vol. 2 No. 4, pp. 651–657.

Blessing, L.T.M. and Chakrabarti, A. (2009), DRM, a Design Research Methodology, DRM, a Design Research Methodology, Springer, https://dx.doi.org/10.1007/978-1-84882-587-1.

Braun, F., Kreimeyer, M., Kopal, B. and Paetzold, K. (2017), “Herausforderungen in der Validierung der Variantenbeschreibung komplexer Produkte”, DFX 2017: Proceedings of the 28th Symposium Design for X, pp. 61–73.

Chan, K.Y., Kwong, C.K. and Hu, B.Q. (2012), “Market segmentation and ideal point identification for new product design using fuzzy data compression and fuzzy clustering methods”, Applied Soft Computing Journal, Vol. 12 No. 4, pp. 1371–1378, https://dx.doi.org/10.1016/j.asoc.2011.11.026.

Davies, D.L. and Bouldin, D.W. (1979), “A cluster separation measure”, IEEE Transactions on Pattern Analysis and Machine Intelligence, IEEE, No. 2, pp. 224–227.

DIN 199-1. (2002), “Technische Produktdokumentation-CAD-Modelle, Zeichnungen und Stücklisten-Teil 1: Begriffe”, Beuth Berlin.

Géron, A. (2017), Hands-on Machine Learning with Scikit-Learn and TensorFlow : Concepts, Tools, and Techniques to Build Intelligent Systems, O’Reilly Media.

Greisel, M., Kissel, M., Spinola, B. and Kreimeyer, M. (2013), “Design for adaptability in multi-variant product families”, Proceedings of the International Conference on Engineering Design, ICED, Vol. 4 DS75-04 No. August, pp. 179–188.

Han, J., Pei, J. and Tong, H. (2012), Data Mining: Concepts and Techniques, doi: https://doi.org/10.1016/C2009-0-61819-5.

Hancock, J.T. and Khoshgoftaar, T.M. (2020), “Survey on categorical data for neural networks”, Journal of Big Data, SpringerOpen, Vol. 7 No. 1, pp. 1–41.

Hochdörffer, J., Laule, C. and Lanza, G. (2018), “Product variety management using data-mining methods - Reducing planning complexity by applying clustering analysis on product portfolios”, IEEE International Conference on Industrial Engineering and Engineering Management, Vol. 2017-Decem, pp. 593–597, https://dx.doi.org/10.1109/IEEM.2017.8289960.

Hu, S.J., Zhu, X., Wang, H. and Koren, Y. (2008), “Product variety and manufacturing complexity in assembly systems and supply chains”, CIRP Annals - Manufacturing Technology, Vol. 57 No. 1, pp. 45–48, https://dx.doi.org/10.1016/j.cirp.2008.03.138.
Jonas, H. (2013), Eine Methode Zur Strategischen Planung Modularer Produktprogramme, Technische Universität Hamburg-Harburg.
Kissel, M.P. (2014), “Mustererkennung in komplexen Produktportfolios”, p. 212.
Krause, D. and Gebhardt, N. (2018), Methoden Zur Entwicklung Modularer Produktfamilien, Methodische Entwicklung Modularer Produktfamilien, Vol. №3, https://dx.doi.org/10.1007/978-3-662-53040-5_6.
Kreimeyer, M., Baumberger, C., Deubzer, F. and Ziethen, D. (2016), “An integrated product information model for variant design in commercial vehicle development”, Proceedings of International Design Conference, DESIGN, Vol. DS 84 No. 1, pp. 707–716.
Kreimeyer, M., Förg, A. and Lienkamp, M. (2013), “Mehrstufige modulorientierte Baukastenentwicklung für Nutzfahrzeuge”, VDI-Berichte, No. 2186, pp. 99–112.
Kubat, M. (2021), An Introduction to Machine Learning, https://dx.doi.org/10.1007/9783030819354.
Kusiak, A., Smith, M.R. and Song, Z. (2007), “Planning product configurations based on sales data”, IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews, Vol. 37 No. 4, pp. 602–609, https://dx.doi.org/10.1109/TSMCC.2007.897503.
Ma, J. and Kim, H.M. (2016), “Product family architecture design with predictive, data-driven product family design method”, Research in Engineering Design, Springer London, Vol. 27 No. 1, pp. 5–21, https://dx.doi.org/10.1007/s00163-015-0201-4.
Mehlstäubl, J., Braun, F., Denk, M., Kraul, R. and Paetzold, K. (2022), “Using Machine Learning for Product Portfolio Management: A Methodical Approach to Predict Values of Product Attributes for Multi-Variant Product Portfolios”, Proceedings of the Design Society, Vol. 2, pp. 1659–1668, https://dx.doi.org/10.1017/pds.2022.168.
Mehlstäubl, J., Braun, F., Gadzo, E. and Paetzold, K. (2023), “Machine Learning to generate Knowledge for Decision-making Processes in Product Portfolio and Variety Management”, 9th International Conference on Research Into Design.
Mehlstäubl, J., Braun, F. and Paetzold, K. (2021), “Data Mining in Product Portfolio and Variety Management – Literature Review on Use Cases and Research Potentials”, 2021 IEEE Technology & Engineering Management Conference-Europe (TEMSCON-EUR), pp. 442–447.
Mehlstäubl, J., Gadzo, E., Atzberger, A. and Paetzold, K. (2022), “Herausforderungen datengetriebener Methoden in der Produktentwicklung/Challenges of data-driven methods in product development”, Konstruktion, Vol. 74 No. 06, pp. 60–66, https://dx.doi.org/10.37544/0720-5953-2022-06-60.
Mitchell, T.M. (1997), Machine Learning, Vol. 1, McGraw-hill New York.
Mortensen, N.H., Yu, B., Skovgaard, H. and Harlou, U. (2000), “Conceptual modeling of product families in configuration projects”, Workshop at the 14th European Conference on Artificial Intelligence, Berlin, Germany, pp. 68–73.
Murphy, K.P. (2012), Machine Learning: A Probabilistic Perspective, MIT press, https://dx.doi.org/10.1109/pes.2005.1489456.
Neis, J. (2015), Analyse Der Produktportfoliokomplexität Unter Anwendung von Verfahren Des Data Mining, Shaker Verlag.
Romanowski, C.J. and Nagi, R. (2004), “A data mining approach to forming generic bills of materials in support of variant design activities”, Journal of Computing and Information Science in Engineering, Vol. 4 No. 4, pp. 316–328, https://dx.doi.org/10.1115/1.1812556.
Rousseeuw, P.J. (1987), “Silhouettes: a graphical aid to the interpretation and validation of cluster analysis”, Journal of Computational and Applied Mathematics, Elsevier, Vol. 20, pp. 53–65.
Samuel, A.L. (1959), “Some Studies in Machine Learning Using the Game of Checkers”, IBM Journal of Research and Development, Vol. 3 No. 3, pp. 210–229, https://dx.doi.org/10.1147/rd.33.0210.
Schmieder, M. and Thomas, S. (2005), Plattformstrategien Und Modularisierung in Der Automobilentwicklung, Shaker.
Schuh, G., Riesener, M. and Jank, M.-H. (2018), “Managing Customized and Profitable Product Portfolios Using Advanced Analytics”, Customization 4.0, pp. 203–216, https://dx.doi.org/10.1007/978-3-319-77556-2_13.
Tucker, C.S., Kim, H.M., Barker, D.E. and Zhang, Y. (2010), “A ReliefF attribute weighting and X-means clustering methodology for top-down product family optimization”, Engineering Optimization, Vol. 42 No. 7, pp. 593–616, https://dx.doi.org/10.1080/03052150903353328.
Wirth, R. and Hipp, J. (2000), “CRISP-DM : Towards a Standard Process Model for Data Mining”, Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining, Vol. 1, Springer-Verlag London, UK, pp. 29–39.
Zhang, Y., Jiao, J. and Ma, Y. (2007), “Market segmentation for product family positioning based on fuzzy clustering”, Journal of Engineering Design, Vol. 18 No. 3, pp. 227–241, https://dx.doi.org/10.1080/09544820600752781.

https://doi.org/10.1017/pds.2023.265 Published online by Cambridge University Press