DATA-DRIVEN DECISION SUPPORT SYSTEMS FOR PRODUCT DEVELOPMENT – A DATA EXPLORATION STUDY USING MACHINE LEARNING

Omsri Kumar Aeddula

Blekinge Institute of Technology
Licentiate Dissertation Series No. 2021:10

Department of Mechanical Engineering

Data-Driven Decision Support Systems for Product Development – A Data Exploration Study Using Machine Learning

Omsri Kumar Aeddula

Blekinge Institute of Technology Licentiate Dissertation Series
No 2021:10

Data-Driven Decision Support Systems for Product Development – A Data Exploration Study Using Machine Learning

Omsri Kumar Aeddula

Licentiate Dissertation in
Mechanical Engineering

Department of Mechanical Engineering
Blekinge Institute of Technology
SWEDEN

2021 Omsri Kumar Aeddula
Department of Mechanical Engineering
Publisher: Blekinge Institute of Technology
SE-371 79 Karlskrona, Sweden
Printed by Exakta Group, Sweden, 2021
ISBN: 978-91-7295-433-5
ISSN: 1650-2140
urn:nbn:se:bth-22322

There is no good and bad time
It's only application and learning

Omsri Aeddula, 2021

ABSTRACT

Modern product development is a complex chain of events and decisions. The ongoing digital transformation of society, increasing demands in innovative solutions puts pressure on organizations to maintain, or increase competitiveness. As a consequence, a major challenge in the product development is the search for information, analysis, and the build of knowledge. This is even more challenging when the design element comprises complex structural hierarchy and limited data generation capabilities. This challenge is even more pronounced in the conceptual stage of product development where information is scarce, vague, and potentially conflicting. The ability to conduct exploration of high-level useful information using a machine learning approach in the conceptual design stage would hence enhance be of importance to support the design decision-makers, where the decisions made at this stage impact the success of overall product development process.

The thesis aims to investigate the conceptual stage of product development, proposing methods and tools in order to support the decision-making process by the building of data-driven decision support systems. The study highlights how the data can be utilized and visualized to extract useful information in design exploration studies at the conceptual stage of product development. The ability to build data-driven decision support systems in the early phases facilitates more informed decisions.

The thesis presents initial descriptive study findings from the empirical studies, showing the capabilities of the machine learning approaches in extracting useful information, and building data-driven decision support systems. The thesis initially describes how the linear regression model and artificial neural networks extract useful information in design exploration, providing support for the decision-makers to understand the consequences of the design choices through cause-and-effect relationships on a detailed level. Furthermore, the presented approach also provides input to a novel visualization construct intended to enhance comprehensibility within cross-functional design teams. The thesis further studies how the data can be augmented and analyzed to extract the necessary information from an existing design element to support the decision-making process in an oral healthcare context.

Keywords: Product Development, Data-driven DSS, Machine Learning, Conceptual Stage, Data Analytics.

Omsri Aeddula, 2021

ACKNOWLEDGEMENTS

The research was carried out at the Department of Mechanical Engineering, Blekinge Institute of Technology, Karlskrona, Sweden. The research work was supervised by Professor Tobias Larsson, Dr Johan Wall, Professor Johan Sanmartin Berglund, and Professor Peter Anderberg.

I would like to express my gratitude to my supervisors for giving me inspiration and support together with knowledge sharing throughout this work. Special thanks to Tobias Larsson for involving me in high-value research projects. I especially want to thank Ryan Ruvald for interesting discussions as well as valuable support at all times and some fun business trips. Thank you, Stefan Renvert for a fruitful collaboration.

I would like to acknowledge the following people for some interesting discussions and support with the work:

Bobbie Frank, Johan Flyborg, Christina Karlsson, Viktoria Bjerström, Erik Resebo, Caroline Svendsen, Martin Frank.

Department of Mechanical Engineering and the Department of Health has created a nice working environment that is highly appreciated. Thanks to Raj Machchhar, Carl Toller, Alexander Barlo. and all my colleagues at BTH. Thanks to Associate professor Prashant Goswami, and Associate Professor Santhosh Jagtap for valuable feedback.

A constant source of energy, inspiration are my parents and grandparents. I would like to thank my family for standing by me and supporting me. Nanna, you are one of the hardest working and coolest people. Amma, you are the kindest and sweetest person. Both of you are true role models to me. Thank you for all your sacrifices!!

Finally, I would like to gratefully acknowledge the Knowledge Foundation and partners via the Model-Driven Development and Decision Support Project for their financial support and also the industrial research partners for their support.

I appreciate all your efforts!! Thank you all.

Omsri Kumar Aeddula,

December 2021.

Omsri Aeddula, 2021

LIST OF PAPERS

The thesis is based on the following studies, referred to in the text by their Roman alphabets.

A. Wall, J., Aeddula, O.K., Larsson, T., 2020. Data Analysis Method Supporting Cause and Effect Studies in Product-Service System Development. Proc. Des. Soc.: Des. Conf. 1, 461–470. https://doi.org/10.1017/dsd.2020.123

B. Aeddula, O.K., Wall, J., Larsson, T., 2021. Artificial Neural Networks Supporting Cause-and-Effect Studies in Product–Service System Development, In: Chakrabarti, A., Poovaiah, R., Bokil, P., Kant, V. (Eds.), Design for Tomorrow—Volume 2, Smart Innovation, Systems and Technologies. Springer Singapore, Singapore, pp. 53–64. https://doi.org/10.1007/978-981-16-0119-4_5

C. Aeddula, O.K., Flyborg, J., Larsson, T., Anderberg P., Johan S.B., Renvert S, 2021. A Solution with Bluetooth Low Energy Technology to Support Oral HealthCare Decisions for Improving Oral Hygiene, In: 2021 5th International Conference on Medical and Health Informatics. Presented at the ICMHI 2021: 2021 5th International Conference on Medical and Health Informatics, ACM, Kyoto Japan, pp. 134–139. https://doi.org/10.1145/3472813.3473179

Omsri Aeddula, 2021

Contents

1. INTRODUCTION 1
2. AIM 4
 2.1. Research Aim 4
 2.2. Research Question 4
3. RESEARCH APPROACH 5
 3.1. Research Methodology 5
 3.2. Data Collection 7
 3.3. Validation Measures 8
4. KNOWLEDGE DOMAINS 9
 4.1. Role of Data in Product Development 9
 4.2. Conceptual Design in Product Development 10
 4.3. Data Driven DSS in Decision Making 12
 4.4. Machine Learning Models in Product Development 14
5. SUMMARY OF APPENDED PAPERS 16
 5.1. Paper A 16
 5.2. Paper B 17
 5.3. Paper C 18
6. SUMMARY OF RESULTS 20
 6.1. Data Augmentation 20
 6.2. Data Analysis 21
 6.3. Visualization Construct 24
 6.4. Validation 26
7. Conclusions 27
 7.1. Future Work 28
REFERENCES 29

1. INTRODUCTION

Increasing global competition and rising demands in innovative solutions pushes organizations to change their product development practices. To be able to manage, organizations should heed the winds of change due to the ongoing digital transformation of society by embracing computer-based tools, such as tools for simulations, modeling, and visualizations to develop better products with changes to methods, tools, product offers, strategies, etc in order to maintain a competitive edge (Morgan and Liker, 2006).

The complexity of engineered systems is increasing dramatically. This development is driven by technological advancements to fulfill product functionality and the transition towards product-service system. Product-service system is a business model that emphasizes the need to co-develop product and service bundles based on product life extension, hence inviting cross-functional design teams while also enforcing a lifecycle perspective in the early stages of design (Bertoni et al., 2019). Complexity of these systems is associated with the number of technologies, subsystems, and components used to fulfill the desired functionality. Complexity is also strongly associated with interdependencies among the technologies, subsystems, and components (Sosa et al., 2004). A common approach to deal with complexity is to transform 'complex' into 'manageable and controllable' through the use of modeling and simulation (ElMaraghy et al., 2012). Furthermore, product development is an iterative decision-making process, characterized by uncertainty and ambiguity (Stacey and Eckert, 2003). Uncertainties in decision-making associated with the lack of knowledge require methods and tools to support complex decision-making processes (Eriksson et al., 2008; Platt and Huettel, 2008). During the early stages of the product development process i.e., conceptual stage, a multitude of decisions needs to be made. Understanding the impacts of these decisions at the early stage can help to increase the value of the engineered system both from a user and a provider perspective. Hence, a major quest in product development is the search for information and subsequent build of knowledge. Increased complexity and uncertain decision scenarios stresses the need for model-based approach that could be used to develop understanding (Eriksson et al., 2008).

The advent of technological development enables, supports, and motivates the use of a model-driven approach to support the decision-making process by gathering product data using sensors, and through simulations.

Simulations in this thesis, is referred to as experimentation with models of the studied system (Neelamkavil, 1987), in this case more specifically mathematical models. To add value in decision scenarios, decision-makers need to be able to access, analyze and interpret these data in order to explore patterns, proceed to several diagnoses and prognostics, and collect insights for decision-making. However, effectively extracting, understanding, and structuring knowledge without information overload and enabling its interpretation is also a challenge (Keim et al., 2006). Thus, researchers increasingly acknowledge the potentiality of data science in product development to support this (Kim et al., 2017). With the ongoing digital transformation, the volume of data will increase in the coming years, and therefore, processing of the data becomes imperative and increasingly valuable but also more complex to understand and translate into knowledge (Bickel et al., 2019; Han and Kamber, 2012; Zaki and Meira, Jr, 2014).

Supporting the decision-making process using data motivates in developing and using data-driven decision support systems in product development. A data-driven decision support system guides cross-functional design teams by helping them to visualize and utilize the data and models to solve design problems. A data-driven decision support system also provides a potential setting for information sharing, communication, understanding, and building knowledge. With a diverse group of decision-makers, structuring and sharing knowledge, values and data is believed to augment decision-making ability in the conceptual stage of product development (Wall et al., 2018). Building on a data-driven approach, decision models can not only predict outcome in the presence of uncertainty or unknown dynamics but can also extract inherent rules or knowledge from the data and thereby support the decision-making process (Lu et al., 2019).

Research Motivation

Given the evidence that the data contain valuable information, research into establishing identification techniques for data exploration would be needed to extract to their full potential. As a result, several data analysis techniques are becoming popular mostly in the detailed design stage, due to their effectiveness and success in supporting the development of products and services (Quinones-Gomez, 2021). However, most of the existing tools lack explicit support for high-level knowledge extraction and/or data augmentation particularly in the conceptual stage of product development, given the complexity and uncertainty of available data in the early stages.

2. AIM

2.1. Research Aim

This research aims to explore data using the machine learning approach in the conceptual stages of product development, where the information is scarce, vague, and potentially conflicting. With this aim, the research focuses on how machine learning techniques can be utilized to support the decision-making process through a data-driven decision support system.

2.2. Research Question

- How can machine learning techniques augment the decision-making capability in the conceptual stage of product development?

3. RESEARCH APPROACH

3.1. Research Methodology

Design research is unique in the scientific field with many different definitions already existing but struggles to find its implementation in practice. Design research is an integration of the development of understanding, and development of support (Blessing et al., 2009). It must be scientific to validate the results in some generic, theoretical as well as practical sense, which requires developing and validating the knowledge systematically. This necessitates the need for design research methodology, which should guide the selection and application of an appropriate approach and methods, as well as encourage reflection on the approach and methods to be used. In order to structure the research process systematically, Design Research Methodology (DRM) was implemented as an approach, DRM entails a set of methods and guidelines to be used in order for the design research to become more effective and efficient (Blessing et al., 2009). This method is particularly relevant in research where the focus is on the data-driven design phenomenon with an aim to understand and improve or develop the support.

DRM consists of four discrete stages, not necessarily sequentially, with iterations between stages taking place for a better understanding of the findings emerging from them. The prescribed stages are Research Clarification, Descriptive Study I, Prescriptive Study, and Descriptive Study II. The overall framework flow adopted from (Blessing et al., 2009), seen in figure 1 below:

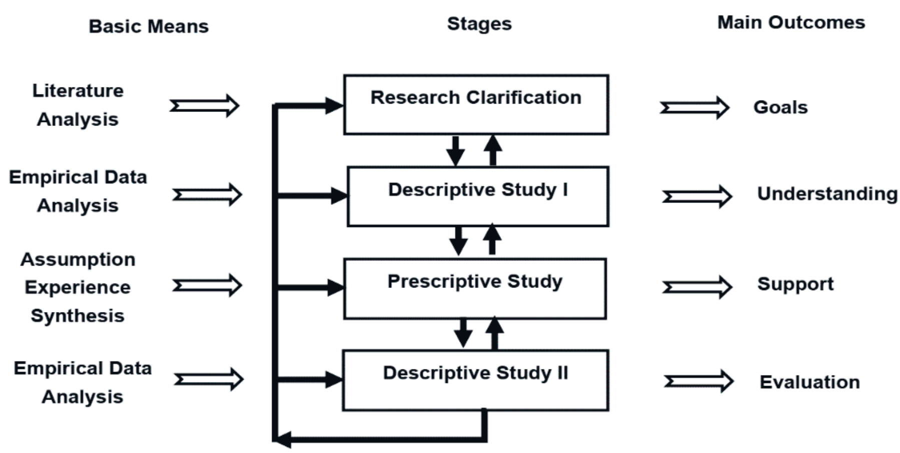


Figure 1. DRM framework, reproduced from (Blessing et al., 2009).

- Research Clarification: At this stage the researchers search the literature for some evidence or at least indications that supports their assumptions in order to formulate the overall realistic research goal and plan.
- Descriptive Study- I: At this stage, the researchers review the literature for additional influencing factors to elaborate the initial description of the existing situation, given that they are having a clear goal and focus. The intention is to focus on the factor(s) which needs to be addressed to improve task clarification as effectively and efficiently as possible.
- Prescriptive Study: At this stage, the researchers having a better understanding of the existing situation, assess and elaborate on their initial description of the desired situation. This description provides reflections on how addressing one or more factors in the existing situation would lead to realization of the desired outcome or improved situation. This stage aims at a systematic development of design support.
- Descriptive Study II: At this stage, the researchers investigate the impact of the developed design support and its ability to realize the desired situation. The study aims to evaluate the applicability and usefulness of the design support.

Many of these stages overlap in this thesis, and the relevance of each paper to each step is depicted in figure 2. Because of the iterative approach, as defined by (Blessing et al., 2009), and that phenomenon to be understood has not been addressed directly in this context by previous literature, the depth of each stage varies between Review-based, Comprehensive, and Initial, as described below:

- Review-based: based only on the review of existing literature.
- Comprehensive study: includes a literature review, as well as a study in which the results are produced by the researcher, i.e., the researcher undertakes an empirical study, develops support, or evaluates support.
- Initial study: show the consequences of the results at a particular stage and prepare the results for use by others.

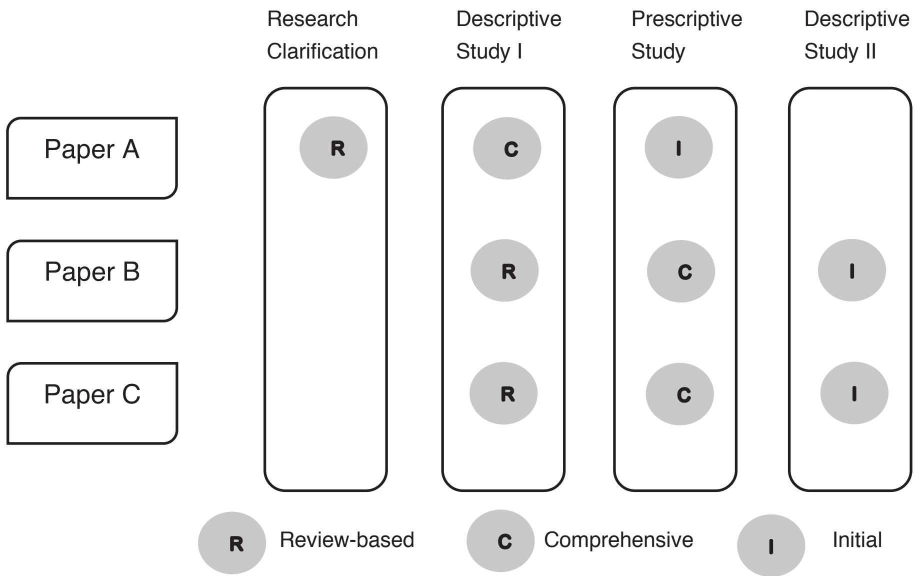


Figure 2. Thesis papers' focus in DRM framework.

3.2. Data Collection

For the empirical case studies in paper A and B, the author was provided with a dataset to analyse. This dataset originates from a design exploration study evaluating proposed concepts of road construction equipment. More information on how this dataset is generated can be found in (Bertoni et al., 2019).

Experimental prototyping was utilized to generate and gather data in the case study of paper C. The prototypes were installed at the participant's homes for a period of 4-6 months while the data was collected. The developed prototype consists of a single core-processor: raspberry pi zero ("Raspberry Pi Zero W,") and an Oral-B braun electric toothbrush ("Oral-B Genius 6000 Rechargeable Electric Toothbrush I Oral-B,").

3.3. Validation Measures

Validation, as used in the thesis, is a measure of usefulness or accuracy of the models to the investigation objective. This is vital for models to be trusted as support for design decisions (Sargent, 2013).

Coefficient of Determination (R^2): Coefficient of determination (R^2) determines how close the predicted data matches to the original data, according to the equation (1) (Kurz-Kim and Loretan, 2007).

$$R^2 = 1 - \frac{\sum(y-Y)^2}{\sum(y-\bar{y})^2} \tag{1}$$

where y represents the actual data values, Y is the predicted data values and $\bar{y}$ is the mean value of y.

Root Mean Square Value: Root Mean Square Error (RMSE) calculated the prediction error rate of the model. The RMSE is calculated as the sum of squared differences of the predicted values and the original values of the regression variable divided by the number of predictions as shown in equation (2) (Witten et al., 2011).

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{n} (\hat{y} - y)^2} \tag{2}$$

where $\hat{y}$ is the predicted value and y is the actual value.

4. KNOWLEDGE DOMAINS

4.1. Role of Data in Product Development

Data science, and industry 4.0 revolution gaining momentum, the world is witnessing a huge surge in data generation and usage across several sectors ranging from education/machine learning to engineering, to economics and medicine (Gandomi and Haider, 2015; Khan and Al-Badi, 2020), thereby increasing the analytical capabilities available for research (Hey, 2009). This increase of data and analytical capabilities has significantly expanded the boundaries of achievable outcomes in industrial practice (Provost and Fawcett, 2013). As a result, the rise of data and data-driven competencies has reshaped both industrial and scientific landscapes (Nielsen, 2012; Nosek et al., 2015). Following the data-driven revolution, engineering design research and applied health technology research also show signs of becoming more data-intensive, fostering its empirical grounding and incorporating various types of data and analysis methods into research methodologies (Blessing et al., 2009; Cash et al., 2016). With this, data is being used more effectively than ever before, and it lies at the heart of decision-making (Wang et al., 2016). However, extracting relevant knowledge to support the decision-making is still a challenge.

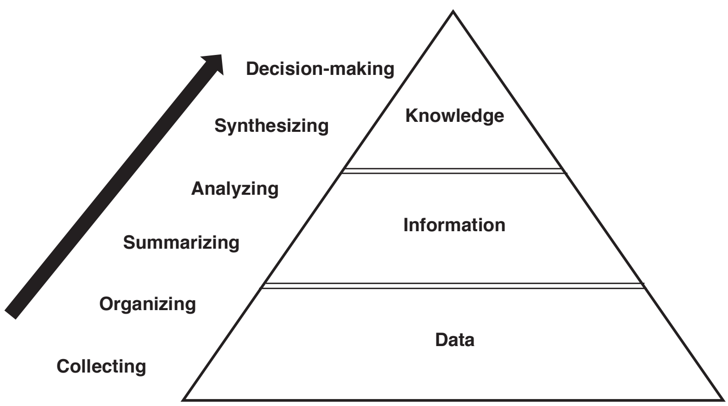


Figure 3. Knowledge Pyramid, reproduced from (Light et al., 2004)

Figure 3 identifies the three-level hierarchy with six steps in transforming data into knowledge to support the decision-making process. Through this process, raw data can be made into meaningful information and knowledge, by being related to the context of the problem. There are also many ways to define knowledge. One way is to relate the term to data, information, using the hierarchy represented by the "knowledge pyramid". For the purpose of this thesis the following descriptions have been adopted (Light et al., 2004).

- **Data** exists in a raw state. It has no understanding or meaning by itself and no relation to anything else.
- **Information** is data that has been given meaning and understanding by relating it to context.
- **Knowledge** is the collection of useful information, eventually a pattern will start to appear, useful to guide the actions. It is created through a sequential process. Data and information are about what has happened before and they form an external source of input, while knowledge is more about understanding the present and is an individual's internal form of input.

With the rising volume of data, a challenge is to associate data with the correct context, in search of information and knowledge, and also in building knowledge together with collaborative efforts of transcending disciplinary teams. The six-step process mentioned in figure 3 highlights that the data becomes more actionable and useful for the decision-makers when turned into information. Similarly, information becomes more actionable and useful when it is converted to knowledge, providing an opportunity for the decision-makers to make well-informed decisions.

4.2. Conceptual Design in Product Development

Product development is a process by which an organization transforms data on market opportunities and technical possibilities into information assets for commercial production. It involves four generic stages: conceptual design, system-level design, detailed design, and system integration (Ulrich and Eppinger, 2004), visualized as figure 4.

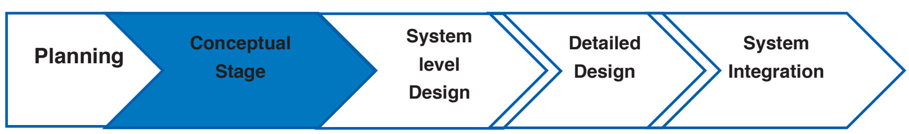


Figure 4. Product Development Process, adopted from (Ulrich and Eppinger, 2004)

Product development is essentially a process of making decisions about a product, while building knowledge along the process. Since the decisions need to be constructed from the available knowledge, the decisions in the conceptual stage of the product development, where available information normally is scarce, and possibly conflicting, must be made based on the fragmented knowledge. At the same time, design freedom is high. The knowledge about the product under development grows along the design process stages while the design freedom shrinks as design gates are passed and investments are made. This is referred to as the design paradox (Ullman, 2018), as in figure 5.

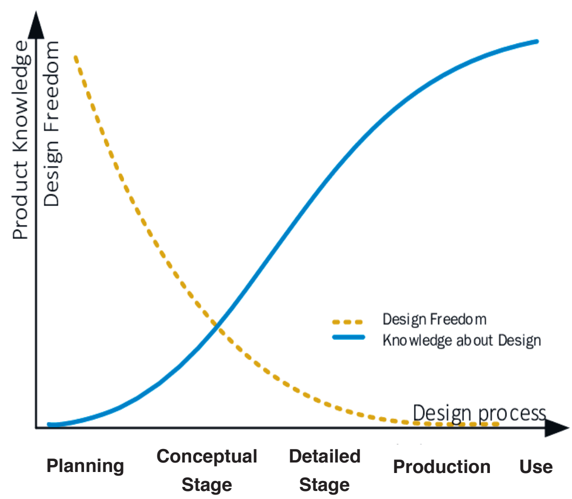


Figure 5. The design process paradox, adopted from Ullman (2018)

The need for research on decision making within product development has been recognized several times in previous research (Dwarakanath and Wallace, 1995; McNally and Schmidt, 2011). During the product development process, the early stage, i.e., conceptual stage, is critical. It is in the conceptual design stage, where the feasibility of the product is addressed, a design concept is developed, and early functional or proof-of-concept prototypes may be built and tested (Ulrich and Eppinger, 2004). It is also in the conceptual stage where the most impact of the final value of the product is possible because of the vast design solution space (Boukhris et al., 2017). However, a high level of uncertainty at this stage challenges the decision-making process, as limited knowledge related to the problem or solution exists initially other than organizational and individual tacit knowledge (Carleton et al., 2008). This highlights the need for further research regarding conceptual stage decision making, and the development of suitable support for those activities, where the conceptual stage can be seen as an opportunity to enhance the results with little effort (Legardeur et al., 2010).

The use of data science for the design of predictive models supporting the transformation of data into knowledge is an established practice in the detailed design stage (Akram et al., 2010; Huang et al., 2011; Jeong et al., 2005; H. Zhao et al., 2007), yet the same cannot be applied in the conceptual stage. However, even the highest standard model of the detailed design cannot compensate for a poor design concept formulated at the conceptual stage as a consequence of imprecise and incomplete knowledge. This emphasizes the need of developing support for the decision-making process at the conceptual stage (Hsu and Liu, 2000). In addition, during the product development process, there is evidence of designers spending 60% of their time in the exploration of the needed information and knowledge from unstructured data resources (Ullman, 2018). Building of knowledge in the conceptual stage is considered as a significant role for innovation; however, it is often challenging, as at this stage the information is scarce and has higher ambiguity (Bertola and Teixeira, 2003). As a result, the decision-making process at the conceptual stage necessitates the use of a huge amount of data, i.e., data-driven decision-making.

4.3. Data-Driven DSS in Decision Making

Decision problems can be categorized into structured, unstructured, and semi-structured. Structured decision problems can be handled by known processes or solutions and thus, do not require decision support. Unstructured decision problems are unique, and rather contain vague, incomplete, and potentially conflicting data. Their solution requires insights, intuition based on the available information and knowledge (Laudon and Laudon, 2006). Thus, design problems at the conceptual stages can be considered as unstructured decision problems. Semi-structured decision problems are in between the structured and unstructured and thus requires decision support, requiring the combination of interaction between the user and analytical methods to develop alternatives based on criteria and optimal solutions (Phillips-Wren, 2013). A decision support system (DSS) might be defined as an interactive computer-based system, which helps the decision-makers to visualize and utilize the data and models to solve unstructured problems, supporting the decision-making process (Sprague, 1980). In general, DSS is a class of computational systems that supports the decision-making process. The decision-makers would be a part of the system, thereby providing an opportunity to do several tasks in the decision-making process such as input selection, investigating the outputs, drilling down explanations, making it an interactive system (Phillips-Wren, 2013). In other words, it improves the decision-making process by facilitating communication, structuring knowledge, analyzing situations. Data-driven decision support system utilizes the data to enhance and improve support for the decision-makers. It has been shown that machine learning techniques, when combined with DSS, provides powerful aids in solving difficult applied problem involving large amounts of data (Phillips-Wren, 2013).

Data-driven Decisions:

As a consequence of the increased availability of data accompanied with constantly enhanced computational capacity, organizations are changing the way of making decisions. As they get more data-centric, more decisions are backed up by data related to decision problems, a process known as data-driven decision making (Lu et al., 2019). The objective of data-driven decision-making is to infer a decision based on deep analysis and learning based on historical and current data related to the decision problem, which differs from traditional model-driven decision-making. The data-driven decision-making process involves the collection, storage, organization, and analysis of huge volumes of data considering the uncertainty and data complexity (Khan and Al-Badi, 2020). Building on a data-driven approach, decision models can not only be predicted in the presence of uncertainty or unknown dynamics, but inherent rules or knowledge can also be extracted from the data and support the decision-making process. As a result, decisions are expected to become more rational and less influenced by intuition, increasing the potentiality of this approach (Brynjolfsson et al., 2011). In the product development process, there is evidence of the positive influences of this approach (Provost and Fawcett, 2013). Inherently, accounting for uncertainty and complexity factors is an indispensable part of the decision-making process. Low-quality predictions, non-representative, ineffectively analyzed data can lead to bad or erroneous decisions. Therefore, how to effectively use these huge volumes of data and as well as the process of data-driven decision-making has become a research hotspot (Quinones-Gomez, 2021; Zhang et al., 2017).

4.4. Machine Learning Models in Product Development

Machine learning is defined as a method that enables computers to think and learn explicitly. It refers to the study of computer algorithms performing intelligent predictions based on the dataset (Mitchell, 1997). Machine learning is among the pioneering drivers of digitalization and industry 4.0. In a product development context, machine learning can support in reducing the design time and efforts needed in a design space exploration study as shown by for example (Y. Zhao et al., 2007). Furthermore, it supports design decision-makers in adopting proper strategies at any stage of new product development especially at its early stages (Soltani-Fesaghandis and Pooya, 2018). The thesis depicts the role of machine learning modeling in three aspects: data augmentation, data analysis, and visualization.

Data augmentation is the method of increasing the amount of data by adding or slightly modifying the already existing data or by creating synthetic data from existing data (Perez and Wang, 2017). It is assumed that the increase in the amount of data provides an opportunity to extract more information from the dataset (Kawakami et al., 2020). The study used artificial intelligence in combination with data analysis, and data augmentation to reduce the development time for manufacturing automatic transmissions in vehicles.

Data analysis is a process of utilizing the data with the goal of discovering useful information from the data to support the decision-making process (Kudyba, 2014). The incorporation of data analysis in decision models is useful because the data model might contain uncertain or unknown information due to the existence of hidden data patterns or data relationships (Lim et al., 2006). Furthermore, machine learning-based data analysis supports discovering useful information or predicting unknown parameters from the data supporting the decision-making processes (Lu et al., 2019) and can be used in the evaluation and support of decision-making process (Feng et al., 2020). However, it might be hard for the diverse group of decision-makers to make sense of the data, even if the outcome of the data analysis stage contains the necessary information to support the decision-making process. An interactive system with exploration capabilities could be a potential aid for a diverse group of decision-makers in understanding the information, a process referred to as visualization. Visualization provides capability to contrast and compares results (Wall et al., 2020).

Benefits of information and knowledge visualization in the decision-making process have been extensively studied in the literature. The studies mainly concern on the dynamics of the decision-making process, what influences people's decisions, and how can the process be enhanced to reduce uncertainty. Visualization can be defined as a communication process of interpreting the information in a graphical form. Knowledge visualization is a computer-based visual representation of insights, assessments, or experts' opinions to augment communication. Interactive visualization tools offer great potential for the improvement of knowledge communication (Eppler, 2004).

A number of studies describe the effectiveness of knowledge and information visualization in the decision process. Interactive visualization facilitates knowledge sharing, thereby increasing individual learning and team performances (Eppler and Platts, 2009). The study also reflected upon the effectiveness of information and knowledge visualization during the decision-making process. Thus, knowledge sharing through visualization is an effective tool to support the decision-making process. Color-based knowledge visualization is a key technique for enhancing knowledge and information visualization in the decision-making process, see for example (Bertoni et al., 2011; So and Smith, 2002). The study shows that color have been found to improve the usefulness of an information display system.

5. SUMMARY OF APPENDED PAPERS

5.1. Paper A

Wall, J., Aeddula, O.K., Larsson, T., 2020. Data Analysis Method Supporting Cause and Effect Studies in Product-Service System Development. Proc. Des. Soc.: Des. Conf. 1, 461–470. https://doi.org/10.1017/dsd.2020.123

Summary

The paper proposes a method based on a supervised machine learning technique called partial least square regression method, to assess cause-and-effect relationships in a design exploration study. In addition, the study also proposed a variable clustering method to create a virtual intermediate level in the variable hierarchy, without losing the significant contributions of other non-cluster design variables.

A design exploration study may generate an abundance of data with an intricate hierarchy of attributes. The proposed variable clustering method may aid design teams in understanding how attributes on different levels are affected by system/subsystem/components or on a more detailed level relates to component features. The method clusters and aggregates the effects of multiple design variables based on the structural hierarchy of the evaluated system. The generated information is used as a structured input to a visualization construct, using CAD geometry as a base. In a model-driven environment for collaborative decision making focused on early phases of engineering design, data analysis and visualization become key enablers in aiding information sharing, communication, understanding, and building of knowledge within the cross-functional team.

Relation to Thesis

The paper explored data ambiguity specifically in the conceptual stage of product development. It introduced a machine learning technique to interpret the data and a variable clustering method to understand the cause-and-effect relationships of design attributes on different levels. The combination of data analysis and visualization construct enables a data-driven decision support system, facilitating a detailed level of understanding of design attributes on a different level, augmenting the decision-making process.

Author's Contribution

The author developed the methodology and wrote part of the text. The author contributed to the definition of the paper. Johan Wall contributed to the theoretical background of the paper and review of the draft versions. Tobias Larsson supported with constructive feedback, ensuring the paper standard.

5.2. Paper B

Aeddula, O.K., Wall, J., Larsson, T., 2021. Artificial Neural Networks Supporting Cause-and-Effect Studies in Product–Service System Development, In: Chakrabarti, A., Poovaiah, R., Bokil, P., Kant, V. (Eds.), Design for Tomorrow—Volume 2, Smart Innovation, Systems and Technologies. Springer Singapore, Singapore, pp. 53–64. https://doi.org/10.1007/978-981-16-0119-4_5

Summary

The objective of the paper was to extend the work presented in paper A. The method presented in paper A is limited in applicability to datasets with linear characteristics. However, design problems often contain non-linear relationships between design variables and design attributes, hence a method able to handle also non-linear relationships is desired. The paper proposes a machine learning method based on artificial neural networks to derive design attribute dependencies on multiple design variables. In addition, the study also proposed a variable clustering method based on a rotated-coordinate system. It is shown that the applied clustering method does not affect the contributions of non-clustered design variables. The approach showed a comparable predictive capability of the clustered data as compared to the original dataset.

Relation to Thesis

The work leading to this paper has created an understanding of regression model inapplicability to datasets involving non-linear characteristics. In continuation, the paper proposed a new artificial neural network method aimed at the conceptual stage of product development to derive the relative importance of design variable contributions. This provides an opportunity to interpret and visualize the data and structure the knowledge on a detailed level by varying the desired design variables for a specific design attribute. The approach helps to understand design attributes cause-and-effect relationships on different structural hierarchy levels, thereby supporting the decision-making process, and the predictive capability of the approach serves as a data-driven decision support system

Author's Contribution

The author conducted the research investigation in developing the methodology. The author defined the structure of the paper, wrote the entire paper with reviews, feedback. Johan Wall and Tobias Larsson supported with major edits, constructive feedback, and ensuring the paper standard.

5.3. Paper C

Aeddula, O.K., Flyborg, J., Larsson, T., Anderberg P., Johan S.B., Renvert S, 2021. A Solution with Bluetooth Low Energy Technology to Support Oral HealthCare Decisions for Improving Oral Hygiene,In: 2021 5th International Conference on Medical and Health Informatics. Presented at the ICMHI 2021: 2021 5th International Conference on Medical and Health Informatics, ACM, Kyoto Japan, pp. 134–139. https://doi.org/10.1145/3472813.3473179

Summary

The paper describes a case study where the aim is to collect and analyze data autonomously to support the decision-making process. In the oral healthcare context, information technologies such as powered toothbrushes with associated mobile health application provides an opportunity to collect data; however, this requires some efforts from the participants and has certain constraints on the data collection. The paper proposes a solution with Bluetooth Low Energy technology paired with the associated toothbrush to support collection of reliable and comprehensible toothbrush data. In addition, a machine learning approach was proposed to convert the raw data into useful information with visualizations. This makes a case for expanded opportunity to plan assistant capacities to protect or improve factors that influence oral wellbeing in individuals with mild cognitive impairments. The proposed method supports determining various additional parameters, making it adaptable and conceivable to execute in various oral care contexts.

Relation to Thesis

The paper presents an approach for data augmentation to support the decision-making process in designing a plan on oral hygiene tailored to each individual. A machine learning approach converts the dataset into actionable information suited for the decision-maker. The approach presented in the paper provides the opportunity to infer additional data as well as meaningful information with visualization to support the decision-making process.

Author's Contribution

The author wrote the full text of the paper and contributed to developing the research approach. The author conducted the research investigation for data collection along with Johan Flyborg. Tobias Larsson and Johan Flyborg contributed with edits in the text and as well as constructive feedback along with the other co-authors.

6. SUMMARY OF RESULTS

The empirical studies focused on identifying what, and how strategies around the use of machine learning techniques in order to support the decision-making process in the conceptual stage of the product development. The literature pointed out that increase in complexity of products led to rapid use of digital technologies, which has increased the volume, and the complexity of available data. This resulting in design engineers spending most of the time on exploring the necessary information and knowledge. The research demonstrated the capabilities of machine learning as a means for exploring conceptual phase ambiguity to support and augment the decision-making process in product development.

The empirical studies [Paper A and B] used a dataset originating from a design exploration study evaluating proposed concepts of road construction equipment. The dataset consists of 700 variants of a proposed vehicle platform, driven by seven design variables and containing 16 design attributes. Design variables are entities that may be changed in an experiment affecting the shape or properties of the studied system. An attribute can be defined as any aspect of the product itself or its use that can be used to compare product alternatives. Paper C collected a dataset from powered toothbrushes to extract information about oral health status. The results from the case studies as presented in the appended papers (A, B, and C) to this thesis are summarized in the following subsections. The following subsections describe the thesis results focusing on the following aspects of machine learning: data augmentation, data analysis, and visualization, supporting the decision-making process in the conceptual stage of product development.

6.1. Data Augmentation

Paper A describes that the design variables in a simulation-based design exploration study might be classified as a two-level hierarchy with design variables at the bottom level and the design attributes at the top level. However, these design variables are often not identifiable as components or subsystems, rather a subset of design variables in a study experiment, combined with parameters and constants drives the design configuration of a particular component or subsystem and hence also the attributes associated with that component or subsystem. These components or subsystems can therefore be viewed as a virtual intermediate level in the structural hierarchy with no observed data to populate and analyze. In the thesis methods are proposed to generate new, clustered variables inferred from the observed design variables to populate this new virtual intermediate level. This is challenging and adding the element of aggregating the variables without losing the significant contributions of other design variables serves to challenge the process even further, as an alteration in variable contributions would affect the accuracy and efficiency of the model. To achieve this context, the intended variables are clustered based on the structural hierarchy using partial least square regression [Paper A] and rotated coordinate system [Paper B]. The dataset also contains information enabling mapping each design variable directly to components or subsystem in the structural hierarchy and the results show that the process is reliable and provide more additional information, supporting the decision-making process.

Another data augmentation challenge addressed in this thesis is to utilize the data for the decision-making process from an already existing design element with a limitation of new data generation capability. Limited data brings a huge barrier to the design decision-makers and data collection is strenuous activity if not fully automated, requiring a lot of effort from the participants. This is particularly difficult if participants have cognitive impairment such as elderly and sick people suffering from mild cognitive impairment, and dementia. Therefore, it is a vital task to automate data collection and synthesize the data to extract the necessary information in the decision-making process. Paper C discusses how a single-core processor supports automated data acquisition and also describes a process to avoid noisy data. This process generates a dataset with information containing all the necessary parameters, previously unattainable using the design element. The results of Paper C show that the process could be reliable and support in acquiring comprehensible data of toothbrush use and propensities that can guide to decision-making process of making a general unique plan on oral hygiene tailored to each individual.

6.2. Data Analysis

To support the decision-making process, it is important to extract, from the current scenario, useful information from the datasets. Furthermore, data analysis assists in exploring and transforming information into knowledge through discovering the patterns in the data or exploring the relationships among the data. To elaborate on this concept, Paper A discusses the application of a supervised machine learning method i.e., partial least square (PLS) regression model in establishing the cause-and-effect relationships. However, the method generates a linear model, supporting only linear design variables. As a result, the dataset containing three categorical variables, which are rendering a non-linear behavior is modified by fixing two out of the three categorical variables to a specific choice and transforming the third one to continuous. The modified dataset was used to establish the cause-and-effect relationships.

The PLS regression model presented in Paper A was extended in a further study [Paper B], which explored the quantification of dependencies including the non-linear data variables. Paper B discusses how the use of artificial neural networks supports the quantification of dependencies between design variables and design attributes. The study showed that artificial neural networks are capable of learning from the data through iterations without any prior knowledge and supports in establishing cause-and-effect relationships. For example, see figure 6, where the relative contributions of the design variable for a design attribute are shown. Similarly, the design variables in the virtual intermediate level are quantified with a design attribute, as shown in figure 7.

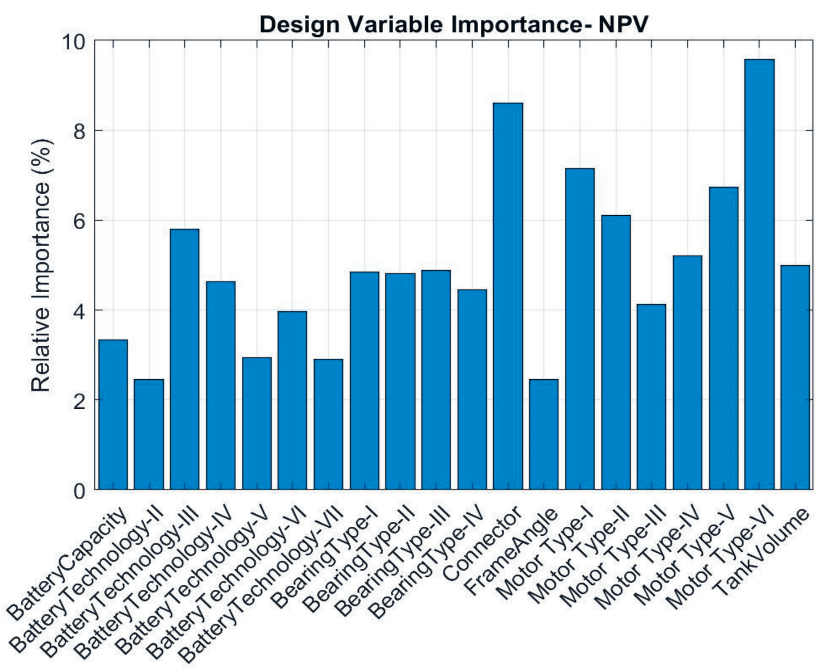


Figure 6. Quantification of cause-and-effect relationship

Comparison of both these figures revealed that the creation of virtual intermediate level does not affect the significance of other non-clustered design variables, consider the example of connector design variable. It has the same contributions in both the structural hierarchical levels.

Establishing the cause-and-effect relationship could be used to predict the values of the design attribute. Paper B validated the established cause-and-effect relationship by deriving the design attribute values, unused in the machine learning model, both at the virtual intermediate level and the component level. From a decision-maker perspective, this builds trust in the predictive capability of the developed model to predict. In the same line, it can be extended to other hierarchical levels. This kind of prediction and understanding supports the decision-makers in testing different design concepts and analyzing the consequences of their design choices.

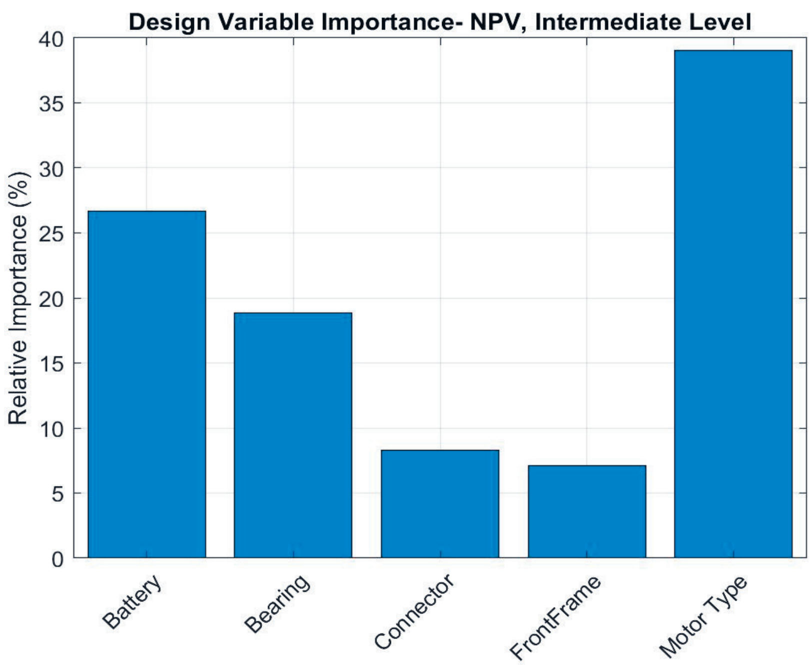


Figure 7. Quantification of cause-and-effect relationship for clustered variables

Similarly, in the process of data exploration, Paper C describes a machine learning method to extract the information from an unreadable (raw) dataset. The raw data contains a mixture of words and numerical values which are unreadable and need to be analyzed. Natural language processing method is applied for mapping the raw data to an excel spreadsheet containing parameters such as date, time, total time of applied pressure in seconds, and total brushing time in seconds. As a result, more useful parameters such as average brushing time per day, week, and month, and number of brushing sessions per day, week, and month can be estimated. In addition, depending on the user requirements more additional parameters can be derived such as total pressure applied in seconds per day, week, and/or month. Utilizing the methods for establishing cause-and-effect analysis, it is also possible to do a feasibility check of quantification of dependency of the total time of brushing to total applied pressure per session.

6.3. Visualization Construct

Estimations from machine learning methods such as regression model [Paper A], and natural language processing [Paper C] could be used as a structured input for visualizations. Figure 8 shows the visualization of useful parameters such as brushing time (blue line) and applied pressure time (red line) in seconds on a day scale from data representing one participant in the study. The results of Paper C revealed that raw data obtained from toothbrushes can be mapped to the parameters using machine learning technique supports in better understanding of the data and therefore making the design concept adaptable and conceivable to execute in various oral care contexts. This kind of additional information

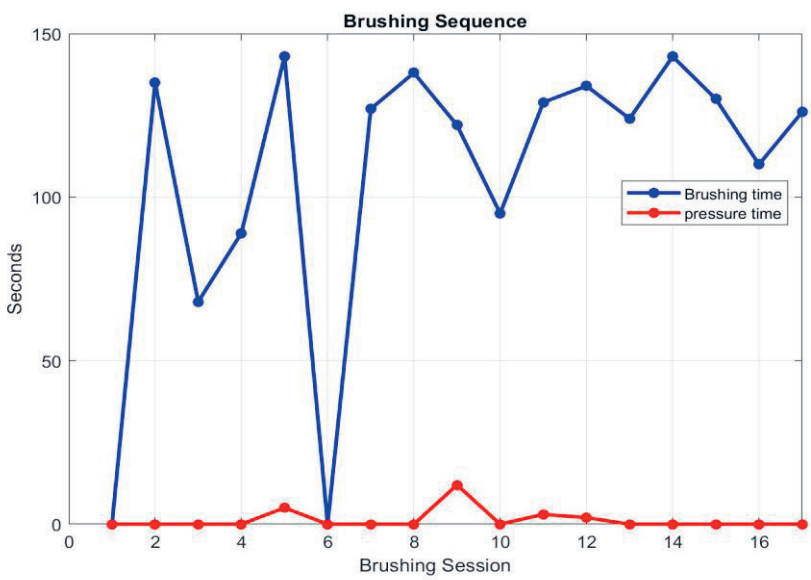


Figure 8. Brushing sequence per day of a participant

provides knowledge and opportunity for the decision-maker to develop aids and methods for improved general and oral health, resulting in increased quality of life i.e., health care decision-making process.

Paper A revealed that estimating the relative contributions of the design variables among the hierarchical level supports in providing a structured input to CAD visualization, as shown in figure 9; This type of visualization makes the data accessible and interpretable for decision-makers, facilitating the understanding of the dependencies on design attributes even without domain-specific knowledge.

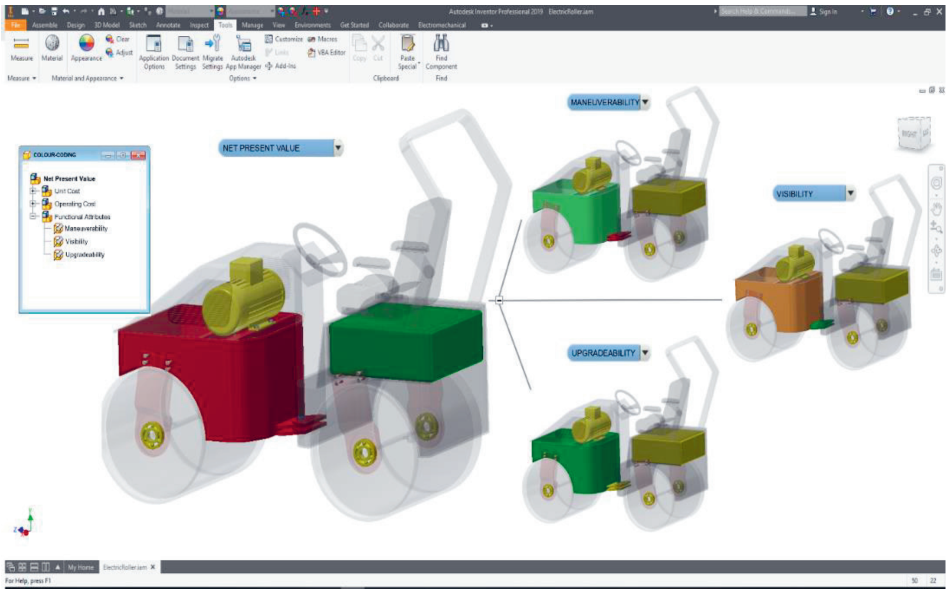


Figure 9. Color-coded CAD model visualization from a structured input

The results of the research presented in Papers A and C revealed that combining data analysis with visualization system i.e., DSS can support the decision-making process for a complex system at the conceptual stage of the product development. They describe a method to associate data and information with identifiable components or subsystems [Paper A] and extract useful parameters data [Paper C] within the studied system in an efficient way to share information, aiding understanding and building of knowledge. Thereby according to DSS literature (see for instance, (Sharda et al., 1988)) it increases the efficiency and effectiveness of the decision-making process and enhances the decision-making ability of the designers. The research also supports the cross-functional teams in a collaborative setting to share knowledge, values, and data for design exploration studies, augmenting the decision-making ability.

6.4. Validation

Validation is an important factor to assess the reliability and accuracy of the proposed models. Higher reliability and accuracy values indicate better predictive capability of the model. Verified predictive capability, improves the trust in the models by the decision-makers in the product development process. Paper A and B validate the machine learning models using the statistical analysis techniques, coefficient-of-determination (R^2), and Root Mean Square Error with good results.

Paper C validates the proposed method using the ground truth values. Figure 10 shows the comparison of ground truth values with the generated data. The blue dots represent the ground truth values and the orange circles represent the data obtained from the proposed method. For every generated data, the corresponding ground truth is compared, as the original dataset lack complete ground truth values. The figure shows the matching of the actual data with the generated data, shows correlation between the datasets, which builds trust with the decision-maker to develop aids for improved oral hygiene.

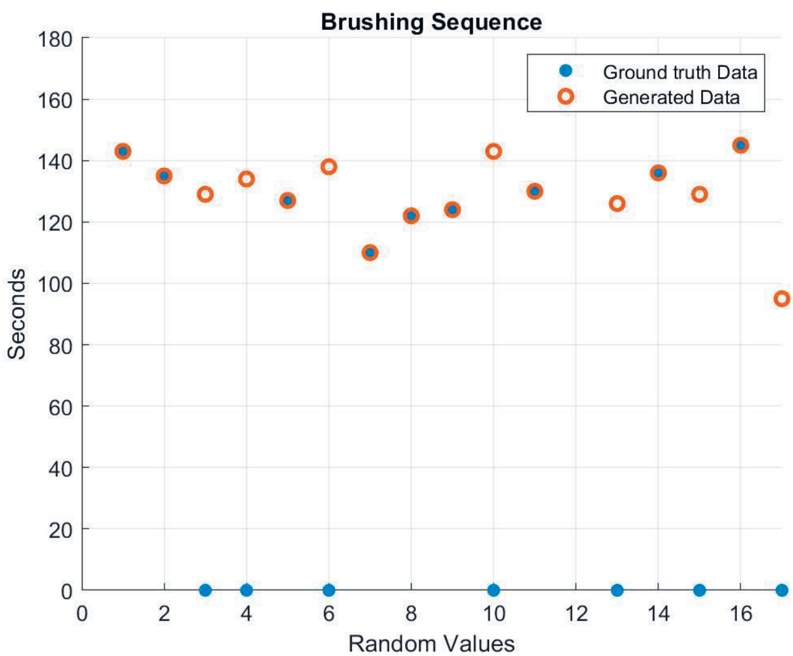


Figure 10. Comparison of ground truth data vs generated data

7. Conclusions

This thesis was aimed at exploring how machine learning techniques in different case studies can support in building of decision support system and augmentation of decision-making process in the conceptual stage of product development. Conceptual design was targeted as decisions made during this stage have a greater impact on the final solution, while available information is normally scarce. The research findings reflect that machine learning techniques have the capability to create a data-driven decision support system for supporting the decision-making process in the conceptual stage of product development.

The data augmentation, data analysis, and visualization were carried out in two case studies with complex systems in different application areas. It was shown that a machine learning approach support in extracting and analyzing the useful information required for the decision-makers. Establishing cause-and-effect relationships between design attributes and design variables guides the decision-makers in the decision-making process. Clustering of design variables for generating a new intermediate level in the structural hierarchy without affecting the significances of other non-clustered design variables provides a detailed understanding on each level of the structural hierarchy. This could be further utilized as a structured input to visualization tools, thereby building on a data-driven decision support system in the conceptual stage of product development. Furthermore, an approach on how to extract useful information from an existing design element with data scarcity was proposed. The exploration of data combined with visualization provides an opportunity for structuring the knowledge necessary to make decisions, providing an opportunity to develop aids and improve oral and general health. Proposed methods are validated using statistical techniques. Good correlations with the observed values are shown.

Combining the above findings, it is concluded that exploring the data through the machine learning approach supports the decision-making process in order to improve product development performance. The thesis also concludes that in a data exploration study having data augmentation, data analysis, visualization aspects provide actionable information and knowledge thereby augmenting the decision-making process. This approach also supports cross-functional teams to share information, knowledge, value and data, augmenting the decision-making ability of the team.

7.1. Future works

The results of the conducted research show the potential of incorporating machine learning techniques in conceptual design. This motivates further research to investigate deeper into the advanced models of machine learning, to make smart decision support systems, within the product development area. This may also be expanded to the product-service system.

Study on data collection through prototype for an existing design element might be extended to integrating the prototype with the design element, as a single usage product. It would be interesting to discover and develop algorithms, exploring more hidden relationship patterns in the data involving interaction terms using methods such as deep learning, and artificial neural networks. Furthermore, analyzing the value of engineered systems both from a user perspective and a provider perspective.

REFERENCES

Akram, F., Prior, M., Mavris, D., 2010. Design Space Exploration of Submerged Inlet Capturing Interaction Between Design Parameters, in: 28th AIAA Applied Aerodynamics Conference. Presented at the 28th AIAA Applied Aerodynamics Conference, American Institute of Aeronautics and Astronautics, Chicago, Illinois. https://doi.org/10.2514/6.2010-4680

Bertola, P., Teixeira, J.C., 2003. Design as a knowledge agent. Design Studies 24, 181–194. https://doi.org/10.1016/S0142-694X(02)00036-4

Bertoni, A., Bertoni, M., Isaksson, O., 2011. Communicating the Value of PSS Design Alternatives using Color-Coded CAD Models, in: Hesselbach, J., Herrmann, C. (Eds.), Functional Thinking for Value Creation. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 51–56. https://doi.org/10.1007/978-3-642-19689-8_11

Bertoni, M., Pezzotta, G., Scandella, B., Wall, J., Jonsson, P., 2019. Life cycle simulation to support cross-disciplinary decision making in early PSS design. Procedia CIRP 83, 260–265 . https://doi.org/10.1016/j.procir.2019.03.138

Bickel, S., Spruegel, T.C., Schleich, B., Wartzack, S., 2019. How Do Digital Engineering and Included AI Based Assistance Tools Change the Product Development Process and the Involved Engineers. Proc. Int. Conf. Eng. Des. 1, 2567–2576. https://doi.org/10.1017/dsi.2019.263

Blessing, L.T.M., Chakrabarti, A., Blessing, L.T.M., 2009. DRM, a design research methodology. Springer, Dordrecht ; London.

Boukhris, A., Fritzsche, A., Möslein, K., 2017. Co-creation in the Early Stage of Product-service System Development. Procedia CIRP 63, 27–32. https://doi.org/10.1016/j.procir.2017.03.316

Brynjolfsson, E., Hitt, L.M., Kim, H.H., 2011. Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance? SSRN Journal. https://doi.org/10.2139/ssrn.1819486

Carleton, T., Cockayne, W.R., Leifer, L., 2008. An Exploratory Study about the Role of Ambiguity during Complex Problem Solving, in: AAAI Spring Symposium: Creative Intelligent Systems.

Cash, P., Stanković, T., Štorga, M. (Eds.), 2016. Experimental Design Research. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-33781-4

Dwarakanath, S., Wallace, K.M., 1995. Decision-making in Engineering Design: Observations from Design Experiments. Journal of Engineering Design 6, 191–206. https://doi.org/10.1080/09544829508907913

ElMaraghy, W., ElMaraghy, H., Tomiyama, T., Monostori, L., 2012. Complexity in engineering design and manufacturing. CIRP Annals 61, 793–814. https://doi.org/10.1016/j.cirp.2012.05.001

Eppler, M., 2004. Facilitating Knowledge Communication through Joint Interactive Visualization. http://www.alexandria.unisg.ch/Publikationen/54794.

Eppler, M.J., Platts, K.W., 2009. Visual Strategizing. Long Range Planning 42, 42–74. https://doi.org/10.1016/j.lrp.2008.11.005

Eriksson, J., Johnsson, S., Olsson, R., 2008. MODELLING DECISION-MAKING IN COMPLEX PRODUCT DEVELOPMENT. DS 48: Proceedings DESIGN 2008, the 10th International Design Conference, Dubrovnik, Croatia 1129–1138.

Feng, Y., Zhao, Y., Zheng, H., Li, Z., Tan, J., 2020. Data-driven product design toward intelligent manufacturing: A review. International Journal of Advanced Robotic Systems 17, 172988142091125. https://doi.org/10.1177/1729881420911257

Gandomi, A., Haider, M., 2015. Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management 35, 137–144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007

Han, J., Kamber, M., 2012. Data mining: concepts and techniques, 3rd ed. ed. Elsevier, Burlington, MA.

Hey, A.J.G. (Ed.), 2009. The fourth paradigm: data-intensive scientific discovery. Microsoft Research, Redmond, Washington.

Hsu, W., Liu, B., 2000. Conceptual design: issues and challenges. Computer-Aided Design 32, 849–850. https://doi.org/10.1016/S0010-4485(00)00074-9

Huang, Z., Wang, C., Chen, J., Tian, H., 2011. Optimal design of aeroengine turbine disc based on kriging surrogate models. Computers & Structures 89, 27–37. https://doi.org/10.1016/j.compstruc.2010.07.010

Jeong, S., Chiba, K., Obayashi, S., 2005. Data Mining for Aerodynamic Design Space. Journal of Aerospace Computing, Information, and Communication 2, 452–469. https://doi.org/10.2514/1.17308

Kawakami, T., Ide, T., Moriyama, E., Hoki, K., Muramatsu, M., 2020. Development of Artificial Intelligence to Classify Quality of Transmission Shift Control Using Deep Convolutional Neural Networks. IEEE Trans. Veh. Technol. 69, 16168–16172 https://doi.org/10.1109/TVT.2020.3032191

Keim, D.A., Mansmann, F., Schneidewind, J., Ziegler, H., 2006. Challenges in Visual Data Analysis, in: Tenth International Conference on Information Visualisation (IV'06). Presented at the Tenth International Conference on Information Visualisation (IV'06), IEEE, London, England, pp. 9–16. https://doi.org/10.1109/IV.2006.31

Khan, A.I., Al-Badi, A., 2020. Emerging Data Sources in Decision Making and AI. Procedia Computer Science 177, 318–323. https://doi.org/10.1016/j.procs.2020.10.042

Kim, H.H.M., Liu, Y., Wang, C.C.L., Wang, Y. (Eds.), 2017. Special Issue: Data-Driven Design (D3). Journal of Mechanical Design 139, 110301. https://doi.org/10.1115/1.4037943

Kudyba, S., 2014. Big Data, Mining, and Analytics: Components of Strategic Decision Making, 0 ed. Auerbach Publications. https://doi.org/10.1201/b16666

Kurz-Kim, J.-R., Loretan, M., 2007. A Note on the Coefficient of Determination in Models with Infinite Variance Variables. SSRN Journal. https://doi.org/10.2139/ssrn.996664

Laudon, K.C., Laudon, J.P., 2006. Management information systems: managing the digital firm, 9th ed. ed. Pearson/Prentice Hall, Upper Saddle River, NJ.

Legardeur, J., Boujut, J.F., Tiger, H., 2010. Lessons learned from an empirical study of the early design phases of an unfulfilled innovation. Res Eng Design 21, 249–262. https://doi.org/10.1007/s00163-010-0090-5

Light, D., Wexler, D., Heinze, J., 2004. How practitioners interpret and link data to instruction: Research findings on New York City Schools' implementation of the Grow Network.

Lim, A.E.B., Shanthikumar, J.G., Shen, Z.J.M., 2006. Model Uncertainty, Robust Optimization, and Learning, in: Johnson, M.P., Norman, B., Secomandi, N., Gray, P., Greenberg, H.J. (Eds.), Models, Methods, and Applications for Innovative Decision Making. INFORMS, pp. 66–94. https://doi.org/10.1287/educ.1063.0021

Lu, J., Yan, Z., Han, J., Zhang, G., 2019. Data-Driven Decision-Making (D ^{3} M): Framework, Methodology, and Directions. IEEE Trans. Emerg. Top. Comput. Intell. 3, 286–296. https://doi.org/10.1109/TETCI.2019.2915813

McNally, R.C., Schmidt, J.B., 2011. From the Special Issue Editors: An Introduction to the Special Issue on Decision Making in New Product Development and Innovation: From the Special Issue Editors. Journal of Product Innovation Management 28, 619–622. https://doi.org/10.1111/j.1540-5885.2011.00843.x

Mitchell, T.M., 1997. Machine Learning, McGraw-Hill series in computer science. McGraw-Hill, New York.

Morgan, J.M., Liker, J.K., 2006. The Toyota product development system: integrating people, process, and technology. Productivity Press, New York.

Neelamkavil, F., 1987. Computer simulation and modelling. Wiley, Chichester [Sussex, England] ; New York.

Nielsen, M.A., 2012. Reinventing discovery: the new era of networked science. Princeton University Press, Princeton, N.J.

Nosek, B.A., Alter, G., Banks, G.C., Borsboom, D., Bowman, S.D., Breckler, S.J., Buck, S., Chambers, C.D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D.P., Hesse, B., Humphreys, M., Ishiyama, J., Karlan, D., Kraut, A., Lupia, A., Mabry, P., Madon, T., Malhotra, N., Mayo-Wilson, E., McNutt, M., Miguel, E., Paluck, E.L., Simonsohn, U., Soderberg, C., Spellman, B.A., Turitto, J., VandenBos, G., Vazire, S., Wagenmakers, E.J., Wilson, R., Yarkoni, T., 2015. Promoting an open research culture. Science 348, 1422–1425. https://doi.org/10.1126/science.aab2374

Oral-B Genius 6000 Rechargeable Electric Toothbrush I Oral-B [WWW Document], n.d. URL https://oralb.com/en-us/oral-b-genius-6000-rechargeable-electric-toothbrush.

Perez, L., Wang, J., 2017. The Effectiveness of Data Augmentation in Image Classification using Deep Learning. arXiv:1712.04621 [cs].

Phillips-Wren, G., 2013. Intelligent Decision Support Systems, in: Doumpos, M., Grigoroudis, E. (Eds.), Multicriteria Decision Aid and Artificial Intelligence. John Wiley & Sons, Ltd, Chichester, UK, pp. 25–44. https://doi.org/10.1002/9781118522516.ch2

Platt, M.L., Huettel, S.A., 2008. Risky business: the neuroeconomics of decision making under uncertainty. Nat Neurosci 11, 398–403. https://doi.org/10.1038/nn2062

Provost, F., Fawcett, T., 2013. Data science for business: what you need to know about data mining and data-analytic thinking.

Quinones-Gomez, J.C., 2021. Creativity Forward: A Framework That Integrates Data Analysis Techniques to Foster Creativity Within The Creative Process In User Experience Contexts. Creativity Studies 14, 51–73. https://doi.org/10.3846/cs.2021.12933

Raspberry Pi Zero W [WWW Document], n.d. . Raspberry Pi. URL https://www.raspberrypi.com/products/raspberry-pi-zero-w/

Sargent, R.G., 2013. Verification and validation of simulation models. J Simulation 7, 12–24. https://doi.org/10.1057/jos.2012.20

Sharda, R., Barr, S.H., McDonnell, J.C., 1988. Decision Support System Effectiveness: A Review and an Empirical Test. Management Science 34, 139–159.

So, S., Smith, M., 2002. Colour graphics and task complexity in multivariate decision making. Acc Auditing Accountability J 15, 565–593. https://doi.org/10.1108/09513570210440603

Soltani-Fesaghandis, G., Pooya, A., 2018. Design of an artificial intelligence system for predicting success of new product development and selecting proper market-product strategy in the food industry. International Food and Agribusiness Management Review 21, 847–864. https://doi.org/10.22434/IFAMR2017.0033

Sosa, M.E., Eppinger, S.D., Rowles, C.M., 2004. The Misalignment of Product Architecture and Organizational Structure in Complex Product Development. Management Science 50, 1674–1689. https://doi.org/10.1287/mnsc.1040.0289

Sprague, R.H., 1980. A Framework for the Development of Decision Support Systems. MIS Quarterly 4, 1. https://doi.org/10.2307/248957

Stacey, M., Eckert, C., 2003. Against Ambiguity. Computer Supported Cooperative Work (CSCW) 12, 153–183. https://doi.org/10.1023/A:1023924110279

Ullman, D.G., 2018. The mechanical design process, Sixth edition. ed. David G. Ullman, Independence, Oregon.

Ulrich, K.T., Eppinger, S.D., 2004. Product design and development, 3rd ed. ed. McGraw-Hill/Irwin, Boston.

Wall, J., Bertoni, M., Larsson, T., 2020. The Model-Driven Decision Arena: Augmented Decision-Making for Product-Service Systems Design. Systems 8, 22. https://doi.org/10.3390/systems8020022

Wall, J., Bertoni, M., Larsson, T.C., 2018. A model-driven decision arena : Augmenting decision making in early design.

Wang, H., Xu, Z., Fujita, H., Liu, S., 2016. Towards felicitous decision making: An overview on challenges and trends of Big Data. Information Sciences 367–368, 747–765. https://doi.org/10.1016/j.ins.2016.07.007

Witten, I.H., Frank, E., Hall, M.A., 2011. Data mining: practical machine learning tools and techniques, 3rd ed. ed, Morgan Kaufmann series in data management systems. Morgan Kaufmann, Burlington, MA.

Zaki, M.J., Meira, Jr, W., 2014. Data Mining and Analysis: Fundamental Concepts and Algorithms, 1st ed. Cambridge University Press. https://doi.org/10.1017/CBO9780511810114

Zhang, Y., Ren, S., Liu, Y., Sakao, T., Huisingh, D., 2017. A framework for Big Data driven product lifecycle management. Journal of Cleaner Production 159, 229–240. https://doi.org/10.1016/j.jclepro.2017.04.172

Zhao, H., Icoz, T., Jaluria, Y., Knight, D., 2007. Application of data-driven design optimization methodology to a multi-objective design optimization problem. Journal of Engineering Design 18, 343–359. https://doi.org/10.1080/09544820601010981

Zhao, Y., Tang, H., Su, N., Wang, W., 2007. Extension-Based Clustering Method: An Approach to Support Adaptable Design of the Product, in: ASME 2007 International Manufacturing Science and Engineering Conference. Presented at the ASME 2007 International Manufacturing Science and Engineering Conference, ASMEDC, Atlanta, Georgia, USA, pp. 841–849. https://doi.org/10.1115/MSEC2007-31205.

Paper A

Wall, J., Aeddula, O.K., Larsson, T., 2020. Data Analysis Method Supporting Cause and Effect Studies in Product-Service System Development. Proc. Des. Soc.: Des. Conf. 1, 461–470. https://doi.org/10.1017/dsd.2020.123

INTERNATIONAL DESIGN CONFERENCE – DESIGN 2020
https://doi.org/10.1017/dsd.2020.123

DATA ANALYSIS METHOD SUPPORTING CAUSE AND EFFECT STUDIES IN PRODUCT-SERVICE SYSTEM DEVELOPMENT

J. Wall ✉, O. K. Aeddula and T. Larsson

Blekinge Institute of Technology,
Sweden
✉ johan.wall@bth.se

Abstract

A data analysis method aiming to support cause and effect analysis in design exploration studies is presented. The method clusters and aggregates effects of multiple design variables based on the structural hierarchy of the evaluated system. The resulting dataset is intended as input to a visualization construct based on colour-coding CAD models. The proposed method is exemplified in a case study showing that the predictive capability of the created, clustered, dataset is comparable to the original, unmodified, one.

Keywords: visualisation, product-service systems (PSS), product development, data analysis

1. Introduction

The act of designing industrial products or services is seldom done by one individual, rather it is normally a group effort (McComb et al., 2015). Collaboration in design can therefore be thought of as joint problem solving, i.e. working with others, sharing a common understanding of goals and available means, attempting to find solutions that are satisfying to all involved stakeholders. The knowledge needed to make well informed decisions rarely resides in single persons, not even departments. In contrast, cross-functional teams willing to share their knowledge and values are needed (Murakami, 2016). Wang et al. (2002) further emphasize the importance of cross-functional teams for rapid and reliable evaluation of different options in conceptual design. This development is fuelled by the since long prevailing concurrent engineering approach in product development, and the transition towards product-service systems as well as increasing product complexity (Sundin et al., 2009).

As a consequence of the evolving product development landscape there is a significant need for collaboration, both internally within an organisation but also externally with suppliers, customers and other stakeholders. This is especially true in the early conceptual phase of product development where information is scarce and potentially conflicting. Hence, there is a clear incentive to involve all stakeholders into the decision process. Gathering the involved, normally diverse, group of stakeholders in a collaborative setting for design exploration exercises, sharing knowledge, values, and data is believed to augment decision making ability in early design. Research interest regarding an interactive group workspace focused on digital workflows has increased in recent years, see for example Nieminen et al. (2013) or Benyon and Mival (2015). Research regarding model-centric interactive workspaces are however not as common, an exception is the decision theatre at Arizona State University, see for example White et al. (2015). At Arizona State focus is on sustainability and how decisions affect the system in study on a macroscale. The need for model-centric environments in engineering design has been presented by Rhodes and Ross (2016). In line with that, Wall et al. (2018) proposes a model-driven environment for collaborative decision making focused on early phases of engineering design and presents initial work on conceptualizing, developing, and testing such an environment. Data analysis and visualisation become key enablers in these environments aiding information sharing, communication, understanding and building of knowledge within the cross-functional team (Wall et al., 2018).

Visualization may augment problem-solving capabilities by enabling the processing of more data without overloading the user. Cognitive tools propel users into far more effective thinkers and computer-based tools with visual interfaces may be the most powerful and flexible cognitive systems (Ware, 2005). Several authors have proposed visualisation constructs intended to support the product development process based on colour-coding the CAD model of the studied system. In colour-coding, system attributes are mapped to a colour scale to highlight components or subsystems that are negatively or positively affected by new designs, schematically exemplified in Figure 1. OstadAhmad-Ghorabi et al. (2009) discusses colour-coding of the CAD model to visualise the environmental impact of components based on a life cycle assessment. Bertoni (2013) developed a lifecycle value representation approach connecting qualitative value scores to the actual CAD representation of the product under analysis. Geromin et al. (2018) proposes to colour-code CAD models to visualize design rationale maturity. Described applications of colour-coded CAD models are either described on a conceptual basis or based on qualitative assessments of attributes to be visualized. However, for colour-coding to be viable in design exploration schemes based on automated simulation setups, methods supporting quantitative analysis linking attributes to specific components in the studied system is necessary.

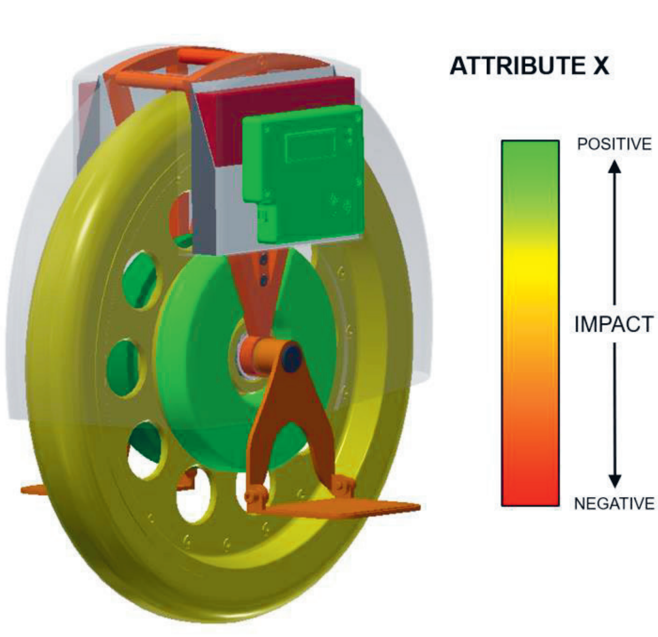


Figure 1. Example of colour-coded visualization

Based on the hypothesis that “Associating data and information with identifiable components and subsystems within the studied system is an efficient way to share information, aiding understanding in a cross-functional team” this paper presents a data analysis method enabling visualisation of quantified cause and effect relationships by colour-coding the CAD representation of the studied system. More specifically the aim is to develop a method able to support quantification of dependencies between design variables (independent variables) and design attributes (dependent variables) on a component or subsystem level in cases where more than one independent variable drives the configuration of the component or subsystem.

The remainder of this paper is organized as follows; In Section 2, the hierarchical system view is presented. Section 3 presents the proposed data analysis method. In section 4 the proposed method is exemplified through a case study. The paper concludes with discussion, conclusions and directions for future research.

2. System view and analysis

Complexity of engineered systems has increased rapidly over the last decades going from essentially mechanical towards mechatronic systems that to a large extent relies on embedded software to fulfil functionality. Demands on a prospective solutions connectedness carried by the rapid development of Internet of Things combined with the conceivable development of communication and information technology further adds to this. There is no concise definition of what a complex system is (Ladyman et al., 2013). In the current work a complex system is defined by the integration of subsystems forming a system that exhibits behaviour not attainable by any of the subsystems alone. The interaction of the subsystem elements often results in emergent behaviour of the complex system that was not originally intended, designed, or desired. To be able to assess such system using a model-based approach, concept description needs to consider these interactions and component interdependencies according to overall system description. A common approach, attempting to make complex systems comprehensible, is to view them as a top-down hierarchy, decomposing the system down to indivisible parts. This decomposition is generally done from a structural point of view. In a structural hierarchy, components reside at the lowest level. At the mid-level, components are assembled into functional units, so-called sub-systems. At the top-level of the hierarchy, an assembly of sub-systems are connected to form a system able to perform a desired function. This way of visualizing systems makes them comprehensible for a wider audience within the development effort of the solution.

Information about a system might be conveyed using other hierarchical descriptions, for example the attribute hierarchy. An attribute can be defined as any aspect of the product itself or its use that can be used to compare product alternatives (Grunert, 1989). As an example, attributes of an automobile might be acceleration or fuel consumption. System attributes might be classified in numerous ways, see for example Crnkovic and Larsson (2004). In the current study attributes are divided into two types, directly composable and derived attributes. A directly composable attribute is a function of and only of the same attribute. Whereas a derived attribute is composed by other attributes of the system. As a derived attribute and the attributes it is derived from has a parent-child relationship, a hierarchy of attributes and associated quantifiable models used to predict them may be defined as exemplified in Figure 2. The structural and attribute hierarchies are directly interrelated as the structural hierarchy in combination with contextual information, such as usage scenarios, are the input to the models used to predict the attributes in the performance and resource space respectively.

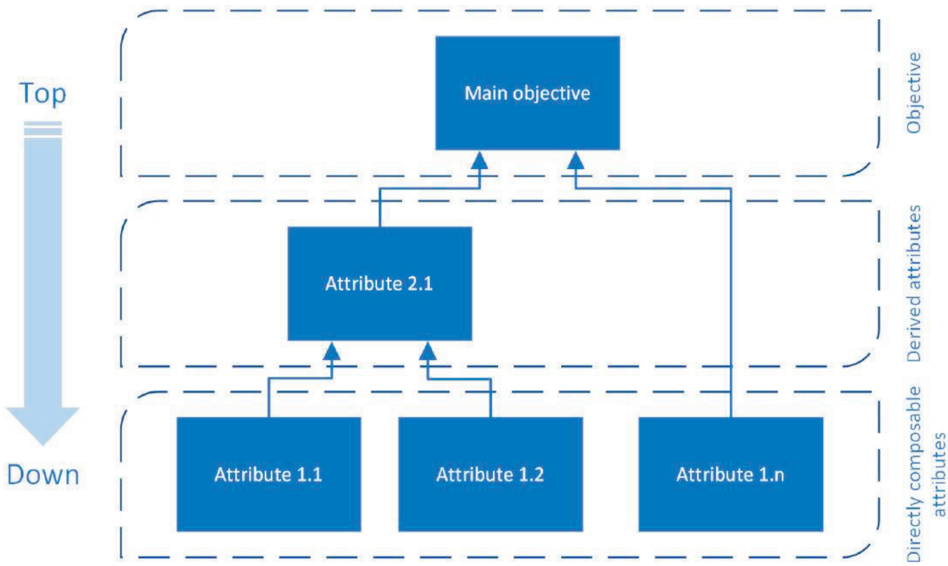


Figure 2. Attribute hierarchy

A hierarchical description of attributes might foster a shared understanding how these attributes are related. As such it might act as a boundary object (Larsson, 2003), mitigating negotiation regarding system design within the cross-functional team. However, to truly understand cause and effect relationships, i.e. how alternation of a proposed design concept affects a specific attribute or the understanding of how the concept should be modified in order to reach a specific target, experimentation is needed. By generating, and through computational methods evaluate and compare all feasible conceptual solutions in the current design space is rarely a practical solution. Furthermore, in an extreme case, if an optimal solution was stumbled upon early on in an exploration endeavour the ability to recognize that is lacking unless comparative studies are conducted. Commonly, datasets to be analysed origins from experiments varying independent variables according to a specified experimental plan, this is referred to as Design-of-Experiment (DOE). DOE is a technique for choosing a limited set of data samples in the design space with the goal of maximizing the amount of information produced (Giunta et al., 2003). When the experiment is executed, associated attributes for all concept variants in the DOE generated experimental plan are predicted and stored in a database. This type of design exploration activities in relation to complex systems may generate an abundance of data containing an intricate hierarchy of attributes. To understand how attributes on different levels is affected by system/subsystem/components or on a more detailed level relates to component features is imperative, making exploration of cause-and-effect relationships through data analysis a vital task.

The independent variables in a design exploration study might, for example, be associated with the proposed mechanical structure, usage scenario or even other attributes of the studied system. The resulting data structure of an experiment, including independent and dependent variables, is classified as a two-level variable hierarchy according to the left part of Figure 3. However, typically these independent variables are not identifiable as components or subsystems. Rather a subset of the independent variables in an experiment, combined with parameters and constants, drive the configuration of a particular component or sub-system and, as a consequence, also attributes associated with that component or sub-system. This might be visualized as an intermediate level in the variable hierarchy of the analysis as exemplified in the right part of Figure 3. As this intermediate level is not directly included in the conducted experiment, data to populate and in the next step analyse it, does not exist. To create this intermediate level, data may be synthesized by clustering the identified subset of independent variables and aggregate their effect on the dependent variable. To fit the intended visualisation scheme, by colour-coding the CAD model as discussed in the introduction, the independent variables are clustered based on the structural decomposition given by the structural hierarchy. A method for how to synthesis data for this intermediate level is described in section 3.

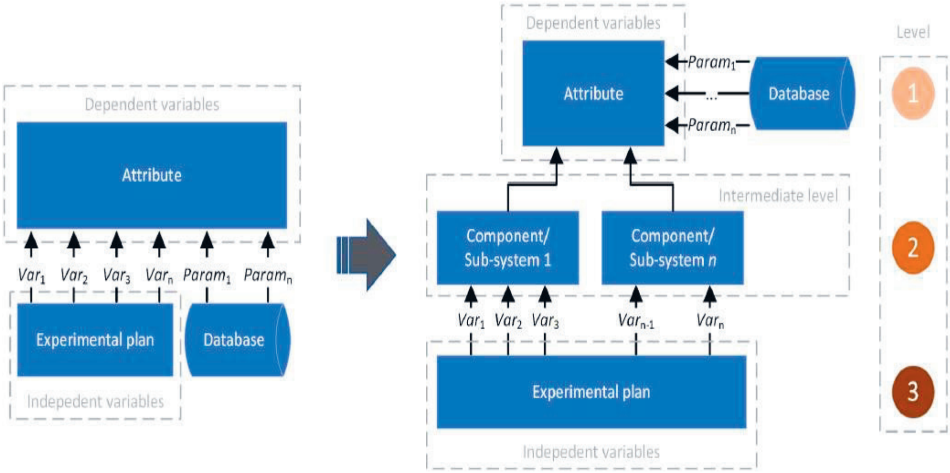


Figure 3. Variable hierarchy

3. Data analysis method

The cause-and-effect relationship between dependent and independent variables may be studied by establishing mathematical relationships between them and analysing the dependent variable dependencies on independent variables from the equation coefficients as is done in for example regression analysis (Allen, 1997). This type of analysis is well established studying relationships between dependent and independent variables that are directly observable. Generating new, clustered, variables that are inferred from the direct variables and establishing the relationship between clustered variables and the dependent variables are however challenging. A data analysis method is proposed to cluster the independent variables according to the structural hierarchy and construct a new intermediate level in the variable hierarchy. The method is designed to work under the assumptions that the variables are continuous and that the relationships between dependent variable and independent variables are linear. The assumption of linearity is a simplification applied in this work aiming at presenting an initial methodology for this type of analysis. Furthermore, the method is designed to analyse relationships between one dependent variable and several independent variables. Figure 4 depicts an overview of the proposed method, schematically also showing its role in the concept evaluation process. In the proposed method, data is partially regressed between dependent and independent variables and the respective independent variable coefficient is utilized to cluster the variables, creating an intermediate hierarchical level. All the variables are standardized to bring down the variables with different metric units to a single scale, to analyse the dependent variable dependencies on the independent variables (Bring, 1994). Standardization reduces the mean value of the variables to “0” and standard deviation to “1”, implying that all the variables are distributed normally on the same scale, which helps to understand which independent variables has greater effect on the dependent variable, when they have different metric units (Devore, 2012). Partial least square regression is applied again to understand the cause and effect relationships for the desired hierarchical levels or between desired dependent and independent variables.

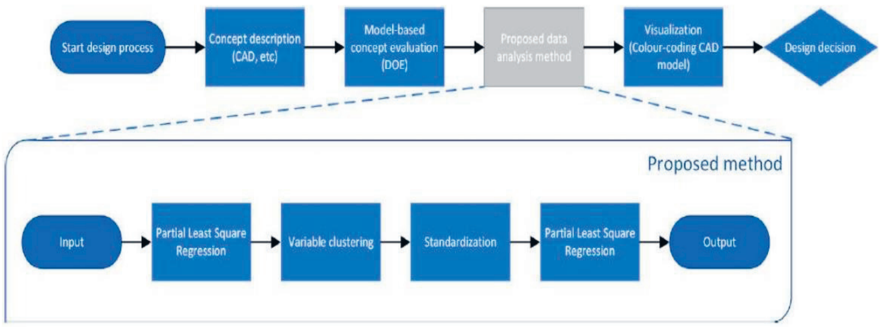


Figure 4. Overview of proposed method

3.1. Partial regression

Partial least square (PLS) regression is a statistical method to find the linear regression model for the desired input and output variables. It generates a regression model even when there are less number of concepts than number of variables (Rännar et al., 1995). The dataset consists of generated independent and dependent variable, where the dependencies are expressed in terms of an equation between them, there can be an infinite number of solutions satisfying the equation i.e., either nonparallel lines or identical lines or intersecting at a common point on independent variables plane. Using PLS, helps to identify the unique solution, which inter-links grouping variables and the output variable. PLS is useful to extract the variables that are not directly observed but are rather being inferred from other variables, in a way it assists to cluster the variables for a new hierarchical level (Pirouz, 2006).

In this context, partial least squares regression is used before standardization to understand the dependencies of the independent variables to be clustered and to generate an intermediate level in the variable hierarchy with respect to the dependent variable. In general, only a subset of existing independent variables are clustered and mapped onto the new intermediate level. The generated variables of the new hierarchy level are created with respect to the dependent variable to keep the contributions of the nonclustered independent variables being fixed with respect to the output at any hierarchy level. If the variables are clustered by dimension reduction method and without considering the dependent variable, then the significance of the contributions of the non-clustered variables are lost during the experimentation. Dependency of dependent variable on the independent variables are analysed using the regression coefficients (β). PLS finds the regression model by projecting the dependent variable (Y) and independent variable matrix (X) to another new space, finding the multidimensional direction in X space that explains the maximum multidimensional variance direction in the Y space. Considering the linear assumption between the dependent and independent variables, according to the Equation (1).

$$Y = \beta X + C \tag{1}$$

where Y is (n x 1) matrix of the dependent variable and X is (n x p) matrix of the independent variables, C is the noise (p x 1) matrix and β is an *(n x 1)* matrix of regression coefficients. PLS decomposes the dependent and independent variables according to Equations (2) and (3) to maximise the covariance of T in X and Y space (Boulesteix and Strimmer, 2006).

$$Y = TQ^T + F \tag{2}$$

$$X = TP^T + E \tag{3}$$

where T is (n x 1) matrix giving the latent components for the variable observations, P (p x 1) matrix and Q (q x 1) matrix are matrices of coefficients and E (n x p) and F (n x q) are random error terms. The latent component T matrix is constructed as a linear function of X, as in Equation (4).

$$T = XW \tag{4}$$

where W is *(p x 1)* matrix of weights. The latent components matrix T is used to predict the regression coefficients and matrix of coefficients Q. Q matrix is obtained as the least square solution of Equation (2).

$$Q^T = (T^T T)^{-1} T^T Y \tag{5}$$

Q^T in Equation (2), transformed to Equation (6)

$$\mathrm{Y} = (XW)Q^T + F \tag{6}$$

Comparing Equation (6) and Equation (1) results in Equation (7)

$$\beta = WQ^T \tag{7}$$

Regression coefficients (β) has a minimum norm satisfying the Equation (1)

Min $\|\beta\|$ such that $Y = \beta X$

Where β is calculated using the Equation (8)

$$\beta = W(T^T T)^{-1} T^T Y \tag{8}$$

The regression coefficients (β) signifies how much the mean of the dependent variable changes when there is a one-unit shift of independent variable while keeping the others fixed. The sign represents the directional change, i.e., when there is a positive coefficient the dependent variable increases, as the particular independent variable increases and vice versa.

3.2. Variable clustering

Variable clustering involves the grouping of independent variables to form a new variable. Initial regression coefficients of the original independent variables are utilized to cluster the variables, according to Equation (9). In this work independent variables are clustered according to referenced structural hierarchy, independent variables belonging to the same component or subsystem are clustered together. Partial least square regression is applied with the dependent variable and independent variables to generate the regression coefficient matrix $[\beta_0, \beta_{i_1}, \beta_{i_2}, \dots \dots]$ and the respective regression coefficient is multiplied with the respective independent variable.

$$X_{new} = (\beta_0 + \beta_{i_1} x_{i_1} + \beta_{i_2} x_{i_2} + \cdots) \tag{9}$$

where X_{new} is new derived variable created by clustering sub-set of the independent variables. β_0 is the intercept term, β_{i_1} the regression coefficient for i_1 independent variable and dependent variable and x_{i_1} is the independent variable to be clustered. The regression coefficients are either negative or positive, when combined with the independent variables according to the Equation (9), it is most likely that there will be a positive regression coefficient for the clustered variable with the dependent variable, in-order to maintain the sign convention with the original independent variables. To avoid this issue, the individual effect of the variables in the clustered group is investigated, whether there is a larger negative regression coefficient or larger positive regression coefficient by summing up all the similar independent variables regression coefficient. If the summation is negative value, then Equation (9) is multiplied by -1 or else the Equation (9) is valid.

4. Case study

To exemplify the proposed method, it was used to analyse an existing dataset originating from a design exploration study evaluating proposed concepts of road construction equipment. In the study different design configurations for the vehicle platform were evaluated considering both performance and resource space. Concept evaluations were done through model-based experimentation (simulation). Simulation is an effective means to enable extensive exploration, so to learn faster (by performing more and earlier iterations) about the characteristics of the best possible design (see: Thomke and Fujimoto, 2000). A Value Driven Design (VDD) approach were applied as Systems Engineering (SE) research has stressed the importance of the value model to frontload engineering design activities (Collopy and Hollingsworth, 2011). This model is expressed as a single objective function that aims at measuring the “goodness” of the design.

In the study 700 variants of the vehicle platform was studied. Variations were driven by seven design variables (independent variables in the study) and 16 functional attributes were assessed. Based on these, derived attributes in resource space was assessed. More information on how this dataset is generated and the applied modelling and simulation scheme can be found in Bertoni et al. (2019). The value model was fed with the output of these models to render a value score for the design configuration under analysis. All aspects of value was quantified in monetary terms, enabling easy trade-off with more traditional requirements. This quantification process was based on the implementation of Net Present Value (NPV) from the VDD literature (Price et al., 2012). The dataset also contained information enabling mapping each independent variable directly to components or subsystems in the structural hierarchy.

The studied dataset contained three categorical variables rendering a non-linear behaviour. As proposed method is valid only for systems showing a linear behaviour, the dataset was linearized by fixing two out of three categorical variables to a specific choice and transforming the third one to continuous, making use of the dummy variables (Yip and Tsang, 2007). Dummy variables are numerical variables representing subgroups of the categorical variable. Number of subgroups for a categorical variable depends on the number of classes it contains. The model was custom trained to identify the type of categorical variable and regress accordingly. To exemplify the method, NPV was chosen as dependent variable, and the effect of the seven design variables were assessed. The regression coefficients for the dependent variable “NPV” are shown in the left part of Figure 5.

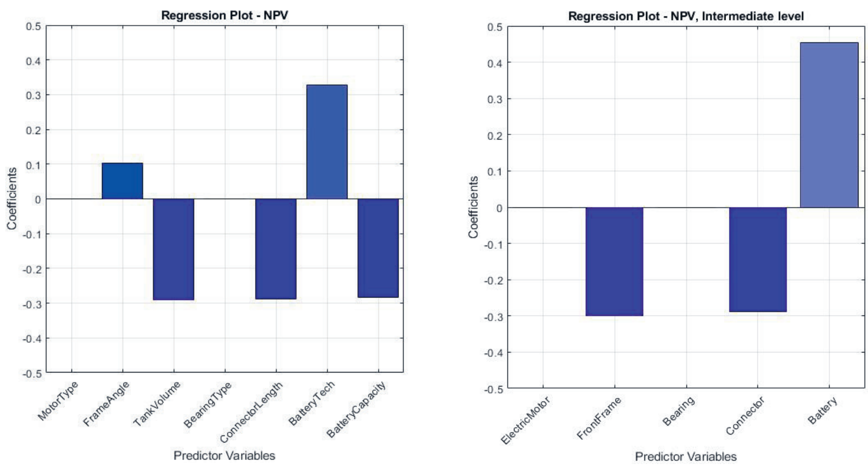


Figure 5. Regression Coefficients, example NPV

In the next step the independent variables were clustered according to the structural hierarchy, reducing them to five at the intermediate level of the variable hierarchy. These five variables are at this stage directly interpretable as components or subsystems of the studied system. The regression coefficients of the second stage are shown in the right part of Figure 5. In Figure 5, it can be seen that nor motor type or bearing type has an effect in this experiment. This is due to the applied linearization procedure where both these variables are set to fixed values.

Regression coefficients from the intermediate level are used to drive colour-coding of associated CAD-model. The data exchange between the simulation environment and the CAD application is enabled by establishing generic communication protocols/functions that exploits the application programming interface of each software. A colour scheme ranging from red to green over yellow is used. Red represents a negative effect on studied attribute whereas green represents a positive effect. Yellow indicates no or little effect on studied attribute. Figure 6 shows a mock-up of an intended implementation visualizing cause and effects within the attribute hierarchy. In this simplified example of visualisation, three out of the assed 16 attributes are shown as well as their aggregation in form of NPV rendered by the value model. Results from Figure 5 accompanied by results from similar results of the in the figure exemplified derived attributes, manoeuvrability, visibility and upgradeability, are shown. Components and subsystems not included in the experiment are shown as transparent.

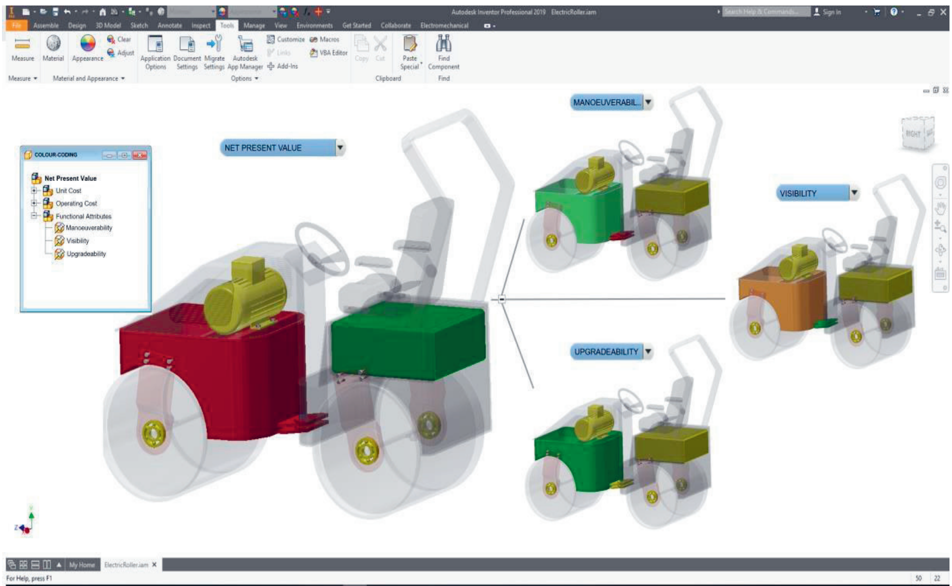


Figure 6. Colour-coded CAD-model in case study (CAD mock-up)

Coefficient of determination (R^2) of a statistical model was used to validate the methodology. It determines how close the predicted data matches to the original data, according to the Equation (10) (Loretan and Kurz-Kim, 2007).

$$R^2 = 1 - \frac{\Sigma(y-Y)^2}{\Sigma(y-\bar{y})^2} \tag{10}$$

where y represents the actual data values, Y is the predicted data values and $\bar{y}$ is the mean value of y. The value of the regression coefficient of determination ranges from [0,1]. As the value of R^2 approaches unity, the predicted values are equal to the actual data values, i.e., the closer the value of R^2 to 1, the greater is the fit of the data, and the closer the value of R^2 to 0, the poorer is the fit of the data. For the NPV approximation, R^2 is 0.6402, in comparison coefficient of determination of assessed derived attributes are 0.95 and above.

5. Discussion and conclusion

A data analysis method that clusters and aggregates effects of multiple design variables, independent variables in a study, is presented. The method aims to support cause and effect analysis supplying structured input to a visualisation construct. The intent is to map the analysed data based on the structural hierarchy of the proposed concept and to visualise the result using the CAD geometry as base.

The proposed data analysis method is based on the partial least square regression method. The regression is conducted in two steps with an intermediate variable clustering through linkage to the proposed concept's structural hierarchy. The proposed method ensures that the significance of the contributions of the non-clustered variables are not lost during the experimentation. The proposed method has no limitation on number of variables it can handle. In practice however this will be an issue of availability of computational resources and time.

The method is exemplified through a case including visualisation by colour-coding the CAD model of the studied concept. Applying a validation scheme based on coefficient of determination analysis it is shown that the data structure using a "virtual" intermediate level performs, considering the non-linearities in the original dataset, comparably to the original dataset when it comes to predicting attributes.

As of now the application of the presented method is limited to problems where relationship between dependent and independent variables may be categorised as linear. During development of the algorithm datasets from multiple test cases were analysed. It was found that non-linear effects may affect the accuracy of the method to such an extent that, depending on the stage of the development project, these effects might need to be considered. Future work aims to extend the feasibility of the method to also include problems featuring non-linear relationships. An enhanced and generalized version of the proposed method, combining the data analysis and the visualisation construct, could be implemented as a design support in CAD software's enabling visualisation of quantified cause and effect relationships and thereby supporting collaborative decision making in design.

Acknowledgement

The research leading to these results has received financial support by the Swedish Knowledge and Competence Development Foundation (Stiftelsen för kunskaps- och kompetensutveckling) through the Model Driven Development and Decision Support research profile at Blekinge Institute of Technology.

References

Allen, M.P. (1997), *Understanding regression analysis*, Plenum Press, New York. https://doi.org/10.1007/b102242

Benyon D. and Mival, O. (2015), "Blended Spaces for Collaboration", *Computer Supported Cooperative Work (CSCW)*, Vol. 24, pp. 223-249. https://doi.org/10.1007/s10606-015-9223-8

Bertoni, A. (2013), "Analyzing Product-Service Systems conceptual design: The effect of color-coded 3D representation", *Design Studies*, Vol. 34 No. 6, pp. 763-793. https://doi.org/10.1016/j.destud.2013.02.003

Bertoni, M. et al. (2019) "Life cycle simulation to support cross-disciplinary decision making in early PSS design", *Procedia*, ISSN 2212 - 8271, EISSN 83, Procedia CIRP, Elsevier. https://doi.org/10.1016/j.procir. 2019.03.138

Boulesteix, A.-L. and Strimmer, K. (2006), "Partial least squares: a versatile tool for the analysis of highdimensional genomic data", *Briefings in Bioinformatics*, Vol. 8 No. 1, pp. 32-44. https://doi.org/10.1093/ bib/bbl016

Bring, J. (1994) "How to Standardize Regression Coefficients", *The American Statistician*, Vol. 48 No. 3, pp. 209-213. https://doi.org/10.1080/00031305.1994.10476059

Collopy, P.D. and Hollingsworth, P.M. (2011), "Value-driven design", *Journal of Aircraft*, Vol. 48 No. 3, pp. 749-759. https://doi.org/10.2514/1.C000311

Crnkovic, I. and Larsson, M. (2004), "Classification of quality attributes for predictability in component-based systems", *Proc. of Workshop on Architecting Dependable System,* IEEE

Devore, J.L. (2012), *Probability and statistics for engineering and the sciences*, 8th ed., Brooks/Cole, Cengage Learning, Australia.

Geromin, A. et al. (2018), "CAD modelling based on knowledge synthesis for design rational", *Procedia CIRP*, Vol. 70, pp. 156-161. https://doi.org/10.1016/j.procir.2018.01.008.

Giunta, A.A., Wojtkiewicz, S.F. and Eldred, M.S. (2003), "Oveview of modern design of experiment methods for computational simulation", *41st Aerospace*

Sciences Meeting and Exhibit, Reno, Nevada. https://doi.org/10.2514/6.2003-649

Grunert, K.G. (1989), "Attributes, attribute values and their characteristics: A unifying approach and an example involving a complex household investment", *In Journal of Economic Psychology*, Vol. 10 No. 2, pp. 229251. https://doi.org/10.1016/0167-4870(89)90021-4

Ladyman, J., Lambert, J. and Wiesner, K. (2013) "What is a complex system?", *European Journal for Philosophy of Science*, Vol. 3 No. 1, pp. 33-67. https://dx.doi.org/10.1007/s13194-012-0056-8.

Larsson, A. (2003) "Making sense of collaboration: the challenge of thinking together in global design teams", in. ACM Press, p. 153. https://dx.doi.org/10.1145/958160.958184.

Loretan, M.S. and Kurz-Kim, J.-R. (2007), "A note on the coefficient of determination in regression models with infinite-variance variables", *Discussion Paper Series 1: Economic Studies*. Available at: https://ideas.repec. org/p/zbw/bubdp1/5574.html (Accessed: 1 November 2019).

McComb, C., Cagan, J. and Kotovsky, K. (2015), "Rolling with the punches: An examination of team performance in a design task subject to drastic changes", *Design Studies*, Vol. 36, pp. 99-121. https://doi.org/10.1016/j.destud.2014.10.001

Murakami, Y. (2016), "Linking knowledge absorption and transmission toward innovation in R & D organizations", *in Proceedings of the 17th European Conference on Knowledge Management*, ECKM 2016. vol. 2016-January, Academic Conferences Limited, pp. 667-675.

Nieminen, M.P., Tyllinen, M. and Runonen, M. (2013), "Digital war room for design", *Proceedings of the 15th international conference on Human Interface and the Management of Information: information and interaction for learning, culture, collaboration and business*, July 21-26, Las Vegas, NV. https://doi.org/ 10.1007/978-3-642-39226-9_39

Ostad-Ahmad-Ghorabi, H., Collado-Ruiz, D. and Wimmer, W. (2009), "Towards Integrating LCA into CAD", *Proceedings of ICED 09, the 17th International Conference on Engineering Design*, Vol. 7, Design for X / Design to X, Palo Alto, CA, USA, pp. 301-310.

Pirouz, D.M. (2006), "An Overview of Partial Least Squares", *SSRN Electronic Journal*. https://doi.org/ 10.2139/ssrn.1631359

Price, M. et al. (2012), “A novel method to enable trade-offs across the whole product life of an aircraft using value driven design”, *Journal of Aerospace Operations*, Vol. 1 No. 4, pp. 359-375. https://doi.org/10.3233/ AOP-120028

Rhodes, D.H. and Ross, A.M. (2016), “A vision for human-model interaction in interactive model-centric systems engineering”, *In 26th Annual INCOSE International Symposium,* Edinburgh. Scotland . https://doi.org/10.1002/inst.12162

Rännar, S. et al. (1995), “A PLS kernel algorithm for data sets with many variables and few objects. Part II: Cross-validation, missing data and examples”, *Journal of Chemometrics,* Vol. 9 No. 6, pp. 459-470. https://doi.org/10.1002/cem.1180090604

Sundin, E. et al. (2009), “Challenges for Industrial Product/Service Systems - Experiences from a learning network of large companies”, *In: Industrial product- service systems (IPS2): proceedings of the 1st CIRP IPS2 Conference*, 01 - 02 April, Cranfield University, UK.

Thomke, S. and Fujimoto, T. (2000), “The effect of ”Front-Loading“ problem-solving on product development performance”, *Journal of Product Innovation Management*, Vol. 17 No. 2, pp. 128-142. https://doi.org/ 10.1016/S0737-6782(99)00031-4

Wall, J., Bertoni, M., and Larsson T. (2018), “A model-driven decision arena: Augmenting decision making in early design”, *Proceedings of NordDesign* 2018, August 16th-18th Linköping.

Wang, L. et al. (2002), “Collaborative conceptual design—state of the art and future trends”, *Computer-Aided Design*, Vol. 34 No. 13, pp. 981-996. https://doi.org/10.1016/S0010-4485(01)00157-9

Ware C. (2005), “Visual Queries: The Foundation of Visual Thinking”, In: Tergan, S.O. and Keller, T. (eds), *Knowledge and Information Visualization. Lecture Notes in Computer Science*, Vol. 3426, Springer, Berlin, Heidelberg. https://doi.org/10.1007/11510154_2

White, D.D. et al. (2015), “Water management decision makers’ evaluations of uncertainty in a decision support system: the case of WaterSim in the Decision Theater”, *Journal of Environmental Planning and Management*, Vol. 58 No. 4, pp. 616-630, https://doi.org/10.1080/09640568.2013.875892

Yip, P.S.L. and Tsang, E.W.K. (2007), “Interpreting dummy variables and their interaction effects in strategy research”, *Strategic Organization,* Vol. 5 No. 1, pp. 13-30. https://doi.org/10.1177/1476127006073512.

PAPER B

Aeddula, O.K., Wall, J., Larsson, T., 2021. Artificial Neural Networks Supporting Cause-and-Effect Studies in Product–Service System Development, In: Chakrabarti, A., Poovaiah, R., Bokil, P., Kant, V. (Eds.), Design for Tomorrow—Volume 2, Smart Innovation, Systems and Technologies. Springer Singapore, Singapore, pp. 53–64. https://doi.org/10.1007/978-981-16-0119-4_5

Artificial Neural Networks Supporting Cause-and-Effect Studies in Product–Service System Development

Omsri Kumar Aeddula, Johan Wall, and Tobias Larsson

Abstract A data analysis method based on artificial neural networks aiming to support cause-and-effect analysis in design exploration studies is presented. The method clusters and aggregates the effects of multiple design variables based on the structural hierarchy of the evaluated system. The proposed method is exemplified in a case study showing that the predictive capability of the created, clustered, dataset is comparable to the original, unmodified, one. The proposed method is evaluated using coefficient of determination, root mean square error, average relative error, and mean square error. Data analysis approach with artificial neural networks is believed to significantly improve the comprehensibility of the evaluated cause-and effect relationships studying PSS concepts in a cross-functional team and thereby assisting the difficult and resource-demanding negotiations process at the conceptual stage of the design.

1 Introduction

Concept evaluation in product development has historically been overly reliant on experts' ambiguous and subjective judgments and qualitative descriptions [1]. Hence, a major quest in engineering design is the search for information and the subsequent analysis and build of knowledge. An efficient way to generate data is through experimentation. This means running some kind of test procedure for various input data and to study how these data affect the output of the procedure. Experimentation is commonly done using models, simplified representations of the studied system, nowadays predominantly mathematical models. Experiment-

O. K. Aeddula (✉) · J. Wall · T. Larsson

Blekinge Institute of Technology, Karlskrona,
Sweden e-mail: omsri.kumar.aeddula@bth.se

53 A. Chakrabarti et al. (eds.), Design for Tomorrow–Volume 2, Smart Innovation, Systems and Technologies 222, https://doi.org/10.1007/978-981-16-0119-4_5

ation with mathematical models is here referred to as simulation. Building and solving mathematical models, considering a diverse set of systems and disciplines, have received significant attention within the research community during the last decades. However, the question regarding how designers should interpret and use these models and solutions in design has been largely neglected [2]. This problem is even more pronounced considering that designing industrial products or services is nowadays normally a group effort and rarely an individual act [3].

The transition away from product view, toward providing product–service systems, increasing product complexity, and the long-prevailing concurrent engineering approach, makes inroads for this development [4]. This emphasizes the need for collaborations, internally within the organizations and externally with suppliers, customers, and other stakeholders in the value chain, especially in the early phases of product development, where the information is scarce, vague, and potentially conflicting.

Supporting cross-functional design teams with data analysis and visualization, helping them to interpret and utilize simulation results, hence become a key success factor, while providing a potential setting for information sharing, communication, understanding, and building of knowledge. Data analysis assists in transforming information into knowledge, exploiting the relationships among the data, to drive the product or service design process [5].

Bititci et al. [6] emphasize exploring the cause-and-effect relationships among the multiple variables involved, as a part to measure the quality of design improvement. Wall et al. [7] propose a data analysis method based on partial least square regression supporting cause-and-effect analysis in design exploration studies for linear and continuous variables. The proposed method clusters and aggregates the effects of multiple design variables based on the structural hierarchy of the evaluated system. According to [8], multiple linear regression analysis is a tool to understand the relationship between one dependent variable and several independent variables. Linear regression methods are easy to implement and simpler in design. The limitation of multiple linear regression analysis is that they are modeled for linear data, owing to which they cannot discern any nonlinear relationships in data [9]. However, design problems often contain nonlinear relationships between design variables and attributes. Hence, a method able to cope also with nonlinear relationships is sought [10].

In recent years, artificial neural networks have been widely employed in modeling dynamic systems, as they are found to accurately model continuous functions and shown to have good predictive performance in simulations of nonlinear dynamic systems, due to their adaptability, flexibility, and optimization capabilities [11]. They have the ability to learn the system behavior from samples by inspection. This paper extends and modifies the method proposed by [7].

An artificial neural network-based data analysis method to understand quantified cause-and-effect relationships of the studied system is proposed. More specifically the aim is to develop a method able to support quantification of dependencies between design variables (independent variables) and design attributes (dependent variables), irrespective of the data linearity, on a component or subsystem level in cases where more than one independent variable drives the configuration of the component or subsystem.

The remainder of this paper is organized as follows; In Sect.5.2, the proposed data analysis method is presented. In Sect.5.3 the proposed method is exemplified through a case study. The paper concludes with discussion, conclusions, and directions for future work.

2 Data Analysis Method

A cause-and-effect relationship may be studied by establishing mathematical equations between design variables (independent variables) and design attributes (dependent variables). Design variables are entities that may be changed in an experiment affecting the shape or properties of the studied system. An attribute can be defined as any aspect of the product itself or its use that can be used to compare product alternatives [12]. As an example, the attributes of an automobile might be acceleration or fuel consumption. Quantification of dependencies between design variables and design attributes is established by analyzing the design variable’s contribution toward the design attribute. Linear regression methods analyze the design attributes dependencies on the design variables from the generated linear mathematical equation coefficients [13]. This type of analysis fits for linear observed variables and limits the applications to data with nonlinearities and for clustered variables.

The variables in a design exploration study might be classified as a two-level hierarchy including design variables and design attributes. Typically, these design variables are not identifiable as components or subsystems. Rather, a subset of design variables in a design exploration drives the configuration of a particular component or subsystem and also attributes associated with that component or subsystem. As this “intermediate level” of components and subsystems are not directly included in the conducted experiment, data to populate, and in the next step to analyze it, does not exist. Rather it is to be inferred from the set of observed design variables by clustering the identified subset of design variables and aggregate their effect on the design attribute. Creating an intermediate level without losing the significant contributions of other non-clustering design variables is a vital task.

Data linearity affects the “goodness” of fit of the intermediate level in the structural hierarchy [7]. This paper proposes a variable clustering method, in combination with an artificial neural network to fit the intermediate level, irrespective of the linearity without altering the significant contributions of other non-clustering design variables. The independent variables are clustered based on the structural decomposition given by the structural hierarchy of the studied system. This inline with the research aim presented in the introduction. Generating new, clustered, variables inferred from the observed design variables is challenging, and adding the element of relationship exploration without losing the significant contributions of other design variables serves to confound the process even further.

Figure 1 depicts an overview of the proposed method, schematically also showing its role in the concept evaluation process. In the proposed method, data is initially analyzed to identify the continuous and categorical variables. A subset of design variables, according to the structural hierarchy, is clustered and each design variable forms a node in the input layer of the designed artificial neural network. The network has a single output node with clustered variable constructs.

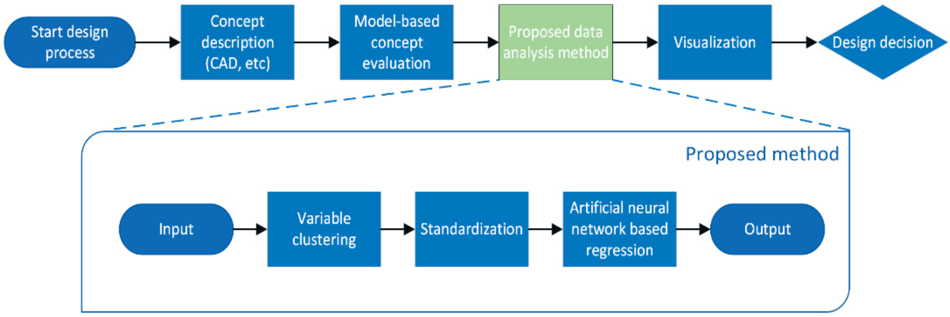


Fig. 1 Overview of the proposed method

Exploration of the neural networks layer weights reveals the contributions of design variables on the design attribute [14]. Design variables and design attributes are normally of different metric units. In this work, they are standardized to scale down all the variables to a common metric scale. Standardization reduces the mean value of the variables to “0” and their standard deviation to “1”, aligning all the variables, distributed normally, on the same scale. This helps to understand the contributions of design variables on a design attribute when they have different metric units [15]. An artificial neural network with designated nodes based on the number of design variables and a predetermined number of hidden layers is employed to understand the cause-and-effect relationships for the desired hierarchical levels or between desired design variables and a design attribute.

2.1 Artificial Neural Networks

An artificial neural network is a mathematical and computational model that simulates the human brain functions of perception, computation, and memory. The ability of neural networks to learn and generalize from the input data makes them a powerful tool to solve numerous real-world applications [16]. An artificial neural network is a system consisting of processing elements namely nodes in each layer with connections (Synapse) between them. Its inherently nonlinear structure is particularly useful in exploring complex relationships of real-world problems. Artificial neural networks learn from the data through iterations without any prior knowledge, capable of handling data with noisy, linear, and nonlinear relationships [17].

In the studied system, the architecture of the artificial neural network is characterized by a single input layer, a single hidden layer with a nonlinear activation function, and a single output layer. A multilayer perceptron, also referred to as a feedforward network, is employed featuring supervised learning. The network maps a set of inputs, design variables, onto a design attribute linking through the hidden layer. The neural network is depicted in Fig. 2. The input layer is a set of neurons with design variables and the output layer contains a single neuron with a design attribute. IW and LW are the weights associated with the input-hidden layer and hidden-output layer, respectively.

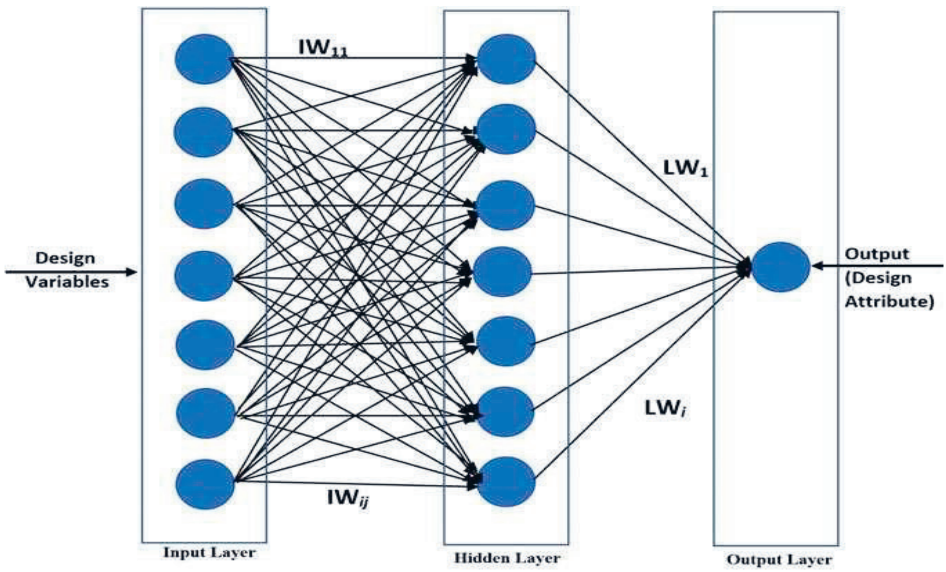


Fig. 2 Artificial neural network architecture

The adapted neural network is described [18], according to Eq. (1):

$$y = \left(\sum_{j=1}^{h} LW_j * f_{inner}\left(\sum_{i=1}^{d} IW_i * x_i + b_0^{(1)}\right) + b_0^{(2)}\right) \quad (1)$$

where *y* is the design attribute; IW_i and LW_j indicate the weights connecting the input layer with hidden layer, and hidden layer with output layer respectively; *h* is the total number of hidden layer neurons; *d* is the number of input layer neurons; $b_0^{(1)}$ and $b_0^{(2)}$ are the biases for the hidden neuron and the output neuron. f_{inner} is a hyperbolic tangent sigmoid activation function. A hyperbolic tangent sigmoid transfer function is given by Eq. (2):

$$f_{inner}(n) = \frac{2}{(1+\exp(-2*n))-1} \tag{2}$$

Network weights are identified using the following iterative method [19], according to Eq. (3) and (4):

$$\{w\}_{i+1} = \{w\}_i - \eta \frac{\partial E}{\partial \{w\}}(\{w\}_i) \tag{3}$$

$$E = \sum_{n=1}^{N}\{t_n - y_n\}^2 \tag{4}$$

where t_n is the actual output value; y_n is the estimated value; and n is the index of the training data.

A Levenberg–Marquardt algorithm is employed to train the adapted network, as the algorithm combines the benefits of gradient descent and Gauss–Newton methods and it speeds up the convergence rate [20]. The algorithm trains the neural network, as long as the weights, net inputs, and transfer functions have a derivative function. According to [21], the connection weights method proved to be better compared to other variable contribution methods such as the stepwise method, profile method, and perturb method. Connection weights method make use of the weight vectors IW and LW in determining the design variables contributions (RI) to the design attribute [22], according to Eq. (6):

$$Q_{(h,i)} = \frac{|(IW_{(h,i)})|*|(LW_h)|}{\sum_{i=1}^{n}|(IW_{(h,i)})|*|(LW_h)|} \tag{5}$$

$$RI_n = \frac{\sum_{h=1}^{n} Q_{(h,i)}}{\sum_{i=1}^{n}\sum_{h=1}^{n} Q_{(h,i)}} \tag{6}$$

where h is the total number of neurons in the hidden layers; and n is the total number of neurons in the input layer.

2.2 Variable Clustering

Design variables clustering involves constructing latent variables, inferred from the observed design variables. In the studied system, design variables are clustered according to the referenced structural hierarchy. Each design variable to be clustered is expressed as a product of two matrices with weights for each variable value and original data with a rotated coordinated system, according to Eq. (7).

$$P = [w] \times Z^T \tag{7}$$

where $[w]$ is the weight matrix of the design variable; and Z is the rotated coordinated system matrix.

The rotated matrix is analyzed to estimate the variance (v) of the rotated matrix. Clustered variables are identified as a sum of the product of total variance of the rotated matrix with the rotated matrix of associated design variables to be clustered, according to Eq. (8):

$$X_{new} = \{(V_1 \times Z_1) + (V_2 \times Z_2) + \cdots + (Vn \times Zn)\} \tag{8}$$

where X_{new} is the new clustered variable in the intermediate hierarchy level, according to variable hierarchy; V_i is the total variance of the rotated matrix of the i^{th} design variable; and n is the number of variables to be clustered.

3 Case Study

To exemplify the proposed method, it was used to analyze an existing dataset originating from a design exploration study evaluating proposed concepts of road construction equipment. In the study, 700 variants of the vehicle platform were studied. Variants driven by seven design variables and 16 design attributes were assessed. The dataset is divided into two sets with 70%of data for the artificial neural network process and 30% for testing the derived observations from the proposed method. The adapted neural network divides the input data into three sets, training, validation, and testing sets, with 70% of input data going to the training set.

A value-driven design (VDD) approach was applied in the design experiment as systems engineering (SE) research has stressed the importance of the value model to front load engineering design activities [23]. This model is expressed as a single objective function that aims at measuring the “goodness” of the design. More information on how this dataset is generated and the applied modelling and simulation scheme can be found in [24]. The model-based experimentation relies on a hybrid model environment evaluating both performance and resource space. The value model was fed with the output of these models to render a value score for the design configuration under analysis.

All aspects of value were quantified in monetary terms, enabling easy trade-off with more traditional requirements. This quantification process was based on the implementation of net present value (NPV) from the VDD literature [25]. The dataset also contained information enabling mapping each design variable directly to components or subsystems in the structural hierarchy.

The studied dataset contained three categorical variables rendering nonlinear behavior. These variables were transformed into continuous variables, generating dummy variables equal to the number of classes in a categorical variable [26]. Dummy variables are continuous variables with logical values representing subgroups of the categorical variable. The model was custom trained to distinguish between categorical and continuous variables and transforming all categorical variables to continuous variables.

To exemplify the method, NPV was chosen as a design attribute, and the relative importance of the continuous design variables was assessed. The relative importance values could be used to understand the cause-and-effect proposition on the magnitude scale, where it represents the relative change in the design variables for a unit change in the design attribute value. The relative importance of the design variables for the design attribute "NPV" is shown in the left part of Fig. 3. The design variables were clustered according to the structural hierarchy, rendering an intermediate level with five clustered design variables. At this stage, these five variables are interpretable as subsystems or components of

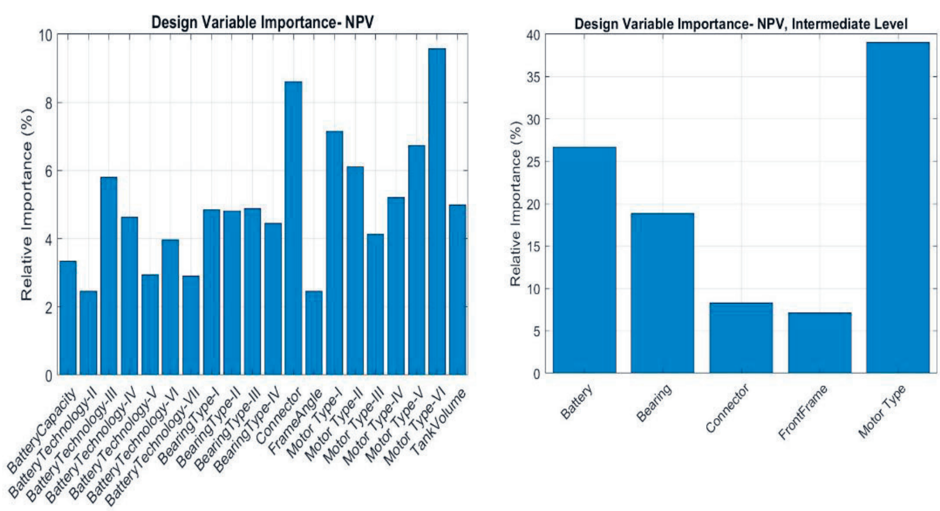


__Fig. 3__ Variable importance

the studied system. The contributions of the clustered design variables are shown in the right part of Fig. 3. The non-clustered variable “connector” has a contribution value of 8.6 both in the left and right part of Fig. 3, indicating that the significance of the non-clustered variable remains the same.

The mathematical equation framing design attribute from design variables, according to Eq. (1), is a nonlinear function, thereby the proportionality is also a nonlinear function. The cause-and-effect relationship between a specific design variable and the design attribute was analyzed by varying the value of a design variable while keeping all other design variables constant. This method of estimation is cumbersome, as the artificial neural networks popularly referred to as “black box” [16] does not provide a direct method to estimate the variable contribution with magnitude and direction directly.

Statistical analysis technique, coefficient of determination (R^2), was used for methodology validation. It determines how close the predicted data matches the original data, according to Eq. (9) [27].

$$R^2 = 1 - \frac{\sum(y-Y)^2}{\sum(y-\bar{y})^2} \tag{9}$$

where y represents the actual data values; Y is the predicted (estimated) data values; and $\bar{y}$ is the mean value of y.

Figure 4 shows visualization of the goodness of fit for the NPV design attribute based on the proposed neural network architecture.

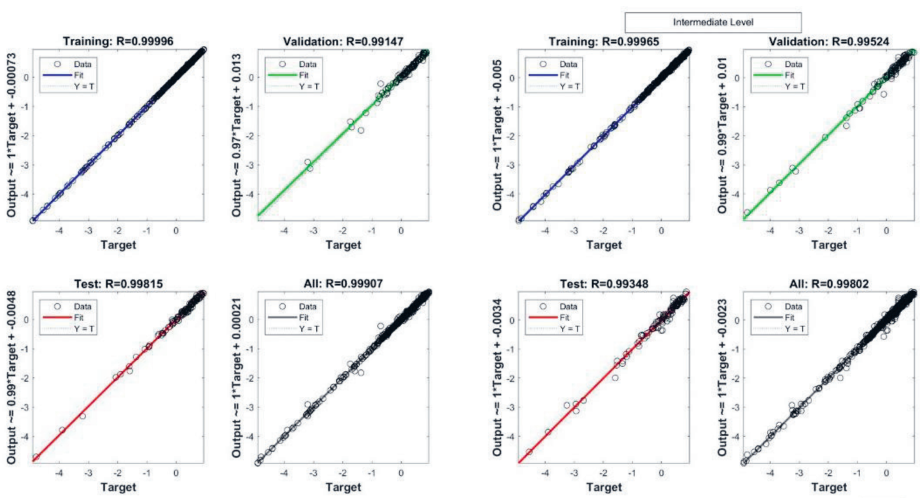


Fig. 4 *Artificial neural network regression plot*

The value of the regression coefficient of determination ranges from [0,1]. As the value of R^2 approaches unity, the predicted values are equal to the actual data values, i.e., the closer the value of R^2 to 1, the greater is the fit of the data, and the closer the value of R^2 to 0, the poorer is the fit of the data. For NPV, R^2 tends out to be 0.9990, in comparison coefficient of determination of assessed attributes are 0.98 and above. Clustered variables were derived from a set of observed design variables, thereby an indirect method was used for validation of intermediate level. Figure 5 shows the predictive capability of the derived neural network model using 30% of the unused design attribute (NPV) data both at the intermediate level and component level. Root mean square error (RMSE) and average relative error (RE) measure the deviation of the predicted values from actual observed values. Prediction of unused design attribute values, along with RMSE and RE, according to Eq. (10) and (11), [15] validates the methodology.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (Y_i - y_i)^2} \tag{10}$$

$$RE = \left| \frac{Y-y}{y} \right| \tag{11}$$

where y represents actual data values; and Y represents predicted data values.

At the intermediate level, RMSE for the estimated design attribute tends out to be 0.0026 and average relative error of 0.0033, similarly, at the component level, RMSE for the design attribute tends to be 0.0023 and average relative error of 0.0029.

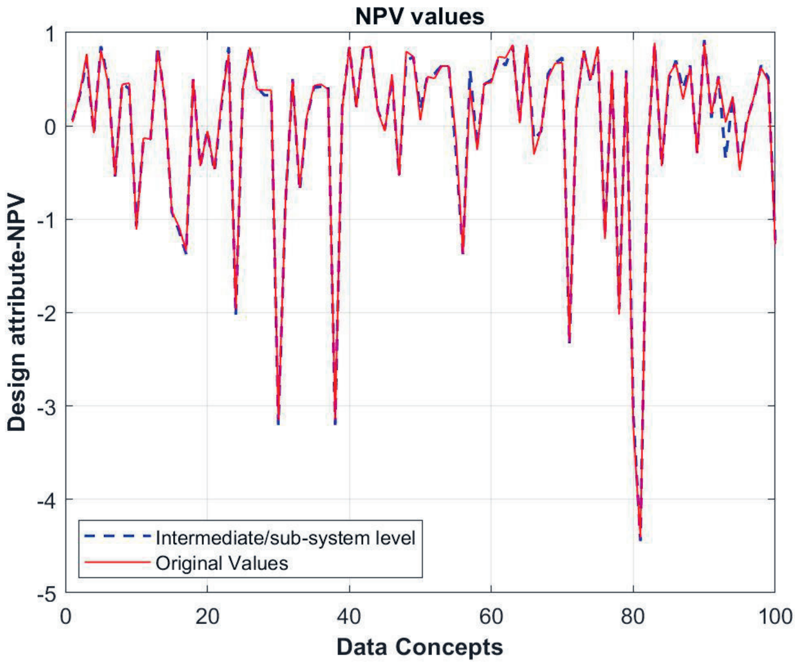


Fig. 5 Estimated data versus observed data

4 Discussion and Conclusion

An artificial neural network-based data analysis method that clusters and aggregates the effects of multiple design variables in a study is presented. The proposed artificial neural network is trained using the Levenberg–Marquardt algorithm, with 70% input data as training data. The method focuses on the exploration of cause-and-effect analysis providing a structured input to a visualization construct. The intention is to generate an intermediate level and map the analyzed data based on the structural hierarchy of the proposed concept.

The proposed method is based on artificial neural networks and matrix decomposition for variable clustering, generating an intermediate level in the structural hierarchy, thereby reducing the dimensionality of the data and hence decrease in computational complexity compared to the original data. The proposed method ensures that the significances of the variable contribution of the non-clustered design variables are not lost during the experimentation. The goodness-of-fit measure in the intermediate level reveals that the adapted neural network model closely fits the data, with clustered variable preserving the original information of the design variable. The proposed method extends the predictive range of linear relationship variable modeling, and the model is trained to solve complex data, overcoming the assumptions in [7]. The proposed method has no limitation to the number of variables, and the intermediate level analysis supports in reducing the computational resources and time compared to the computation of all design variables.

The method is exemplified through a case study including visualization of cause and-effect relationships by bar graphs of the studied concept. Applying a validation scheme based on the coefficient of determination, testing of non-trained data, RMSE, and relative error, it is shown that the data structure using a “virtual” intermediate level performs, considering the nonlinearities in the original dataset, comparably to the original dataset when it comes to predicting attributes .Future works aim to extend the proposed method by implementing a cascaded network for clustering based on unsupervised learning and comparing it with other aggregating variable methods such as the Karhunen–Loeve method.

Acknowledgements The research leading to these results has received financial support by the Swedish Knowledge and Competence Development Foundation (Stiftelsen för kunskaps-och kompetensutveckling) through the Model Driven Development and Decision Support research profile at Blekinge Institute of Technology.

References

1. Zhang, Z., Gong, L., Jin, Y., Xie, J., Hao, J.: A quantitative approach to design alternative evaluation based on data-driven performance prediction. Adv. Eng. Inform. 32, 52–65 (2017). https://doi.org/10.1016/j.aei.2016.12.009
2. Komoto, H., Masui, K.: Classification of design parameters with system modeling and simulation techniques. CIRP Ann. 63(1), 193–196 (2014). https://doi.org/10.1016/j.cirp.2014. 03.098

3. McComb, C., Cagan, J., Kotovsky, K.: Rolling with the punches: an examination of team performance in a design task subject to drastic changes. Des. Stud. 36, 99–121 (2015). https:// doi.org/10.1016/j.destud.2014.10.001
4. Sundin, E., Sandström, G.Ö., Lindahl, M., Rönnbäck, A.Ö., Sakao, T., Larsson, T.C.: 'Challenges for industrial product/service systems: experiences from a learning network of large companies', p. 7
5. VanHorn, D.,Olewnik, A.,Lewis, K.:Design analytics: capturing,understanding, and meeting customer needs using big data. In: Volume7: 9th International Conference on Design Education; 24th International Conference on Design Theory and Methodology, Chicago, Illinois, USA, Aug 2012, pp. 863–875. doi: https://doi.org/10.1115/DETC2012-71038
6. Bititci, U., Nudurupati, S.: Driving continuous improvement. Manuf. Eng. 81(5), 230–235 (2002). https://doi.org/10.1049/me:20020506
7. Wall, J., Aeddula, O.K., Larsson, T.: Data analysis method supporting cause and effect studiesin product-service system development. Proc. Des. Soc. Des. Conf. 1, 461–470 (2020). https:// doi.org/10.1017/dsd.2020.123
8. Hair, J.F. (ed.): Multivariate Data Analysis with Readings, 4th edn. Prentice Hall, Englewood Cliffs, N.J (1995)
9. Jobson, J.D.: Multiple Linear Regression. In: Applied Multivariate Data Analysis. Springer, New York, NY, pp. 219–398 (1991)
10. Tenenbaum, J.B.: A global geometric framework for nonlinear dimensionality reduction. Science 290(5500), 2319–2323 (2000). https://doi.org/10.1126/science.290.5500.2319
11. Haykin, S.S.: Neural Networks: A Comprehensive Foundation, 2nd edn. Prentice Hall, Upper Saddle River, N.J (1999)
12. Grunert, K.G.: Attributes, attribute values and their characteristics: a unifying approach and an example involving a complex household investment. J. Econ. Psychol. 10(2), 229–251 (1989). https://doi.org/10.1016/0167-4870(89)90021-4
13. Understanding Regression Analysis. Springer US, Boston, MA (1997)
14. Olden, J.D., Jackson, D.A.: Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks. Ecol. Model. 154(1–2), 135–150 (2002). https://doi.org/10.1016/S0304-3800(02)00064-9
15. Devore,J.L.: Probability and Statistics for Engineering and theSciences,8th edn. Brooks/Cole, Cengage Learning, Boston, MA (2012)
16. Hassoun, M.H.: Fundamentals of Artificial Neural Networks. MIT Press, Cambridge, Mass(1995)
17. Baughman, D.R., Liu, Y.A.: Neural Networks in Bioprocessing and Chemical Engineering. Academic Press, San Diego (1995)
18. Bishop, C.M.: Neural Networks for Pattern Recognition. Clarendon Press, Oxford; Oxford University Press, New York (1995)

19. Marwala, T., SpringerLink (Online service): Artificial Intelligence Techniques for Rational Decision Making. Springer International Publishing, Cham, Springer, Imprint (2014)
20. Gavin, H.P.: The Levenberg-Marquardt method for nonlinear least squares curve-fitting problems c © (2013)
21. Olden, J.D., Joy, M.K., Death, R.G.: An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. Ecol. Model. 178(3–4), 389–397 (2004). https://doi.org/10.1016/j.ecolmodel.2004.03.013
22. Gevrey, M., Dimopoulos, I., Lek, S.: Review and comparison of methods to study the contribution of variables in artificial neural network models. Ecol. Model. 160(3), 249–264 (2003). https://doi.org/10.1016/S0304-3800(02)00257-0
23. Collopy, P.D., Hollingsworth, P.M.: Value-driven design. J. Aircr. 48(3), 749–759 (2011). https://doi.org/10.2514/1.C000311
24. Bertoni, M., Pezzotta, G., Scandella, B., Wall, J., Jonsson, P.: Life cycle simulation to support cross-disciplinary decision making in early PSS design. Procedia CIRP 83, 260–265 (2019). https://doi.org/10.1016/j.procir.2019.03.138
25. Price, M., Soban, D., Mullan, C., Butterfield, J., Murphy, A.: A novel method to enable trade-offs across the whole product life of an aircraft using value driven design. J. Aerosp. Oper. 1(4), 359–375 (2012). https://doi.org/10.3233/AOP-120028
26. Yip, P.S.L., Tsang, E.W.K.: Interpreting dummy variables and their interaction effects in strategy research. Strateg. Organ. 5(1), 13–30 (2007). https://doi.org/10.1177/147612700607 3512
27. Kurz-Kim, J.-R.,Loretan, M.:A note on the coefficient of determination in models with infinite variance variables. SSRN Electron. J. (2007). https://doi.org/10.2139/ssrn.996664

Paper C

Aeddula, O.K., Flyborg, J., Larsson, T., Anderberg P., Johan S.B., Renvert S, 2021. A Solution with Bluetooth Low Energy Technology to Support Oral HealthCare Decisions for Improving Oral Hygiene,
In: 2021 5th International Conference on Medical and Health Informatics. Presented at the ICMHI 2021: 2021 5th International Conference on Medical and Health Informatics, ACM, Kyoto Japan, pp. 134–139. https://doi.org/10.1145/3472813.3473179

A Solution with Bluetooth Low Energy Technology to Support Oral HealthCare Decisions for Improving Oral Hygiene

Aeddula Omsri Kumar[1], Flyborg Johan[1], Larsson Tobias[1], Anderberg Peter[1], Sanmartin Berglund Johan[1], Renvert Stefan[1,2]

1 Blekinge Institute of Technology, Sweden

2 Kristianstad University, Sweden

ABSTRACT

The advent of powered toothbrushes and associated mobile health applications provides an opportunity to collect and monitor the data, however collecting reliable and standardized data from large populations has been associated with efforts from the participants and researchers. Finding a way to collect data autonomously and without the need for cooperation imparts potentiality to build large knowledge banks. A solution with Bluetooth low energy technology is designed to pair a powered toothbrush with a single-core processor to collect raw data in a real-time scenario, eliminating the manual transfer of powered toothbrush data with the mobile health applications. Associating powered toothbrush with a single-core processor is believed to provide reliable and comprehensible data of toothbrush use and propensities can be a guide to improve individual exhortation and general plans on oral hygiene quantifies that can prompt improved oral wellbeing. The method makes a case for an expanded chance to plan assistant capacities to protect or improve factors that influence oral wellbeing in individuals with mild cognitive impairment. The proposed framework assists with determining various parameters, which makes it adaptable and conceivable to execute in various oral care contexts.

KEYWORDS: Dental device, Oral health, Oral hygiene, Oral health information system

ACM Reference Format:

Aeddula Omsri Kumar, Flyborg Johan, Larsson Tobias, Anderberg Peter, Berglund Johan Sanmartin, and Renvert Stefan. 2021. A Solution with Bluetooth Low Energy Technology to Support Oral HealthCare Decisions for Improving Oral Hygiene. In *2021 5th International Conference on Medical and Health Informatics (ICMHI 2021), May 14–16, 2021, Kyoto, Japan.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3472813.3473179

1 INTRODUCTION

Life expectancy has increased over the past century due to major medical research results such as the discovery of antibiotics, various vaccines, and also social policy work for a better standard of living for the population [1]. As a result, the prevalence of individuals with mild cognitive impairments (MCI) and dementia has also increased [2]. It is well documented that cognitive impairment and dementia lead to an increased incidence of oral diseases, resulting in rapid deterioration of oral health and quality of life (QoL) [3], [4]. At the same time more elderly people in Sweden, as in many other higher-income countries, retain their teeth for a longer period of life [5]. The oral hygiene management of people suffering from cognitive impairment and dementia is critical and important to maintain good oral health as an integral part of their general health [6]. Oral hygiene constitutes the primary preventive measure of oral disease, and there exists a connection between oral hygiene and QoL in older adults [7]. Decreasing oral health in this group leads to individual suffering and increasing financial and resource demands on society. Finding aids to maintain and improve good oral hygiene for people with MCI and dementia is of great importance to reduce pain and illness and maintain a high level of Quality of life [8].

Information and communication technology (ICTs) have led to many significant transformations in the health care systems, particularly mobile health applications that have improved the quality of services provided to the people [9]. Patil et al., underlined mobile health application’s effectiveness for improving oral hygiene and oral health outcomes in orthodontic patients [10]. The widespread popularity of mobile health applications for wireless devices provides an opportunity to influence people’s behavior [11]. Accordingly, Oral-B [Procter & Gamble, Cincinnati, OH, United States] has combined state-of-the-art wireless technology with clinically proven powered toothbrushes in an innovative mobile platform to increase cleaning performance, increase patient motivation, and create better daily oral hygiene habits through real-time instant feedback and tracking. The interactive Oral-B® Professional Care 6000®with Bluetooth®4.0 connectivity allows for two-way communication between the toothbrush and the mobile application to enable instant feedback around brushing force and session length [12]. The limitation of the existing model is the limited storage capacity. This entails linking the powered toothbrush with the mobile health application, frequently transferring the data between the powered toothbrush and the mobile health application. This can be difficult to carry out both motorically and cognitively for elderly and sick individuals. Lack of continuity in the data transmission leads to loss of data and impairs usability. The existing model and its associated mobile application compiles and present a few possible parameters such as brush session length and brush power limits access to obtain more information from the powered toothbrush useful for decision-making. To register individuals', use of the powered toothbrush and, constructing from that, prepare individual oral hygiene advice and instructions and follow up on the results, extended functionality is required. It shall not require active manual data transfer and must present data regarding date, time, average brushing time and number of brushings per selected period. Collecting and storing data on oral health and oral hygiene in modern information systems provides knowledge and the opportunity to develop aids and methods for improved general and oral health, resulting in increased quality of life. Health care information systems can discover the required data about oral health and provide support in the healthcare decision-making process [13].

In this paper, a solution with Bluetooth low energy technology system is proposed to support oral healthcare decisions for improving oral hygiene will be explained and discussed. The remainder of this paper is organized as follows; In part 2, the proposed system is presented. In part 3 the proposed method is exemplified, and the paper concludes with discussion, conclusions, and directions for future work.

2 MATERIAL AND METHODOLOGY

2.1 *System Architecture*

The proposed system comprises two main components: hardware, and software. Integrating both the components supports in accomplishing the data acquisition process. Hardware components are essential for providing a medium for communication, and for the data gathering process, whereas software components are essential for establishing one-way data transfer between the hardware components.

2.1.1 Hardware Component. The hardware components consist of a powered toothbrush and a single-core processor. In the study, Oral-B® Professional Care 6000®was used as a powered toothbrush and Raspberry Pi Zero W (RPiZ) was used as a single-core processor. Oral-B® Professional Care 6000®is a rechargeable powered toothbrush with Bluetooth connectivity, which allows for connection to the smartphone to store the brushing data [14]. Raspberry Pi Zero W is a single-core processor extending the family of pi with added Bluetooth connectivity [15].

Powered Toothbrush. The study was carried out with OralB® Professional Care 6000®as the powered toothbrush, however, the proposed system works for any powered toothbrush with Bluetooth connectivity feature. The employed powered toothbrush transmits or advertises data such as average brushing time, and a total time of applied pressure from the pressure sensor through Bluetooth packets. Each powered toothbrush contains a universally unique identifier (UUID) [14]. The transmitted data or the real-time data can be visualized in the developed Oral-B mobile application, however, each version of the mobile application differs in the amount of visualized data.

Single-core Processor. Raspberry Pi Zero W is a single-core processor, popularly known as a minicomputer extending the original Pi family with added functionalities such as Bluetooth connectivity, and wireless LAN. It requires a 1.2A power supply and can be supplied using a micro-USB charging case and consists of Bluetooth low energy technology and the other functionalities as the original raspberry pi. The single-core processor needs to flash write with a separate raspberry pi operating system and the system works on Linux operating environment [15]. A python programming in a Linux environment with the library modules supports a one-way data transmission process in a real-time scenario.

2.1.2 Software Component. The software component contains four main sub-components namely Bluetooth low energy, generic attribute profile (GATT), bluepy, and UUID in the testing process. As mentioned above, the main functionality of this component is to establish a connection between the hardware components. The data transmission mode for the system is initialized using the Bluetooth low energy connectivity. While Bluetooth low energy devices essentially communicate on a generic attribute profile (GATT). A python library module bluepy supports the process of initiating the GATT communication.

Bluetooth Low Energy. Bluetooth low energy is a low power wireless communication technology for control and monitoring applications, that can be used for shorter distance communications enabling smart devices to receive or transmit information [16]. RPiZ communicates with a powered toothbrush through Bluetooth technology.

Generic Attribute Profile. A GATT communication profile defines the way a powered toothbrush transmits or advertise data packets to RPiZ. GATT works under the concept of server and client relationship, Information uncovered by a server is introduced in a GATT profile which is a progressive design of traits permitting the exchange of data among server and client [16]. where RPiZ acts as the GATT server, which holds the attribute protocol lookup data and power toothbrush acts as GATT client, which sends data to the server.

Bluepy: Bluepy is the python module for establishing communications with Bluetooth low energy devices built on raspberry pi, internally working on the BlueZ library. BlueZ is an official Linux Bluetooth stack that provides support for core Bluetooth layers and protocols. The module consists of a peripheral, scanner, scan entry, default delegate, UUID, service, and characteristics classes for specific communication functionalities [17]. A peripheral class encapsulates a connection to Bluetooth low energy peripheral, once the connection is established services and characteristics offered by powered toothbrush can be discovered and read. Scanner object scans the within-range available Bluetooth devices. The information obtained from scanner object Bluetooth devices can be obtained from the Scan entry class. UUID class is an object value given by the user to connect to the powered toothbrush. Service class represents a collection of characteristics and descriptors which are all related to one particular function of the peripheral. Bluetooth LE characteristic class represents a short data item that can be read or written [17].

UUID. A 128-bit UUID supports custom services and characteristics. As mentioned earlier, the UUID class is an object value given by the client to interface with the force toothbrush. “BLESCAN (Android)” or “LightBlue (IOS)” mobile application can be used to determine the UUID’s of the powered toothbrush.

2.2 *Research Approach*

The research presented in this paper can be framed as one of the activities conducted during the Descriptive Study, I phase of the Design Research Methodology (DRM) proposed by Blessing and Chakrabarti [18]. The methodology was based on the work [19, p. 6] focusing on the data transmission between Bluetooth low energy devices. In this respect, the usage of RPiZ was identified as a highly suitable method to determine various parameters, extending the opportunity to plan assistant abilities to protect or improve factors.

2.3 *Testing Process*

The participants are recruited from a sample of 170 individuals who participate in an ongoing study aimed to investigate the effect of introducing a powered toothbrush in people with MIC. Participants have undergone the Mini-Mental State Examination (MMSE) and have values between 20 and 28 [20]. They have also suffered from memory difficulties in the past six months. The participants were instructed in handling the powered toothbrush and recommended to use the brush for at least two minutes every morning and evening, with the system installed in their homes for a period of four to six months. Figure 1 shows the proposed system of communication where the brush data is transmitted to the RPiZ controller in a real-time scenario. The entire setup is executed in a python programming environment with suitable library modules. The Python module “bluepy” supports the process of initiating the GATT communication for data

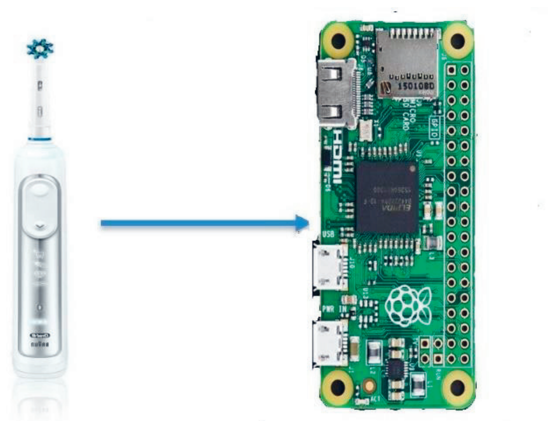


Figure 1: The proposed system.

transmission. RPiZ is paired up with the specific power toothbrush using the UUID value or the raspberry pi zero scans locally for the powered toothbrush which advertises Bluetooth packets. Once the connection is established, the data from the powered toothbrush can be stored locally in the raspberry pi or network-based storage either as a text file or as a spreadsheet. The system is pre-installed with a virtual networking computing server to virtually operate the system at the participant's home and update the UUID of the powered toothbrush.

3 RESULTS

The proposed system is exemplified by installing the designed systems in the homes of the above-mentioned participants. Virtual network computing server assisted in associating the powered toothbrush with RPiZ using UUID. Associating the RPiZ with the powered toothbrush eliminates the problem of noisy data, where data from other powered toothbrushes here is referred to as noisy data. Figure 2 shows two instances of the proposed system installation in the participant's homes. The left part of Figure 2 shows raspberry pi zero w in a protective case, and the right part of Figure 2 shows the RPiZ connected to a power supply.

Figure 3 and Figure 4 show the text and a participant's spreadsheet format data, respectively. Text files represent raw data collected from the raspberry pi zero controller and the spreadsheet data is a processed data synthesized from raw data. The spreadsheet contains different estimated parameters as columns. The raw data from the text file was analyzed and transformed into meaningful information. The participant ID was a unique number assigned to each participant, and each assigned system name was replaced with the participant ID respectively. Python library module "DateTime" imports the date and time into

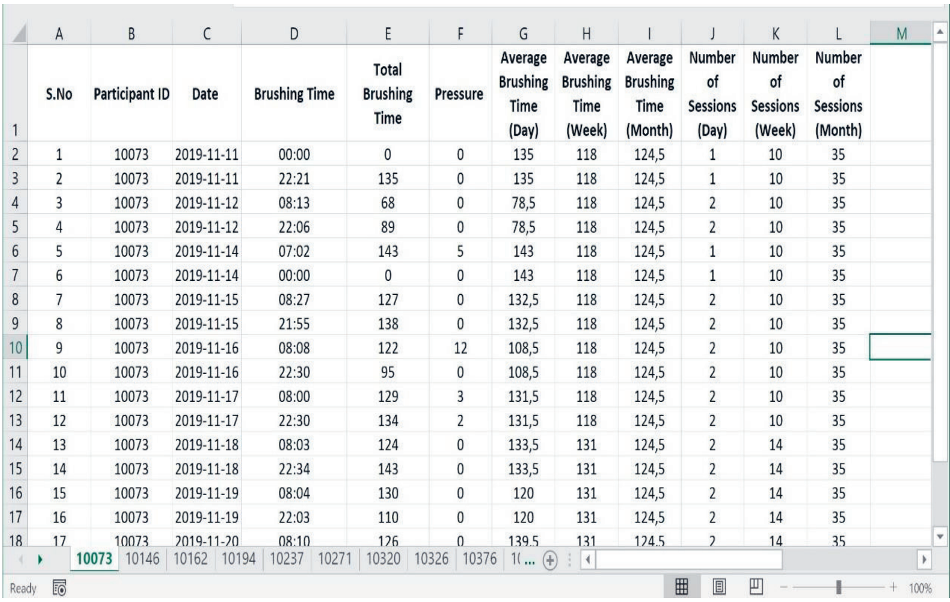
Figure 2: Installation of the system in the participant's home.

the text file according to the brushing session, and the module “date finder” extracts the date and time from the text file [16]. These modules assist in extracting the date and time information. Extracting the digits between the strings helps to identify the total brushing time and similarly, the sensor data from the text was analyzed to determine the total applied pressure time per brushing session. These input parameters from the spreadsheet could be further used to determine other useful parameters such as average brushing time per day, week, month, and a total number of brushing sessions per day, week, month. Average brushing time per day was calculated based on an average of two rows in the spreadsheet and estimating the week number averaged the brushing time per week and similarly the average brushing time per month. The same logic applied to calculate the total number of sessions per day, week, and month. Different parameters could be visualized as useful for decision-making processes as shown in figure 5 and figure 6. Figure 5 shows the visualization of the total brushing time of each session in seconds corresponding to participant ID 10073 and figure 6 shows the graph of average brushing time per day, week, and month with blue bar represents the average brushing time per day, orange and grey represents the average brushing time per week and per month respectively. Only one participant data with few days data was shown in the below figures for better


data-10073 - Notepad
File Edit Format View Help
Date 2019-11-11 22:21:10.326789 Recorded time is 1 Secs
Date 2019-11-11 22:21:11.739939 Recorded time is 2 Secs
Date 2019-11-11 22:21:12.422798 Recorded time is 3 Secs
Date 2019-11-11 22:21:13.348766 Recorded time is 4 Secs
Date 2019-11-11 22:21:14.323982 Recorded time is 5 Secs
Date 2019-11-11 22:21:15.494058 Recorded time is 6 Secs
Date 2019-11-11 22:21:16.371408 Recorded time is 7 Secs
Date 2019-11-11 22:21:17.493001 Recorded time is 8 Secs
Date 2019-11-11 22:21:18.321510 Recorded time is 9 Secs
Date 2019-11-11 22:21:19.345325 Recorded time is 10 Secs
Date 2019-11-11 22:21:20.369093 Recorded time is 11 Secs
Date 2019-11-11 22:21:21.441675 Recorded time is 12 Secs
Date 2019-11-11 22:21:22.417302 Recorded time is 13 Secs
Date 2019-11-11 22:21:23.343127 Recorded time is 14 Secs

Figure 3: Text file data of a participant (raw data).

understanding. These visualizations support knowledge about factors leading to good oral hygiene, help develop aids and improve oral and general health.



S.No	Participant ID	Date	Brushing Time	Total Brushing Time	Pressure	Average Brushing Time (Day)	Average Brushing Time (Week)	Average Brushing Time (Month)	Number of Sessions (Day)	Number of Sessions (Week)	Number of Sessions (Month)
1	10073	2019-11-11	00:00	0	0	135	118	124,5	1	10	35
2	10073	2019-11-11	22:21	135	0	135	118	124,5	1	10	35
3	10073	2019-11-12	08:13	68	0	78,5	118	124,5	2	10	35
4	10073	2019-11-12	22:06	89	0	78,5	118	124,5	2	10	35
5	10073	2019-11-14	07:02	143	5	143	118	124,5	1	10	35
6	10073	2019-11-14	00:00	0	0	143	118	124,5	1	10	35
7	10073	2019-11-15	08:27	127	0	132,5	118	124,5	2	10	35
8	10073	2019-11-15	21:55	138	0	132,5	118	124,5	2	10	35
9	10073	2019-11-16	08:08	122	12	108,5	118	124,5	2	10	35
10	10073	2019-11-16	22:30	95	0	108,5	118	124,5	2	10	35
11	10073	2019-11-17	08:00	129	3	131,5	118	124,5	2	10	35
12	10073	2019-11-17	22:30	134	2	131,5	118	124,5	2	10	35
13	10073	2019-11-18	08:03	124	0	133,5	131	124,5	2	14	35
14	10073	2019-11-18	22:34	143	0	133,5	131	124,5	2	14	35
15	10073	2019-11-19	08:04	130	0	120	131	124,5	2	14	35
16	10073	2019-11-19	22:03	110	0	120	131	124,5	2	14	35
17	10073	2019-11-20	08:10	126	0	139,5	131	124,5	2	14	35

Figure 4: Data visualization of total brushing time.

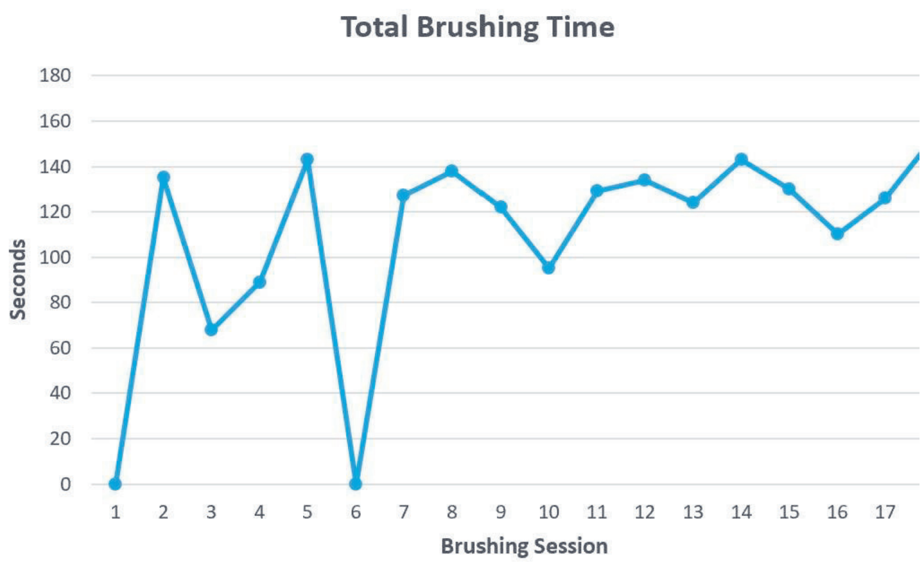


Figure 5: Spreadsheet file data of a participant.

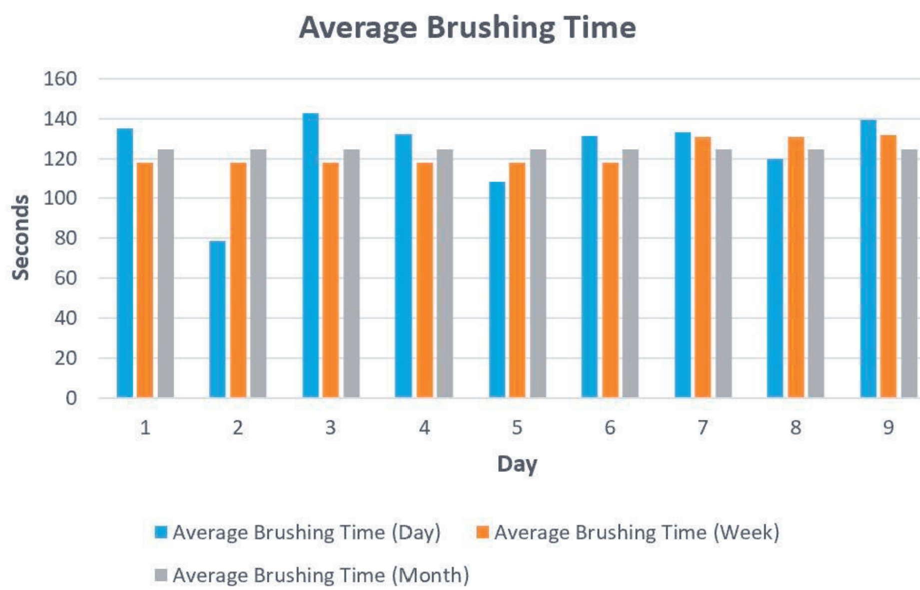


Figure 6: Data visualization of average brushing time.

4 DISCUSSION AND CONCLUSION

The proposed system was designed to collect data from the powered toothbrush. The purpose was to create a base for increased opportunity to design auxiliary functions to preserve or improve factors that affect oral health in people with MCI. The proposed system was trained using a python environment involving library modules to collect the associated toothbrush data by establishing a GATT communication through Bluetooth connectivity between the powered toothbrush and single-core processor RPiZ. The methodology focuses on receiving and decoding the advertised Bluetooth packets to more meaningful information that is useful for the decision-making process to improve oral hygiene. The proposed system helps determine different parameters, making it flexible and possible to implement in different oral care contexts. The proposed method receives the advertised Bluetooth packets in real-time signifying that the powered toothbrush can be designed without having any memory slot. This provides the cost efficiency and simultaneously ensures the data transfer. Associating the UUID of the powered toothbrush ensures that obtained data would be noise-free. In conclusion, the ability to access reliable and comprehensive data on toothbrush use and habits can help improve individual advice and general plans on oral hygiene measures that can lead to improved oral health. Future works include integrating the features of the system with the powered toothbrush as a single usage product and providing a user interface for the proposed system to support users.

ACKNOWLEDGMENTS

The research leading to these results has received financial support by the Swedish Knowledge and Competence Development Foundation (Stiftelsen för kunskaps- och kompetensutveckling) through the Model Driven Development and Decision Support research profile at Blekinge Institute of Technology.

REFERENCES

[1] E. M. Crimmins, 'Lifespan and Healthspan: Past, Present, and Promise', Gerontologist, vol. 55, no. 6, pp. 901–911, Dec. 2015, doi: 10.1093/geront/gnv130.
[2] M. Prince, R. Bryce, E. Albanese, A. Wimo, W. Ribeiro, and C. P. Ferri, 'The global prevalence of dementia: A systematic review and metaanalysis', Alzheimer's & Dementia, vol. 9, no. 1, pp. 63-75.e2, Jan. 2013, doi: 10.1016/j.jalz.2012.11.007.
[3] S. Delwel, T. T. Binnekade, R. S. G. M. Perez, C. M. P. M. Hertogh, E. J. A. Scherder, and F. Lobbezoo, 'Oral health and orofacial pain in older people with dementia: a systematic review with focus on dental hard tissues', Clin Oral Investig, vol. 21, no. 1, pp. 17–32, Jan. 2017, doi: 10.1007/s00784-016-1934-9.
[4] L. J. M. van de Rijt *et al.*, 'The Influence of Oral Health Factors on the Quality of Life in Older People: A Systematic Review', Gerontologist, vol. 60, no. 5, pp. e378–e394, Jul. 2020, doi: 10.1093/geront/gnz105.
[5] F. Schwendicke, J. Krois, T. Kocher, T. Hoffmann, W. Micheelis, and R. A. Jordan, 'More teeth in more elderly: Periodontal treatment needs in Germany 1997–2030', J Clin Periodontol, vol. 45, no. 12, pp. 1400–1407, Dec. 2018, doi:10.1111/jcpe.13020.
[6] S. S. Gao, C. H. Chu, and F. Y. F. Young, 'Oral Health and Care for Elderly People with Alzheimer's Disease', Int J Environ Res Public Health, vol. 17, no. 16, Aug. 2020, doi: 10.3390/ijerph17165713.
[7] T. Willumsen, B. Fjaera, and H. Eide, 'Oral health-related quality of life in patients receiving home-care nursing: associations with aspects of dental status and xerostomia: Oral quality of life, teeth and saliva', Gerodontology, vol. 27, no. 4, pp. 251–257, Dec. 2010, doi: 10.1111/j.1741-2358.2009.00344.x.
[8] T. de S. Rolim *et al.*, 'Evaluation of patients with Alzheimer's disease before and after dental treatment', Arq Neuropsiquiatr, vol. 72, no. 12, pp. 919–924, Dec. 2014, doi: 10.1590/0004-282X20140140.

[9] J. A. Powell, M. Darvell, and J. A. M. Gray, 'The Doctor, The Patient and the World-Wide Web: How the Internet is Changing Healthcare', J R Soc Med, vol. 96, no. 2, pp. 74–76, Feb. 2003, doi: 10.1177/014107680309600206.
[10] S. Patil *et al*., 'Effectiveness of mobile phone applications in improving oral hygiene care and outcomes in orthodontic patients', J Oral Biol Craniofac Res, vol. 11, no. 1, pp. 26–32, Mar. 2021, doi: 10.1016/j.jobcr.2020.11.004.
[11] B. Tiffany, P. Blasi, S. L. Catz, and J. B. McClure, 'Mobile Apps for Oral Health Promotion: Content Review and Heuristic Usability Analysis', JMIR Mhealth Uhealth, vol. 6, no. 9, p. e11432, Sep. 2018, doi: 10.2196/11432.
[12] C. Erbe *et al.*, 'A comparative assessment of plaque removal and toothbrushing compliance between a manual and an interactive power toothbrush among adolescents: a single-center, single-blind randomized controlled trial', BMC Oral Health, vol. 18, no. 1, p. 130, Aug. 2018, doi: 10.1186/s12903-018-0588-1.
[13] P. E. Petersen, D. Bourgeois, D. Bratthall, and H. Ogawa, 'Oral health information systems–towards measuring progress in oral health promotion and disease prevention', Bull World Health Organ, vol. 83, no. 9, pp. 686–693, Sep. 2005, doi: /S0042-96862005000900014.
[14] 'Oral-B Genius 6000 Rechargeable Electric Toothbrush I Oral-B'. https://oralb.com/en-us/oral-b-genius-6000-rechargeable-electric-toothbrush/
[15] 'Buy a Raspberry Pi Zero W – Raspberry Pi'. https://www.raspberrypi.org/products/raspberry-pi-zero-w/ .
[16] S. I. G. Bluetooth, 'Specification of the Bluetooth System-Covered Core Package version: 4.0', Bluetooth Special Interest Group, 2010.
[17] 'bluepy - a Bluetooth LE interface for Python — bluepy 0.9.11 documentation'. http://ianharvey.github.io/bluepy-doc/ .
[18] L. T. M. Blessing, A. Chakrabarti, and L. T. M. Blessing, DRM, a design research methodology. Dordrecht; London: Springer, 2009.
[19] J. Campos, S. Colteryahn, and K. Gagneja, 'IPv6 transmission over BLE Using Raspberry PI 3', in 2018 International Conference on Computing, Networking and Communications (ICNC), Maui, HI, Mar. 2018, pp. 200–204, doi: 10.1109/ICCNC.2018.8390350.
[20] M. F. Folstein, S. E. Folstein, and P. R. McHugh, "'Mini-mental state"', Journal of Psychiatric Research, vol. 12, no. 3, pp. 189–198, Nov. 1975, doi: 10.1016/00223956(75)90026-6.

ABSTRACT

Modern product development is a complex chain of events and decisions. The ongoing digital transformation of society, increasing demands in innovative solutions puts pressure on organizations to maintain, or increase competitiveness. As a consequence, a major challenge in the product development is the search for information, analysis, and the build of knowledge. This is even more challenging when the design element comprises complex structural hierarchy and limited data generation capabilities. This challenge is even more pronounced in the conceptual stage of product development where information is scarce, vague, and potentially conflicting. The ability to conduct exploration of high-level useful information using a machine learning approach in the conceptual design stage would hence enhance be of importance to support the design decision-makers, where the decisions made at this stage impact the success of overall product development process.

The thesis aims to investigate the conceptual stage of product development, proposing methods and tools in order to support the decision-making process by the building of data-driven decision support systems. The study highlights how the data can be utilized and visualized to extract useful information in design exploration studies at the conceptual stage of product development. The ability to build data-driven decision support systems in the early phases facilitates more informed decisions.

The thesis presents initial descriptive study findings from the empirical studies, showing the capabilities of the machine learning approaches in extracting useful information, and building data-driven decision support systems. The thesis initially describes how the linear regression model and artificial neural networks extract useful information in design exploration, providing support for the decision-makers to understand the consequences of the design choices through cause-and-effect relationships on a detailed level. Furthermore, the presented approach also provides input to a novel visualization construct intended to enhance comprehensibility within cross-functional design teams. The thesis further studies how the data can be augmented and analyzed to extract the necessary information from an existing design element to support the decision-making process in an oral healthcare context.

ISSN: 1650-2140

ISBN: 978-91-7295-433-5

2021:10