INTERNATIONAL DESIGN CONFERENCE – DESIGN 2020
https://doi.org/10.1017/dsd.2020.4

DATA-DRIVEN DESIGN IN CONCEPT DEVELOPMENT: SYSTEMATIC REVIEW AND MISSED OPPORTUNITIES

A. Bertoni ✉

Blekinge Institute of Technology, Sweden
✉ alessandro.bertoni@bth.se

Abstract

The paper presents a systematic literature review investigating definitions, uses, and application of data-driven design in the concept development process. The analysis shows a predominance of the use of text mining techniques on social media and online reviews to identify customers' needs, not exploiting the opportunity granted by the increased accessibility of IoT in cyber-physical systems. The paper argues that such a gap limits the potential of capturing tacit customers' needs and highlights the need to proactively plan and design for a transition toward data-driven design.

Keywords: data-driven design, conceptual design, engineering design, data science, data mining

1. Introduction

The increasing computational capabilities and the sensible reduction of the technological barriers for data collection and analysis have been two of the main enablers for the digitalization process that has highly affected ways of life and industrial businesses in recent years. The potential of data science in the design and development of new products and systems is increasingly acknowledged by practitioners and researchers (Kim et al., 2017). In this context, the term Data-Driven Design (DDD) is often associated with the use of any kind of data science algorithm supporting in some way a specific phase of the product development process. While the use of data science algorithms to create surrogate models is an established practice in detailed design since several years (see for Jeong et al., 2005; Zhao et al., 2007; Akram et al., 2010; Huang et al., 2011), the same cannot be said for the other stages of the product development process. This paper presents a systematic review of the scientific publications proposing the use of DDD methods in the 'concept development' stage of the product development process as described by Ulrich and Eppinger (1995). The research question addressed in the paper is: "How are Data-Driven Design methods currently used in the concept development stage, and what are the major data sources and data science algorithms applied in this context?" To answer this question, research publications have been analyzed with particular attention to the source of the data and the data science algorithm proposed, in order to provide a complete picture of the research contributions in DDD, highlighting established areas of contributions and emerging trends. Based on the results of the analysis, emerging research opportunities for DDD in concept development have been described, together with the proposal a narrower definition of what should be considered a DDD method.
The paper begins by describing the method applied for the systematic literature review, concerning a first screening of research contributions applying DDD methods in product development and the consequent categorization to identify those applied in conceptual design. Section 3 presents a

summary of the papers identified in the review categorizing them into the different stages of the concept development process. Section 4 discusses the status and trends for DDD in conceptual design and section 5 draws the final conclusions.

2. Research approach

This work is based on a systematic literature review in the field of DDD, based on the procedures described by Kitchenham et al. (2009). This was followed by a process of data analysis based on results categorization and trend analysis. The systematic review was conducted first identifying the contributions in the whole product development process, to focus on a second step on selecting those related to the conceptual design stage. The SCOPUS database was selected as being one of the widest and most complete databases of peer-review scientific literature. The following research string was used in the search engine: TITLE-ABS-KEY ("data driven" OR "data intense" OR "big data" OR "machine learning" OR "data mining" AND "product development" OR "engineering design" OR "service development" OR "concept design" OR "design space exploration" OR "fuzzy front end" OR "preliminary design" OR "concept development") AND (LIMIT-TO (SUBJAREA, "ENGI")) AND (LIMIT-TO (DOCTYPE, "cp") OR LIMIT-TO (DOCTYPE, "ar")) AND (LIMIT-TO (LANGUAGE, "English"). The string was defined with the intent to include a large number of synonyms identified by investigating the more recurrent keywords related to data science and engineering design. In order to confine the research to a specific area of contribution and avoid an excessively heterogeneous set of data to analyze, the search was deliberately limited to conference papers and journal articles written in English and in the subject areas of "Engineering". The sample, therefore, excluded all the publications in the field of data science, business management, and marketing which did not have any explicit connection to the engineering topic. An exception to the last rules was made by including the proceeding of the International Design Conference to the dataset. The choice was made based on the author's consideration of the conference to be highly relevant in the engineering design field despite being indexed in the subject area "Mathematics" in the search engine SCOPUS. The first round of search in the database was run at the end of August 2019 and rendered 591 articles, of which 588 from the search string and 3 from the International Design Conference. The 591 articles were screened by reading titles and abstracts and those outside the scope of the research were excluded. To be qualified as a relevant paper each article needed to deal with a topic in the engineering design field and to feature an application, or a proposed application, of a data science algorithm. Double hits happened on two occasions and duplicates were eliminated. The screening of relevant articles rendered a selection of 116 papers, which was expanded to 127 after the recommendation received during the review process.

2.1. Categorization and mapping of the selected articles

The analysis of the resulting 127 articles happened in two stages. In the first stage, each full article was read and the content briefly summarized in one or two sentences containing the main objectives of the work. Each article was categorized on the basis of the stage of the product development process in which it was contributing. The product development process as proposed by Ulrich and Eppinger (1995) was adopted as a reference, thus the article was categorized either as "planning", "concept development", "system-level design", "detailed design", "production and ramp-up". An additional category named "overall process or framework" was added to address those papers not specifically referring to a defined stage, rather taking a more high-level perspective on DDD approaches. Additionally, the "concept development" phase was expanded by adding the following subcategories "identify customers' needs", "establish target specification", "generate product concept", "select concept", "test concept", "set final specification", "plan downstream development", "early design space exploration". The latter subcategory, not present in the Ulrich and Eppinger's model, was added to reflect a trend in recent articles to run early design space exploration in concept development. The second stage of the analysis consisted of repeating the reading of the full text of the articles to verify the correctness of the initial categorization. Additionally, each paper was mapped on a table indicating the source of the data used in the DDD application and the method used for the data analysis.

https://doi.org/10.1017/dsd.2020.4 Published online by Cambridge University Press

3. Results from the review

This section provides an overview of the categorization of the 127 publications obtained as a result of the systematic review process. Figure 1 shows the number of publications falling in each category. Concept development proved to be the most discussed product development phase featuring 52 papers. The second most discussed category, 29 papers, concerned the development of overall processes and framework for DDD, but not proposing any specific application of data collection and analysis.

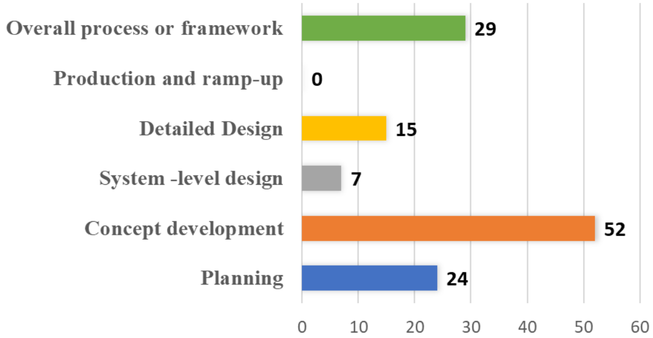

Figure 1. Focus on the product development process of the 127 research contributions identified in the systematic review

The planning stage and the detailed design stage featured respectively 24 and 15 papers, while system-level design only 7. It has to be noted that no paper was categorized as production and rump-up.
Figure 2 presents the number of publications for each category adding a time perspective. The term Data-Driven Design (DDD) was first coined in 1995 by Domazet et al. (1995) in their work defining event-condition-action rules to create dynamic product models in databases. Five years later the academic discussion about the role of data science in engineering design started with the work by Kusiak and Tseng (2000) describing the first prototyped application of a data mining algorithm to generate, select and validate a solution. Subsequent works by Agard and Kusiak (2004) and by Kusiak and Smith (2007) proposed respectively a data mining methodology for the analysis of functional requirements in product families and an overall analysis of the potential areas of investigation for data mining in product and production development. Concurrently, Menon et al. (2004, 2005) described the potential of using text mining on databases for product development. However, despite this early interest on DDD and some scattered contribution along the years (e.g. Rönher et al., 2010; Cheung et al., 2011), the research focus on DDD has radically increasing in the last four year as evident in Figure 2 (it has to be noted that the figure shows partial data for the year 2019). It is of particular interest to highlight the increased efforts in developing processes and framework for DDD in the last years (the green area in Figure 2), while the research focus on DDD for system-level design and detailed design shows to be quite stable in the last ten years.

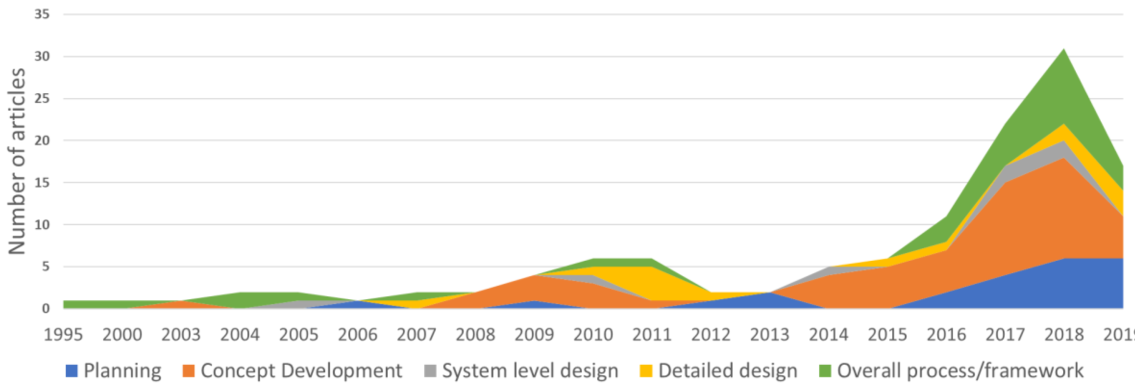

Figure 2. Distribution of the selected papers based on the area of contribution and year of publication (data up to the end of August 2019)

https://doi.org/10.1017/dsd.2020.4 Published online by Cambridge University Press

The publications focusing on concept development shows a constantly increasing trend since 2014. DDD for product planning shows occasional publications along the years with an increase to 6 publications in 2018 and 6 in 2019 (partial data).
A further level of analysis concerned the understanding of which type of data have been used to build DDD models. Each paper was analyzed to define the source of data on the basis of the data analysis. The analysis presented in this paper focuses on the concept development stage and the following subsection narrows the analysis considering the different substage of concept development.

3.1. Data-driven design for concept development

The concept development is the phase in which the needs and the target markets are identified, different product concepts are generated and assessed, and decisions are made to promote one or more concepts to further development (Ulrich and Eppinger, 1995). As shown in Figure 1, the concept development stage is the phase in which most of the research contributions on the DDD topic were found. For each contribution, a sub-category has been assigned as described in section 2.1. A summary of such division is shown in Figure 3, highlighting that the majority of contributions concerned the identification of customers' needs. The main contributions to costumers' needs identification are here described in section 3.1.1, while the main contributions in the other sub-categories are described in section 3.1.2.

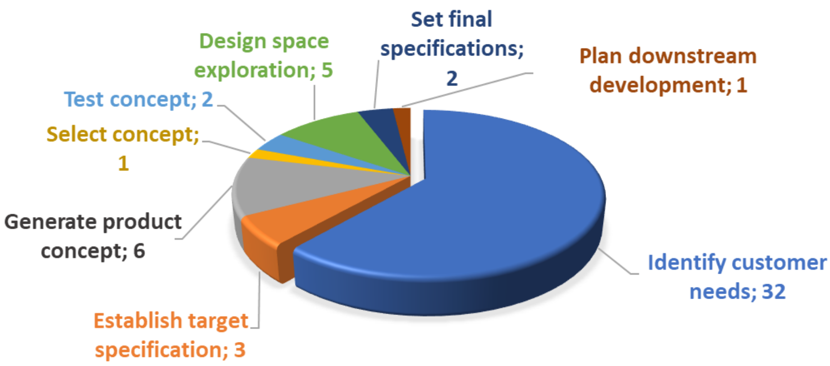

Figure 3. Number of applications of DDD in the different sub-phases of concept development

3.1.1. Data-driven design to identify customers' needs

The identification of customers' needs is the focus of 32 papers identified during the systematic review spanning from 2008 to 2019. The analysis revealed a predominant use of text mining techniques (19 over a total of 32) applied of social media and online reviews (17 over a total of 32) with very little focus on historical data on real data collected by sensors. This trend becomes even more evident focusing the analysis on the use of specific data source and analysis methods by unique authors, that means, counting multiple contributions of a single main author using the same type of analysis on the same type of data source as a unique contribution. This analysis, although not giving credits to authors publishing multiple works in the same field, gives a picture of the diffusion and popularity of the different approaches in the research community. Figure 4 shows, on the left-hand side, a summary of the source of data used for DDD application, and, on the right-hand side, which methods have been used for the analysis. Clear trends are also present analyzing the literature from a time perspective. The use of A-priori algorithms is central in the first contributions exploring the use of data science for needs identification. Professor Shu-Hsien Liao, together with a number of co-authors, published between 2008 and 2010 five papers in which A-priori algorithms were used to investigate customers' needs based on customers' purchase data in a variety of situations (e.g Liao et al., 2008a, 2008b). The same approach for analysis was later used by Zhang et al. (2010) to mine historical design data and by Bae and Kim (2011) on the design of a digital camera, combining the results with the application of a C5.0 algorithm creating decision trees based on customers' surveys. From 2014 researchers increasingly focused on extracting needs from online reviews and social media. The application of text mining on social media data has been initially proposed to automatically identify innovative customers (Taurob and Tucker 2014) as well as to quickly identify lead users (e.g. Pajo et al., 2014, 2015).

https://doi.org/10.1017/dsd.2020.4 Published online by Cambridge University Press

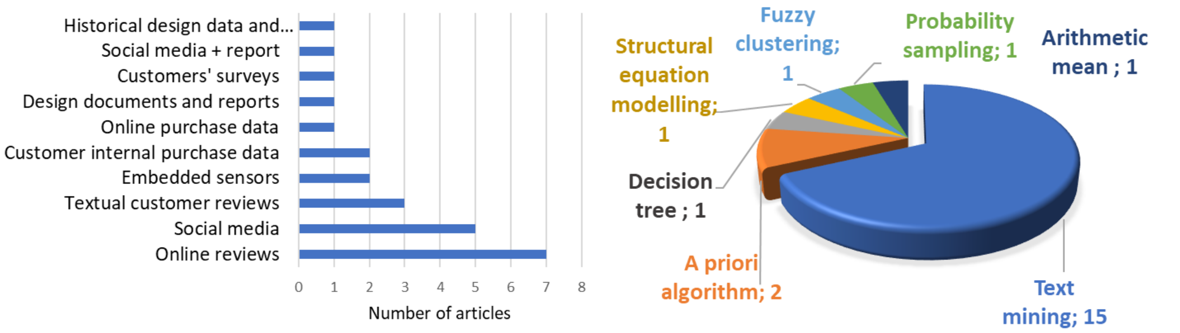

Figure 4. Sources of data (on the left side) and method of analysis (on the right side) for DDD application for needs identification applying the concept of unique contribution

Online reviews were first used by Wu et al. (2014) to analyze product usability, and increasingly became the data source of customers' needs in the most recent years (e.g. Zhang et al., 2019). Building on the increasing diffusion of cyber-physical systems, Ghosh et al. (2017) applied a framework for cyber-emphatic design, collecting through embedded sensors the data of the interaction between the product and the users to later analyze them through structural equation modeling. A second example of using usage data collected through sensors comes from Lützenberger et al. (2016) focusing on better understanding how customers use washing machines in order to improve the design requirements for the next generation of designs.

3.1.2. Data-driven design in the other stages of concept development

In the other stages of concept development, including establishment of target specification, generate of product concepts, select concepts, concept testing, set final specification, plan downstream development, and early design space exploration, the source of data considered, and the type of method applied for data analysis changes sensibly compared to needs identification. Concepts descriptions and the results from early simulations are the main sources of data, while a lot of different techniques for data analysis are employed, with the classification through decision trees being the most frequently used (see Figure 5).

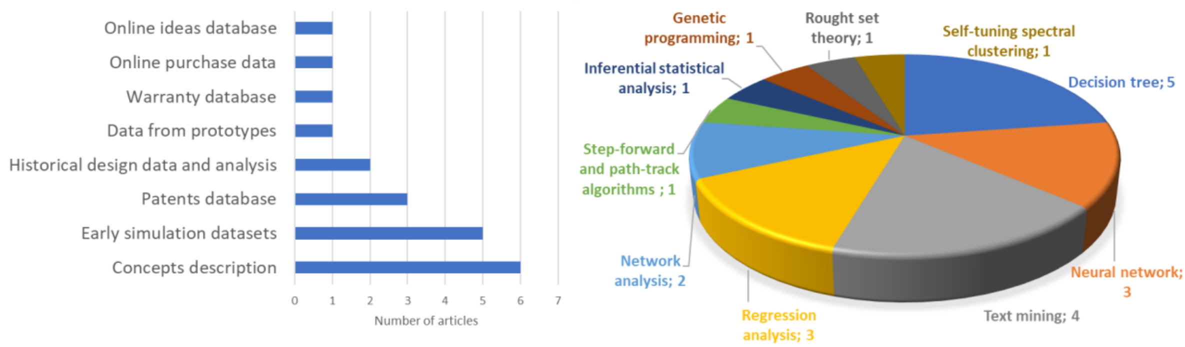

Figure 5. Sum of the type of source of data (on the left side) and method of analysis (on the right side) for DDD application in concept development (excluding needs identification)

Keivanpour and Ait Kadi (2018) and Li and Wang (2010) used historical design data to established target specifications thought respectively neural network and rough set theory and decision tree. A different approach was taken by Chowdhery and Bertoni (2018) mining the second-hand online purchase data to set targets for the new machine to be designed. The descriptions of previous product concepts is a frequently used source of data for the generation and selection of new concepts, either through applying mathematical modeling (Zhang et al., 2017), step-forward and path-track (Chen et al., 2019), or text mining (Georgiou et al., 2016). The latter method of analysis is also used by Wodehouse et al. (2018) on a database of patents. Similarly, Song et al. (2017) and Venkataraman et al. (2017) applied network analysis on patens databases to be used as stimuli on concept ideation. Finally, Ahmed

https://doi.org/10.1017/dsd.2020.4 Published online by Cambridge University Press

and Fuge (2018) also applied bisociative information networks on an online database of ideas to foster concept generation. To test new concepts Vale and Shea (2003) applied inferential statistical analysis to the description of concepts, while Garg and Tai (2014) collected data from prototypes and analyzed them through genetic programming and artificial neural network. Early simulations have been run to collect data for a preliminary design space exploration using decision trees to build surrogate models for concept development (Tucker and Kim, 2009; Bertoni et al., 2018; Du and Zhu, 2018; Bang and Selva, 2016). Alternatively, such data have been also analyzed by self-tuning spectral clustering by Morris and Seepersad (2018). Finally, to set the final concept specifications Alkahtani et al. (2019) applied text mining and artificial neural network to warranty database, while Georgiou et al. (2016) applied regression analysis from data derived from the descriptions of the concepts.

4. Data analysis and discussion

In their editorial on the Journal of Mechanical Design, Kim et al. (2017) have described DDD as an opportunity granted by the increased diffusion of cyber-physical systems and concurrent reduction of technological barriers for data collection and communication. The application of data science in cyber-physical systems in not new in the engineering field. Smart devices and sensors have been successfully deployed in a variety of contexts: from predicting failures, to running preventive maintenance, to providing driving assistance, and performing fleet management (see for instance Murthy et al., 2002; Tango and Botta, 2013). Such applications consist of introducing data collection and monitoring capabilities in existing systems to optimize the system performances or output.

However, the analysis of the articles identified in this review shows that the integration of data science as a discipline supporting the development of new concepts has not reached the same level of maturity. The analysis of the literature presented in section 3 shows a clear predominance of publications integrating data science in the needs' identification phase of concept development. Among those, the majority make use of text mining on data collected from social media and online reviews, apparently in contrast with the argumentations positioning DDD as a response to improved access to IoT in cyber-physical systems. The predominance of analysis based on social networks and online customers' reviews can be justified by the relatively easiness to access the data: those are often freely created by the customers thus their analysis, mostly based on text mining, marginally impacts the costs for a company developing products. However, literature on needs identification (e.g. Bergström et al., 2008) shows that using an available dataset that has already been elaborated by the customers does not allow to elicit "tacit" needs, not directly perceived as relevant by the customers. Collecting and analyzing data from the product, rather than from the customers' perception of its use, could grant the discovery of unexpected behaviors, allowing the identification of needs of which the customer is unaware (such as proposed by from Lützenberger et al., 2016, and Ghosh et al., 2017). However, the regular collection of data from a product or a system, either in real-time or at defined intervals, requires a proactive initiative from the manufacturer, facing setup costs, maintenance cost, and cost for dealing with customers' privacy. In other words, it is necessary to proactively plan and design a system to allow DDD implementation. The analysis presented shows that such applications are still far to be a consistent reality for collecting needs in concept development. As highlighted by Parraguez and Maier (2017) and Bertoni et al. (2017), there is a need to make visible and evident the value that such integration could have in the product development process.

Such a situation is not only limited to needs identification. As shown in Figure 5, in the other sub-phases, where the contact with the customers is not present, internal concept descriptions and results from early simulations are often the major sources of data. Such data are internally produced by the company, but not for the specific goal of using them in DDD models. In other words, the large majority of the applications of DDD in these contexts are based on data, and documents, that would be produced anyway. Exceptions to this can be found in the increased focus (5 papers between 2016-2018) on creating early design simulations for multidisciplinary value investigation, rather than as optimization method, and in the systematic investigation of patents database to support concept ideation.

From the perspective of engineering practices, although not claiming to have an evident direct impact, this literature review highlights the need for engineering practitioners to have access to a structured approach to proactively plan for the implementation of DDD methods to make the best use of data and

https://doi.org/10.1017/dsd.2020.4 Published online by Cambridge University Press

information communication technology in concept development. This is in agreement with the increasing number of publications focusing on the development of DDD processes and framework as shown in Figure 2 (e.g. Arnarsson et al., 2017).
From the perspective of contributing to the product development theory, the analysis raises the need for a clearer definition of what shall be considered as a DDD approach. This implies both the consideration of the nature and origin of the set of data used for the analysis, and of the type of analysis applied to them, either based on data science or on traditional statistics. About the latter, the analysis shows that the border between the two fields is often difficult to define in practice, with data science commonly referring to multi-disciplinary applications to support decision making and statistics referring to the use of quantifiable models to represent data. The author argues that the novelty and originality of DDD rely on more than the statistical representation in quantifiable models of a set of data, but implies the consideration of the multi-disciplinary perspective of decision making in engineering design. In practice, the development of DDD models should account for decision-making dynamics including the need for data visualization, exploration, and trade-off. Furthermore, the increasing development of cyber-physical systems raises the need to consider the unexploited opportunities of DDD approaches to support design decisions through the constant acquisition and analysis of data from the usage phase of a product, enabled by the integration of information communication technologies. The review showed a limited number of tentative applications of DDD models based on data acquisitions from sensors on product and prototypes, opening for a wider range of opportunities for the design process, and calling for further development of DDD models. About these opportunities, literature has highlighted how product-service systems might be keener to the application of DDD methods than traditional products because granting a higher possibility to monitor the usage during the lifecycle (Bertoni and Bertoni, 2019). Although recognized as relevant, the role of product-service systems is not discussed in the frame of this paper which intended to limit the discussion on the cross-pollination between engineering design and data science, so to contribute in building the theoretical background of the DDD concept in engineering design. Additionally, the paper does not claim to provide a 360-degree picture of the research directions in the field of DDD, rather such reflections are focused on the opportunities introduced by the falling technological barriers for data collection and analysis of behaviors and performances of cyber-physical systems.

5. Conclusion

The paper has presented the results of a systematic literature review investigating definitions, uses, and application of data-driven design in the concept development process. The review has identified 127 scientific contributions that have been categorized based on their use in different product development stages. The analysis has focused on the concept development stage, systematically describing the DDD applications by investigating the source of data and the analysis methods used in the different approaches. The results of the analysis showed a predominance of the use of text mining techniques on social media and online reviews to identify customers' needs, and a variety of approaches applied in the other stages of concept development. The results appear to be not aligned with the theory describing DDD approaches as a natural response to the opportunity granted by the increased diffusion of cyber-physical systems and increased accessibility of data science and information communication technology. The paper argues that the presence of such a gap in engineering concept development limits the potential of capturing tacit customers' needs. It highlights the need for product development engineers to proactively plan and design for a transition toward DDD that encompasses the direct collection and analysis of data for the usage phase of products, rather than applying traditional analysis methods on already available data. The work presented is part of a wider research effort investigating trends and opportunities for DDD in the product development process but, due to practical reasons, the analysis presented in the paper has been limited to the concept development stage.

Acknowledgment

The research has received financial support from the Swedish Knowledge and Competence Development Foundation through the Model-Driven Development and Decision Support research profile at Blekinge Institute of Technology.

References

Ahmed, F. and Fuge, M. (2018), “Creative exploration using topic-based bisociative networks”, *Design Science*, 4.

Agard, B. and Kusiak, A. (2004), “Data-mining-based methodology for the design of product families”, *International Journal of Production Research*, Vol. 42 No. 15, pp. 2955-2969. https://doi.org/10.1080/00207540410001691929

Akram, F., Prior, M. and Mavris, D. (2010), “Design Space Exploration of Submerged Inlet Capturing Interaction between Design Parameters”, In *28th AIAA Applied Aerodynamics Conference*, p. 4680. https://doi.org/10.2514/6.2010-4680

Alkahtani, M. et al. (2019), “A decision support system based on ontology and data mining to improve design using warranty data”, *Computers & Industrial Engineering*, Vol. 128, pp. 1027-1039. https://doi.org/10.1016/j.cie.2018.04.033

Arnarsson, Í.Ö. et al. (2017), “Design analytics is the answer, but what questions would product developers like to have answered?”, In *DS 87-7 Proceedings of the 21st International Conference on Engineering Design (ICED 17) Vol 7: Design Theory and Research Methodology*, Vancouver, Canada, 21-25.08. 2017, pp. 071-080. ISBN: 978-1-904670-95-7

Bae, J.K. and Kim, J. (2011), “Product development with data mining techniques: A case on design of digital camera”, *Expert Systems with Applications*, Vol. 38 No. 8, pp. 9274-9280. https://doi.org/10.1016/j.eswa.2011.01.030

Bang, H. and Selva, D. (2016), “December. iFEED: interactive feature extraction for engineering design”, In *ASME 2016 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. American Society of Mechanical Engineers Digital Collection. iFEED: interactive feature extraction for engineering design*.

Bergström, M. et al. (2008), “Needs as a basis for design rationale”, In *DS 48: Proceedings DESIGN 2008, the 10th International Design Conference*, Dubrovnik, Croatia, pp. 281-288.

Bertoni, A. and Bertoni, M. (2019), “Modeling ‘ilities’ in early Product-Service Systems design”, In *11th CIRP Conference on Industrial Product-Service Systems*. https://doi.org/10.1016/j .procir.2017.04.009

Bertoni, A. et al. (2018), “Model-based decision support for value and sustainability assessment: Applying machine learning in aerospace product development. In *15th International Design Conference*, Dubrovnik, Vol. 6, pp. 2585-2596. The Design Society. https://doi.org/10.21278/idc.2018.0437

Bertoni, A. et al. (2017), “Mining data to design value: A demonstrator in early design”, In *DS 87-7 Proceedings of the 21st International Conference on Engineering Design (ICED 17) Vol 7: Design Theory and Research Methodology*, Vancouver, Canada, 21-25.08. 2017, pp. 021-029. ISBN: 978-1-904670-95-7

Chen, L. et al. (2019), “An artificial intelligence based data-driven approach for design ideation”, *Journal of Visual Communication and Image Representation*, Vol. 61, pp. 10-22. https://doi.org/10.1016/j.jvcir.2019.02.009

Cheung, W.M. et al. (2011), “Data-driven through-life costing to support product lifecycle management solutions in innovative product development”, *International Journal of Product Lifecycle Management*, Vol. 5 No. 2/3/4, pp. 122-142. https://doi.org/10.1504/IJPLM.2011.043184

Chowdhery, S.A. and Bertoni, M. (2018), “Modeling resale value of road compaction equipment: a data mining approach”, *IFAC-PapersOnLine*, Vol. 51 No. 11, pp. 1101-1106. https://doi.org/10.1016/j.ifacol.2018.08.457

Domazet, D.S. et al. (1995), “Active data-driven design using dynamic product models”, *CIRP annals*, Vol. 44 No. 1, pp. 109-112. https://doi.org/10.1016/S0007-8506(07)62286-0

Du, X. and Zhu, F. (2018), “A new data-driven design methodology for mechanical systems with high dimensional design variables”, *Advances in Engineering Software*, Vol. 117, pp. 18-28. https://doi.org/10.1016/j.advengsoft.2017.12.006

Garg, A. and Tai, K. (2014), “An ensemble approach of machine learning in evaluation of mechanical property of the rapid prototyping fabricated prototype”, *In Applied mechanics and materials*, Vol. 575, pp. 493-496. https://doi.org/10.4028/www.scientific.net/AMM.575.493.

Georgiou, A. et al. (2016), “Advanced phase powertrain design attribute and technology value mapping”, *Journal of Engineering, Design and Technology*, Vol. 14 No. 1, pp. 115-133, ISSN: 1726-0531.

Georgiou, A. et al. (2016), “Attribute and technology value mapping for conceptual product design phase”, *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, Vol. 230 No. 11, pp. 1745-1756. https://doi.org/10.1177/0954406215585595

Ghosh, D. et al. (2017), “Cyber-Empathic Design: A data-driven framework for product design”, *Journal of Mechanical Design*, Vol. 139 No. 9, p. 091401. https://doi.org/10.1115/1.4036780

Kim, H.H.M. et al. (2017), “Data-Driven Design (D3)”, *Journal of Mechanical Design*, Vol. 139 No. 11, p. 110301. https://doi.org/10.1115/1.4037943

https://doi.org/10.1017/dsd.2020.4 Published online by Cambridge University Press

Kitchenham, B. et al. (2009), “Systematic literature reviews in software engineering–a systematic literature review”, *Information and software technology*, Vol. 51 No. 1, pp. 7-15. https://doi.org/10.1016/j.infsof.2008.09.009
Huang, Z. et al. (2011), “Optimal design of aeroengine turbine disc based on kriging surrogate models”, *Computers & structures*, Vol. 89 No. 1, pp. 27-37. https://doi.org/10.1016/j.compstruc.2010.07.010
Jeong, S., Chiba, K. and Obayashi, S. (2005), “Data Mining for Aerodynamic Design space”, *JACIC*, Vol. 2 No. 11, pp. 452-469. https://doi.org/10.2514/1.17308
Keivanpour, S. and Ait Kadi, D. (2018), “Strategic eco-design map of the complex products: toward visualisation of the design for environment”, *International Journal of Production Research*, Vol. 56 No. 24, pp. 7296-7312. https://doi.org/10.1080/00207543.2017.1388931
Kusiak, A. and Smith, M. (2007), “Data mining in design of products and production systems”, *Annual Reviews in Control*, Vol. 31 No. 1, pp. 147-156. https://doi.org/10.1016/j.arcontrol.2007.03.003
Kusiak, A. and Tseng, T.I. (2000), “Data mining in engineering design: a case study”, In *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No. 00CH37065) Vol. 1*, pp. 206-211. IEEE. https://doi.org/10.1109/ROBOT.2000.844060
Li, J.R. and Wang, Q.H. (2010), “A rough set based data mining approach for house of quality analysis”, *International Journal of Production Research*, Vol. 48 No. 7, pp. 2095-2107. https://doi.org/10.1080/00207540802665907
Liao, S.H., Chen, C.M. and Wu, C.H. (2008a), “Mining customer knowledge for product line and brand extension in retailing”, *Expert systems with Applications*, Vol. 34 No. 3, pp. 1763-1776. https://doi.org/10.1016/j.eswa.2007.01.036
Liao, S.H., Hsieh, C.L. and Huang, S.P. (2008b), “Mining product maps for new product development”, *Expert Systems with Applications*, Vol. 34 No. 1, pp. 50-62. https://doi.org/10.1016/j.eswa.2006.08.027
Lützenberger, J. et al. (2016), “Improving Product-Service Systems by Exploiting Information from the Usage Phase”, *A Case Study. Procedia CIRP*, Vol. 47, pp. 376-381. https://doi.org/10.1016/j.procir.2016.03.064
Menon, R., Tong, L.H. and Sathiyakeerthi, S. (2005), “Analyzing textual databases using data mining to enable fast product development processes”, *Reliability Engineering & System Safety*, Vol. 88 No. 2, pp. 171-180. https://doi.org/10.1016/j.ress.2004.07.007
Menon, R. et al. (2004), “The needs and benefits of applying textual data mining within the product development process”, *Quality and reliability engineering international*, Vol. 20 No. 1, pp.1-15. https://doi.org/10.1002/qre.536
Morris, C. and Seepersad, C.C. (2018), “Efficient Identification of Promising Regions in High-Dimensional Design Spaces with Multilevel Materials Design Applications”, In *ASME 2018 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. American Society of Mechanical Engineers Digital Collection. https://doi.org/10.1115/DETC2018-85273
Murthy, D.N.P., Atrens, A. and Eccleston, J.A. (2002), “Strategic maintenance management”, *Journal of Quality in Maintenance Engineering*, Vol. 8 No. 4, pp. 287-305. https://doi.org/10.1108/13552510210448504
Pajo, S. et al. (2014), “Lead User Identification through Twitter: Case Study for Camera Lens Products”, In *DS 81: Proceedings of NordDesign 2014, Espoo*, Finland 27-29th August 2014. ISBN: 978-1-904670-58-2
Pajo, S. et al. (2015), “Towards automatic and accurate lead user identification”, *Procedia engineering*, Vol. 131, pp. 509-513. https://doi.org/10.1016/j.proeng.2015.12.445
Parraguez, P. and Maier, A. (2017), “Data-driven engineering design research: opportunities using open data”, In *DS 87-7 Proceedings of the 21st International Conference on Engineering Design (ICED 17)* Vol 7. ISBN: 978-1-904670-95-7
Röhner, S., Gruber, G. and Wartzack, S. (2010), “Concept for the Architecture of a Self-Learning Engineering Assistance System”, In *DS 61: Proceedings of NordDesign 2010, the 8th International NordDesign Conference*, Göteborg, Sweden, 25.-27.08. 2010, pp. 205-216.
Song, B., Srinivasan, V. and Luo, J. (2017), “Patent stimuli search and its influence on ideation outcomes”, *Design Science*, 3.
Tango, F. and Botta, M. (2013), “Real-time detection system of driver distraction using machine learning”, *IEEE Transactions on Intelligent Transportation Systems*, Vol. 14 No. 2, pp. 894-905. https://doi.org/10.1109/TITS.2013.2247760
Tuarob, S. and Tucker, C.S. (2014), “Discovering next generation product innovations by identifying lead user preferences expressed through large scale social media data”, In *ASME 2014 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, pp. V01BT02A008-V01BT02A008. American Society of Mechanical Engineers. https://doi.org/10.1115/DETC2014-34767

Tucker, C.S. and Kim, H.M. (2009), “Data-driven decision tree classification for product portfolio design optimization”, *Journal of Computing and Information Science in Engineering*, Vol. 9 No. 4, p. 041004. https://doi.org/10.1115/1.3243634

Ulrich, K. and Eppinger, S. (1995), *Product Design and Development*, McGraw Hill. Inc., New York. ISBN 9780071137423

Vale, C.A. and Shea, K. (2003), “A machine learning-based approach to accelerating computational design synthesis”, In *DS 31: Proceedings of ICED 03, the 14th International Conference on Engineering Design*, Stockholm pp. 183-184.

Venkataraman, S. et al. (2017), “Investigating effects of stimuli on ideation outcomes”, In *DS 87-8 Proceedings of the 21st International Conference on Engineering Design (ICED 17) Vol 8: Human Behaviour in Design*, Vancouver, Canada, 21-25.08. 2017, pp. 309-318.

Wodehouse, A. et al. (2018), “Realising the affective potential of patents: a new model of database interpretation for user-centred design”, *Journal of Engineering Design*, Vol. 29 No. 8-9, pp. 484-511. https://doi.org/10.1080/09544828.2018.1448056

Wu, M. et al. (2014), “An approach of product usability evaluation based on Web mining in feature fatigue analysis”, *Computers & Industrial Engineering*, Vol. 75, pp. 230-238. https://doi.org/10.1016/j.cie.2014.07.001

Zhang, C. et al. (2017), “Concept Clustering in Design Teams: A Comparison of Human and Machine Clustering”, *Journal of Mechanical Design*, Vol. 139 No. 11, p. 111414. https://doi.org/10.1115/1.4037478

Zhang, L., Chu, X. and Xue, D. (2019), “Identification of the to-be-improved product features based on online reviews for product redesign”, *International Journal of Production Research*, Vol. 57 No. 8, pp. 2464-2479. https://doi.org/10.1080/00207543.2018.1521019

Zhang, Z., Cheng, H. and Chu, X. (2010), “Aided analysis for quality function deployment with an Apriori-based data mining approach”, *International Journal of Computer Integrated Manufacturing*, Vol. 23 No. 7, pp. 673-686. https://doi.org/10.1080/0951192X.2010.492840

Zhao, H. et al. (2007), “Application of data-driven design optimization methodology to a multi-objective design optimization problem”, *Journal of Engineering Design*, Vol. 18 No. 4, pp.343-359. https://doi.org/10.1080/09544820601010981

https://doi.org/10.1017/dsd.2020.4 Published online by Cambridge University Press