A Crowdsourcing Data-Driven Approach for Innovation

Hannah Forbes[1]*, Ji Han[1] and Dirk Schaefer[1]

[1] Division of Industrial Design, School of Engineering, University of Liverpool, Liverpool, Merseyside, United Kingdom

*Corresponding author, E-mail: Hannah.forbes@liverpool.ac.uk

Abstract

Creativity is an essential element of innovation, but producing creative ideas is often challenging in design. Many computational tools have been developed recently to support designers in producing creative ideas that are new to individuals. As a common feature, most of the tools rely on the databases employed, such as ConceptNet and the US Patent Database. However, the limitations of these databases have constrained the capabilities of the tools. Thereby, new computational databases for supporting the generation of ideas that are new to a crowd or even history are needed. Crowdsourcing outsources tasks conventionally performed in-house to a crowd and uses external knowledge to solve problems and democratize innovation. Social media is often employed in crowdsourcing for a crowd to create and share knowledge. A novel approach employing social media to crowdsource knowledge from a crowd for constructing crowd knowledge databases is proposed in this paper. The crowd knowledge database is expected to be used by the current computational tools to support designer producing highly creative ideas, which are new to the crowd, in new product design, and ultimately leading to innovation. Challenges of employing this approach are discussed to provide insights and potential directions for future research.

Keywords: Creativity, crowdsourcing, data-driven design, innovation, social media

1. Introduction

Creativity is connected to innovation via design (Han et al., 2018a), while creativity is often associated with idea generation. Idea generation, also known as ideation, is the process of coming up with ideas during the early phases of design. It has been considered the foundation of innovation (Cash & Štorga, 2015; Sarkar & Chakrabarti, 2011), which is also a significant element in business success (Howard et al., 2011). Therefore, generating creative ideas is essential for achieving innovation. However, it is always challenging for individuals to produce creative ideas, due to limited knowledge, many existing ideas, time pressure and lack of creative mind (Han et al., 2018a). Knowledge is a significant resource in supporting innovation (Bertola & Teixeira, 2003) but it is difficult and time-consuming to collect information and knowledge for assisting idea generation. Ullman (2010) indicated that design engineers spend 60% of the time during the design process to explore the information and knowledge needed. Therefore, in order to support designers in creativity and leading to innovation, relevant knowledge or a database containing the needed knowledge needs to be provided to designers.

There is a growing interest in using computational tools for supporting designers in generating creative ideas in recent decades. Databases, containing knowledge for supporting design, are often employed by the tools. Various databases are used, for instance, design repositories, ConceptNet, biological and engineering systems in structure-behaviour-function forms, and customised ones. However, some databases involve a limited amount of knowledge, some are not suitable for design, and some mainly contains past knowledge. Besides, new knowledge emerges rapidly in nowadays fast developing world. In order to produce creative ideas for developing nowadays innovative products, up-to-date knowledge is needed. Thereby, it is needed to explore how to employ rapidly emerged knowledge to support designers in creativity and innovation. Crowdsourcing is a model where many solutions are generated by answering open calls. Goucher-Lambert and Cagan (2019) have shown the use of crowdsourcing to generate inspirational stimuli to support idea generation. Social media is described as 'a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content' (Kaplan & Haenlein, 2010). Thus, social media, such as Twitter and Facebook, are considered platforms which are often used by crowds for creating knowledge. Taking the advantages of crowdsourcing and social media, databases containing up-to-date knowledge created by the crowd could be constructed.

The authors of this paper aim to explore a crowdsourcing data-driven approach to construct crowd knowledge databases for innovation through supporting creative idea generation. In the approach, social media will be used as platforms to crowdsource knowledge for producing the databases. The crowd knowledge databases are intended to be employed in existing computational creativity tools for improving the tools' performances and capabilities. This will benefit the generation of creative ideas and lead to innovative products. Creativity in design is investigated in the next section. Crowdsourcing and related frameworks are explored in section 3 and 4, respectively. Based on the explorations, the crowdsourcing data-driven approach is proposed in section 5. Challenges involved in this approach are discussed in section 6, and the paper is concluded in section 7.

2. Design Creativity

Creativity is considered a significant element in design, which is defined as the process of producing something judged to be creative (Amabile, 1983). Han, Forbes and Schaefer (2019) have indicated that novelty, surprise, and usefulness are the three core factors of creativity in design. Idea generation involves the process of creating developing and communicating ideas, where ideas are fundamental elements of thoughts in visual, concrete and abstract forms (Jonson, 2005). Idea generation has been considered essential to innovation (Cash & Štorga, 2015; Sarkar & Chakrabarti, 2011). However, idea generation, especially generating creative ideas, is a challenging process in new product design and development.

Creativity tools and methods are thereby developed and used to support designers in creative idea generation during the early stages of design. There exist two categories of tools for supporting creative idea generation, non-computational and computational tools. Non-computational tools, such as TRIZ (Altshuller, 1984), design-by-analogy (Goldschmidt, 2001) and the 77 design heuristics (Yilmaz et al., 2016), provides designers with guidelines and instructions for producing creative ideas. However, some of the tools rely heavily on designers' knowledge, while some others are challenging to master.

In recent years, computational tools which involve the use of computational techniques for supporting idea generation have been explored. These tools could produce creative prompts and provide relevant knowledge to support designers in creative idea generation more effectively and efficiently. The Retriever (Han et al., 2018b) prompts designers in generating creative ideas through constructing new ontologies to support reasoning by employing real-world data. The database employed in the tool is the ConceptNet (Speer et al., 2017), which is a machine-understandable knowledge network. The knowledge contained is mainly common-sense knowledge, which has limited the Retriever in constructing highly novel ontologies for supporting idea generation. Analogy Finder (McCaffrey & Spector, 2017) provide users with adaptable analogous ideas for solving technical problems by conducting searches using the US patent database. However, the tool requires the users to have strong expertise and knowledge to adapt the ideas retrieved from the US patent database employed for solving problems. Idea Inspire 4.0 (Keshwani & Chakrabarti, 2017) designers in generating creative ideas for solving problems via analogical design. A searchable knowledge base is employed in the tool containing a limited number of biological systems. An automated approach has been proposed by Keshwani and Chakrabarti (2017) for populating the database.

Creativity has been classified into two main categories, H-creativity and P-creativity (Boden, 2004). H-creativity refers to historical-creativity, which indicates generating ideas that are new in history. P-creativity, also known as psychological-creativity, indicates producing ideas that are new to the person who produced the idea. Comparing with the design creativity studies at P-creativity levels, fewer studies focus on H-creativity levels. Therefore, there is a need to explore design creativity at H-creativity levels, investigating how to produce ideas that are new to a group of people, a crowd and ultimately history. From a group perspective, studies, such as the ones conducted by Paulus and Dzindolet (2008) and Nijstad and Stroebe (2006) have shown that collaboration has positive effects on creativity. Paulus, Dzindolet, and Kohn (2012) have revealed collaborative creativity could produce better outcomes than individual creativity. This indicates that using groups could produce ideas that are better than the ones produced by individuals. Ideas produced by a group are new to the group, which are beyond P-creativity and close to H-creativity. Thereby, employing an even larger number of people, such as a crowd, could potentially lead to the generation of ideas that belong to the H-creativity category.

As illustrated above, databases play a significant role in nowadays computational tools. However, the databases employed by the tools have various limitations, which have negative impacts on the tools' capabilities. Besides, the use of a crowd in supporting design creativity, especially creative idea generation could yield superior results. A crowd could be employed to produce ideas or provide knowledge for solving design problems. The ideas produced and knowledge provided by the crowd could be constructed into a crowd knowledge database to support designers in producing creative ideas to solve the design problem. Thus, a new approach to create crowd knowledge databases for computational tools to support designers in creative idea generation needs to be explored.

3. Crowdsourcing for Innovation

Crowdsourcing is described as a web-based creative problem-solving model, in which "a distributed network of individuals produces solutions to an open call for proposals" (Brabham, 2008). In the context of design, Forbes and Schaefer (2018) suggest that crowdsourcing is most suited to evaluation and ideation, as shown in Figure 1. Later design phases require a higher skill level and are therefore are harder to "open to the crowd". The suitability for ideation and other early design stages, therefore, is as a consequence of the inverse relationship between the size of the qualified crowd and the level of skill for contribution. For example, in concept generation, "ideas are not scrutinised on their technical rigor or feasibility" (Daly et al., 2012; Forbes et al., 2019). The number of those qualified to make these contributions is higher than later design phases and therefore the crowd available in this phase is large. This is, however, founded on the assumption that a larger number of contributions results in a more successful crowdsourcing initiative. Panchal (2015) discusses several "modes of failure" for crowdsourcing initiatives, including "a lack of submissions" but also the result of "numerous poor-quality submissions". It is important to consider, therefore, that while we make the assumption that higher number of submissions is preferable, it is possible that too many submissions can be detrimental to the success of the crowdsourcing initiative.

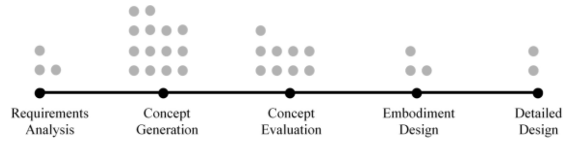


Fig. 1 Current literature's exploration of crowdsourcing in each product development phase (One grey dot represents one source) (Forbes & Schaefer, 2018)

Examples of initiatives that use crowdsourcing for idea generation includes Goucher-Lambert & Cagan (2019) who have used crowdsourcing techniques to "obtain inspirational stimuli" to support designers in ideation. "Connect and Develop" from Procter and Gamble, is another example described as an "organisation partnership" with "the world's most innovative minds". As part of Connect and Develop, Procter and Gamble encourage the crowd to submit product ideas and suggestions according to a theme most relevant to their organisation at the time (Dodgson et al., 2006). Since using crowdsourcing for idea generation, Procter and Gamble's R&D productivity increased 60% and 45% of new initiatives had elements discovered externally (Dodgson et al., 2006; Forbes et al., 2019). A final example is the DARPA crowdsourcing initiative which awarded one million dollars to a design team, external to the organisation, for the creation of an "innovative marine tank drive train" designed to significantly improve efficiency of tank movement (Ackerman, 2013). Crowdsourcing has therefore been demonstrated as a success in many idea generation initiatives (Forbes et al., 2019). Including the crowdsourcing process as an element of a data driven approach for design creativity, whereby formalising this process, could therefore prove useful to designers.

There are two types of crowdsourcing; active crowdsourcing and passive crowdsourcing. Active crowdsourcing is leveraged when the crowd actively participates in a contest or call for submissions. There are four types of active crowdsourcing initiative; crowdsourcing contests, open calls with direct rewards, open calls with direct rewards and micro-tasking. Table

1 below gives definitions and examples of these crowdsourcing initiatives.

Table 1 Active Crowdsourcing Initiatives (Panchal, 2015)

Initiative	Example	Description
Crowdsourcing contests	Gold Corp "Global Search Challenge" (Brabham, 2008)	Participants from around the world were encouraged to examine geologic data from Goldcorp's Red Lake Mine and submit proposals identifying potential targets where the next 6 million ounces of gold will be found. $500,000 in prize money was offered to the 25 top finalists who identified the most gold deposits. (Brabham, 2008; Corp, 2001)
Open calls with direct rewards	Procter & Gamble's Connect & Develop (Dodgson et al., 2006)	Procter & Gamble advertise their research and develop needs through a crowdsourcing initiative, called Connect & Develop, in classified categories. Anyone who is interested or has a solution within the advertised categories could propose their ideas by submitting through the Connect & Develop website. The ideas get assessed by a specialized team and reward payments can range from $10,000 to $100,000 (Dodgson et al., 2006)
Open calls with in- direct benefits	Dell Idea Storm	In a similar setup to Connect & Develop, Dell Idea Storm seeks ideas on their website from a community of non-experts. Contributors, however, are not rewarded financially and instead benefit indirectly from the company's implementation of the ideas in their products (Di Gangi & Wasko, 2009)
Micro-tasks	Amazon Mechanical Turk	Amazon Mechanical Turk is a website that allows businesses to hire participants "to perform discrete on-demand tasks that computers are currently unable to do." (Buhrmester et al., 2011)

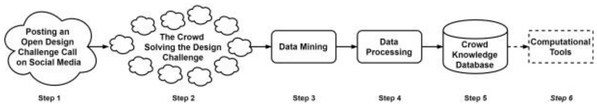
Passive crowdsourcing, on the other hand, uses information from the crowd that is in the public domain, or that has been collected with permission from the crowd (Charalabidis et al., 2014). How the information is used is dependent completely on the methodology applied by the data collectors and is not influenced by the content of the data. An example of passive crowdsourcing is Netflix's use of customer choices, to supply film and TV recommendations.

Using crowd data to populate computational creativity tools is a hybrid crowdsourcing approach using both active and passive crowdsourcing. An open call with indirect rewards, an active crowdsourcing initiative, is used to encourage the crowd to share their ideas. A set method is then used to process the data for use in a computational creativity tool, representative of a passive crowdsourcing approach. Several other authors have implemented hybrid active and passive crowdsourcing approaches. For example, Janssen et al. (2017) use a hybrid approach to crowdsourcing for policy making. They state that "synergy can be created by combining both approaches. The results of passive crowdsourcing can be used for guiding active crowdsourcing to avoid asking users for similar types of input". Similarly, Charalabidis et al. (2014) uses a hybrid approach for policy making by "exploiting the extensive political content continuously created in numerous Web 2.0 [technologies]". Finally, Akshay et al. (2018) use passive and active crowdsourcing for monitoring video for critical events stating that this approach "increases the feasibility of deploying continuous real-time crowdsourcing systems in real-world settings". There is therefore evidence of using crowdsourcing and an active-passive crowdsourcing approach for innovation, in several fields of research.

Despite evidence of similar successful uses of crowdsourcing, some crowdsourcing initiatives are more effective than others (Panchal, 2015). Ineffective crowdsourcing initiatives may invite inadequate submissions that fail to reach the required quality. A crowdsourcing initiative can also become ineffective if the expense of running the initiative exceeds the cost of an in-house team (Brabham, 2008; Panchal, 2015). As a consequence, there is a need to frame crowdsourcing processes. In the following section, existing crowdsourcing frameworks are presented.

4. Crowdsourcing Frameworks

Crowdsourcing has emerged with the birth of the internet and with the ability to share information quickly and easily, worldwide. Social media has been a



catalyst in this growth by facilitating and supporting users to create, share and edit information, as well as build relationships through interaction and collaboration (Mount & Martinez, 2014). Kemp (2019) reported that there are 3.48 billion social media users in 2019, which leads to millions of posts every minute (Forbes et al., 2019). When an open call crowdsourcing initiative is launched on social media, therefore, potential participants can be reached, and ideas can be submitted quickly and easily. Preventing crowdsourcing failure, when leveraging social media, requires a methodical approach. Before presenting a new crowdsourcing social media framework for computational creativity, the authors explored existing research in this area.

Crowdsourcing frameworks are most prevalent in the field of product design and development. Niu et al. (2019) present a framework for the application of crowdsourcing in product development, guiding the user through important crowdsourcing decisions. Panchal (2015) also presents a framework for the use of crowdsourcing in product development, providing a four-step approach to crowdsourcing application. This framework includes three key steps; selecting crowdsourcing initiatives, making design decision and incentive design. Panchal also provides further detail regarding "incentive design" by presenting a game-theoretic model for managing crowd participation. Similarly, Abrahmason et al. (2013) present an "Incentives Mix Framework" for understanding crowd participation and Cullina et al. (2016) and Gerth et al. (2012) provide in depth research on finding the "qualified crowd" in crowdsourcing contests. Finally, Kittur et al. (2011) consider the crowdsourcing of Human Intelligence Tasks (HITs) and "provide a systematic and dynamic way to break down tasks into subtasks and manage the flow and dependencies between them".

In other fields, few authors have presented a crowdsourcing framework for their domain. To and Shahabi (2018) propose a crowdsourcing framework for "protecting worker location privacy in spatial crowdsourcing", Liu (2014) present a "crisis crowdsourcing framework" for "designing strategic configurations of crowdsourcing for the emergency management domain" and Chen et al. (2009) present a "QoE evaluation framework for multimedia content". These authors represent the scarcity of crowdsourcing frameworks and demonstrates the relative youth of this research topic. By creating a crowdsourcing framework for creativity, and specifically computation creativity, is therefore a significant contribution in an emerging literature sector. Furthermore, existing crowdsourcing frameworks are, in general, at a low-level of abstraction, addressing and guiding small aspects of the crowdsourcing process as opposed to offering high-level support. For example, Cullina et al. (2016) discusses the need to understand crowd motivation in contests which is a single factor contributing to the successful implementation of crowdsourcing. By presenting a high-level, crowdsourcing framework for computational creativity, the authors are offering more holistic guidance for crowdsourcing application.

5. The Crowdsourcing Data-driven Approach

Fig. 2 The crowdsourcing data-driven approach of creating a crowd knowledge databased

As illustrated above, crowdsourcing initiatives allow varied and numerous data points to be collected from the crowd. They are particularly effective in early design phases as the prerequisite skill level for participation in these phases is reduce, In this section, it is demonstrated how crowdsourcing could acquire knowledge from a crowd to support creative design activities in new product design and development, such as idea generation and evaluation, by partnering crowdsourcing with computational creativity tools. A novel approach using social media to crowdsource design knowledge for creating crowd knowledge databases is proposed, as shown in Figure 2. In step 1, an open design challenge call is posted on social media, such as Twitter and Facebook. A dedicated hashtag is involved in the open call post. The hashtag will help the crowd identify the open call on social media, as well as be used as a target to support the later data mining process. In step 2, an active crowdsourcing method is used to encourage the crowd to generate ideas using descriptive text for solving the design challenge in the open call. The ideas generated are posted back on social media containing the dedicated hashtag. Data mining is conducted in the next step to retrieve posts containing the dedicated hashtag only. This will help to discard noise data which are irrelevant to the open call. In step 4, the retrieved data are processed by using natural language processing tools to extract useful words and phrases. The extracted data are then used to construct crowd knowledge databases for supporting creativity and innovation in step 5. In the last step, the crowd knowledge databases constructed will be used by exiting computational design creativity tools to enhance the capabilities of the tools in supporting idea generation. For example, the databases could be employed by the Combinator (Han et al., 2018a) to produce combinational prompts associating knowledge produced by the crowd.

6. Discussion

Having presented the approach, this section considers the hurdles and challenges for implementation. There are three key phases of the approach that require attention. This includes, firstly, how participation will be encouraged and managed. Secondly, how the submitted responses will be processed is significant in determining the value of ideas generated from this crowd-knowledge database. Finally, it is important to determine how the submitted responses are included as part of the computational creativity tool and whether this should differ from other databases. The third phase, regarding use of the database, is managed by existing computational creativity tools but the first and second phases are included in the discussion (Forbes et al., 2019).

6.1 Managing Participation on Social Media

When considering the management of participation, social media allows access to the largest number of people possible which makes it an effective medium for hosting both passive and active crowdsourcing initiatives. The difficulty, however, is gaining active participation in on these platforms. “Social media is used extensively and constantly to attract attention and users can often be overwhelmed with online content” (Forbes et al., 2019). Enticing submissions therefore requires strategic thinking. In addition, high numbers are important but high variety is also important for generating innovative ideas (Howe, 2006). Organisations use crowdsourcing initiatives because they recognise a need to involve other perspectives beyond those of their in-house teams. Effort must therefore be made to increase exposure of the hashtag but while limiting the “echo chamber effect” that can reduce heterogeneity of the responses (Colleoni et al., 2014; Forbes et al., 2019). There is a need to manage how the hashtag is exposed to potential crowdsourcing participants to ensure text-based responses from users are effective for generating creative ideas.

Within crowdsourcing and creativity research domains, solutions to this challenge are limited. The authors therefore considered other research domains such as digital marketing to offer an understanding of how organisations can compete for social media attention while running a crowdsourcing initiative. To correspond with the required traits of captured data, the authors were interested in solutions to capture diverse information and solutions to capture numerous data. With regards to managing diversity, existing literature on the impact of social media on political preference, offered insight. Ensuring a heterogenous dataset, meant limiting the impact of “social media bubbles” or “echo chambers” (Zhan et al., 2016; Romero et al., 2011) which is of significant interest in the current political climate. Garimella et al. (2017) offer a solution that could be applicable to the use of crowdsourcing for computational creativity. They suggest when “exposing information” to users, a “symmetric difference function” could be “optimized” to limit the dominance of one piece of information in the case of two competing instances of information. In the context of ensuring diverse submissions, engaging a “symmetric difference function” could ensure that a single submission on the social platform would not influence subsequent submissions. Dubois and Blank (2018) also propose another solution which suggests the ownness is on the user to limit their vulnerability to polarising online content. They demonstrate that users with “diverse interests” on social media platforms are significantly less susceptible to exposure to polarising content. A solution to ensure heterogeneous submissions for a crowdsourcing activity, therefore, could be target users with connections with a range of interests and political viewpoints.

The authors were also interested in learning how a crowdsourcing initiative could “compete for attention” on social media platforms (Romero at al., 2011). Feng, et al., (2015) suggest garnering attention on “busy” social media platforms, information sharers need to understand how and when users become “overload with information” and respond accordingly. They show how information spread on social media can be represented by a fractional susceptible infected recovered (FSIR) model. In this case bacteria spread is analogous to information spread and infection presents information overload (Feng et al., 2015). Using this model, Feng et al. (2015) suggest spreading information early in the day and early in an “social information cycle” which they describe in detail. Iyer and Zsolt (2015) suggest that to compete for attention on social media platforms, information sharers must consider the incentives users respond to for social media use in general. They then suggest embedding these incentives, such as the ability to connect with others, into the mechanism they use to spread information (Iyer & Zsolt, 2015). Each of these existing solutions can be considered when implementing the crowdsourcing data-driven approach.

6.2 Processing a variety of information types

How the submitted responses will be processed is significant in determining the value of ideas generated from this crowd-knowledge database. Using texts to provoke the designers' mind in producing creative ideas has been demonstrated in a number of previous studies, but in various forms (Forbes et al., 2019). For example, Shi et al. (2017) employed network-based texts, while Han et al. (2018a) used combinational texts. However, the presentation form of the crowd knowledge, the solutions generated by the crowd and processed by computational means in this study, still needs to be explored (Forbes et al., 2019).

Collection of social media data differs from data (text) used in previous studies. Crowd data may include sentimental as well as emotionality aspects. This means that the process of natural language process must include a measurement of sentiment to determine the positivity, as well as negativity, of the whole text. Overall, emotionality needs to be calculated on individual text segments to indicate positive and negative text segments. Emotionality could support designers in decision-making by ensuring they have a greater understanding and further context of crowd data. For example, designers might need to avoid the design aspects related to negative knowledge and enhance design features related to positive knowledge (Forbes et al., 2019). This might also help the computational tools in a better comprehension of the crowd knowledge database employed.

The way social media users communicate has developed beyond just text-based, which should be considered, further, to processing emotional and sentimental aspects of participant responses, "Emojis", "GIFs" and "memes" are frequently and extensively used on social media to communicate ideas. Their use means either they must be filtered and removed, or "translated" for inclusion in a crowd database. One approach to this, as shown in Figure 2, includes the use of key words to identify the key idea communicated in participant responses. It could be the case, however, that the key idea is communicated in a text-based caption with an image accompaniment to bolster, as opposed to convey, the idea. How this varying use of video and image-based content is managed should be taken into consideration.

Twitter and other social media platforms are purposefully designed to encourage collaboration and interaction between users. This results in functionality elaborating and "commenting" on other responses that is considered integral to the design of these online platforms. As a result, however, the processing of participant involvement needs to recognise not only individual responses including the hashtag but "clusters" or responses that all represent one idea (Forbes et al., 2019). As an example, one participant may include the "crowdsourcing hashtag" to present an idea which initiates an online conversation, with further responses elaborating on or supporting the initial idea. Some of these comments may be new ideas but others could be minor alterations of additions to the original submission. This means that including every response involved in the conversation and weighting them equally could disrupt the value of crowd data. An understanding of how collaboration occurs on social media is therefore fundamental to procuring valuable results for idea generation (Forbes et al., 2019).

Utilising crowd knowledge from social media shows great potential for supporting creativity and innovation. There are, however, several research challenges such as participation management and data processing, to overcome. Furthermore, the way social media users communicate has and will change to incorporate more media-based content. Further research is needed to solve these research challenges and recognise new opportunities in the applications of this crowdsourcing data-driven approach. The key next research step is to conduct a case study of using the crowd knowledge from a specific social media platform to solve a design challenge. The authors hope to provide more insights on this new and novel data-driven computer-aided innovation approach.

7. Conclusion

Generating ideas, especially creative ones, is significant to innovation. However, it is challenging to produce creative ideas. Many computational support tools are thereby developed to assist this process, but the current solutions are constrained by available databases. Lacking knowledge in terms of quantity and variety is one of the main issues of the databases. Besides, knowledge collection has been considered a time-consuming and frustrating activity. Crowdsourcing is a model for creative problem-solving which uses the knowledge produced by a distributed network of individuals also known as a crowd. Social media, which allows creating and exchanging contents created by users, is often employed to generate and share knowledge.

Thus, the authors of this paper have proposed a novel data-driven approach utilising social media to crowdsource knowledge to construct databases for computational tools in supporting creative idea generation, and ultimately leading to innovation. The databases constructed are called the crowd knowledge databases, which are populated by providing and distributing open design challenge calls with responses using unique hashtags for identification. Data mining and natural language processing are used in the construction process to retrieve and extract data, respectively. The crowd knowledge databases can then be implemented into existing as well as future computational tools to enhance their performances. Using the Combinator (Han et al., 2018a) as an example, the tool could associate crowd knowledge from the database to produce new combinational prompts, which are new to the crowd, for stimulating users creative mind. The data-driven approach proposed has implied its value of utilising some of the most used and data-rich platforms available to achieve innovation.

However, a number of challenges need to be solved to realize the crowdsourcing data-driven approach. In this paper, how to manage participation on social media and how to process a variety of information types are discussed. Several participation management methods, such as information spreading and incentives, as well as several information processing issues, such as sentiment measurements and collaboration understands, are indicated. Further research is required to explore these challenges and to overcome them, in order to fully employ the proposed crowdsourcing data-driven approach in computational support tools for innovation. This paper has thereby shown a new research direction in using crowdsourcing data to support innovation, contributing to the computer-aided innovation research area. The authors have planned to conduct a case study of solving a design challenge using the crowd knowledge from a specific social media platform, such as Twitter, in their next study to provide more valuable insights.

References

Abrahamson, S., Ryder, P., & Unterbeg, B. (2013). *Crowdstorm: the future of innovation, ideas, and problem solving*. John Wiley & Sons.

Ackerman, S. (2013). This Is the Million-Dollar Design for Darpa's Crowdsourced Swimming Tank. Retrieved from Wired: https://www. wired. com/2013/04/darpafang-winner

Akshay, R. K., & Walter, S. L. (2018). Plexiglass: Multiplexing passive and active tasks for more efficient crowdsourcing. *Sixth AAAI Conference on Human Computation and Crowdsourcing*.

Altshuller, G. S. (1984). *Creativity as an exact science: The theory of the solution of inventive problems*. Amsterdam, Netherlands: Gordon and Breach Publishers.

Amabile, T. M. (1983). *The social psychology of creativity*. New York, USA: Springer-Verlag.

Bertola, P., & Teixeira, J. C. (2003). Design as a knowledge agent: How design as a knowledge process is embedded into organizations to foster innovation. *Design Studies*, 24(2), 181-194.

Boden, M. A. (2004). *The creative mind: Myths and mechanisms* (Vol. 2). London, UK: Routledge.

Brabham, D. C. (2008). Crowdsourcing as a Model for Problem Solving: An Introduction and Cases. *Convergence*, 14(1), 75-90.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's mechanical turk: A new source of inexpensive, yet high-quality, data?". *Perspectives on psychological science*, 6(1), 3–5.

Cash, P., & Štorga, M. (2015). Multifaceted assessment of ideation: using networks to link ideation and design activity. *Journal of Engineering Design*, 26(10-12), 391-415.

Charalabidis, Y. N., Loukis, E., Androutsopoulou, A., Karkaletsis, V., & Triantafillou, A. (2014). Passive crowdsourcing in government using social media. *Transforming Government: People, Process and Policy*, 8(2), 283-308.

Chen, K. T., Wu, C. C., Chang, Y. C., & Lei, C. L. (2009). A crowdsourceable QoE evaluation framework for multimedia content. *Proceedings of the 17th ACM international conference on Multimedia* (pp. 491-500). ACM.

Colleoni, E., Rozza, A., & Arvidsson, A. (2014). Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. *Journal of communication*, 64(2), 317-332.

Corp, G. (2001). The goldcorp challenge winners. Retrieved July 2019, from Gold Corp Challenge: http://www. goldcorpchallenge.com/challenge1/winnerslist/chaleng2.pdf.

Cullina, E., Conboy, K., & Morgan, L. (2016). Choosing the right crowd: An iterative process for crowd specification in crowdsourcing initiatives. *System Sciences (HICSS) 49th Hawaii International Conference* (pp. 4355-4364). IEEE.

Daly, S. R., Yilmaz, S., Christian, J. L., Seifert, C. M., & Gonzalez, R. (2012). Design heuristics in engineering concept generation. *Journal of Engineering Education*, 101(4), 601-629.

Di Gangi, P. M., & Wasko, M. (2009). Steal my idea! organizational adoption of user innovations from a user innovation community: A case study of dell ideastorm. *Decision Support Systems*, 48(1), 303–312.

Dodgson, M., Gann, D., & Salter, A. (2006). The role of technology in the shift towards open innovation: the case of Procter & Gamble. *R&D Management*, 36(3), 333-346.

Dubois, E., & Blank, G. (2018). The echo chamber is overstated: the moderating effect of political interest and diverse media. *Information, Communication & Society*, 21(5), 729-745.

Feng, L., Yanqing, H., Baowen, L. H., Eugene, S., Shlomo, H., & Braunstein, L. A. (2015). Competing for attention in social media under information overload conditions. *PloS*, 10(7).

Forbes, H. L., & Schaefer, D. (2018). Crowdsourcing in product development: Current state and future research directions. *Proceedings of the DESIGN 2018 15th International Design Conference*. Dubrovnik, Croatia: The Design Society.

Forbes, H., Han, J., & Schaefer, D. (2019). Exploring a Social Media Crowdsourcing Data-Driven Approach for Innovation. Proceedings of the International Conference on Systematic Innovation.

Garimella, K., Gionis, A., Parotsidis, N., & Tatti, N. (2017). Balancing information exposure in social networks. *Advances in Neural Information Processing Systems*, 4663-4671.

Gerth, R. J., Burnap, A., & Papalambros, P. (2012). Crowdsourcing: A primer and its implications for systems engineering. Michigan University Ann Arbor.

Goldschmidt, G. (2001). Visual Analogy - a Strategy for Design Reasoning and Learning. In G. Goldschmidt, *Design Knowing and Learning: Cognition in Design Educatio*n (pp. 199-219). Oxford: Elsevier Science.

Goucher-Lambert, K., & Cagan, J. (2019). Crowdsourcing inspiration: Using crowd generated inspirational stimuli to support designer ideation. *Design Studies*, 61(1), 1-29.

Han, J., Forbes, H., & Schaefer, D. (2019). An Exploration of the Relations between Functionality, Aesthetics and Creativity in Design. *Proceedings of the Design Society: International Conference on Engineering Design*. 1 (1), pp. 259-268. The Design Society.

Han, J., Shi, F., Chen, L., & Childs, P. R. (2018). A computational tool for creative idea generation based on analogical reasoning and ontology. *Artificial Intelligence for Engineering Design*, 32(4), 462-477.

Han, J., Shi, F., Chen, L., & Childs, P. R. (2018). The Combinator – a computer-based tool for creative idea generation based on a simulation approach. *Design Science*, 4, 11.

Howard, T. J., Culley, S., & Dekoninck, E. A. (2011). Reuse of ideas and concepts for creative stimuli in engineering design. *Journal of Engineering Design*, 22(8), 565-581.

Howe, J. (2006). The rise of crowdsourcing. *Wired Magazine*, 14(6), 1-4.

Iyer, G., & Zsolt, K. (2015). Competing for attention in social communication markets. *Management Science*, 62(8), 2304-2320.

Janssen, M., Konopnicki, D., Snowdon, J. L., & Ojo, A. (2017). Driving public sector innovation using big and open linked data (BOLD). *Information systems frontiers*, 19(2), 189-195.

Jonson, B. (2005). Design ideation: the conceptual sketch in the digital age. Design Studies, 26(6), 613-624.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, 53(1), 59-68.

Kemp, S, Hootsuite. (2019). Digital Report 2019: Essential Insights into How People Around the World Use the Internet Mobile Devices. Retrieved from *Social Media and E-Commerce*.

Keshwani, S., & Chakrabarti, A. (2017). Towards automatic classification of description of analogies into SAPPhIRE constructs. *International Conference on Research into Design*.

Kittur, A., Smus, B., & Khamkar, S. K. (2011). Crowdforge: Crowdsourcing complex work. *Proceedings of the 24th annual ACM symposium on User interface software and technology* (pp. 43-52). ACM.

Liu, S. B. (2014). Crisis crowdsourcing framework: Designing strategic configurations of crowdsourcing for the emergency management domain. *Computer Supported Cooperative Work (CSCW)*, 23(4-6), 389-443.

McCaffrey, T., & Spector, L. (2017). An approach to human–machine collaboration in innovation. *Artificial Intelligence for Engineering Design*, 1-15.

Mount, M., & Martinez, M. G. (2014). Social Media: A Tool for Open Innovation. *California Management Review*, 56(4), 124-143.

Nijstad, B. A., & Stroebe, W. (2006). How the Group Affects the Mind: A Cognitive Model of Idea Generation in Groups. *Personality and Social Psychology Review*, 10(3), 186-213.

Niu, X. J., Qin, S. F., Vines, J., Wong, R., & Lu, H. (2019). Key crowdsourcing technologies for product design and development. *International Journal of Automation and Computing*, 16(1), 1-15.

Panchal, J. H. (2015). Using crowds in engineering design–Towards a holistic framework. *Proceedings of the 20th International Conference on Engineering Design (ICED 15), 8: Innovation and Creativity*, pp. 041-050. Milan, Italy.

Paulus, P. B., & Dzindolet, M. (2008). Social influence, creativity and innovation. *Social Influence*, 3 (4), 228-247.

Paulus, P. B., Dzindolet, M., & Kohn, N. W. (2012). Chapter 14 - Collaborative Creativity: Group Creativity and Team Innovation . In *Handbook of Organizational Creativity*. San Diego: Academic Press,.

Romero, Mauricio, D., Meeder, B., Barash, V., & Kleinberg, J. (2011). Maintaining ties on social media sites: The competing effects of balance, exchange, and betweenness. *Fifth International AAAI Conference on Weblogs and Social Media.*

Sarkar, P., & Chakrabarti, A. (2011). Assessing design creativity. *Design Studies*, 32(4), 348-383.

Shi, F., Chen, L., Han, J., & Childs, P. (2017). A Data-Driven Text Mining and Semantic Network Analysis for Design Information Retrieval. *Journal of Mechanical Design*, 139(11), 111402-111414.

Siddharth, L., & Chakrabarti, A. (2018). Evaluating the impact of Idea-Inspire 4.0 on analogical transfer of concepts. *Artificial Intelligence for Engineering Design*, 32(4), 431-448.

Speer, R., Chin, J., & Havasi, C. (2017). ConceptNet 5.5: an open multilingual graph of general knowledge. *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*. San Francisco, California.

To, H., & Shahabi, C. (2018). Location privacy in spatial crowdsourcing. In *Handbook of Mobile Data Privacy* (pp. 167-194). Springer, Cham.

Ullman, D. G. (2010). *The mechanical design process: Part 1*. McGraw-Hill.

Yilmaz, S., Daly, S. R., Seifert, C. M., & Gonzalez, R. (2016). Evidence-based design heuristics for idea generation. *Design Studies*, 46(C), 95-124.

Zhan, L., Yongqiang Sun, N. W., & Xi, Z. (2016). Understanding the influence of social media on people’s life satisfaction through two competing explanatory mechanisms. *Journal of Information Management*, 68(3), 347-361.

AUTHOR BIOGRAPHIES

Hannah Forbes is a 4th year PhD Candidate and Co-Chair of the Systems Realization Laboratory at The University of Liverpool. Hannah's research focuses on the democratization of product design and development. Her most recent work is on crowdsourcing and crowdfunding, and how to support SMEs in applying crowdsourcing initiatives. Hannah is also the Founder of The Funding Crowd, a crowdfunding consultancy, and was named as a Rising Star in Finance by Innovate Finance and was awarded a Women in Finance Award 2020 by Finance Monthly Magazine.

Dr Ji Han holds an Assistant Professor position in Industrial Design at the School of Engineering, the University of Liverpool. He is the founding Programme Director of Product Design Engineering BEng/MEng (with Year in Industry) and the Programme Director of Industrial Design BEng/MEng. Ji's research focuses on design engineering, AI in design, design neuro-cognitions, design creativity, computational design creativity, and data-driven design. His research addresses various topics relating to design and creativity, and places a strong emphasis on exploring new design theories and developing advanced design support tools.

Professor Dirk Schaefer holds the Chair in Industrial Design at the University of Liverpool in the United Kingdom. He is the Director of the Systems Realization Laboratory, and an international thought leader in Cloud-Based Design and Technology for Digital Manufacturing in the context of Industry 4.0. His accomplishments comprise more than 180 technical publications, including ten books. He has delivered more than 120 talks at events all over the world. Prof. Schaefer is a registered Chartered Engineer, Chartered Technological Product Designer, Chartered Scientist, and a Chartered IT-Professional in the UK.