Combining Humans and Machine Learning: A Novel Approach for Evaluating Crowdsourcing Contributions in Idea Contests

Abstract. The creative potential from innovative contributions of the crowd constitutes some critical challenges. The quantity of contributions and the resource demands to identify valuable ideas is high and remains challenging for firms that apply open innovation initiatives. To solve these problems, research on algorithmic approaches proved to be a valuable way by identifying metrics to distinguish between high and low-quality ideas. However, such filtering approaches always risk missing promising ideas by classifying good ideas as bad ones. In response, organizations have turned to the crowd to not just for generating ideas but also to evaluate them to filter high quality contributions. However, such crowd-based filtering approaches tend to perform poorly in practice as they make unrealistic demands on the crowd. We, therefore, conduct a design science research project to provide prescriptive knowledge on how to combine machine learning techniques with crowd evaluation to adaptively assign humans to ideas.
Keywords: idea contests, design-science research, machine learning, idea filtering

1 Introduction

Firms increasingly engage in open innovation efforts to leverage the creative potential of a huge and diverse crowd of contributors [1-2]. Therefore, one popular approach is to solve innovative problems by starting an open call to a crowd with heterogeneous knowledge and diverse experience via a web-based innovation platform (e.g., BrightIdea, Salesforce, and Ideascale). Individual members of the crowd then contribute creative ideas to solve such problems and the firm rewards the best contribution in a contest approach [3]. This novel way to solicit ideas from online communities is a powerful mechanism to utilize open innovation.

However, the creative potential that arises from the innovative contributions of the crowd constitutes some critical challenges. The quantity of contributions and the demands on expertise to identify valuable ideas is high and remains challenging for firms that apply crowdsourcing [4]. Famous examples illustrate these novel phenomena. For instance, during the IBM “Innovation Jam” in 2006 more than 150,000 users from 104 countries generated 46,000 product ideas for the company [5]. Moreover, Google launched a crowd-innovation challenge in 2008 to ask the crowd ideas that have the potential to change the world in their “Project 10^100”. After receiving over 150,000 submissions, thousands of Google employees reviewed the ideas to pick a winner, which took nearly two years and tens of thousands of dollars [6]. As previous research suggests only about 10–30% of the ideas from crowdsourcing engagements are considered valuable. Furthermore, screening this vast amount of

Multikonferenz Wirtschaftsinformatik 2018,
March 06-09, 2018, Lüneburg, Germany

contributions to identify the most promising ideas is one of the toughest challenge of crowdsourcing to date [7].

To solve these problems, different streams of research emerged that attempt to filter ideas [8]. First, expert evaluations, which use executives within the firm to screen ideas, were identified as costly and time consuming [9]. Second, research on algorithmic approaches proved to be a valuable way by identifying metrics to distinguish between high- and low-quality ideas [10-12]. However, such filtering approaches always risk missing promising ideas by identifying "false negatives" (classifying good ideas as bad ones) and are rather capable to cull low quality ideas than identifying valuable ones, which is a task that demands human decision makers. In response to this, the third approach to screen ideas is crowd-based evaluation [8-9, 13]. Organizations have turned to the crowd to not just for generating ideas but also to evaluate them to filter high quality contributions. This way has in fact shown to be of same accuracy such as expert ratings, if the members of the crowd have suitable domain knowledge [14]. However, this approach frequently fails in practice, when facing huge amounts of ideas. Crowd-based filtering approaches tend to perform poorly as they make unrealistic demands on the crowd regarding their expertise, time, and cognitive effort [8].

To address the aforementioned shortcomings, we propose the following research question: *How should idea filtering approaches be designed that improve the identification of valuable ideas in large scale crowdsourcing engagements?*

By combining algorithmic machine learning approaches with human evaluation to adaptively assign crowd members that have the required domain knowledge to ideas, we propose a semi-automatic approach that leverages the benefits of both approaches and overcomes limitations of previous research. We thus propose that a hybrid approach is superior to sole crowd-based and computational evaluation for two reasons: First, various research suggests that computational models (or machines) are better at tasks such as information processing and provide valid results [15], while human decision makers are cognitively constrained or biased [16]. Additionally, previous research shows the importance of human decision makers in the context of innovation [17]. In this highly uncertain and creative context, decision makers can rely on their intuition or gut feeling [18].

Following a design science approach [19-20], we so far identified awareness of real-world problems in the context of filtering crowdsourcing contributions and derived design principles for such systems, which we evaluated with experts on crowdsourcing and requirement engineering. We then explain our further progress of research and how we plan to implement and evaluate our proposed filtering approach ex post.

We, therefore, intend to extend previous research on idea filtering in crowdsourcing engagements through combining algorithmic and crowd-based evaluation. This research therefore will contribute to both descriptive and prescriptive knowledge [21-22], which may guide the development of similar solutions in the future.

2 Related Work

Idea Contests

In general, crowdsourcing denotes a mechanism that allows individuals or companies, who face a problem to openly call upon a mass of people over the web to provide potentially valuable solutions. One instantiation of crowdsourcing that seems to be particularly interesting from both a practical and a research perspective are idea contests [1,23-25]. Idea contests are usually conducted via platforms that allow companies to collect ideas from outside the organization. The output (i.e. the ideas) of such contests are usually artefact ideas that can take on different forms such as plain text, plans, designs and predictions from both experts and lay crowds [25]. The basic idea behind idea contests is thereby for companies to expand the solution space to a problem and thereby increasing the probability to obtain creative solutions to said problem [3,25]. The effectiveness of idea contests is also underpinned by research showing that only under certain conditions users are willing, as well as capable to come up with innovative ideas [24,26]. Thus, by providing various incentives such as monetary rewards, firms increase the number of contributions and the probability to receive a creative submission [27]. In simple terms attracting larger crowds leads to a more diverse set of solutions [28-29].

Previous Approaches to Identify Valuable Ideas

Such idea contests lead to a high number of ideas that cannot be efficiently processed by current approaches. Thus, successful idea contests often lead to a flood of contributions that must be screened and evaluated before they can be moved to the next stage and further developed [7]. To identify valuable contributions that are worth implementing, one important task is filtering the textual contributions in such idea contests. Existing filtering approaches to separate valuable from bad contributions in crowdsourcing mainly apply two content-based filtering approaches to evaluate the creative potential of ideas: computational, algorithmic evaluation approaches and crowd-based evaluation approaches [8].

Computational Evaluation Approaches

One current approach to evaluate textual contributions in the context of crowdsourcing is computational evaluation, wherein algorithms are used to filter ideas based on metrics for idea quality such as word frequency statistics [10]. Within the approaches for computational evaluation, two dominant approaches are emerging to support the decision making of the jury, which reviews the ideas to identify the most valuable ones.

First, clustering procedures examine how the vast amount of textual data from crowdsourcing contributions can be organized based on topics [10] or domain-independent taxonomy for idea annotation [11]. Second, machine learning approaches can be used to filter ideas based on rules that determine the value of the content [12, 30]. This approach is particularly useful if training data sets are available. Previous research in this context uses variables for contextual (e.g., length, specificity, completeness, writing style) or representational (e.g., readability, spelling mistakes) characteristics [12] as well as crowd activity (e.g., likes, page views, comments), and behavior of the contributor of the idea (e.g. date of submission, number of updates) [30] to determine the value of crowdsourcing contributions.

Crowd-based Evaluation Approaches

The second approach to evaluate crowdsourcing contributions is applying crowd-based evaluation approaches. In this context, members of the crowd evaluate contributions individually and the results are aggregated [8-9]. Such users might include other users of the contest, or even paid crowds on crowd work platforms [31] that are asked to evaluate ideas from the crowdsourcing engagement.

Previous research on crowd-based evaluation examined the applicability of one or multiple criteria [32] in voting mechanism (where users vote for valuable ideas) [33], ranking approaches (where members of the crowd rank submissions) [34], and rating mechanisms (where the crowd score ideas) [9]. Moreover, prediction markets can be used where users trade ideas by buying and selling stocks to identify the most valuable idea by aggregating this trades as a stock price [35]. Depending on the context of evaluation settings, these approaches proved to be equally accurate compared to the evaluation of experts [14].

3 Methodology

For resolving the above-mentioned limitations, we conduct a design science research (DSR) project [19-20, 22] to design a new and innovative artifact that helps to solve a real-world problem. To combine both relevance and rigor we use inputs from the practical problem domain and the existing body of knowledge (rigor) for our research project [36]. Abstract theoretical knowledge thus has a dual role. First, it guides the suggestions for a potential solution. Second, the abstract learnings from our design serve as prescriptive knowledge to develop other artefacts that address similar problems in the future [21]. To conduct our research, we followed the iterative design research cycle methodology interpretation of [37] as illustrated in figure 1.

So far, we analyzed the body of knowledge on collective intelligence, idea contests, and crowd-based evaluation as well as computational filtering approaches and identified five theory-driven problems of current idea filtering approaches that adversely affects evaluation accuracy. These problems represent the starting point for our solution design. Based on deductive reasoning, we derived five design principles for a potential solution that we evaluated in an ex-ante criteria-based evaluation with experts in the field of community- and service -engineering [38]. In the next steps, we will develop a prototype version of the novel filtering technique and implement it within the context of an idea contest. By conducting an A/B-test to compare the accuracy of our filtering approach against current filtering approaches [30], we intend to evaluate our proposed design. This also constitutes our summative design evaluation [38]. We will, therefore, use a consensual assessment of experts as baseline [9]. Finally, the abstract learning from our design will provide prescriptive knowledge in the form of principles of form and function for building similar artefacts in the future [21].

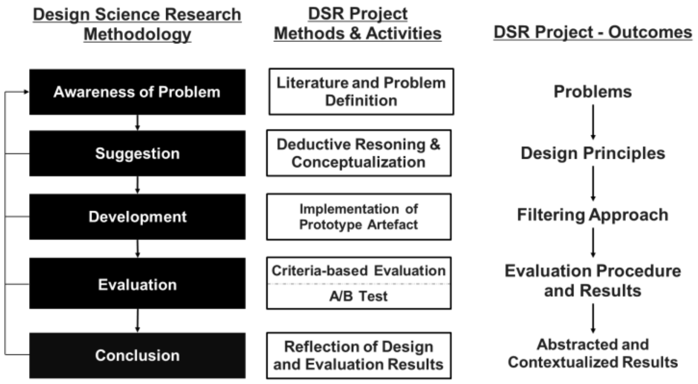

Figure 1. Research Design

4 Awareness of Limitations of Computational and Crowd Approaches

One solution that is currently employed in idea contests is shortlisting. Shortlisting can be considered as an algorithmic solution with the aim to shortlist the best ideas. In doing so shortlisting algorithms often face a tradeoff between specificity and sensitivity. Thus, if such algorithms are not balanced out (i.e. they are too specific, or they are too sensitive) this may lead to ideas being shortlisted that are not innovative (i.e. the algorithm might include false positives) or to promising ideas not being shortlisted (i.e. the algorithm might favor false negatives). In both cases this might lead to unfavorable results such as ideas that are labelled as innovative when in fact they are not truly innovative ideas (**Problem 1**).

One limitation of previous crowd-based evaluation approaches is the cognitive load associated with the volume and variety of idea contributions in crowdsourcing [8]. As cognitive load increases, users in the crowd may become frustrated [39] make low quality decisions [9] or simply deny evaluating ideas. Such load may arise due to the complexity of the evaluation mechanism itself (e.g., prediction markets) and the increasing time and cognitive complexity demands for the raters. Moreover, the information overload in which cognitive processing capacity is exceeded by the volume and diversity of the crowdsourcing contributions makes it difficult for the crowd to evaluate each idea especially when the proposals are complex, such as in the context of innovation problems. Thus, users need to judge manifold, diverse, maybe even paradox ideas with a high degree of novelty. This cognitive load renders previous approaches of crowd-based evaluation problematic for use in practice, where the number of contributions is large (**Problem 2**).

Furthermore, contributions will vary in their textual representation such as writing style, schema, or language which accelerates the cognitive demands on the crowd. Consequently, in practice only a small number of contributions are evaluated. These contributions and their (positive) evaluations then create an anchoring effect [40-41] and will socially influence other decision makers in the crowd [42]. Generally, the ones that are presented on the top of the page and have been positively evaluated by peers a priori, which creates (potentially negative) information cascades [8] (**Problem 3**).

Another major problem in crowd-based evaluation methods so far is that not all users in an idea contest are necessarily capable to evaluate ideas. Therefore, the crowd-based evaluation results might not be a proxy for expert ratings, if users do not have the required expertise for being a "judge" [14], [43-44]. This is particularly problematic when crowdsourcing contributions are complex and diverse. Although previous research highlighted the requirements on the crowd for evaluating ideas, the bottleneck of domain expertise is almost neglected in both theory and practice. To be appropriate for identifying valuable ideas and improving decision quality and predictions in idea filtering, a user should also be an expert in the field [25,45]. Therefore, the crowd should combine both problem knowledge as well as solution knowledge [4,46], which are crucial in the evaluation of innovation. While knowledge about the problem domain might be assumed for users that contribute an idea to a specific problem call, the variety of submitted solutions might be enormous as each diverse solver within the crowd deeply know different parts of the potential solution landscape [24]. Therefore, not every user in the crowd is equally appropriate to evaluate a certain idea due to limited domain knowledge of each part of the solution space submitted, which represents a major weakness of previous approaches in crowd-based evaluation (**Problem 4**).

5 Suggestion and Development of Design Principles for a Hybrid Filtering Approach

To overcome the limitations of previous approaches and to define objectives for a potential solution, we combine algorithmic approaches from machine learning with crowd-based evaluation approaches rather than treat them as substitutes. This approach enables our solution to support the human judge by using machine learning algorithms that identify the expertise of a crowd user, the expertise requirements for evaluating a specific crowdsourcing contribution, and match both to gather more reliable results in identifying valuable contributions. Our proposed design principles (DP)mainly focuses on improving the idea evaluation phase in innovation contests (see Figure 1).

First, the expertise requirements for each textual contribution needs to be identified to match it with suitable members of the crowd [44]. Therefore, the hybrid filtering approach should extract topical features (i.e. latent semantics) to identify the knowledge requirements for potential judges. Thus, we propose:

DP1: *Filtering crowdsourcing contributions should be supported by approaches that extract solution knowledge requirements from textual idea contributions within an idea contest by identifying relevant themes.*

In the next step, the hybrid filtering approach needs to consider the expertise of a crowd participant [47]. One source of such expertise description is the user profile, which includes the self-selected proficiency of a participant. Thus, we propose:

DP2: *Filtering crowdsourcing contributions should be supported by approaches that screen user profiles to extract expertise.*

Apart from the expertise description in the users´ profile (i.e. static), crowd participants gain ability through their activity (i.e. dynamic) in idea contests over time. Users constantly learn through their own contributions [48]. This needs to be additionally considered for the hybrid filtering approach. Moreover, this offers the possibility to ensure that users have really expertise in a topic as they proved it by making contributions. In contrast, expertise descriptions in user profiles might be biased due to overconfidence. Thus, we propose:

DP3: *Filtering crowdsourcing contributions should be supported by approaches that extract solution expertise from users´ prior textual idea contributions across idea contests by identifying relevant themes.*

Idea contest are highly dynamic [7]. To match crowd participants with suitable ideas for evaluation, the expertise profiles of each user need to be dynamic [49]. This means it should constantly update the expertise of a user through dynamically updating the abstract user profile based on the input and contributions of a crowd participant. Contributions include both past idea proposals, as well an idea quality indicator (i.e. the corresponding idea rating) Thus, we propose:

DP4: *Filtering crowdsourcing contributions should be supported by approaches that create adaptive user profiles containing expertise extracted from the user profile and prior contributions.*

As the evaluation quality of the crowd is highly dependent on the ability of each individual member of the crowd [44], in the last step the hybrid filtering approach needs to match crowdsourcing contributions with suitable users. Previous work on such select crowd strategies in the field of psychology suggests that approximately five to ten humans are required to benefit from the aggregated results of evaluation [43-44]. This sample size is most suitable for leveraging the error reduction of individual biases as well as the aggregation of diverse knowledge. Thus, we propose:

DP5: *Filtering crowdsourcing contributions should be supported by approaches that match solutions with users that have the required expertise and assign textual contributions to this user for evaluation.*

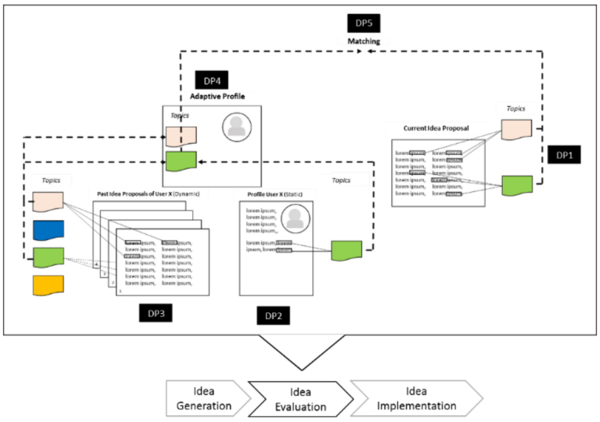

Figure 2. Proposed Hybrid Filtering Approach

6 Ex-Ante Evaluation of Design Principles

For the criteria-based ex-ante evaluation of our artifact (DP1 -DP5) we conducted an online-survey with experts on community and systems engineering. The criteria for our evaluation were derived from [38]. Specifically, we made use of the following criteria: completeness, understandability, fidelity with real world, applicability, level of detail, internal consistency, clarity. Hence, experts were asked to evaluate each of the proposed design principles based on the aforementioned criteria. Table 1 provides the results to our criteria-based ex-ante evaluation.

Table 1: Results of ex-ante design principles evaluation

criteria	*N*	*DP1 Mean (SD)*	*p*	*DP2 Mean (SD)*	*p*	*DP3 Mean (SD)*	*p*	*DP4 Mean (SD)*	*p*	*DP5 Mean (SD)*	*p*
completeness	13	4.85 (1.07)	.0017	5.38 (.96)	.000111	5.32 (.93)	.00022	5.23 (0.83)	.00009	4.69 (.75)	.003055
understand-ability	13	5.31 (.85)	.000066	5.15 (1.14)	.001694	4.92 (.76)	.000448	4.77 (1.17)	.017381	5.38 (1.12)	.000396
fidelity	13	5.23 (1.17)	.001242	4.15 (1.07)	.001074	5.54 (1.05)	.000097	4.77 (1.30)	.027277	5.15 (.80)	.000111
applicability	13	5.15 (1.21)	.002493	5.31 (0.58)	.000066	5.0 (1.0)	00179	5.00 (1.21)	.007045	4.85 (1.21)	.013704
level of detail	13	5.15 (1.14)	.001694	5.32 (1.17)	.001242	5.08 (.95)	.000777	5.08 (1.19)	.003352	5.23 (.93)	.00022
internal consistency	13	5.58 (.79)	.00001	5.46 (0.97)	.000074	5.46 (1.27)	.000661	4.85 (0.69)	.000411	5.23 (1.09)	.000791
clarity	13	5.15 (.90)	.00029	5.46 (1.27)	.000661	5.0 (1.08)	.002944	5.15 (1.34)	.004681	5.33 (.98)	.000262

Our results suggest that the majority of our design principles score relatively high in terms of the proposed criteria (i.e. the means of our criteria to evaluate DP1-DP5 are found on the upper bound of a seven-point Likert-scale). This is also supported by the p-values indicating that the scores of our criteria are all significantly different from the mean. Based on these results, we conclude that our design principles are clear and concise to warrant further development and refinement of our idea filtering approach.

7 Further Work and Summative Evaluation

As we proceed, we will develop and implement our hybrid filtering approach within the context of an idea contest on an existing crowdsourcing platform. To identify required solution knowledge and users´ expertise, we will design a machine learning algorithm based on probabilistic topic modelling [51]. Topic modeling is a text mining approach that uses Latent Dirichlet Allocation (LDA) [50] as unsupervised statistical learning method to discover abstract "topics" in text documents.

We then automatically match users to current idea proposals based on the proximity of topics extracted from the static user profiles and their past idea contributions [51-52]. The developed filtering approach will then be evaluated. We will therefore conduct two A/B-tests to compare our filtering approach against current filtering approaches. Our first comparison will include our hybrid filtering approach (A) against a computational filtering approach (B1). Our second comparison will include our hybrid filtering approach against a crowd-based filtering approach (B2) [30]. For our filtering approach, we will match each idea with approximately five users and then ask for evaluation on a rating scale [13] and then combine the evaluations to derive the average (i.e. mean) [44]. The performance of both filtering approaches will then be evaluated against a baseline. For constructing our baseline, we used the most commonly approach of expert evaluation through consensual assessment technique, which combines the consensus-based classification of a crowdsourcing contribution through several domain experts [9]. Figure 3 displays our planned evaluation procedure.

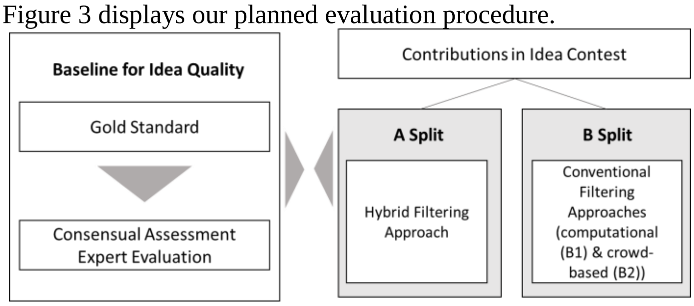

Figure 3. Proposed Evaluation Procedure

The performance of filtering approaches will be assessed through its accuracy. Therefore, we calculate the true positive (correctly identified high quality idea related to baseline) and the false positive (incorrectly identified high quality idea related to baseline) for each filtering approach. The receiver operating characteristic (ROC) plots the true positive rate against the false positive rate. The area under this curve then provides a measure for accuracy. A perfectly accurate filter approach would have an area of 1.0 [8].

8 Conclusion

Our research introduces a novel filtering approach that combines the strengths of both machines and humans in evaluating creative ideas by using machine learning approaches to assign the right user with the required solution knowledge to a corresponding idea. To this end, we propose tentative design principles that we validated in the field with experts on crowdsourcing and system engineering. To the best of our knowledge, this is the first study that takes this topic into account. Our research offers a novel and innovative solution for a real-world problem and contribute to the body of knowledge on idea filtering for open innovation systems by considering the required expertise of crowd evaluations [43-44]. We, therefore, intend to extend previous research on idea filtering in crowdsourcing engagements through combining algorithmic and crowd-based evaluation.

References

1. Leimeister, J.M., Huber, M., Bretschneider, U., Krcmar, H.: Leveraging crowdsourcing: activation-supporting components for IT-based ideas competition. Journal of Management Information Systems 26, pp. 197–224 (2009)
2. Afuah, A., Tucci, C. L.: Crowdsourcing as a solution to distant search. Academy of Management Review 37, pp. 355–375 (2012)
3. Lakhani, K. R., Jeppesen, L. B.: Getting unusual suspects to solve R&D puzzles. Harvard Business Review 85, pp. 30–32 (2007)
4. Toubia, O., Florès, L.: Adaptive idea screening using consumers, Marketing Science 26, pp. 342–360 (2007)
5. Bjelland, O. M., Wood, R. C.: An inside view of IBM's' Innovation Jam'. MIT Sloan Management Review 50, p. 32 (2008)
6. Buskirk, E. V.: Google Struggles to Give Away $10 Million. Wired Magazine (2010)
7. Blohm, I., Leimeister, J.M., Krcmar, H.: Crowdsourcing: how to benefit from (too) many great ideas. MIS Quarterly Executive 12, pp. 199–211 (2013)
8. Klein, M., Garcia, A.C.B.: High-speed idea filtering with the bag of lemons. Decision Support Systems 78, pp. 39–50 (2015)
9. Blohm, I., Riedl, C., Füller, J., Leimeister, J.M.: Rate or trade? identifying winning ideas in open idea sourcing. Information Systems Research 27, pp. 27–48 (2016)
10. Walter, T.P., and Back, A.: A text mining approach to evaluate submissions to crowdsourcing contests. Proceedings of the Hawaii International Conference on Systems Science (2013)
11. Westerski, A., Dalamagas, T., Iglesias, C.A.: Classifying and comparing community innovation in Idea Management Systems. Decision Support Systems 54, pp. 1316–132 (2013)
12. Rhyn, M., Blohm, I.: A Machine Learning Approach for Classifying Textual Data in Crowdsourcing. Proceedings of the WI 2017 (2017)

13. Riedl, C., Blohm, I., Leimeister, J.M., Krcmar, H.: The effect of rating scales on decision quality and user attitudes in online innovation communities. International Journal of Electronic Commerce (17:3), pp. 7–36 (2013)
14. Magnusson, P.R., Wästlund, E., Netz, J.: Exploring users' appropriateness as a proxy for experts when screening new product/service ideas. Journal of Product Innovation Management 33, pp. 4–18 (2016)
15. Nagar, Y., Malone, T.: Making business predictions by combining human and machine intelligence in prediction markets. Proceedings of the International Conference on Information Systems (2011)
16. Kahneman, D., Tversky, A.: Variants of uncertainty. Cognition 11, pp. 143–157 (1982)
17. Kornish, L.J., Ulrich, K.T.: The importance of the raw idea in innovation: Testing the sow's ear hypothesis, Journal of Marketing Research 51, pp. 14–26 (2014)
18. Huang, L., Pearce, J.L. Managing the unknowable: The effectiveness of early-stage investor gut feel in entrepreneurial investment decisions. In: Administrative Science Quarterly 60, pp. 634–670 (2015)
19. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design science in information systems research. In: MIS Quarterly 28, pp. 75–105 (2004)
20. Peffers, K., Tuunanen, T., Rothenberger, M.A., and Chatterjee, S.: A design science research methodology for information systems research. Journal of management information systems 24, pp. 45–77 (2007)
21. Gregor, S., Jones, D.: The anatomy of a design theory. Journal of the Association for Information Systems 8, p. 312 (2007)
22. Gregor, S., Hevner, A.R. Positioning and presenting design science research for maximum impact. MIS Quarterly 37, pp. 337–355 (2013)
23. Boudreau, K.J., Lakhani, K.R.: Using the crowd as an innovation partner. Harvard Business Review 91, pp. 60–69 (2013)
24. Jeppesen, L.B., Lakhani, K.R.: Marginality and problem-solving effectiveness in broadcast search. Organization Science 21, pp. 1016–1033 (2010)
25. Terwiesch, C., Xu, Y.: Innovation contests, open innovation, and multiagent problem solving. Management science 54, pp. 1529–1543 (2008)
26. Poetz, M.K., Schreier, M.: The value of crowdsourcing: can users really compete with professionals in generating new product ideas? Journal of Product Innovation Management 29, pp. 245–256 (2012)
27. Bayus, B.L.: Crowdsourcing new product ideas over time: An analysis of the Dell IdeaStorm community. Management science 59, pp. 226–244 (2013)
28. Malone, T.W., Laubacher, R., Dellarocas, C.: Harnessing crowds: Mapping the genome of collective intelligence. MIT Working Paper Series (2009)
29. Surowiecki, J.: The wisdom of crowds: Anchor. (2005)
30. Nagar, Y., Boer, P. de, Bicharra Garcia, A.C.: Accelerating the Review of Complex Intellectual Artefacts in Crowdsourced Innovation Challenges. Proceedings of the International Conference on Information Systems (2016)
31. John, T.: Supporting Business Model Idea Generation Through Machine-generated Ideas: A Design Theory. Proceedings of the International Conference on Information Systems (2016)
32. Dean, D.L., Hender, J.M., Rodgers, T.L., Santanen, E.: Identifying good ideas: constructs and scales for idea evaluation (2006)
33. Bao, J., Sakamoto, Y., Nickerson, J.V.: Evaluating Design Solutions Using Crowds, Proceedings of the Americas Conference on Information Systems (2011)
34. Salganik, M.J., Levy, K.E.C. Wiki surveys: Open and quantifiable social data collection (2012)

35. Soukhoroukova, A., Spann, M., & Skiera, B.: Sourcing, filtering, and evaluating new product ideas: An empirical exploration of the performance of idea markets. Journal of Product Innovation Management 29, pp. 100-112 (2012).
36. Hevner, A. R. A three-cycle view of design science research. Scandinavian Journal of Information Systems 19, p. 4 (2007)
37. Wagenknecht, T., Crommelinck, J., Teubner, T., Weinhardt, C.: When Life Gives You Lemons: How rating scales affect user activity and frustration in collaborative evaluation processes. Proceedings of the WI 2017 (2017)
38. Vaishnavi, Vijay K., and William Kuechler. Design science research methods and patterns: innovating information and communication technology. Crc Press, (2015)
39. Sonnenberg, C., & vom Brocke, J.: Evaluations in the science of the artificial–reconsidering the build-evaluate pattern in design science research. Design Science Research in Information Systems. Advances in Theory and Practice, pp. 381-397 (2012).
40. Tversky, A., Kahneman, D.: Availability: A heuristic for judging frequency and probability. Cognitive Psychology 5, pp. 207–232 (1973)
41. Larrick, R.P., Mannes, A.E., Soll, J.B., Krueger, J.I.: The social psychology of the wisdom of crowds. Social Psychology and Decision Making, pp. 227–242 (2011)
42. Deutsch, M., Gerard, H.B.: A study of normative and informational social influences upon individual judgment. The Journal of Abnormal and Social Psychology 51, p. 629. (1955)
43. Mannes, A.E., Soll, J.B., Larrick, R.P.: The wisdom of select crowds. Journal of Personality and Social Psychology 107, p. 276 (2014)
44. Keuschnigg, M., Ganser, C.: Crowd Wisdom Relies on Agents' Ability in Small Groups with a Voting Aggregation Rule. Management Science 63, pp. 818–828 (2017)
45. Amabile, T.: Creativity in context: Westview Press (1996)
46. Hippel, E. von.: "Sticky information" and the focus of problem solving: implications for innovation. Management Science 40, pp. 429–439 (1994)
47. Kulkarni, A., Narula, P., Rolnitzky, D., & Kontny, N.: Wish: Amplifying creative ability with expert crowds. In Second AAAI Conference on Human Computation and Crowdsourcing 2014.
48. Faraj, S., Krogh, G. von, Monteiro, E., Lakhani, K.R.: Special Section Introduction—Online Community as Space for Knowledge Flows. Information Systems Research 27, pp. 668–684 (2016)
49. Yuan, A., Luther, K., Krause, M., Vennix, S. I., Dow, S. P., & Hartmann, B.: Almost an expert: The effects of rubrics and expertise on perceived value of crowdsourced design critiques. In Proceedings of the 19th ACM Conference on CSCW & Social Computing, pp. 1005-1017 (2016)
50. Blei, D.M.: Probabilistic topic models. Communications of the ACM 55, pp. 77–84 (2012)
51. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. Journal of Machine Learning Research 3, pp. 993–1022 (2003)
52. Shi, Z.M., Lee, G.M., Whinston, A.B.: Toward a Better Measure of Business Proximity: Topic Modeling for Industry Intelligence. MIS Quarterly 40, pp. 1035–1056 (2016)