creativity and innovation management

How good are ideas identified by an automatic idea detection system?

Journal:	*Creativity and Innovation Management*
Manuscript ID	Draft
Manuscript Type:	Article
Keywords:	Idea quality, Text mining, Machine learning, Natural language processing, Big data, Online communities, Support vector machines, Crowdsourcing

How good are ideas identified by an automatic idea detection system?

Abstract

Online communities are an attractive source of potential ideas for products and process'. Recent advances in machine learning have made it possible to screen the vast amounts of information in online communities and automatically detect user-contributed ideas. However, it is still uncertain whether the ideas identified by such a system will also be regarded as sufficiently novel, feasible and valuable by firms who might decide to develop them further. A validation study is reported in which 200 posts were extracted from an online community using the automatic idea detection system by Christensen, Nørskov, Frederiksen and Scholderer (2016; DOI: 10.1111/caim.12202). Two company professionals evaluated the posts in terms of idea content and idea quality. The results suggest that the automatic idea detection system is sufficiently valid to be deployed for the harvesting and initial screening of innovation ideas and that the profile of the identified ideas (in terms of novelty, feasibility and value) follows the same pattern identified in studies of user ideation in general.

Introduction

The digitalisation of business life is progressing: more and more tasks can be solved by automated systems. Whilst in the past, these were predominantly tasks of a mundane and repetitive nature, recent advances in artificial intelligence have also made it possible to solve complex problems. A common problem during the introduction of such systems is that they can be intransparent to their prospective users. Whilst the traditional business processes they are intended to rationalise have often been in use for many years and are implicitly trusted by management and staff, newly introduced automated systems lack such a track record. Scepticism and reactance can be the consequence.

To earn the trust of prospective users, automated systems have to enable superior performance. Benchmarked against the traditional business processes they are intended to rationalise, they should lead to increases in effectiveness or efficiency. This is easily demonstrated in application areas such as sales forecasting or inventory control where commonly accepted and routinely measured performance criteria exist. Such criteria rarely exist in more complex and creative areas such as innovation management. The aim of the research presented here is to show how the performance of automated systems in such areas can be evaluated. We will demonstrate this in the context of a particular type of task: the automated detection of ideas for product and process innovations in the contributions to an online developer forum.

Online communities as idea reservoirs

Firms need a continuous stream of ideas to fuel their innovation processes (Van de Ven, 1986; Ekvall, 1997; Vandenbosch, Saatcioglu, & Fay, 2006; van den Ende, Frederiksen, & Prencipe, 2015). Ideas do not have to originate from the creative mind of the firm's employees but can also originate from the users of its products, services and technologies

(Kristensson, Gustafsson, & Archer, 2004; Magnusson, 2009; von Hippel, Ogawa, & PJ de Jong, 2011; Poetz & Schreier, 2012; Majchrzak & Malhotra, 2013; Magnusson, Wästlund, & Netz, 2014). Online communities where users exchange experiences and discuss potential improvements are a particularly rich reservoir of ideas for product and process innovations.

Prominent examples include the user communities hosted by Dell (di Gangi, Wasko, & Hooker, 2010; Poetz & Schreier, 2012), Lego (Antorini, 2007; Antorini, Muñiz, & Askildsen, 2012; Nørskov, Antorini, & Jensen, 2015), Propellerhead (Jeppesen & Frederiksen, 2006) and IBM (Mahr & Lievens, 2012). Firm-hosted communities such as these have the advantage that the hosting firm can retain a certain degree of control. The communities are typically based on software that allows registered users to post ideas, comment on and vote for ideas posted by other users in a highly structured manner. The downside of this approach is that it requires an extensive base of committed product users or firm-loyal customers who have an intrinsic interest in suggesting ideas to the firm.

However, users do not only gather in firm-hosted communities. A vast amount of online communities exists that are firm-free (Füller, Bartl, Ernst, & Mühlbacher, 2006; Füller, Jawecki, & Mühlbacher, 2007). The most prominent cases include open-source software development communities such as those responsible for the Linux kernel, R and Python. These are examples of firm-free "products" that have been developed in a distributed manner, utilising online collaboration platforms such as GitHub and Sourceforge. The fact that the resulting products are now perfectly able to compete with their commercial counterparts (such as the products ranges of the SAS Institute or Microsoft) is a clear demonstration of the potential of such communities (von Krogh, Spaeth, & Lakhani, 2003; von Krogh & von Hippel, 2006)

The problem with firm-free communities is that they, unlike most firm-hosted communities, are usually *not* based on a crowdsourcing architecture that would enable easy harvesting and collaborative filtering of the community-generated ideas. Assigning employees to manual monitoring of community contributions is often the only viable solution if firms want to benefit from the ideas generated in firm-free communities. This is time-consuming and expensive; online communities may contain several hundred thousand posts and comments. The sheer amount of information in which the ideas are hidden is a practical barrier to finding the ideas and utilising them for innovation (Lin, Hsieh, & Chuang, 2009; Thorleuchter & Van den Poel, 2013).

Automatic idea detection

A new and efficient way of solving the needle-in-a-haystack problem is to use classification algorithms that can screen arbitrary amounts of community posts and comments and identify those that are likely to contain ideas. Using natural language processing and machine learning methods, Christensen, Nørskov, Frederiksen, & Scholderer (2016) develop such an algorithm and demonstrate its classification performance and efficiency for the case of extracting new product ideas from an online community related to Lego. Christensen et al. (Submitted manuscript) show that the same principles can be applied to extract ideas for innovations from a community related to craft brewing.

The authors argue that their method is applicable across different technological areas and product categories because most people use a specific set of words and expressions when they communicate ideas to each other. Since the presence of such linguistic markers can easily be detected in a given post or comment, it can also be exploited for the screening of arbitrarily large collections of posts, comments or other types of semi- or unstructured text. Implemented as a screening tool in a firm's R&D or marketing department, it can significantly

reduce the labour costs that would arise if R&D staff were assigned to manual monitoring of community activity.

How good are automatically detected ideas?

A crucial question is whether the ideas detected by such an automated system would also be seen as sufficiently novel, feasible and valuable by the R&D or marketing staff who would have to decide if the ideas should be taken further (e.g., developed into concepts or prototypes). Ideas identified by the Christensen et al. (2016) method, for example, have not yet been evaluated by company-internal R&D or marketing staff. The aim of the present paper is to fill this gap. Specifically, we would like to contribute in two respects to the literature:

- Our first contribution is to assess whether ideas from an online community, identified by an artificial intelligence system such as the one described by Christensen et al. (2016), will also be perceived as ideas by company-internal staff.

- Our second contribution is to investigate if the ideas that are detected by the system will also be perceived as *good* ideas by company-internal staff.

These issues reflect potential acceptance problems that were in the innovation literature initially seen as general barriers for the uptake of user-contributed ideas by companies. Since then, many studies have demonstrated that user-contributed ideas can often compete with the ideas generated by company-internal staff (e.g., see (Kristensson et al., 2004; Magnusson, 2009; Poetz & Schreier, 2012; Magnusson et al., 2014) and therefore deserve to be given a fair chance. As a consequence, dedicated crowdsourcing systems have gained widespread acceptance in the business community. Our study extends this question to the *mode* of idea harvesting: can user-contributed ideas identified by an artificial intelligence system reach sufficient recognition among company professionals? An online community

related to craft brewing was used as the idea base for our study. Employees of Norwegian craft brewery *Nøgne Ø* evaluated the automatically extracted ideas.

Method

Machine learning for idea detection

The machine learning system we employed is described in detail in Christensen et al. (2016) and Christensen et al. (Submitted manuscript). Although the technical properties of the system are not the central focus of the present paper, we will give a brief description of the system and how it was employed in our study. The machine learning system takes as input idea texts and non-idea texts that have been identified by human raters. The texts used for this study originate from *alt.beer.home-brewing*, a Usenet-based online community related to craft brewing. In this community people from all over the world discuss brewing-related issues. We expected ideas for product and process to be available in this community. At the time the texts were extracted, the community contained altogether 10582 posts. 3000 of these were extracted for the development of the training of the system. Those that contained ideas were identified by via crowdsourcing, using the *CrowdFlower* platform (a service similar to Amazon's *Mechanical Turk*). Five raters were assigned to each text and instructed to label the text as an idea text if it contained at least one idea.

Before the texts could be used for machine learning, several text pre-processing steps were performed. In this process the raw text content was turned into a row-column format, where each text was represented as a row and each term (i.e., each unique word or expression) as a column. In this process, all numbers, punctuation marks and stop words were removed. Uni-grams, bi-grams and tri-grams were generated. All terms that did not occur in at least 0.2% of the texts were omitted from the analysis. This process resulted in a dataset consisting of 10514 terms representing 10582 texts.

The 3000 training texts were separated from the remaining 10582 texts. From the 3000 training texts, we excluded all texts where not all five CrowdFlower raters had agreed on the

class membership. After excluding these, the new training set contained 1393 texts. 405 of the texts were idea texts and 988 were non-idea texts. The training texts were partitioned at random into three separate data sets: a training set (consisting of 70% of the texts), a validation set (15% of the texts) and a hold-out or test set (15% of the texts). Such a partition is essential for the tuning of the machine learning system (in the validation set) and for an unbiased evaluation of its performance in the context of previously unseen data (hold-out set). Based on the training set, validation set and hold-out, the automatic idea detection system was trained and tested. The system was based on a linear support vector machine classifier (SVM; for details, see Christensen et al., 2016). Performance statistics are reported in Table 1.

--- Table 1 ---

From the remaining 7582 texts which had not been involved in the training, validation and testing of the system in the study by Christensen et al. (Submitted manuscript), another 200 were extracted for the present study. Using the SVM classifier, the texts were scored as to how likely they were to contain an idea. A histogram of the resulting posterior probabilities is shown in Figure 1. These 200 texts were then used in the present study as the idea and non-idea texts to be classified and rated by two brewing professionals.

--- Figure 1 ---

Measuring idea quality

The perceived quality of an idea can depend on the perspective of the person evaluating the idea. This topic has received much attention in the creativity and innovation management literature. In principle, idea quality could be measured on a "good idea" to "bad idea" scale, but in most research it is decomposed into several attributes that represent conceptually distinct dimensions of quality. Dean, Hender, Rodgers, & Santanen (2006)

provide a comprehensive review of the idea quality literature published between 1990 and 2005. Based on the altogether 90 identified studies, they suggest that four dimensions of idea quality can be distinguished: novelty, workability, relevance and specificity. An idea is novel if it contains something that is new. An idea is workable if it is easy to implement and does not violate known constraints. An idea is relevant if it satisfies pre-defined goals. An idea is specific if it has been worked out in detail.

Comparable sets of sub-dimensions have been suggested in the user innovation literature. Kristensson, Gustafsson and Archer (2004) compared the ideation performance of ordinary users, expert users and professionals. They used three quality attributes: originality (comparable to the novelty dimension suggested by Dean et al., 2006), realisability (comparable to the feasibility dimension) and value (comparable to the relevance dimension). In a similar study, Magnusson (2009) compared the ideation performance of professionals, technically skilled users, ordinary users, consulting users and creativity-trained ordinary users. He used the quality attributes originality (comparable to novelty), producibility (comparable to feasibility) and user-value (comparable to relevance). Using the same attributes, Magnusson et al. (2014) compared technically skilled users with technically naïve users. Poetz & Schreier (2012) compared the ideas of users and professionals in terms of the attributes novelty, feasibility and customer benefit (comparable to value). Based on the four studies that have a product user ideation focus, we chose novelty, feasibility and value as the quality attributes for our study.

Procedure

We established contact with Norwegian craft brewery *Nøgne Ø*. The brewery was founded in 2002 by two Norwegian home brewers and is nowadays part of Norwegian brewery group Hansa Borg Bryggerier. In 2015, *Nøgne Ø* produced 30 different styles of ales

and exported to more than 40 markets. Two company professionals were recruited as expert raters. Expert 1 was 29 years old, female and had a business school background. Her responsibilities at *Nøgne Ø* were sales and logistics. At the time the study was conducted, she had been working for the brewery for 12 years. Expert 2 was 40 years old, male and had an engineering background. His responsibilities at *Nøgne Ø* were related to marketing and the web shop. At the time the study was conducted, he had been working for the brewery for 4.5 years.

The experts evaluated the 200 texts one-by-one and independently from each other. First, the experts were instructed to read the respective text carefully. Then, they were asked *"Please evaluate if you think that the text contains one or more ideas"* and to respond on a binary *"yes"* versus *"no"* scale. If the expert had responded *"yes"*, three rating scales were presented on which the expert was asked to evaluate the quality of the idea in terms of the three attributes novelty, feasibility and value. The scales were horizontally aligned ranging from very low (1) to very high (10). The instruction for the novelty attribute was: *"Please valuate the novelty of the idea(s) in the text (by this we mean: to what degree does the idea suggest something new)"*. The instruction for the feasibility attribute was: "*Pleas evaluate the feasibility of the idea(s) in the text (by this we mean: to what degree is it possible to implement the idea)"*. The instruction for the value attribute was: *"Please evaluate the value of the idea(s) in the text (by this we mean: to what degree does the idea solve the underlying problem)"*.

Inter-rater reliability

To assess the inter-rater reliability of the idea/non-idea classification task, we calculated Cohen's kappa, normalised for differences between raters in their marginal distributions (Cohen, 1960; Landis & Koch, 1977; von Eye & von Eye, 2008). The normalised version of kappa takes on values between 0 and 1 where a value of 0 stands for chance-level agreement

and a value of 1 for the theoretical maximum of agreement, given the marginal distributions of the raters. Expert 1 identified 41 texts as containing ideas and 159 as not containing ideas. Expert 2 identified 87 texts as containing ideas and 113 as not containing ideas. They agreed on 35 texts as containing ideas and 107 as not containing ideas (see Table 2 for examples). These counts correspond to a normalised kappa of 0.74, suggesting that there was substantial agreement between the two experts as to whether a given text did or did not contain an idea.

--- Table 2 ---

To assess the inter-rater reliability of the idea quality rating task, we calculated reliability measures based on generalisability theory (Cronbach, Gleser, Nanda, & Rajaratnam, 1972; Brennan, 2001). Only the 69 texts which the machine learning classifier had classified as an idea and which at least one of the brewery professionals had identified as an idea were included in the analysis. The design was a two-facet crossed design with tasks (the three quality attributes) and raters (the two brewery professionals) treated as fixed effects. The reliability (generalisability coefficient) of the averaged rating of a randomly picked idea text on the three attributes by the two raters was $E\rho^2 = 0.71$.

Results

Presence of ideas

Since our two company professionals had not perfectly agreed with each other on the presence or absence of ideas in the texts, we defined two validation criteria: a lenient criterion (Boolean OR: at least one professional had identified the respective text as containing an idea) and a strict criterion (Boolean AND: both professionals had identified the respective text as containing an idea).

Using the lenient criterion as a gold standard (where 47% of the 200 texts would be defined as true idea texts), the automatic idea detection system performed well. The classifier agreed with the company professionals in 77% of the cases as to whether a text did or did not contain an idea (accuracy). 75% of the texts which the classifier had identified as idea texts were also identified as idea texts by the company professionals (precision, also referred to as positive predictive value in the literature). The classifier correctly identified as idea texts 74% of the texts the professionals had identified as ideas (recall, also referred to as sensitivity or true positive rate in the literature). Since precision and recall always represent a trade-off, we also calculated their harmonic mean, the F_1 measure, as a compromise. Using the lenient criterion, it reached a very respectable value of $F_1 = 0.75$. Classification accuracy statistics are reported in Table 3.

Using the strict criterion as a gold standard (where only 18% of the 200 texts would be defined as containing ideas), the automatic idea classification system still agreed with the company professionals in 67% of the cases as to whether a text did or did not contain an idea (accuracy). Due to the much stricter criterion as to what defined an idea text, the precision of the classifier was lower: only 33% of the texts which the classifier had identified as idea texts were also identified as idea texts by the company professionals. For the same reason, recall

was higher: the classifier correctly identified as idea texts 86% of the texts the professionals had identified as ideas. The F_1 measure, as a compromise between precision and recall, reached a value of 0.47.

Taken together, the criterion validity of the automatic idea detection system can be regarded as satisfactory as long as it is used for the screening of potential ideas. Deployed in a company as a tool for filtering out candidate ideas for product and process innovations, it may significantly reduce the time and effort that would otherwise have to be spent by company staff on manual screening and preliminary evaluation of a number of user contributions in potentially relevant online fora.

--- Table 3 ---

Quality of automatically detected ideas

Figure 2 shows the distribution of the quality ratings of the ideas (i.e., those texts that had been identified as ideas by the automatic idea detection system and which had been also been identified as ideas by at least one of the two company professionals). For texts which both company professionals had classified as an idea, the values on the novelty, feasibility and value attributes are the averaged ratings of both company professionals. For texts which only one of the company professionals had identified as an idea, the values are the ratings given by that professional. The overall quality values were calculated as unweighted averages of the ratings on the novelty, feasibility and value attributes.

--- Figure 2 ---

The distribution of the novelty ratings was concentrated in the lower range of the response scale (which had a minimum of 1 and a maximum of 10), the distribution of the feasibility ratings in the upper range of the response scale, and the distributions of the value

ratings and overall quality in the middle of the response scale. The results suggest that, on average, the ideas which the automatic idea detection system extracted from the *alt.beer.home-brewing* community appeared rather feasible to brewery professionals, were not particularly novel, but had medium value and medium overall idea quality.

Discussion and conclusion

The first aim of the present study was to investigate if ideas for product and process innovations detected by an artificial intelligence system (in this case, the one developed by Christensen et al., 2016) would also be regarded as ideas by company-internal staff who will be responsible for taking the ideas further in the innovation process. Our results suggest that this is to a considerable extent the case: the performance of the system can be regarded as sufficient for an initial screening of potential ideas. Deployed in a company as a tool for selecting candidate ideas for product and process innovations, it can significantly reduce the time and effort that would otherwise have to be spent by company staff on wading through a large number of user contributions in potentially relevant online communities.

The exact level of criterion-related validity that our system could achieve depended on several factors. The most important of these are (a) the definition of the "gold standard" against which the predictions are validated and (b) the cut-off used for transforming the continuous posterior probability score generated by the system into a binary prediction. In our analysis, we used two of the possible gold standards: a lenient criterion (at least one of the company professionals had rated the respective text as containing an idea) and a strict criterion (both company professionals had rated the text as containing an idea). The lenient criterion led to an implied base rate of 47% for the target event (i.e., the probability that a randomly chosen text from among the 200 used in the present study would contain an idea), whereas the strict criterion reduced the implied base rate to 18%. It is not possible to define on purely statistical grounds what the right base rate should be. This is complicated by the fact that the two company professionals who served as experts in our study did not have the same base rates in their individual classifications: Expert 1 appeared to use a more conservative

standard of judgment, rating 21% of the 200 texts as containing ideas, whilst Expert 2 appeared to use a more liberal standard, rating 44% of the texts as containing ideas.

Since the two experts also differed in terms of their functional responsibilities in the company, it might not even be appropriate to look for perfect agreement—after all, a company's ability to integrate different functional perspectives is one of the strongest predictors of innovation success (e.g., see Evanschitzky, Eisend, Calantone, & Jiang, 2012). Whether it makes more sense for a given company to use a stricter or more lenient criterion for further filtering of the automatically identified ideas may depend more on strategy and available resources: a lenient criterion may be more appropriate if a company wants to cast its net wide and thereby reduce the risk of missing certain ideas which might not yet be able to achieve full cross-functional consensus. However, the company would also have to be prepared to assign the necessary resources for dealing with the larger number of ideas that would enter the innovation funnel. If, on the other hand, a company wants to limit its resource expenditure and focus on ideas that can already in the early phases achieve cross-functional consensus, a stricter criterion would be appropriate.

A similar objective can be achieved by tuning the cut-off value of the SVM classifier underlying the Christensen et al. (2016) system. The algorithm yields a posterior probability score that is continuous on the (0,1) interval. A traditional way of transforming the posterior probability score into a binary classification is use the value 0.50 as a cut-off such that a text is classified as an idea text if the probability that the text contains an idea, given the support vectors, is larger than 0.50, and classified as a non-idea text otherwise. However, the traditional way of setting the cut-off value may not always be the most useful way. Another heuristic that is typically more useful is to set the cut-off equal to one minus the base rate of the target even, either on the posterior probability scale or on the empirical percentile scale. This heuristic would match the prior probability of classifying a text as an idea to the base rate

of the event. A third way of setting the cut-off is to estimate how many additional ideas a company would be able to absorb into its innovation funnel and to use an appropriate absolute cut-off, selecting the right number of ideas from the top of the posterior probability ranking.

The second aim of the present study was to investigate if the automatic idea detection system developed by Christensen et al. (2016) would extract *good* ideas from the online community that served as an example here. For the online community under investigation, our answer is a qualified yes: the distribution of the overall idea quality score, calculated as the average rating of each idea on the three quality attributes (novelty, feasibility, value) by the two company professionals, was concentrated in the middle of the response scale (mean = 4.8, 25th percentile = 3.8, 50th percentile = 5, 75th percentile = 5.7) and ranged from a minimum of 1 (the lower end of the response scale) to a maximum of 8 (two points below the maximum of the response scale). Overall, the ideas extracted by the automatic detection system appear to have made a reasonable impression on the company professionals.

An interesting detail is that the identified ideas tended to be regarded as more feasible and valuable by our company professionals than they were regarded as novel. This finding reflects results obtained by Kristensson et al. (2004) and Magnusson (2009) for user ideation in general. However, as already observed, agreement between our experts was not perfect here either. As an example, consider the text shown in Table 4: a community member suggests a new mead recipe. Overall, the idea was rated as one of the best by the two company professionals. Expert 1 assigned a rating of 2 on the novelty attribute, 7 on feasibility and 4 on value. Expert 2 rated it 9 on novelty, 9 on feasibility and 9 on value. In the additional, qualitative responses we obtained from the two professionals, it became clear that Expert 1 evaluated the idea in terms of its quality as an idea for process innovation whereas Expert 2 evaluated it in terms of its quality as idea for product innovation. Different perspectives, either due to the functional specialisation of our company professionals or due

to their different levels of experience with the product category, seem to have led to different standards of judgment.

--- Table 4 ---

The results presented here are an evaluation of a particular automatic idea detection system (the one developed by Christensen et al., 2016) to a particular case (the craft brewing community *alt.beer.home-brewing*), evaluated from the point of view of two brewing professionals connected to a particular craft brewing company (*Nøgne Ø*). Naturally, this poses limits to the generalisability of our findings. The ideas detected by an automated system can only be as good as the ideas voiced by the users in the online community under investigation. Furthermore, the 200 texts we selected for evaluation were only a sample and therefore unlikely to reflect the whole range of ideas discussed in the community. It is an open question whether similar results will be achieved when automatic idea detection systems are applied to other technology domains or product categories.

This question can only be answered by follow-up research. However, we do believe that we have demonstrated the potential of automatic idea identification systems: they can be a powerful technique for the harvesting and initial screening of user ideas from online fora that do not conform, and are not limited to, the highly restrictive architecture and user basis of dedicated crowdsourcing systems. We hope that studies such as ours can also make a contribution to a wider discussion: which business tasks of a more complex nature can credibly be solved by artificial intelligence-based systems? We are convinced that the answer does not only lie in what is technically possible but also in what is acceptable to the prospective users of the information generated by such systems. More user evaluations of the performance of artificial intelligence-based systems are needed.

References

Antorini, Y. M. (2007). *Brand Community Innovation: An Intrinsic Case Study of the Adult Fans of LEGO Community*. Copenhagen Business School, Frederiksberg: Center for Europaforskning,.

Antorini, Y. M., Muñiz, J., Albert M., & Askildsen, T. (2012). Collaborating With Customer Communities: Lessons from the Lego Group. *MIT Sloan Management Review*, *53*(3), 73–95.

Brennan, R. L. (2001). *Generalizability theory*. New York: Springer-Verlag.

Christensen, K., Liland, K. H., Kvall, K., Risvik, E., Biancolillo, A., Scholderer, J., … Næs, T. (Submitted manuscript). Mining online community data: The nature of ideas in online communities. *Food Quality and Preference*.

Christensen, K., Nørskov, S., Frederiksen, L., & Scholderer, J. (2016). In search of new product ideas: Identifying ideas in online communities by machine and text mining. *Creativity and Innovation Management (Available Online)*.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements*. New York: Wiley.

Dean, D. L., Hender, J. M., Rodgers, T. L., & Santanen, E. L. (2006). Identifying quality, novel, and creative Ideas: Constructs and scales for idea evaluation. *Journal of the Association for Information Systems*, *7*(1), 646–698.

di Gangi, P. M., Wasko, M. M., & Hooker, R. E. (2010). Getting customers' ideas to work for you: Learning from Dell how to succeed with online user innovation communities. *MIS Quarterly Executive*, *9*(4), 213–228.

Ekvall, G. (1997). Organizational conditions and levels of creativity. *Creativity and Innovation Management*, *6*(4), 11.

Evanschitzky, H., Eisend, M., Calantone, R. J., & Jiang, Y. (2012). Success factors of product innovation: An updated meta-analysis. *Journal of Product Innovation Management*, *29*(S1), 21–37.

Füller, J., Bartl, M., Ernst, H., & Mühlbacher, H. (2006). Community based innovation: How to integrate members of virtual communities into new product development. *Electronic Commerce Research*, *6*(1), 57–73.

Füller, J., Jawecki, G., & Mühlbacher, H. (2007). Innovation creation by online basketball communities. *Journal of Business Research*, *60*(1), 60–71.

Jeppesen, L. B., & Frederiksen, L. (2006). Why do users contribute to firm-hosted user communities? The case of computer-controlled music instruments. *Organization Science*, *17*(1), 45–63.

Kristensson, P., Gustafsson, A., & Archer, T. (2004). Harnessing the creative potential among users*. *Journal of Product Innovation Management*, *21*(1), 4–14.

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, *33*(1), 159.

Lin, F.-R., Hsieh, L.-S., & Chuang, F.-T. (2009). Discovering genres of online discussion threads via text mining. *Computers & Education*, *52*(2), 481–495.

Magnusson, P. R. (2009). Exploring the Contributions of Involving Ordinary Users in Ideation of Technology-Based Services*. *Journal of Product Innovation Management*, *26*(5), 578–593.

Magnusson, P. R., Wästlund, E., & Netz, J. (2014). Exploring Users' Appropriateness as a Proxy for Experts When Screening New Product/Service Ideas: Exploring Users as a Proxy for Expert Judges. *Journal of Product Innovation Management*, *33*(1), 4–18.

Mahr, D., & Lievens, A. (2012). Virtual lead user communities: Drivers of knowledge creation for innovation. *Research Policy*, *41*(1), 167–177.

Majchrzak, A., & Malhotra, A. (2013). Towards an information systems perspective and research agenda on crowdsourcing for innovation. *The Journal of Strategic Information Systems*, *22*(4), 257–268.

Nørskov, S., Antorini, Y. M., & Jensen, M. B. (2015). Innovative brand community members and their willingness to share ideas with companies. *International Journal of Innovation Management*.

Poetz, M. K., & Schreier, M. (2012). The value of crowdsourcing: Can users really compete with professionals in generating new product ideas? *Journal of Product Innovation Management*, *29*(2), 245–256.

Thorleuchter, D., & Van den Poel, D. (2013). Web mining based extraction of problem solution ideas. *Expert Systems with Applications*, *40*(10), 3961–3969.

Van de Ven, A. (1986). Central problems in the management of innovation. *Management Science*, *32*(5), 590–607.

van den Ende, J., Frederiksen, L., & Prencipe, A. (2015). The Front End of Innovation: Organizing Search for Ideas. *Journal of Product Innovation Management*, *32*(4), 482–487.

Vandenbosch, B., Saatcioglu, A., & Fay, S. (2006). Idea management: a systemic view. *Journal of Management Studies*, *43*(2), 259–288.

von Eye, A., & von Eye, M. (2008). On the marginal dependency of Cohen's κ. *European Psychologist*, *13*(4), 305–315.

von Hippel, E., Ogawa, S., & PJ de Jong, J. (2011). The age of the consumer-innovator.

von Krogh, G., Spaeth, S., & Lakhani, K. R. (2003). Community, joining, and specialization in open source software innovation: a case study. *Research Policy*, *32*(7), 1217–1241.

von Krogh, G., & von Hippel, E. (2006). The Promise of Research on Open Source Software. *Management Science*, *52*(7), 975–983.

Table 1 - Performance of the automatic idea detection system used by Christensen et al. (Submitted Manuscript)

Partition	True positives (TP)	True negatives (TN)	False positives (FP)	False negatives (FN)	Classification accuracy	Precision	Recall	F_1
Validation set	27%	70%	1%	2%	0.97	0.97	0.92	0.94
Hold-out set	25%	70%	1%	3%	0.96	0.96	0.88	0.92

Table 2 - Example of an idea text and a non-idea text on which both raters agreed

Idea text	Non-idea text
´Buckwheat has been used as an adjunct for a long time in a few beers. It also is used to make gluten free beers. It has a high gelatinisation temp so need to be boiled first. Extract potential is about 1.032. Can be used lightly roasted to add colour to gluten free beers, or use Kasha (a roasted buchwheat). I think Rogues make a buckwheat ale´	´Thanks for the help. My internet is screwy or I would have replied sooner. I re- pitched and it is going crazy. a load off my mind! now i can concentrate on getting another cider and a wit going. Anyone have any suggestions for a good belgian style ale l ike duvel? I am an extract with specialty grains level brewer, so whole grain is out for now. Thanks again for all the help!´

Table 3 - Presence of ideas: classification accuracy of the automatic idea detection system, validated against the judgments of two company professionals

Validation criterion	True positives (TP)	True negatives (TN)	False positives (FP)	False negatives (FN)	Classification accuracy	Precision	Recall	F_1
Lenient criterion:								
Classified as idea by Expert 1 OR Expert 2	35%	42%	12%	12%	0.77	0.75	0.74	0.75
Strict criterion:								
Classified as idea by Expert 1 AND Expert 2	15%	52%	31%	3%	0.67	0.33	0.86	0.47

Table 4 - Idea text identified by classifier, Expert 1 and Expert 2

´T've made several batches. Below is my recipe The love of my life I love Mead as you can probably tell. Please note, this is Mead but I do not use any water. I use apple juice as the base. You can use water but I find the apple juice makes it a bit nicer for those of you who love apples and like a high alcohol content. No citric acid needed. This is called Apple- Honey Melonomel Meade You will need... 1 Package Red Star wine yeast 4 Gallons apple juice from concentrate 2-5 pounds of pure honey, the more the better. This shit is expensive though. 1 cup table sugar 5 Fuji apples Siphen hose, any small tube will work. A 5 gallon carboy or tub 1 balloon Step one, crush your apples or use a blender. Step two, boil apples in large pot with apple juice. Step three, set aside to cool Step Four, boil honey in large pot of apple juice Step five, set aside to cool. Step six, dump mixture into large 5 gallon carboy and add activated yeast. Step six, allow the mead to ferment for 3-4 weeks, once fermentation begins to slow prime with table sugar by dilluting the 1 cup of table sugar in 1/2 gallon of apple juice then pour this directly into the carboy. A balloon can be placed over the mouth of the carboy to monitor the fermentation. Simply peirce a small hole in the baloon to allow CO2 to escape. Once the Meade has cleared (meaning you can read a newspaper through it) transfer it into a secondary (Save the sediment for use as the Yeast in your next batch of Meade) and let it clarify for 2-3 weeks. After this bottle the meade and let fermintation finish off. Total process about 70 days and its ready to drink. This will burn going down but is smooth as a whistle. Enjoy....´

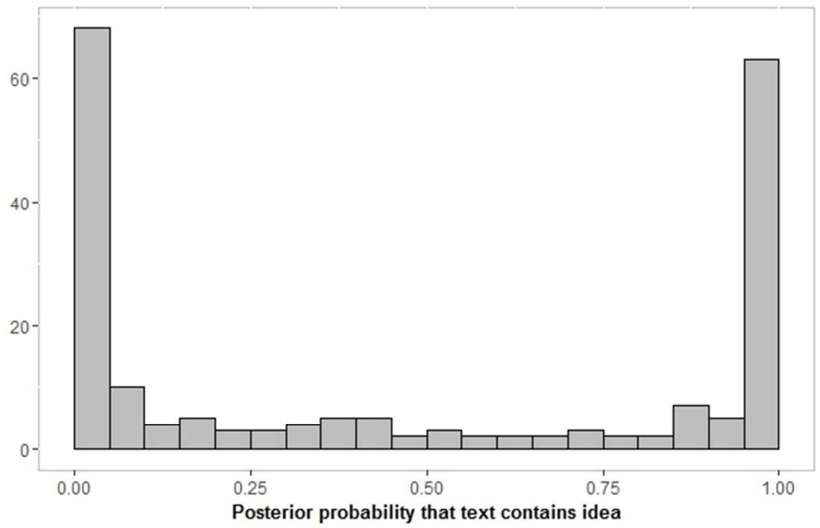

Figure 1 - Histogram of the posterior probability scores generated by the SVM-based automatic idea detection system for the 200 texts used in the present study

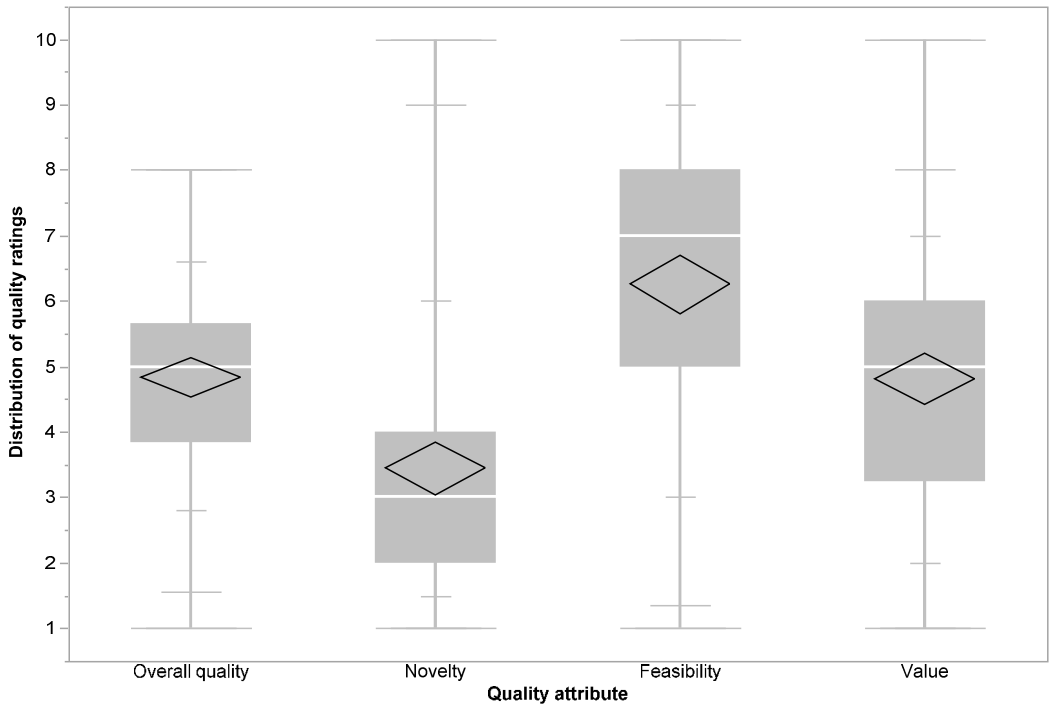

Figure 2 - Box plots of the distribution of quality ratings (overall quality = unweighted average of novelty, feasibility and value; diamonds represent 95% confidence intervals around distribution means)