Decision Support Systems 78 (2015) 39–50

Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

High-speed idea filtering with the bag of lemons

Mark Klein [a,c,*], Ana Cristina Bicharra Garcia [b]

[a] Massachusetts Institute of Technology, University of Zurich, 7.7 Massachusetts Avenue, NE25-754, Cambridge 02139, MA, USA
[b] Universidade Federal Fluminense, Rua Passos da Patria, 156, Niteroi, RJ, Brazil
[c] University of Zurich, Binzmühlestrasse 14, CH-8050 Zürich Switzerland

ARTICLE INFO

Article history:
Received 11 September 2014
Received in revised form 29 June 2015
Accepted 30 June 2015
Available online 4 July 2015

Keywords:
Collective intelligence
Open innovation
Social computing
Idea filtering

ABSTRACT

Open innovation platforms (web sites where crowds post ideas in a shared space) enable us to elicit huge volumes of potentially valuable solutions for problems we care about, but identifying the best ideas in these collections can be prohibitively expensive and time-consuming. This paper presents an approach, called the "bag of lemons", which enables crowd to filter ideas with accuracy superior to conventional (Likert scale) rating approaches, but in only a fraction of the time. The key insight behind this approach is that crowds are much better at eliminating bad ideas than at identifying good ones.

1. The challenge: idea filtering in open innovation

"Open innovation" is the concept of going outside your organization (e.g. to customers, suppliers, stakeholders, and other interested parties) to get ideas for how to solve challenging problems [18]. Increasingly, organizations are using web-based open innovation[1] software platforms (such as Spigit, Imaginitik, Nosco, BrightIdea, Salesforce, and Ideascale) as a powerful tool to solicit ideas from open communities [3,8,17,43,67, 30,48,22,25,37,45,65,63]. In domains ranging from government to industry to NGOs, they have been finding that crowds are willing and able to volunteer ideas, for questions they care about, in vast volumes. The six-day IBM "Innovation Jam" in 2006, for example, involved over 150,000 participants from 104 countries in identifying 46,000 product ideas for the company [8]. Dell's ongoing Ideastorm website [21] has received, to date, over 20,000 suggestions for improved Dell products and services. In the early weeks of his first term, President Obama asked citizens to submit questions on his web site change.gov, and promised to answer the top 5 questions in each category in a major press conference [47]. Over 70,000 questions were submitted. Google's 10 to the 100th project received over 150,000 suggestions on how to channel Google's charitable contributions [15], while the 8000 participants in the 2009 Singapore Thinkathon generated 454,000 ideas [41]. This kind of engagement thus gives organizations access, at very low cost, to a much broader selection of ideas, increasing the likelihood that they will encounter truly superior "out-of-the-box" solutions [36].

This very success has, however, raised a new dilemma: screening this outpouring of ideas to identify the ones most worth implementing [62]. Open innovation engagements tend to generate idea corpuses that are large, highly redundant, and of highly variable quality [49,57,66,9,8, 21]. Previous research suggests that about 10–30% of the ideas from open innovation engagements are considered, by the customers, as being high quality [9]. Convening a group of experts to identify the best ideas, from these corpuses, can be prohibitively expensive and slow. Nearly 100 senior executives at IBM, for example, had to spend weeks sifting through the tens of thousands of postings generated by their Web Jam [8]. Google had to recruit 3000 Google employees to filter the unexpected deluge of ideas for the 10 to the 100th project, a process that put them 9 months behind schedule [15]. The change.gov website, finally, had to be shut down prematurely because the huge volume of contributions overwhelmed the staff's ability to meaningfully process it. It has been estimated that it takes about $500 and four hours to evaluate one idea in a Fortune 100 company [51].

In response to this, organizations have turned to crowds to not just generate ideas, but also filter them, so only the best ideas need be considered by the decision makers. It has in fact been shown that crowds, under the right circumstances, can solve classification problems like that with accuracy equal to or even better than that of experts [61]. But this approach, in practice, has been no panacea. As we will see below, existing filtering approaches, when faced with large idea corpuses, tend to fare poorly in terms of accuracy, and can make unrealistic demands on crowd participants in terms of time and cognitive complexity.

In this paper we will present a novel crowd-based idea filtering technique for meeting this important challenge. We will review the strengths and shortcomings of existing idea filtering techniques, describe our own approach to the problem, present an empirical evaluation

* Corresponding author.
E-mail addresses: M_KLEIN@MIT.EDU (M. Klein), BICHARRA@IC.UFF.BR (A.C.B. Garcia).

[1] These are also sometimes referred to as "idea management" or "social ideation" systems.

http://dx.doi.org/10.1016/j.dss.2015.06.005
0167-9236/

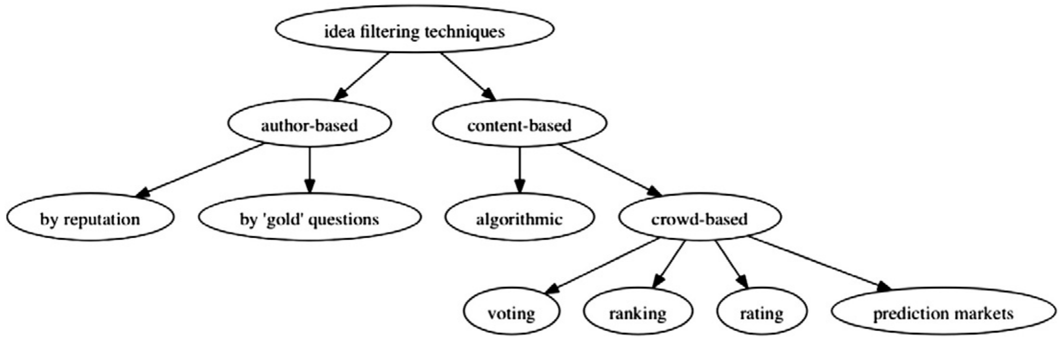

Fig. 1. A taxonomy of idea filtering techniques.

conducted as part of a "real-world" open innovation engagement, and discuss the contributions and future directions for this work.

2. Related work

To provide context for our approach, we review existing techniques that can be applied to idea filtering in open innovation settings. These techniques fall into several categories (Fig. 1).

Author-based filtering filters out ideas based on *who* contributed them. Authors can be excluded, for example, based on their previous behavior (i.e. based on their reputation) [32]. This requires substantial prior knowledge about the authors, however, and may result in low recall, since good ideas can often come from unexpected quarters. Authors can also be excluded based on "gold questions", wherein contributors are asked, before submitting an idea, to perform a simple task with a known answer in order to assess whether or not they have a basic level of competence [46]. This approach may help filter out some of the worse content but has only, as far as we are aware, been applied to estimating quality in crowd-sourced micro-tasks, rather than for filtering ideas from open innovation engagements.

Content-based filtering distinguishes among ideas based on their content, rather than their author. One approach is *algorithmic*, wherein we use software to derive metrics for idea quality based on such features as word frequency statistics. Walter et al., for example, measure the presence of rarely-used words to estimate the creativity of a contribution [64]. Westerski derived metrics based on manually- as well as machine-generated idea annotations (e.g. concerning what triggered the idea) [66]. Such techniques are also fundamentally limited by the fact that current natural language processing algorithms have only a shallow understanding of natural language, and thus can be easily fooled. In the Westerski work, for example, the automatically-generated idea quality metrics only achieved a correlation of 0.1 with the quality scores given by human experts. We can also use machine learning to define idea filtering rules, if we have examples of desired and non-desired content. A learning-based approach has proven useful in contexts, such as movie recommendations or email spam filtering, where very large training sets are readily available [14,1,16]. Finding such training sets is problematic for open innovation engagements, however, because creating them requires exhaustive human evaluation of large idea corpuses, and the rules learned for one particular idea corpus and set of evaluation criteria may not apply equally well in other contexts.

For this reason, much attention has been given to *crowd-based filtering*, where human participants are asked to select the top ideas, since they can potentially understand the ideas much more deeply than software.

This can be done in many ways, including (Fig. 2):

- *Voting*, where participants vote for which ideas should be selected
- *Rating*, where participants give ideas a numeric quality score
- *Ranking*, where participants place the ideas into a full or partial order, and
- *Prediction markets*, where users buy and sell stocks representing predicted winners, knowing they will receive a payoff if they own stocks that are eventually selected as winners: the stock prices then represent the crowd's assessment of the likelihood of the associated prediction

Voting systems ask crowd members to vote for the ideas that they think merit adoption. In an idea-filtering context, a *multi-voting* [31] approach is typically taken, where users are asked to allocate a budget of N votes to the best ideas in the corpus e.g. as in [5]. Voting systems are simple to use but face well-known practical as well as theoretical limitations, especially when applied to large option sets [2,33].

Rating systems [38,56] can gather useful feedback with moderate numbers of options, but are prone to several challenging dysfunctions. One is that rating systems tend to elicit mainly average scores from raters, and thus tend to do a poor job of distinguishing between good and excellent ideas [5]. Another is that rating systems tend to lock into fairly static and arbitrary rankings with large option sets: people do not have time to rate all the options and thus tend to consider only those that have already received good ratings, creating positive feedback loops [54,50,8,70]. This problem can be alleviated somewhat by using algorithms that adaptively assign ideas to raters (e.g. focusing

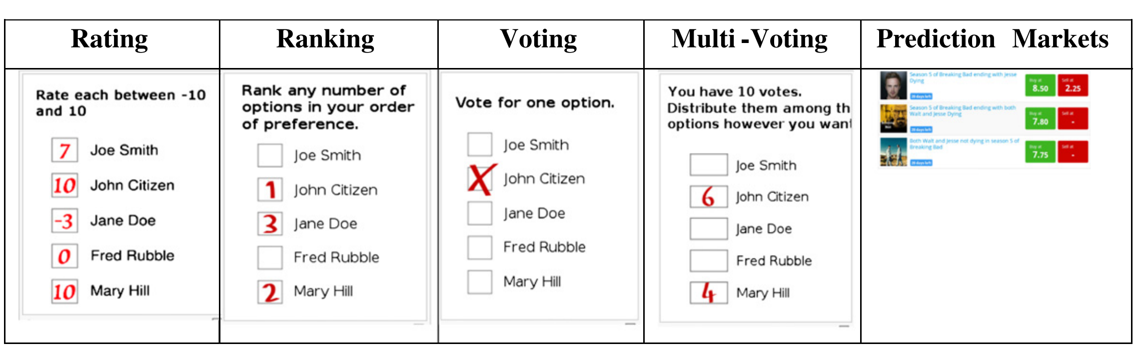

Fig. 2. Examples of crowd-based filtering methods.