CROWDSOURCING "BLOCKBUSTER" IDEAS: A DYNAMIC STRUCTURAL MODEL OF IDEATION

Complete Research Paper

Yan Huang
Heinz College
Carnegie Mellon University
yanhuang@andrew.cmu.edu

Param Vir Singh
Tepper School of Business
Carnegie Mellon University
psidhu@cmu.edu

Kannan Srinivasan
Tepper School of Business
Carnegie Mellon University
kannans@andrew.cmu.edu

Abstract

Crowdsourcing initiatives are becoming an increasing popular tool for new idea generation for firms. Although such initiatives are widely adopted in many different industries, the number of ideas generated often decline over time and the implementation rates (percentage of posted ideas that are implemented by the firm) are very low. Critics of crowdsourcing often attribute these observations to users' restrictive view about firms' products leading to contribution of mainly niche ideas, and limited knowledge about firms' cost structure leading to contribution of mostly infeasible ideas. To investigate these criticisms in detail and to devise policies for firms to alleviate these concerns, we propose a structural model to capture users' idea contribution dynamics. We estimate the model using a rich dataset obtained from Ideastorm.com, which is a crowdsourcing website affiliated with Dell. Using the peer voting score we are able to infer out the true potential of ideas, whereas a firm's costs of implementation are indirectly imputed from the idea implementation data. We find that individuals tend to significantly underestimate firm's costs of implementation of their ideas but overestimate the potential of their ideas in the beginning of the their idea contribution history. Therefore, the "idea market" is overcrowded by ideas which are less likely to be implemented. However, individuals learn about their abilities to come up with high potential ideas and the cost structure of the firm through a Bayesian fashion from peer voting on their ideas and firm's response to contributed ideas. We find that the individuals learn very quickly about their abilities to come up with high potential ideas but the learning regarding the firm's cost structure is quite slow. As a result of the learning process, contributors of low potential ideas eventually drop out, and high potential idea contributors remain active. Over time, the average potential of generated ideas increases, while the number of ideas created decreases, and the firm can reduce the cost of screening ideas without losing high potential ideas. Through a policy simulation, we show that the firm can significantly increase the learning rate of the users regarding their abilities to contribute "blockbuster" ideas by running short term promotions where the users whose ideas are implemented are given very high incentives.

Keywords: Crowdsourcing, User-generated-content, Structural Modeling, Dynamic Learning, Econometric analyses, Economics of information systems, Utility

Introduction

The importance of business innovation has been recognized by both practitioners (Farrell, 2008; Google, 2011) and academic researchers (Fagerberg, 2005). One of the most important aspects of innovation is new product design. The traditional way that firms generate new product ideas is in-coursing, i.e. firms purely rely on their internal professionals to design new products. Typically, the professionals design new products according to their market analyses, past experiences, as well as the firms' high level strategies (Amabile et al. 2005, Goldenberg, et al. 2001, Majchrzak et al. 2004, Schulze and Hoegl 2008). Firms may also conduct market analysis (e.g. conjoint analysis) to test the potential of new products, and then decide whether to implement an idea or not.

The advances in information technology have allowed firms to enhance their communication with customers. Another approach of generating ideas emerged as a result of this enhanced communication between firms and their customers. Jeff Howe (2006) named this new approach Crowdsourcing, and he defined crowd as "the new pool of cheap labor: everyday people using their spare cycles to create content, solve problems, even do corporate R & D." Crowdsoucing initiatives provide users a platform to express their ideas, which are typically generated from their experience or their observations. The ideas that come from the customer crowds can reveal rich information about customers' preferences and concerns. Typical crowdsourcing platforms also allow other customers to promote or demote ideas of their peers, which help a firm in gauging the potential of an idea. Firms can potentially obtain a large number of novel and profitable ideas, at relatively low costs from such initiatives. This ideation approach has been widely applied in various industries, including Retail, Agencies, Education, Public Sector, Local Government, Consultancies, Charities, Media, Construction, Event Management and Membership organizations. Early adopters of this approach are firms such as Dell, Threadless, Starbucks, Adidas, BBC, BMW, Ducati, Muji and Sears.

Although crowdsourcing initiatives have become very popular in a variety of industries, the potential of this new approach is still under debate. Critics of such initiatives raise three important concerns. First, they argue that the users might be too accustomed to current consumption conditions and their own specific needs, and, hence, are more likely to suggest niche ideas with little market potential. Second, unlike the internal R&D teams, customers of the firm are unaware of the internal cost structure of the firm and, hence, are quite likely to suggest ideas which are not viable (Schulze and Hoegl 2008). Third, the firms typically have to invest significant amount of effort to screen through all ideas most of which are low potential and generally infeasible. The low implementation rate of ideas, the decline in the number of ideas posted and the shutdown of a number of crowdsourcing websites observed in practice also seem to be consistent with the arguments against crowdsourcing. However, there is no systematic research which has investigated these issues in depth. In this study, we present an empirical framework to investigate the customers' abilities to contribute viable "blockbuster" ideas. We also highlight ways in which a firm can help its customers in contributing such ideas.

In this paper, we develop an empirical framework to analyze the learning dynamics in users' participation in crowdsourcing websites. We then apply our model to a rich dataset collected from one of the most popular crowdsourcing websites, Ideastorm.com. Our proposed model considers users' learning on firms' cost structure and the potential of their ideas. Users generate utility when their ideas are implemented. At the same time, they incur some cognitive cost when developing and posting ideas. We allow the payoffs that individuals receive when their ideas are implemented to vary across idea categories. The classification of ideas is adopted from Ideastorm.com. At the beginning of the posting history, users have little idea about a firm's cost structure and the potential of their own ideas. That is, they have very flat prior for the beliefs of their ideas' potential and firm's costs. As they participate on the website, they learn about firm's cost structure of implementing different categories of ideas from observing what ideas are implemented, and the potential of their own ideas from observing the voting scores of their ideas. Users update their beliefs in a Bayesian fashion.

Our results show that initially contributors significantly underestimated firm's costs of implementing their ideas and overestimated the potential of their own ideas. However, as they learn (update their beliefs) about the firm's cost structure and the potential of their ideas, users with low ability to generate "blockbuster" ideas drop out. The remaining active contributors tend to have higher ability to generate high potential ideas. This is consistent with our observation that although the number of ideas generated is decreasing over time, the fraction of posted ideas that are implemented is increasing over time significantly. These findings show that over time the crowds can deliver viable "blockbuster" ideas, which addresses the major concerns about crowdsourcing initiatives.

The results also show that users learn about the potential of their ideas faster compared to the firm's cost structure. Since users learn about the potential of their ideas from observing the peer votes their ideas receive, which happen only when they post ideas themselves, one way to accelerate users' learning about the potential of their ideas is to encourage them to post more ideas initially. To test whether this policy is effective, we run a policy experiment where users receive an additional reward in the first 4 weeks if their ideas are implemented. We find that by introducing this temporal additional incentive for posting ideas, firms can filter out users with low potential quickly. As a result, the average potential of ideas posted increases and stabilizes faster.

Our study makes several contributions. First, this is the first study that proposes a structural utility driven model of ideation. Our model allows us to infer a firm's cost structure for implementing crowd generated ideas as well as an idea's potential. We are also able to estimate users' perceptions about the cost structure and their idea potentials. The results show that the difference between the true and perceived cost structure and the idea potential can explain important dynamics observed on crowdsourcing platforms. Our study shows that the criticism raised by the critics of crowdsourcing is only partially valid during the early time of such initiatives and that crowdsourcing can provide viable "blockbuster" ideas in the long term. Second, we conduct an important policy experiment which can significantly reduce a firm's cost of screening all ideas by filtering out low potential ideas quickly. Finally, our paper also has a methodological contribution. In previous Bayesian learning literature, individuals only learn one attribute (e.g. quality), due to the identification of the models. By assuming the potential is linearly correlated with the vote points, we are able to model the learning in ideas' potential independent of the rest of the model. Then, we imply individuals' learning about firms' cost structure from the posting decisions users made. Our model provides a way to incorporate learning about multiple attributes within one model, which significantly extends the application of the Bayesian learning models.

Literature Review

Our paper is related to multiple streams of literature. First, it is related to the emerging literature on crowdsourcing. Lakhani, et al. (2007) study the effect of openness and information sharing on scientific problem solving outside the laboratory context and show that disclosure of problem information to a large group of outside solvers is an effective mean of solving scientific problems. Problem-solving success was found to be associated with the ability to attract specialized solvers with range of diverse scientific interests. Several studies on open source have shown that users are able to create commercially very successful products (Singh et al 2011a, Singh 2010). von Hippel (2005) show that users often innovate for themselves and that many of those user innovations are characterized by high commercial attractiveness. Shah et al. (2000,2003, 2006) conduct a series of studies in sporting products and find that the most commercially important equipment innovations tend to be developed by users. Some other studies independently support the idea that commercially attractive products are often developed by "lead users", who are leading important marketplace trends and expect significant benefit from firms' innovation (Urban and von Hippel, 1988, Morrison et al., 2000; Franke et al., 2006 and Olson and Bakke, 2001). Poetz and Schreier (2010) focus on baby product market and compare the executives' evaluation on customer generated ideas and professional generated ideas. They find that user generated ideas score significantly higher in terms of novelty and customer benefit, but relatively lower in feasibility. They also conclude that user generated ideas might be more successful when the knowledge needed to come up with successful ideas is closely linked to aspects of use experience. Franke and Klausberger (2009) analyze the role of perceived fairness in crowdsourcing communities, which partially answer the question of how a crowdsourcing incentive system affects the outcome. Bayus's study (2010) shows that individual creativity is positively related to current effort, but negatively related to past success and he emphasize the need for a better understanding of the reward and feedback mechanisms in crowdsourcing systems.

This paper is also related to the literature on user generated contents and social media. Ghose and Han (2009) examine multimedia content creation and consumption behavior using mobile phone dataset and find that there exists a negative temporal interdependence between the content generation and usage behavior for a given user. And in a separate paper, Ghose and Han (2010) find evidence of dynamic learning in two-sided mobile Internet forums. Albuquerque et al. (2010) analyze the print-on-demand service of user-created magazines and find that content price and content creator marketing actions have significant impact on purchases. Kumar and Sun (2009) and Lu et al. (2011) empirically model different categories of inter-temporal tradeoffs that the users of online communities have to make. The former paper focus on why users contribute to connected goods in social networking sites and the latter emphasize how users' online social network affects their willingness to share knowledge with peers. Huang et al. (2010) investigate the incentives for the employees to participate in enterprise blogging activities. Aggarwal et al. (2010) study how negative posts by its employees can actually benefit a firm. Lu et al (2010) study the emergence of opinion leaders in online review communities. Singh et al (2010) and Sahoo et al (2011) investigate the switching behavior in social media content consumption of employees.

Finally, our paper is related to the literature in consumer Bayesian learning. The Bayesian learning models are different from classical organizational learning models (Singh et al. 2011b; Mukhopadhyay et al. 2011). Bayesian learning models are widely applied to analyses of consumers' choices under uncertainty. In these learning models, consumers learn brand qualities from multiple resources, such as past experience, advertisement, and price (Erdem and Keane, 1996; Mehta, 2003; Crawford and Shum, 2005; Erdem et al., 2008). The key idea in Bayesian learning models is that individuals' perception of the attributes of the products (quality in these cases) evolves over time. Therefore, their choices could be time variant. Including this component in consumer choice model can help understand interesting dynamic in consumers' choice behavior. This can provide potentially important managerial implications about the long term effects of firms' marketing actions. The applications of Bayesian learning models are mostly concentrated on the learning of quality or quality-related attributes. There are very few other applications. In our paper, we intend to apply the Bayesian learning model to users' learning of both idea potential and firms' cost structure, to better understand the dynamics in users' idea posting behavior.

Research Context

The research context is a crowdsourcing website, Ideastrom.com, which is operated by Dell. Dell launched this website in February 2007 as a way to talk directly to its customers. Ideastorm.com was created to give a direct voice to Dell's customers and an avenue to have online brainstorming sessions to allow the customers to share ideas and collaborate with one another and Dell. The goal of this initiative was to hear what new products or services Dell's customer's would like to see Dell develop.

The structure of Ideastrom.com is quite simple yet effective. Any individual can register on the website to participate in the initiative. Once registered a user can then post any relevant idea. Dell assign 500 Dell points to the contributor for each idea she posts. Once an idea is posted, all the other users can vote on the idea. They can either promote the idea, which will result in a 10 points increase on the idea's votes, or demote the idea, which will result in a 10 points decrease on the idea's votes. Users are also allowed to comment on ideas to express their opinion in better detail. Dell uses the peer voting and comments to gauge the potential of contributed ideas. Dell assigns web managers to maintain the website, and their job is to pass the ideas generated by the users to the corresponding groups in the company for review. After the decision is made the users are updated about the decision and provide more details regarding the decision through comments or blogs. All the users can see how many peer votes any idea received as well as which ideas are implemented by the firm. In our modeling framework, we would allow the users to learn from these two observations. This is the most common structure in crowdsourcing ideation applications. My Starbucks Idea (Starbucks), My Sears Community Ideas (Sears), Best Buy Ideax (Best Buy) and some other crowdsourcing websites affiliated to well-known companies are all using the same structure. There are also companies that provide this kind of platform and services, such as salesforce.com, crowdwork.com, etc. My Starbucks Idea and IdeaStorm are both supported by salesforce.com.

Dell broadly categorizes all the ideas into three categories: Product ideas, Dell Ideas, and Topic Ideas (see Table 1). Each idea could be related to up to three sub-categories. When a user posts an idea on Ideastorm he/she selects the category as well as the sub-categories to which the idea belongs. In our dataset, most of the ideas fall in the first two categories, and very few ideas are categories 3 ideas (less than 10% of the numbers of ideas in Category 1 and 2, see Table 2). Therefore, our analysis focuses on the first two Categories of ideas.

Table 1. Idea Classification

Categories	Sub Categories
Product Idea	Accessories (Keyboards, etc.); Adamo; Alienware; Broadband and Mobility; Desktops; Dimension; Inspiron; Laptop Power; Laptops; Latitude; Linux; Mobile Devices; Monitors and Displays; Desktops and Laptops; Netbooks; New Product Ideas; Operating Systems; OptiPlex; Precision Workstations; Printers and Ink; Servers and Storage; Software; Studio; Vostro; XPS
Dell Idea	Advertising and Marketing; Dell; Dell Community; Dell Web Site; IdeaStorm; Retail; Sales Strategies; Service and Support
Topic Idea	Digital Nomads; Education; Enterprise; Environment; Gaming; Healthcare and Life Sciences; PartnerStorm; Small Business; Storm Session Topics; Women's Interest

Model

Utility Function

In our model, the utility a user can get from posting an idea include the expected benefit they can get when their ideas are implemented, or their needs are satisfied, (Franke and von Hippel, 2003, Kuan 2001, Lakhani and von Hippel 2003), as well as the reputation gain, which is 500 Dell points in our specific context. Specifically, a user's utility function is given by the following equation

$$U_{ijt} = c + r + \theta_j P_{ijt} + \varepsilon_{ijt} \tag{1}$$

In equation (1), P_{ijt} represent the probability of individual i's idea of category j, posted in period t, being implemented. We adopt the classification on the website and set idea categories to be Product ideas and Dell ideas and use 1 and 2 to represent the two categories respectively. The parameter, c, represents the hassle cost associated with constructing and posting an idea in category j, and r is the reputation gain people can get from the 500 Ideastorm points. It is obvious that we cannot identify c and r at the same time as they enter linearly in the utility function, therefore, we are only able to estimate their sum, i.e. $\theta_0 = c + r$. The parameter θ_j, measures individuals' utility gain from the implementation of his/her Category j idea. The error term ε_{ijt} captures the individual choice specific random shock in period t. Then, Equation (1) can be reduced to

$$U_{ijt} = \theta_0 + \theta_j P_{ijt} + \varepsilon_{ijt} \tag{2}$$

Firm's Decision Rule to Implement Ideas

The firm selectively implements ideas generated by users. In general, the firm will consider the potential (market demand) of the ideas, as well as the costs of implementing the ideas. Assume that a firm only implements those ideas that provide it with positive net profit. Let the net profit be represented by π_{ijt} *as*

$$\pi_{ijt} = Q_{it} + C_{ijt}$$

where Q_{it} represents the potential of the idea, and C_{ijt} represents the firm's cost associated with implementing the idea. Then, the probability that an idea will be implemented is

$$P_{ijt} = Pr(\pi_{ijt} > 0)$$

C_{ijt} is observed by the firm, but not observed to econometricians. Therefore, the likelihood that an idea with observed potential Q_{it}, is eventually implemented is

$$P_{ijt} = Pr\big(Q_{it} + C_{ijt} > 0 \,|Q_{it}\big) = 1 - \Phi(\frac{Q_{it}+C_j}{\sigma_\gamma}) \tag{3}$$

where σ_γ represents the true standard deviation of the cost for the firm to implement ideas in the same category.

Users Learning Process

When an individual is making idea contribution decision, he/she does not know a priori the potential of his/her idea as well as the costs the firm may incur to implement the idea (C_{ijt}). However, users can learn about the two components of their utility function from their experience and observations in a Bayesian manner (Erdem & Keane, 1996).

Learning about the Firm's Cost Structure

Suppose that at the moment when the website is launched, users' prior belief of the firm's cost of implementing an idea of category *j* is

$$C_{j0} \sim N(C_0, \sigma^2_{C_0}) \tag{4}$$

In equation (4), C_0 is the prior mean of the cost of implementing an idea in category *j*; and $\sigma^2_{C_0}$ measures prior belief about the variation of the cost associated with the implementation of different ideas within category *j*. Users learn the firm's cost structure by observing the implementation of contributed ideas, including their own ideas, and their peers' ideas. Whenever one idea is implemented, all users receive a noisy signal about the cost the firm incurs. C_{kjt} in the Equation (5) denotes cost signal all users receives when one category *j* idea is implemented in period *t*. And the difference between a particular signal and its mean is captured through the parameter μ_{jt}, which is a zero mean normal random variable, and the variance of it, σ^2_μ, measures the precision of the signals. We assume that all users receive the same cost signal when an idea is implemented.

$$C_{kjt} = C_j + \mu_{jt} \tag{5}$$

$$\mu_{jt} \sim N(0, \sigma^2_\mu)$$

If there are k_{Cjt} category *j* ideas implemented in period *t*, then the cumulative signal that users receive is C_{sjt}. C_{sjt} is simply the average of the k_{Cjt} signals ($C_{1ijt,\dots,}C_{k_{Cjt}ijt}$), and it has the following distribution

$$C_{sjt} \sim N(C_j, \frac{\sigma^2_\mu}{k_{Cjt}}) \tag{6}$$

Let C^e_{jt} denote users' prior mean of C_j in the beginning of period *t*, which is by definition conditional on the cumulative information users have received by the beginning of period *t*. Users update C^e_{jt} using the following Bayesian rule (DeGroot, 1970)

$$C^e_{jt} = C^e_{jt-1} + (C_{sjt} - C^e_{jt-1})\frac{\sigma^2_{C_{jt-1}}}{\sigma^2_{C_{jt-1}} + \frac{\sigma^2_\mu}{k_{cjt}}} \tag{7}$$

$$\sigma^2_{C_{jt}} = \frac{1}{\frac{1}{\sigma^2_{C_{jt-1}}} + \frac{k_{Cjt}}{\sigma^2_\mu}} \tag{8}$$

The prior in period t=0 is $C_{j0}^{e} = C_0$, $\sigma_{C_{j0}}^2 = \sigma_{C_0}^2$.

Learning about the Potential of Own Ideas

IdeaStorm.com allows individuals to vote on their peers' ideas. Voting score is used as a measure of potential of ideas. High voting score means many customers would like to see this ideas being implemented, while low voting score means the idea is probably a niche idea. We assume individuals' ability of generating different categories of ideas is the same. This assumption is required for the model identification, and it is also reasonable because a pair wise two sample test proves that the difference in the votes received by the two categories of ideas posted by the same individual is not significant. Let Q denotes the mean potential of ideas, and then Q_{kit}, the potential of an idea posted by individual i in period t is

$$Q_{kit} = Q + \delta_{it} \tag{9}$$

$$\delta_{it} \sim N(0, \sigma_\delta^2)$$

δ_{it} is the deviation of the potential of a specific idea posted by individual i in period t from the average potential of his/her idea. Note that individuals learn about their potential by observing the voting scores that their ideas receive. We assume that the natural logarithm of votes (V_{kit}) that an idea receives is linearly correlated with the potential of the idea

$$V_{kit} = cons + \varphi Q_{kit} \tag{10}$$

Plugging (9) to (10) we get

$$V_{kit} = V + \xi_{it} \tag{11}$$

$$V = cons + \varphi Q$$

$$\xi_{it} = \varphi \delta_{it}$$

$$\xi_{it} \sim N(0, \sigma_\xi^2)$$

$$\sigma_{\xi_i}^2 = \varphi^2 \sigma_\delta^2$$

where V is the mean value of the logarithm of votes that ideas receive and ξ_{it} is its deviation from the mean. Again, if individual i posts k_{Qit} ideas in period t, then the cumulative signal that she receives is Q_{sit}. Q_{sit} is simply the average of the k_{Qit} signals ($Q_{1it,\dots,}Q_{k_{Qit}it}$), and it has the following distribution

$$Q_{sit} \sim N(Q, \frac{\sigma_\delta^2}{k_{Qit}}) \tag{12}$$

At the moment when the website is launched, individuals' prior beliefs of the potential of their ideas and the log voting scores their ideas may receive are

$$Q_{i0} \sim N(Q_0, \sigma_{Q_0}^2) \tag{13}$$

$$V_{i0} \sim N(\varphi Q_0, \varphi^2 \sigma_{Q_0}^2)$$

Similar to the learning process of firm's cost structure, individuals update their beliefs about V_{it}^e and Q_{it}^e together when they post an idea and observe the voting scores their ideas receive. The updating rules for V_{ijt}^e and Q_{ijt}^e are (Erdem, Keane and Sun, 2008)

$$V_{it}^e = V_{it-1}^e + (V_{sit} - V_{it-1}^e) \frac{\sigma_{V_{it-1}}^2}{\sigma_{V_{it-1}}^2 + \frac{\sigma_\xi^2}{k_{Qit}}} \tag{14}$$

$$Q_{it}^e = Q_{it-1}^e + (V_{sit} - V_{it-1}^e) \frac{\varphi \sigma_{Q_{it-1}}^2}{\varphi^2 \sigma_{Q_{it-1}}^2 + \frac{\sigma_\xi^2}{k_{Qit}}} \tag{15}$$

where

$$\sigma_{V_{it}}^2 = \frac{1}{\frac{1}{\sigma_{V_{it-1}}^2} + \frac{k_{Qit}}{\sigma_\xi^2}} \tag{16}$$

$$\sigma_{Q_{it}}^2 = \frac{1}{\frac{1}{\sigma_{Q_{it-1}}^2} + \frac{k_{Qit}\varphi^2}{\sigma_\xi^2}}$$

In addition, we denote the priors for potential and log-votes at the moment that the website was launched to be $Q_{i0}^e = Q_0$, $\sigma_{Q_{i0}}^2 = \sigma_{Q_0}^2$, and $V_{i0}^e = \varphi Q_0$, $\sigma_{V_{i0}}^2 = \varphi^2 \sigma_{Q_0}^2$.

Users' Decision Making Problem

In each period, users make decisions on whether to post an idea in a category or not based on their expectation on the utility they can get from each choice. We normalize the utility associated with not posting to be 0 and assume that users' decisions on posting ideas to be independent of each other across categories. Let $\widetilde{U}_{ijt} = E(U_{ijt}|I(t))$ be the expected utility individual i can get from posting category j idea in period t, which can be express as

$$\widetilde{U}_{ijt} = \theta_0 + \theta_j Pr(Q_{it} + C_{ijt}|I(t) > 0) + \varepsilon_{ijt} \tag{17}$$

where $Q_{it} + C_{ijt}|\, I(t) \sim N(Q_{it}^e + C_{jt}^e, \sigma_{Q_{it}}^2 + \sigma_{C_{jt}}^2)$. Then, the probability that individual i will post a category j idea in period t will take a standard logit form.

$$Pr\ (i\ posts\ a\ category\ j\ idea\ in\ period\ t) = \frac{\exp(\widetilde{U}_{ijt})}{1 + \exp(\widetilde{U}_{ijt})}$$

Data and Preliminary Analysis

As mentioned earlier, we collected data from one of the most popular crowdsourcing website Ideastorm.com, which is operated by Dell. Our dataset expands from the initiation of Ideastorm.com in early 2007 to the end of 2010. By the end of 2010, more than 12,000 ideas had been posted, and more than 400 had been implemented. However, we only use the data from the initiation of Ideastorm.com to September, 2008, because Dell changed their reward policy at the beginning of December, 2008. Before the policy change, users could earn 500 IdeaStorm points when they posted an idea, however, after the policy change, users earn points only when their ideas are adopted. To avoid the undesirable effect that the policy change imposes on users' participation behavior, we exclude data after September 2008. The reason why the truncation starts two months before the policy change is to make sure that the policy change will not affect the voting scores of the ideas remaining in the sample. After this elimination, we have 86 weeks of data. We exclude data on the first two weeks, because the number of ideas contributed is extremely small (≤ 5) possibly due to public's unawareness of the website, and most of them are announcements made by Dell's employees. After the elimination, we have 84 weeks of data (Week 3 to Week 86).

Table 2. Summary Statistics by Category

Category	1	2	3
Category Name	Product idea	Dell idea	Topic idea
# Posted	5337	4243	392
# Implemented	100	110	10
% Implemented	1.87	2.59	2.55
Average log (votes)	4.626	4.580	4.352
SD of log (votes)	2.160	2.147	2.720

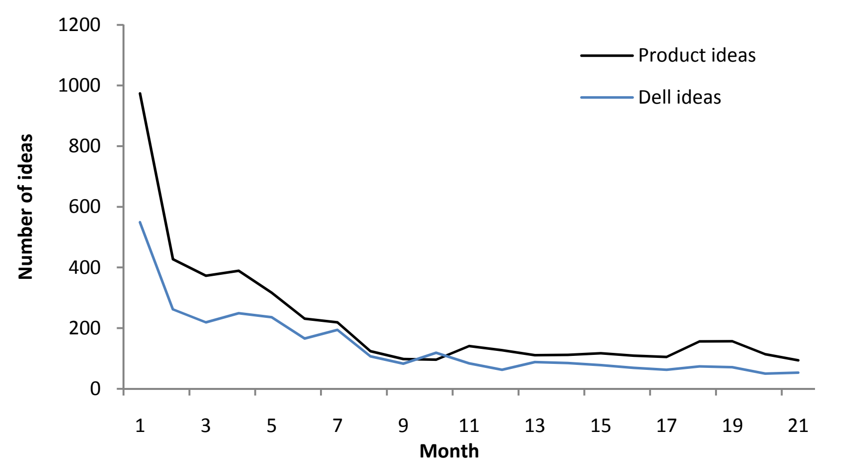


Figure 1. Numbers of Ideas Generated in Each Month

We select users with moderate usage of the website (post at least 2 ideas in the period we are looking at) to estimate the model, because we assume that in each period, users observe the changes in the idea status update their beliefs accordingly. If casual users are included, this assumption may fail. We eliminate such users from our data, and then 490 individuals are left in the sample. Our estimation is based on the behavior of these 490 representative users. If we look at the idea posting behavior over time, we can see clearly that the numbers of the two categories of ideas posted were very high early on. The numbers declined quickly over time and then stabilized (Figure 1). In addition, if we look at the implementation rates of different categories of ideas (Figure 2), we can see that the implementation rates of both category 1 and category 2 ideas increase over time.

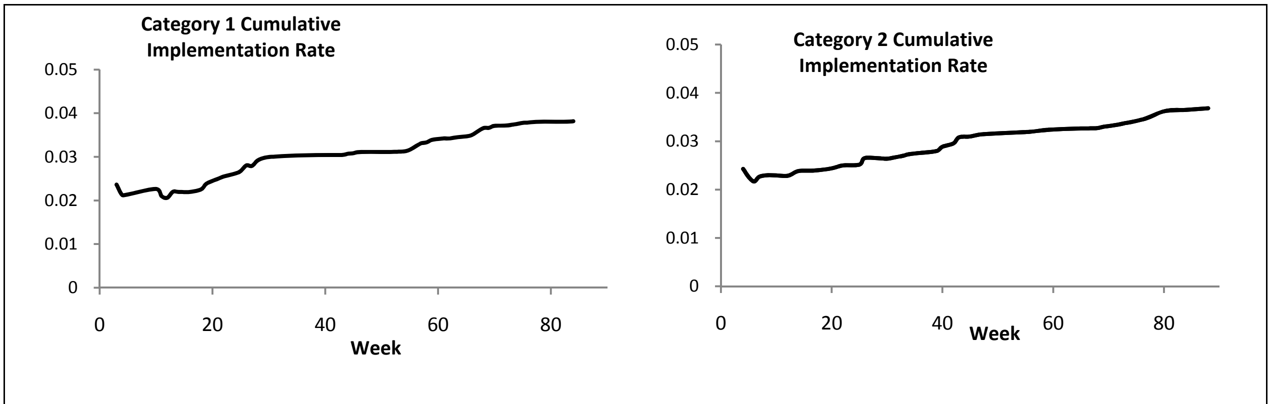


Figure 2. Cumulative Implementation Rate

Estimation

Likelihood function

A user will post a category j idea when the utility calculated from (17) is higher than the utility of no actions. If ε_{ijt} is assume to be type 1 extreme value distributed, then the likelihood of observing action A_{ijt} can be expressed as

$$L(A_{ijt}) = \left(\frac{\exp(\tilde{U}_{ijt})}{1+\exp(\tilde{U}_{ijt})}\right)^{A_{ijt}} \left(\frac{1}{1+\exp(\tilde{U}_{ijt})}\right)^{(1-A_{ijt})} \tag{18}$$

We further assume that the decisions on different categories of ideas are independent, then the joint log-likelihood of observing series of actions for all individuals in the sample is

$$l(A) = \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{2} \ln\ L(A_{ijt}) \tag{19}$$

In addition, let I_{ijt} denote the decision the firm makes on the Category j idea posted by individual i in period t, with value 1 indicating the idea is implemented and 0 otherwise, the likelihood that we observe I_{ijt} given Q_{it}, C_j and σ_γ^2 is

$$L(I_{ijt}) = \Phi\left(\frac{Q_{it}+C_j}{\sigma_\gamma}\right)^{(I_{ijt}-1)} \left(1-\Phi\left(\frac{Q_{it}+C_j}{\sigma_\gamma}\right)\right)^{I_{ijt}} \tag{20}$$

And then $l(I)$ is defined as

$$l(I) = \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{2} \ln\ L(I_{ijt}) \tag{21}$$

Finally, we observe the votes for individuals' own ideas. Given Q we can also calculate the probability densities of observed votes, and this will help identify φ, Q and σ_δ^2. Then the final log-likelihood function is

$$l = l(A) + l(I) + \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{J} p\left(\frac{V_{it}-V}{\sigma_{\xi}}\right) \quad (22)$$

where p denotes the log of standard normal probability density.

Estimation Strategy

If we plug Equation (17) in to Equation (18), we can see that the joint likelihood is a function of users' beliefs about the firm's costs and the potential of their own ideas, both means and variances. Maximum likelihood estimation cannot be directly used here, because we do not observe the actual cost signals, but only know the distribution of them (Erdem and Keane 1996). The high dimensional integration makes calculating the closed form likelihood function infeasible. Therefore, we apply method of simulated maximum likelihood (SML) (Pakes 1987, McFadden 1989). In each iteration in the maximization process, we draw a total of R=1000 sets of random numbers for the cost shocks for each individual, each period and each category of idea. We check that once R reaches 1000, the estimation is not sensitive to further increase in R. To reduce the computational challenge, we keep R=1000. After that, we can derive C_{jt}^{e} and $\sigma_{C_{jt}}^{2}$ using the updating rules discussed above, and then evaluate the simulated likelihood function numerically. Table 3 provides the summary of the key parameters in the model.

Table 3. Summary of the Parameters in the Model

Notation	Explanation
θ_0	Hassle cost incurred by users when posting a category idea (fixed to -5)
θ_j	Utility users generate when users' Category j ideas are eventually implemented
C_0	Users' prior belief of the mean costs of implementing each category of ideas
$\sigma_{C_0}^{2}$	Users' prior belief about the variance of the costs of implementing each category of ideas (set to 50, assume prior is uninformative)
C_j	Mean costs of implementing category j ideas
σ_{μ}^{2}	Variance of cost signals(common across categories)
Q_0	Users' prior belief of the mean potential of each category of ideas
$\sigma_{Q_0}^{2}$	Users' prior belief about the variance of the potential of each category of ideas (set to 50)
Q	Mean potential of all ideas
σ_{δ}^{2}	Variance of ideas' potential
$cons$	Intercept of linear function between log votes and the potential
φ	Slope coefficient between log votes and qualities
σ_{r}^{2}	The variance of true distribution of the costs for the firm to implement ideas in the same category

Identification

Before we estimate the model, we need to make sure that the model is identifiable. Here, we provide some intuition about how the parameters in our model are identified. It is obvious that we cannot identify Q and C_j at the same time, because if we add a constant to Q and then subtract the same constant from all $C_j's$, we will obtain exactly the same likelihood value. For identification purpose, we fixed C_1= -6. As a result, the estimated values of C_2 and Q should be interpreted as relative to C_1. We observe individuals' actions, from which we can infer individuals' utility derived from posting different categories of ideas. Once we know $\widetilde{U}_{ijt}$, we can infer parameters in the utility function. However, we can hardly identify θ_0 and θ_j simultaneously, because if we observe individuals post very frequently, especially when they learn sufficiently about the potential of their ideas, we cannot tell whether it is because they have low cost of posting an idea, or because they have higher payoffs when her ideas are implemented. Therefore, θ_0 is fixed to -5, θ_j should then be interpreted as the value individuals get from the implementation of their idea, relative to the cost they incur when posting an idea. The first implementation of Category 1 ideas happened in Week 11 and the first implementation of Category 2 ideas happened in Week 7. Since the potential of ideas posted is the same for both categories, the systematic difference in the number of ideas of each category generated before Week 7 tells us the difference between θ_1 and θ_2. Q can be identified from the likelihood of ideas' implementation given the votes and the behavior of "well-informed" individuals, whose perception about the firm's cost structure and their quality of ideas is very close to the true value. C_2, can then be identified through the probabilities that individuals post Category 2 ideas, as well as the firm's decision on Category 2 ideas, given the votes each idea receives. Once C_1, C_2 and Q are identified, σ_r^2 can be easily identified from the likelihood for one idea to be implemented. C_0 can be identified through the probability of posting in the first 7 weeks. and Q_0 can be identified through the probability of posting for the late comers throughout the whole observation period. Given all individuals' Q and observed V_{it}, $cons$ and φ determine a linear curve that approximate the relation between $Q$and V_{kit} the best. Finally, the variance parameters σ_μ^2 and σ_δ^2 are both identified from the dynamics of individuals' posting behavior over time.σ_μ^2 is identified through the dynamics of the choice probabilities at the population level. Similarly, the learning speed of the potential of the ideas is affected by both σ_δ^2 and the slope parameter φ. Once we control for φ, σ_δ^2 can then be identified.

Estimation Results

The estimates of the parameters are presented in Table 4. Comparing the estimate of C_2 with C_1 (fixed to -6), we can see that C_2 is slightly smaller in terms of absolute value. This means the cost that the firm incurs when implementing Category 2 ideas is lower than the cost of implementing Category 1 ideas. This result is reasonable because the summary statistics show us that the average log votes for ideas in both categories are similar, while the implementation rate of Category 2 idea is higher. The estimate for C_0 is much higher than both C_1 and C_2, which tells us that individuals initially underestimate the firm's implementation costs . It is consistent with the observation that the numbers of ideas created decrease over time at a population level. The estimate of $log\ (\sigma_\mu^2)$ is 6.272, which is equivalent to say σ_μ^2 is exp(6.272)= 530. This variance is extremely large compared to the absolute values of C_1 and C_2. It means that the implementation cost signals the firm provides to individuals are very imprecise and so individuals cannot learn very fast about firm's implementation costs. Remember that $exp(\sigma_\mu^2)$ is the variance of one signal and there are cases where quite a few ideas are implemented within a week. In those weeks, the variance of the cumulative signal individuals receive will be $exp(\sigma_\mu^2)$ divided by number of ideas implemented in each week and the learning regarding the implementation could still be significant. Relative to the variance of the signal, the estimate of $log\ (\sigma_r^2)$ is much smaller (1.340, i.e. $\sigma_r^2 = 3.819$). Q_0 is also higher than the estimate of Q, indicating that individuals tend to overestimated their ideas' potential before their ideas get voted on. $cons$ and φ determine the linear relation between log votes and potential. The slope coefficient is 1.350, meaning when the potential of the idea increases by 1, the log of one idea's vote increases by 1.350.

Table 4. Pooled Parameter Estimates

Notation	Parameter Estimates	Standard Deviation
C_0	-1.170	0.038
$\sigma_{C_0}^2$	50	-- (Fixed)
C_1	-6	--(Fixed)

C_2	-5.212	0.097
$\log(\sigma_\mu^2)$	6.272	0.093
$\log(\sigma_r^2)$	1.340	0.121
Q_0	3.556	0.329
$\sigma_{Q_0}^2$	50	--(Fixed)
cons	1.269	0.037
φ	1.350	0.019
Q	2.813	0.162
$\log(\sigma_\delta^2)$	-0.597	0.078
θ_1	4.064	0.204
θ_2	3.691	0.201

In addition, the mean of potential of ideas is significantly lower than the cost of implementing both categories of ideas. This is consistent with the low implementation rate we observe in the data. In addition, the variance of potentials of ideas is very low, which equals to exp(-0.597)=0.550. This result shows that the potentials of ideas posted by the same person are relatively consistent. This variance also implies individuals' learning speed about their ideas' potential. The small variance shows that on average, individuals learn about their ideas' potential quickly. When the website was launched, many individuals, i.e. idea providers, entered the market. As they learn about their own ideas' potential, as well as the cost for the firm to implement their ideas, individuals who believe that they have lower ability will drop out, and the "idea market" became efficient in a short time. In other words, the crowdsourcing mechanism is very effective in filtering idea providers, and the "idea market" reaches efficiency very quickly. The payoffs individuals receive when their Category 1 ideas are implemented is slightly higher than when their Category 2 ideas are implemented. This is consistent with the numbers of ideas of these two categories in the first few weeks. It is also intuitive because ideas that fall in Category 1 are more about product improvement and the ideas in Category 2 are the more about customer services and marketing strategies. It is not surprising that individuals receive more payoffs when the firm improves the product design the way individuals suggest, than when the firm provides service and communicates with their customers as they suggested. Note that in our estimation, we assume that all the users are participating from the first period, since we have no information on when the individuals registered. This assumption will lead to underestimation of Q_0, because under current assumption, those who start posting ideas in later part of our observation period did not post in the beginning, which will result in a lower estimate of Q_0. Since even when we underestimate Q_0, Q_0 is still higher than Q, the assumption will not alter the nature of the results. Another effect this assumption may have is that it could mitigate the learning effect by making the difference between Q_0 and Q smaller and the length of the learning period longer. If we can correctly define the time the each users join the community, the effect of learning will be reinforced.

From the estimation result, we can see that users' prior beliefs about the firm's cost structure and the potential of their ideas are imprecise. They learn about the firm's cost structure and the potential of their ideas gradually. Since the users learn about the implementation costs extremely slowly, providing more precise cost signal seems to be a good policy to improve users' learning. However, there are other costs associated with revealing too much of the firm's cost structure. A safer policy is to improve users' learning about their own ability, under which users whose ideas generally have lower potential will drop out more quickly. To do this, firms can provide a short term increase in the reward given to idea posters when their ideas are implemented. A higher reward will increase θ_j in user's utility function, and thus will also increase the probability for the users to post ideas. More idea posting for few periods will accelerate the users' learning about their potential, and thus low potential users will be filtered out quickly. After the filtering process is done, the firm can adjust the reward to the normal level. This is quite beneficial for the firm as it incurs a high cost of screening all ideas a lot of which are infeasible and of low potential. If the firm can filter out all the low ability users from idea posting, they can significantly reduce the screening costs they incur as well as direct more incentives toward high potential idea generator.

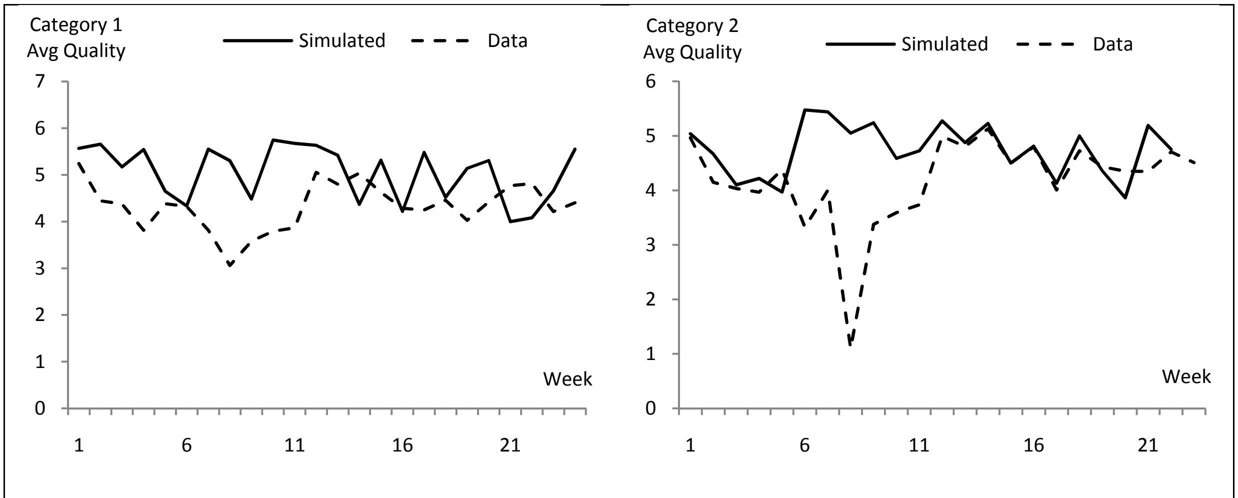


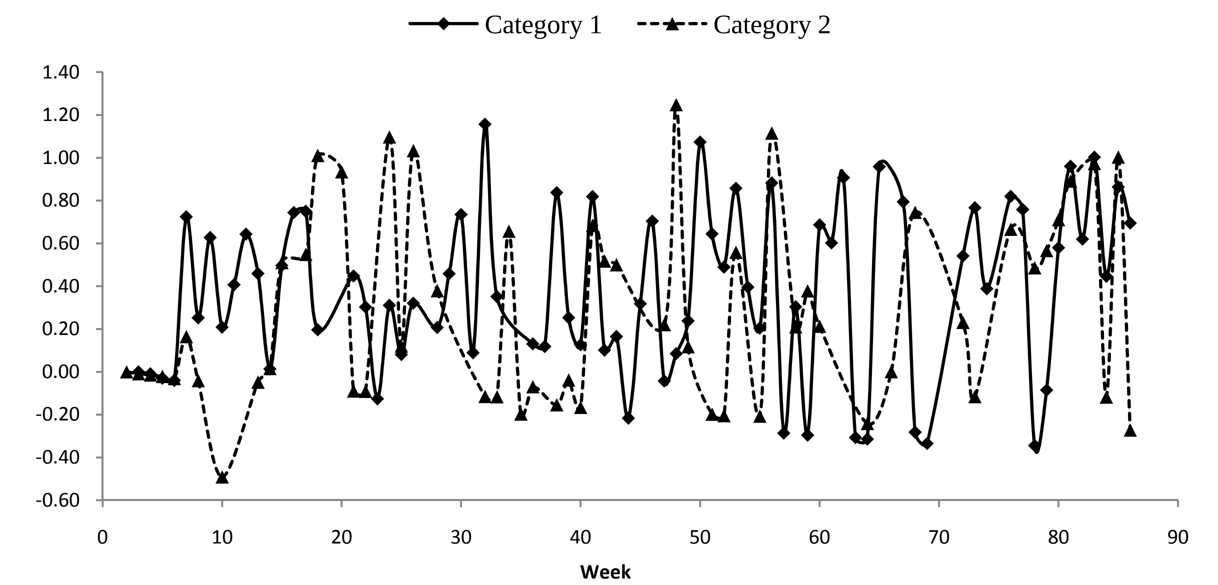
Figure 3. Four Weeks of Promotion vs. Current Policy

To test the effectiveness of such policy, we run a policy experiment in which the firm provides an additional reward when an idea is implemented. We set the reward at a level that the opportunity and cognitive cost can be fully compensated (i.e. the $\theta_1 = \theta_2 = 5$ under the new policy). The additional award is only applied in the first four weeks. We find that with the temporary incentive increase, the average potential of category 1 and category 2 ideas posted increase faster in the promotion periods, which means that the filtering is expedited (Figure 3). This result has good managerial implication. As mentioned in previous sections, it is very costly for the firm to review all the ideas and provide feedback to the idea creators. An acceleration of the learning can reduce the number of low potential ideas, without losing high potential ideas. The firm can still enjoy the profit that it can generate from implementing good ideas, and at the same time, they can reduce the cost of screening and reviewing a large number of ideas. Since the firm only provides the increased reward for very short time, the cost of implementing the policy will not be very high. However, this short term policy will benefit the firm for a longer period. In practice, most of the firms use a single reward policy from the initiation of crowdsourcing website to present without any changes; we believe that policies similar to the one we demonstrated in the experiment would be beneficial for the firms.

Discussion and Conclusion

Why Number of Contributed Ideas Decrease Over Time?

Our results show that initially users not only overestimate the potential of their ideas, but also underestimate the cost of implementing the ideas. Hence, the users tend to overestimate the probability that their idea will be implemented. Therefore they tended to post a lot of ideas. As individuals learn about the true cost structure of the firm as well as their ideas' potential, few users may find an expected net positive utility from posting an idea. Therefore, over time, a lot of user may stop contributing ideas to this website.



* The y axis is the average of perceived potential of the users who post ideas – the average of perceived potential of all users in the sample

Figure 4

Why Fraction of Ideas that are Implemented is Increasing Over Time?

Figure 4 shows the difference between the average of perceived idea potential of the users who post ideas and the average of perceived idea potential of all users in the sample. From the figure we can see that most of the observations are above 0. It tells us that users with higher self-perceived potential are more likely to post ideas. Since potential perception is updated by the observation of number of votes that the users' past ideas receive, users with higher perception about their own potential are by nature those who are proven to have better ability of generating "blockbuster" ideas. In other words, as people learn their potential from their experiences, the users with low perceived potential tend to post fewer ideas, and the remaining active idea creators are those who are of higher ability. This explains the increase in implementation rate as displayed in Figure 2. That is, users learn about the potential of their own ideas, and low potential users drop out while high potential users remain active, the overall potential of ideas are improved.

As crowdsourcing initiatives are becoming an increasing popular tool for new idea generation, a deeper understanding of users' behavior dynamic is needed. Although firms with crowdsourcing applications often appear on the media headlines for the adoption of the customer-oriented idea generating approach, whether crowdsourcing can provide the firms the result they are expecting, and in what kind of circumstances crowdsourcing will be the most effective are still unclear. Criticism about crowdsourcing is also raised when the decline in the number of ideas generated and the low implementation rates are observed. Firms' failure in crowdsourcing is often attributed to users' restrictive view of firms' products and limited knowledge about firms' cost structure.

However, an implicit assumption that these issues are based on is: the user group is static in their participation in the crowdsourcing websites. One important aspect of users' behavior that people pay little attention to is users' learning dynamics. With learning, users can obtain more precise perception of firms' cost structure, and better understanding of the potential of their own ideas. The expected utility of users of higher ability will be higher, and the expected utility of users of lower ability will be lower. Therefore, this learning process performs a filtering function. High ability users will be more willing to contribute new ideas, while low ability users will be less likely to post new ideas. As a result, the total number of ideas generated will decrease, as users of lower ability drop out, but the average quality of the ideas will increase over time. In the long run, the profit that the firms make by implementing the high quality ideas will increase; and at the same time, the cost of screening ideas will decrease. In summary, firms' utility will increase over time.

In this paper, we propose a structure model to capture users' learning dynamics in their participation in idea generation and estimate the model using a rich dataset obtain from Ideastorm.com. We find that the cost for the firm to implement category 1 ideas (product ideas) is higher, compared to category 2 ideas (customer service ideas). In addition, our results also show that when the crowdsourcing website was launched, users seem to underestimate firms' cost structure, and overestimate the potential of their ideas, which is consistent with our observation that in the beginning of the operations of crowdsourcing sites, users tend to post many infeasible ideas. However, as their perceptions about firms' cost structure and their ideas' potential become closer to the truth, users will post only when they believe that the potential of the idea has a good chance to outweigh the cost for the firm to implement it. It is consistent with our observation of the increase in implementation rate over time. In our policy experiment, we simulate users' behavior under a temporary promotion in the first four weeks of the operation of the website. Our simulation result shows that short term policies that accelerate users' learning process can help the firm filtering out low potential users and thus reduce the cost of the firm to review large number of ideas, without losing high potential ideas. Therefore, this policy simulation sheds a light on how firms can effectively crowdsource "blockbuster" ideas at low cost.

Our paper also has some limitations. First, we have not included the text content of the ideas and text comments into our model. These text contents may help explain the heterogeneity in the qualities of ideas within the same idea category. Second, we have not considered the interaction among users in this model. The distribution of the voting scores may be determined by not only the potential of the idea, but also users' social relation in this community. It would be interesting to incorporate the network aspects of such kind of web applications. Third, we have not incorporated individual level heterogeneity in the current model. In our model, we assume that the payoffs individuals receive when their ideas are implemented are the same across individuals. However, individuals may actually value the implementation of their ideas differently. If the correlation between the individuals' idea potential and their valuation of the implementation is positive, the actual learning effect will be smaller than what we our estimated; if the correlation is negative, the actual learning effect will be larger than what we our estimated. In addition, our analysis of users' potential is based on perceived potential, but not the true potential. We hope our work can pave the way for future research on this important area.

References

Aggarwal, R., Gopal, R., Sankaranarayanan, R., Singh, P. V. 2010. "Blog, Blogger and the Firm: Can Negative Posts by Employees Lead to Positive Outcomes?" *Information Systems Research*, forthcoming.

Albuquerque, P., Pavlidis, P., Chatow, U., Chen, K., Jamal, Z., Koh, K. 2010. "Evaluating Promotional Activities in an Online Two-Sided Market of User-Generated Content," *Working Paper*, University of Rochester.

Amabile, M. T., Barsade, G. S., Mueller, S. J., Staw, M. B. 2005. "Affect and Creativity at Work," *Administrative Science Quarterly* (50), pp. 367-403.

Bagozzi, R., Dholakia, U. 2006. "Open Source Software User Communities: A Study of Participation in Linux User Groups," *Management Science* (52:7), pp.1099-1115.

Brabham, D. 2008. “Moving the Crowd at iStock Photo: The Composition of the Crowd and Motivations for Participation in a Crowdsourcing Application,” *First Monday* (13:6-2), http://firstmonday.org /htbin/cgiwrap/ bin/ojs /index.php /fm/article/ viewArticle/2159/1969.

Bayus, B. L. 2010. “Crowdsourcing and Individual Creativity Over Time: The Detrimental Effects of Past Success,” *Working Paper*.

Crawford, G., Shum, M. 2005. "Uncertainty and Learning in Pharmaceutical Demand,” *Econometrica* (73:4), pp.1137–1173.

DeGroot, M. H. 1970.*Optimal Statistical Decisions*, New York, McGraw-Hill.

Erdem, T., Keane, M. P. 1996. “Decision-Making under Uncertainty: Capturing Dynamic Choice Processes in Turbulent Consumer Goods Markets,” *Marketing Science* (15:1), pp.1-20.

Erdem, T., Keane, M. P., Sun, B. 2008. “A Dynamic Model of Brand Choice When Price and Advertising Signal Product Quality,” *Marketing Science* (27:6), pp.1111-1125.

Fagerberg, J. 2005."Innovation: A Guide to the Literature", *The Oxford Handbook of Innovation*. Oxford.

Farrell. M. 2008. "How To Unlock Your Company's Creativity", http://www.forbes.com/2008/04/01/ innovation-netflix-blockbuster-ent-manage-cx_mf_0401innovation.html.

Franke, N., Katharina, K. 2009. “The Role of Perceived Fairness in Company-centred Crowdsourcing Communities,” *Working Paper*. Vienna: Vienna University of Economics and Business Administration.

Franke, N., Shah, S. K. 2003. “How Communities Support Innovative Activities: An Exploration of Assistance and Sharing Among End-Users,” *Research Policy* (32:1), pp. 157-178.

Franke, N., von Hippel, E. 2003. “Satisfying Heterogeneous User Needs via Innovation Toolkits: the Case of Apache Security Software,” *Research Policy* (32:7), pp. 1199-1215.

Franke, N., von Hippel, E, Schreier, M. 2006. “Finding Commercially Attractive User Innovations: A Test of Lead User Theory,” *Journal of Product Innovation Management* (23:4), pp.301-315.

Füller, J. 2010. “Refining Virtual Co-Creation from a Consumer Perspective,” *California Management Review* (52:2), pp. 98-122.

Ghose, A., and Han, S. 2009. “An Empirical Analysis of User Content Generation and Usage Behavior in Mobile Digital Media,” *Working Paper*.

Ghose, A., and Han, S. 2010. “A Dynamic Structural Model of User Learning in Mobile Media Content,” *Working Paper*.

Goldenberg, J., Lehmann, D. R., Mazursky D. 2001. "The Idea Itself and the Circumstances of Its Emergence as Predictors of New Product Success,” *Management Science* (47:1), pp.69-84.

Google. 2011."Patents and Innovation," http://googleblog.blogspot.com/2011/04/patents-and-innovation .html

Howe, J. 2006. “The Rise of Crowdsourcing,” *Wired* (14:6), www.wired.com/wired/archive/14.06/ crowds.html.

Huang, Y., Sing, P.V., Ghose, A. 2010. “Show Me the Incentives: A Dynamic Structural Model of Employee Blogging Behavior," *International Conference of Information Systems*, Saint Louis.

Kuan, J. 2001. "Open Source Software As Consumer Integration Into Production,” *Working Paper*.

Kumar. V., Sun, B. 2010. “Why do Consumers Contribute to Connected Goods? A Dynamic Game of Competition and Cooperation in Social Networks,” *Working Paper*.

Lakhani, K. R., Jeppesen, L. B., Lohse, P, A., Panetta J, A. 2007. "The Value of Openness in Scientific Problem Solving," *Working Paper*. Boston, MA: Harvard Business School.

Lakhani, K. R., von Hippel, E. 2003. “How Open Source Software Works: "Free" User-to-user Assistance,” *Research Policy* (32:6), pp. 923-943.

Leonard, D., Rayport, J. F. 1997. “Spark Innovation Through Empathic Design,” *Harvard Business Review* (75:6), pp.102-113.

Lu, Y., Singh, P. V., Sun, B. 2011. “Learning from Peers on Social Media Platforms”, *International Conference of Information Systems*, Shanghai.

Lu, Y., Jerath, K., and Singh, P.V. 2010. "Emergence of opinion Leaders in Online Review Communities," *Statistical Challenges in Electronic Commerce Research*, Austin 2010.

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

Majchrzak, A., Cooper, L. P., Neece, O. E. 2004. "Knowledge Reuse for Innovation", *Management Science* (50:2), pp. 174-188.

McFadden, D. 1989. “A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Intergration,” *Econometrica* (57), pp.995-1026.

Mehta, N., Rajiv, S., Srinivasan, K. 2003. "Price Uncertainty and Consumer Search: A Structural Model of Consideration Set Formation," *Marketing Science* (22:1), pp. 58-84.

Morrison, P. D., Roberts, J. H., von Hippel, E. 2000. "Determinants of User Innovation and Innovation Sharing in a Local Market," *Management Science* (46:12), pp.1513-1527.

Mukhopadhyay, T., Singh, P.V., and Kim, S. 2011. "Learning Curves of Agents with Diverse Skills in Information Technology Enabled Physician Referral Systems, " *Information Systems Research*, September.

Ogawa, S., Piller, F. 2006. "Collective User Commitment: Reducing the Risks of New Product Development," *Sloan Management Review* (47:2), pp.65-72.

Olson, E. L., Bakke, G. 2001. "Implementing the Lead User Method in a High Technology Firm: A Longitudinal Study of Intentions versus Actions", *The Journal of Product Innovation Management* (18:6), pp. 388-395.

Pakes, A. 1987. "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks," *Econometrica* (54), pp. 755-784.

Poetz, M., Schreier M. 2010. "The Value of Crowdsourcing: Can Users Really Compete with Professionals in Generating New Product Ideas?" *Journal of Product Innovation Management*, forthcoming.

Sahoo, N., Singh, P.V., and Mukhopadhyay, T. 2011. A Hidden Markov Model of Collaborative Filtering. *Utah Winter IS Conference*.

Schulze, A. and Hoegl M. 2008. "Organizational Knowledge Creation and the Generation of New Product Ideas: A Behavioral Approach, " *Research Policy* (37:10), pp.1742-1750.

Shah, S. K. 2000. "Sources and Patterns of Innovation in a Consumer Products Field: Innovations in Sporting Equipment," *Working Paper*, MIT Sloan School of Management.

Shah, S. K. 2006. "From Innovation to Firm Formation: Contributions by Sports Enthusiasts to the Windsurfing, Snowboarding & Skateboarding Industries", *The Engineering of Sport* (6:1), pp. 29-34.

Singh, P.V. 2010. "The Small World Effect: The Influence of Macro Level Properties of Developer Collaboration Networks on Open Source Project Success," *ACM Transactions of Software Engineering and Methodology*, 20(2), pp 6:1-6:27

Singh, P.V., Tan, Y., and Mookerjee, V. 2011a. "Networks Effects: The Influence of Structural Social Capital on Open Source Software Projects," *MIS Quarterly* forthcoming.

Singh, P.V., Tan, Y., and Youn, N. 2011b. " A Hidden Markov Model of Developer Learning Dynamics in Open Source Software Projects," *Information Systems Research* forthcoming.

Singh, P.V., Sahoo, N., and Mukhopadhyay, T. 2010. Seeking Variety: A Dynamic Structural Model of Blog Reading Dynamics. *Workshop on Information systems and Economics*, Saint Louis.

Ulrich, K. T. 2007. *Design: Creation of Artefacts in Society*, Pontifica Press.

Ulrich, K. T., Eppinger, S. D. 2008. *Product Design and Development*, New Jersey: McGraw-Hill.

Urban, G. L., von Hippel, E. 1988. "Lead User Analyses for the Development of New Industrial Products," *Management Science* (34:5), pp. 569-582.

von Hippel, E. 2005. *Democratizing Innovation*, Cambridge, MA: MIT Press.