Scalable Crowd Ideation Support through Data Visualization, Mining, and Structured Workflows

Victor Girotto
Arizona State University
Tempe, AZ 85281, USA
victor.girotto@asu.edu

Erin Walker
Arizona State University
Tempe, AZ 85281, USA
erin.a.walker@asu.edu

Winslow Burleson
New York University
New York, NY 10012, USA
wb50@nyu.edu

CSCW '17 Companion, February 25 - March 01, 2017, Portland, OR, USA
ACM 978-1-4503-4688-7/17/02.
http://dx.doi.org/10.1145/3022198.3026349

Abstract
As the size of innovation communities increases, methods of supporting their creativity need to scale as well. Our research proposes the integration of three scalable techniques into a crowd ideation system: 1) data visualization, 2) structured microtask workflows, and 3) data mining, with the goal of supporting users in convergent and divergent ideation processes. In addition, these techniques do not work in isolation, but instead support each other. Our vision is to create a system that intelligently supports users' ideation in a crowd context while maintaining their agency and facilitating exploration and decision-making.

Author Keywords
Crowdsourcing; ideation; creativity; data visualization; data mining; microtasks;

ACM Classification Keywords
H.5.3. Group and Organization Interfaces: Computer-supported cooperative work.

Introduction & Related Work
Collaboration is occurring at an ever-increasing scale. As communities are formed and grow around shared passions, the possibility of generating innovation

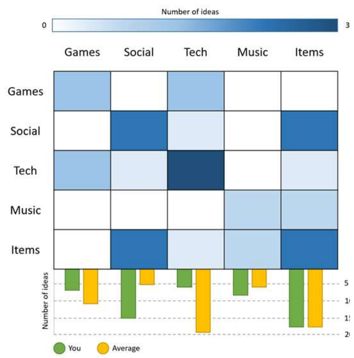

Figure 1: Matrix visualization of the solution space. Each column and row in the matrix represents a different idea category. The darker the color, the more ideas have been developed within a category. Below, a bar chart compares the performance of the user (green) vs. the average performance of the crowd (yellow) in a given category.

through them also increases. For example, crowds are used as a source of ideas ranging from T-shirts (https://www.threadless.com) to tough innovation challenges (https://www.innocentive.com). While in individual or smaller groups a commonly used technique for generating such ideas is brainstorming, it may not necessarily scale well to these large communities, as the sheer number of ideas generated can overload those who are generating or selecting promising ideas.

Currently, however, there are well-known techniques for dealing with large datasets: 1) data mining, which allows for automatic identification of patterns in the data; 2) data visualization, which facilitates visual exploration of the data; and 3) microtask workflows, which breaks down bigger tasks into smaller, more manageable chunks. All three approaches bring forth complementary strengths in dealing with large datasets. Therefore, we seek to integrate them naturally into a system that supports the task of idea generation and selection.

There have been several attempts at supporting group or crowd ideation, such as employing human facilitators [3] or using carefully structured processes [10]. Closer to our approach, the IdeaHound system integrated classification tasks into the interface affordance of spatially organizing ideas into clusters [9]. This allowed the system to calculate similarity metrics between the ideas, enabling three forms of support interventions: supplying the user with diverse examples, similar examples, and a visualization of the solution space.

Our approach differs from most of the aforementioned research in that it seeks to employ the same group of users for both idea and inspiration generation. In this sense, we are in agreement with the goals of the IdeaHound system of being near real-time and self-sustainable [9]. We also similarly seek a hybrid human-computer approach for scalable support. However, this approach differs from theirs in that it makes microtasks explicit rather than implicit, contextualizing them as a form of inspiration and contribution to the ideation process. There is also a greater focus on supporting exploration as well as cognitive and social processes through focused coordinated visualizations rather than accurate model generation. Finally, we hope to use the techniques described here not only for ideation, but also for iteration and selection of ideas.

This research will contribute to the advancement of crowd ideation systems by investigating the effects of data visualization and microtasks in crowd idea generation, improvement, and selection, as well as modeling and adaptive support of crowd ideators.

Approach

Data visualization

The most visible form of support to the user will be a set of coordinated visualizations designed for highlighting important aspects of the ideation process according to best practices established by creativity research. One inspiration for this is research on supporting serendipity through visualization (e.g. [1]).

The first proposed visualization is one of the solution space (Figure 1). Such a visualization could answer questions such as: how many and which idea categories have been developed so far? Which categories have received more attention than others? Are there ideas that overlap two categories? This should help users in different ways. Seeing categories of ideas different than those explored by him/herself, a user can generate more ideas than without that support [7]. Seeing how much attention some categories have received can help users direct their efforts to areas that have not yet been fully explored, perhaps decreasing redundant ideas. Finally, making overlaps between categories explicit can direct users to the process of idea combination (e.g. [10]).

One could also elicit social comparison processes through a visualization, answering user questions such as: how is my performance in comparison to the rest of the group? Which areas have I contributed the most to? Am I one of the strongest contributors to a particular category of ideas? This way, one could avoid issues such as social loafing, while promoting a healthy upwards comparison [8]. This information could be conveyed very simply through bar charts comparing the user's and average crowd's performance.

Microtasks
A common way of inspiring ideators is to show them ideas generated by other people. We propose that instead of simply showing other ideas, we also embed a microtask with them, such as rating their originality. This could have the effect of improving attention to the idea, which could increase the likelihood of it actually inspiring the user [7]. In initial studies, we have found some evidence supporting this hypothesis. By contextualizing tasks as a form of inspiration, users may be motivated to do them.

At the same time that users may be inspired by the ideas they see, they are also contributing with more information on each idea. In the previous example, this extra information would be an originality rating for the idea, which could help when the crowd starts the process of converging into the best ideas. If users performed similarity comparisons, the end result could be a semantic model of the ideas, as done in the IdeaHound system [9].

Finally, microtasks can also help improve the underlying models used in both visualization and data mining (as explained below). For example, the first visualization suggested a taxonomy of ideas. While there are ways of generating this taxonomy automatically, they will not be perfect. Microtasks inspired by previous taxonomy-creation workflows (e.g. [4]) could improve these models.

Data Mining
Underlying the previous two items is the notion of automatic categorization of ideas as the basis of the visualization and of idea suggestions. Techniques such as LDA can be used to generate topics based on a text corpus. While the result has some noise, there are ways of employing visualizations similar to the one proposed here to aid users in understanding the models [5]. Microtasks could then be used to clean them up. Additionally, user models could also be constructed, measuring statistics such as likelihood of user being stuck, fixated, or in other important mental states, allowing the system to intervene accordingly. We plan on leveraging advancements in Intelligent Tutoring Systems (ITS) [6] to model support for ideators.

While microtasks could improve attention to ideas, it is important that the ideas they present are helpful for users. If they are far from their knowledge or interests, for example, they may not be able to benefit from

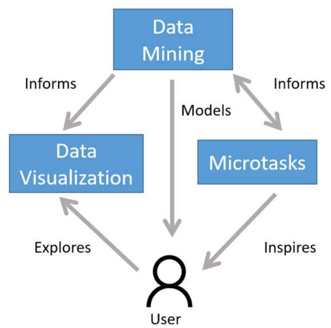

Figure 2: The different components described in this paper and their interactions. The user interacts directly with the visualization (for exploration) and the microtasks (which inspire the user, while also contributing to fine-tuning the data mining component. Meanwhile, the data mining component is modelling the user based on his or her performance, and is improved based on the microtasks. It also informs the visualization (e.g. providing it with the idea categories) and the microtask selection (which ideas or tasks should be shown to a particular user at a given moment?)

them. Research on recommender systems could help to alleviate this issue. For example, a collaborative filtering model [2] could be built based on the categories users contributed to, thus allowing new categories of ideas to be suggested to users who share similar patterns.

Conclusion

Modern techniques to handle large scale data seem ripe to intersect with known creativity enhancing practices in crowd contexts. While no single technique seems to solve all of the issues, a synergy between techniques seems promising. Figure 2 describes the interactions between these different components. This works aims to build on current research to further understand how they can be integrated in order to fully develop the crowd innovation potential.

Acknowledgments

This research was funded by the CAPES Foundation, Ministry of Education of Brazil, Brasília - DF 70040-020, Brazil.

References

1. Eric Alexander, Joe Kohlmann, Robin Valenza, Michael Witmore, and Michael Gleicher. 2014. Serendip: Topic model-driven visual exploration of text corpora. In *Visual Analytics Science and Technology (VAST), 2014 IEEE Conference on*, 173–182.
2. J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez. 2013. Recommender systems survey. *Knowledge-Based Systems* 46: 109–132.
3. Joel Chan, Steven Dang, and Steven P Dow. 2016. Improving Crowd Innovation with Expert Facilitation. 1221–1233.
4. Lydia B. Chilton, Greg Little, Darren Edge, Daniel S. Weld, and James A. Landay. 2013. Cascade: Crowdsourcing taxonomy creation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1999–2008.
5. Jason Chuang, Christopher D. Manning, and Jeffrey Heer. 2012. Termite: Visualization techniques for assessing textual topic models. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, 74–77.
6. Michel C. Desmarais and Ryan S. J. d. Baker. 2012. A review of recent advances in learner and skill modeling in intelligent learning environments. *User Modeling and User-Adapted Interaction* 22, 1–2: 9–38.
7. Bernard A. Nijstad, Wolfgang Stroebe, and Hein FM Lodewijkx. 2002. Cognitive stimulation and interference in groups: Exposure effects in an idea generation task. *Journal of experimental social psychology* 38, 6: 535–544.
8. Paul B. Paulus and Vincent R. Brown. 2003. Enhancing ideational creativity in groups: Lessons from research on brainstorming. In *Group creativity: Innovation through collaboration*. Oxford University Press.
9. Pao Siangliulue, Joel Chan, Steven P. Dow, and Krzysztof Z. Gajos. 2016. IdeaHound: Improving Large-scale Collaborative Ideation with Crowd-Powered Real-time Semantic Modeling. 609–624.
10. Lixiu Yu and Jeffrey V. Nickerson. 2011. Cooks or cobblers?: crowd creativity through combination. In *Proceedings of the SIGCHI conference on human factors in computing systems*, 1393–1402.