Motivations of Citizen Scientists – A Quantitative Investigation of Forum Participation

Ramine Tinati
University of Southampton
Web and Internet Science
rt506@ecs.soton.ac.uk

Markus Luczak-Roesch
University of Southampton
Web and Internet Science
mlr1m12@soton.ac.uk

Elena Simperl
University of Southampton
Web and Internet Science
e.simperl@soton.ac.uk

Nigel Shadbolt
University of Southampton
Web and Internet Science
nrs@soton.ac.uk

Categories and Subject Descriptors

H.4.m [**Information Systems Applications**]: Miscellaneous

Keywords

Citizen science; online communities; discussion forums

1. INTRODUCTION

In recent years, the Web has fostered an emergent activity which has enabled individuals to take part and contribute to scientific discovery without the need for specific knowledge or expertise. These activities, which have been labelled as *Citizen Science* [1, 3, 4], use crowdsourcing and human-computation techniques[6] as a means to complete scientific tasks. Whilst the primary focus of citizen science is to achieve computationally complex tasks there has been growth of online citizen science community activity[5]. Online discussion forums and other forms of communication mechanisms such as social media are offering citizen scientists to talk, engage, share knowledge, and in several cases, achieve citizen-led scientific discovery [1].

In this paper we build upon previous qualitative studies of investigating the motivations of citizen science volunteers [8, 7] and analyse the Galaxy Zoo discussion forum[1]. We explore the motivations of citizen scientists participation and develop a model based on a number of features extracted from an individual's participation within the discussion forum. The purpose of this study is to further develop our current understanding of the characteristics of citizen science participation the factors that affect their participation and help better understand the engagement of citizen scientists.

In summary, we identified 9 clusters of participants which reflect the different features of forum interaction. By contextualising the features, we show that the interactions of the clustered participants may represent the motivations identified in previous studies.

---

[1]Galaxy Zoo Forums http://www.galaxyzooforum.org

Feature	Feature viable	Feature selected
# of posts created by a user	x	x
# of boards a user posted to	x	x
# of threads a user posted to	x	x
active lifetime of user account	x	
average length of a user's posts	x	
Frequency of user posts	x	x
# of personal messages sent	x	x
# of personal messages received	x	
# of voting polls created by a user	x	
# of votes cast by a user	x	x
# of sticky threads created by a user	x	
# of views received for user threads	x	x
# of responses received user threads	x	x
# of thread update notifications	x	x
# of board update notifications	x	
median position of a user's first posts in threads	x	
median position of a user's last posts in threads	x	
# of threads created by a user	x	x
# of threads that the user has provided the first reply	x	x
median # of posts a user contributes a thread	x	
# of questions asked by a user		
# of questions answered by a user		
# of quality posts created		

Table 1: Overview of our extension of the user contribution model and the final set of features selected.

2. RESULTS AND ANALYSIS

The analysis in this paper uses a dataset containing a snapshot of the Galaxy Zoo forum, containing all the posts made between 25th July 2007 (the first entry on the forum) to 31st January 2012. In order to investigate the relationship between the activity of a participant and the motivations of citizen scientists identified in previous studies within the Galaxy Zoo forum [8, 7], we propose a set of features that relate to a participant's activity within a forum based on 'user contribution management' literature [9, 2]. Table 1 lists the user contribution features which we use in this analysis.

We applying the Expectation Maximisation (EM) method as a model-based clustering approach which does not assume a fixed set of clusters and also allows for the detection of over fitting. By refining the feature set and fitting the best suited model to be applied we also gain insight into how many clusters represent the best partitioning of our data. We also apply a k-means approach iteratively for a minimum k=2 up to a maximum of k=15 (k is incremented by 1 in each iteration). This allows for estimating the optimal number of clusters for a given data set by determining significant drops in within groups sum of squares plotted against k.

The first iteration of the EM clustering method indicated that the 20 features result in overfitting. We iteratively eliminated features which did not perform well in terms of differentiating participant types until overfitting was reduced. The model-based approach and the k-means clustering suggest that good results can be achieved by 9 clusters. This result is the first indicator that not all of the 12 motivation categories of citizen scientists can be mapped to participant participation profiles in the Galaxy Zoo forum. Based on the results, we analyse the characteristics of each of the nine clusters and summarise the nine different participant types as follows:

Cluster 1: This group of highly active participants contributes a significantly high proportion of first replies to other participants' posts in a wide variety of threads. Since the amount of thread update notifications is rather low, we suggest that this represents question answering participants.

Cluster 2: The most active forum participants; they produce a very high number of posts in a wider variety of boards and threads, start threads and also provide initial replies. This seems to be the core group of participants, which feature a domestic behaviour on the Galaxy Zoo forum, so that they do not tend to exploit the thread notification feature for example.

Cluster 3: Similar feature characteristics as those in cluster 1 but differentiate from the by the high amount of thread notifications and low amount of first replies. These participants are likely to start new threads and show a significant interest in ongoing discussions without the motivation to contribute answers. Hence, they can be regarded as the followers asking new questions occasionally driven by a learning motivation.

Cluster 4: Participants who produce a high number of initial replies in the widest variety of boards. Since thread notifications are not used by this group, we suggest that these participants answer questions as the participants in cluster 1 but which are typically more trivial and do not result into real community discourse.

Cluster 5: Showing a very similar feature characteristic as cluster 3, this cluster is bordered by the lower personal message activity and amount of thread notifications. Both features indicate that these participants are also followers (as those in cluster 3) but with a lower involvement into the community and most likely also not driven by the motivation to learn from experienced participants and particular contents.

Cluster 6: This largest group of participants can be described as typical forum reader being active over a longer period of time due to a general interest in the forum thread. No feature shows significant peaks or dales.

Cluster 7: Characterised by kicking of a majority of threads in a wide variety of boards, participants in this cluster represent the long tail of only few participants contributing the majority of the content to a forum.

Cluster 8: People classified into this group participate in polls fairly frequent and spread a high number of posts across a wide variety of threads and boards. Also the low number of thread notifications and the high amount of initial replies stick out but most noteworthy is the number of responses the threads created by these participants obtain. We suggest, that these participants are highly engaged in real discussions and dispose fair domain knowledge they share with the community.

Cluster 9: Participants asking individual questions only, with a small number of posts, almost no personal message and poll activity and a low number of first replies to posts of other participants. The latter may indicate that participants within this group ask for help in the forum but do not have any other recognisable motivation for contributing forum content.

3. CONCLUDING REMARKS

The findings in this paper have show how there are several groups of individuals which are highly active, posting multiple times, within many different boards, and are eager respond. Alongside these highly motivated individuals are the the 'lurkers' on the forums that do not engage much. From a citizen science perspective the clusters identified support previous studies of citizen science motivations and relate interaction characteristics to these motivations [8, 7]. These findings provide further insight regarding the motivations of citizen scientists, as well as the study of online communities.

Future work in this area involves taking the features identified in this study and applying it to other citizen science discussion systems as well as communities using social media platforms. By applying the feature model developed in this paper to these systems, we wish to compare the characteristics of participation, and whether the same clusters, thus motivations can be identified.

4. ACKNOWLEDGEMENTS

5. REFERENCES

[1] R. Bonney, C. B. Cooper, J. Dickinson, S. Kelling, T. Phillips, K. V. Rosenberg, and J. Shirk. Citizen science: a developing tool for expanding science knowledge and scientific literacy. *BioScience*, 59(11):977–984, 2009.

[2] K. Chai, V. Potdar, and E. Chang. User contribution measurement model for web-based discussion forums. *2009 3rd IEEE International Conference on Digital Ecosystems and Technologies*, pages 347–352, June 2009.

[3] S. A. Gray, K. Nicosia, and R. C. Jordan. Lessons learned from citizen science in the classroom. a response to" the future of citizen science.". *Democracy and Education*, 20(2):14, 2012.

[4] A. Irwin. *Citizen science: A study of people, expertise and sustainable development*. Psychology Press, 1995.

[5] G. Mugar, C. Ø sterlund, and K. Hassman. Planet Hunters and Seafloor Explorers: Legitimate Peripheral Participation Through Practice Proxies in Online Citizen Science. *CSCW'14*, 2014.

[6] A. J. Quinn and B. B. Bederson. Human computation: a survey and taxonomy of a growing field. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1403–1412, 2011.

[7] J. Raddick, C. Lintott, and S. Bamford. Galaxy Zoo: Motivations of Citizen Scientists. *Astronomy Education Review*, pages 1–41, 2013.

[8] M. J. Raddick, G. Bracey, P. L. Gay, C. J. Lintott, P. Murray, K. Schawinski, A. S. Szalay, and J. Vandenberg. Galaxy zoo: Exploring the motivations of citizen science volunteers. *Astronomy Education Review*, 9:010103, 2010.

[9] Y. Wang and D. Fesenmaier. Understanding the motivation of contribution in online communities: An empirical investigation of an online travel community. *TTRA Annual Conference*, (2000), 2003.