

On the Value of CrowdRE in Research and Practice

Fabiano Dalpiaz
RE-Lab, Utrecht University
Utrecht, The Netherlands
f.dalpiaz@uu.nl

Abstract—Crowd-based Requirements Engineering, or CrowdRE, is an emerging requirements engineering paradigm in which a crowd of users plays a central role in carrying out activities such as elicitation and prioritization. In this keynote, I analyze the value of CrowdRE, focusing on its effectiveness and impact in research and in practice. Starting from research carried out in collaboration with industrial partners, I make the case for the need of conducting in vivo research for CrowdRE. This type of research is crucial to gather evidence for the practical effectiveness and to ultimately establish CrowdRE as a prominent paradigm for the software industry.

Index Terms—CrowdRE, requirements elicitation, case studies

I. INTRODUCTION

User involvement is a key ingredient of requirements elicitation, which has been shown to contribute positively to system success [1]. In Crowd-Based Requirements Engineering (CrowdRE) [2], user involvement typically involves gathering and analyzing user feedback. Software companies may obtain such feedback in two ways [2]: (i) *pull*: by explicitly requesting the crowd to express their needs, often via an online platform such as a user forum [3], [4]; or (ii) *push*: by analyzing unsolicited inputs that are voluntarily provided by the users, such as reviews in app stores [5], [6].

In both scenarios, significant human effort is required from both the product team and the users. After establishing a crowd, the motivation of the participating users needs to be sustained, the collected inputs need to be processed and combined with the product's roadmap, and the users should be made aware that their feedback has been addressed.

Consequently, a core question for the CrowdRE research community is: *How to reduce the required human effort for analyzing user feedback so to increase CrowdRE's value for practice?*

II. VALUE FOR RESEARCH: AUTOMATION

Thus far, most researchers have answered this question by arguing that the solution lies in *automation*. Several semi-automated approaches exist that aim to reduce human effort when analyzing large quantities of user feedback. These approaches often employ machine and deep learning (ML/DL) to classify and to summarize thousands or millions of user reviews. For example, Maalej and colleagues [7] determined that supervised machine learning algorithms outperform keyword-based approaches for classifying user reviews into bug reports, feature requests, user experience, and ratings. Dhinakaran *et al.* [8] showed how employing active learning to perform

similar classification tasks may reduce human effort while providing superior prediction accuracy. Guzman and Maalej [9] investigated the effectiveness of automated sentiment analysis of app store reviews.

These are some examples of the many papers (for a comprehensive picture, see [6], [10], [11]) that perform automated analyses on app store reviews, focusing on push feedback. This growing body of literature testifies the academic relevance and value of CrowdRE. In order to argue for automation's value in CrowdRE practice, the researchers generally rely on optimizing the automation's performance, which is reported using metrics such as precision, recall, and F-score.

III. VALUE FOR PRACTICE?

Less attention has been paid to the actual value in practical settings. Does the value for the research community (i.e., published papers and high citation counts) lead to comparable value for a development team that wishes to adopt CrowdRE techniques? Some studies investigated this question.

Push-feedback. Recent work examined the relative performance of automation (natural language processing and ML/DL) against human performance. Taking as a baseline a crowdsourced annotation of user reviews (relevant vs. irrelevant) [12], Mekala *et al.* showed how deep learning algorithms could outperform crowd annotators [13] in terms of precision, recall, and F-score. Although preliminary, these results seem to indicate a better cost-benefit ratio than crowdsourcing. Yet, do high precision and recall translate to value for practitioners? Are the classes (e.g., relevant vs. irrelevant, or bug vs. feature request) the most relevant ones? And how much value is lost by mis-classifying X% of the data?

Pull-feedback. Some case studies offer preliminary evidence of the effectiveness of CrowdRE feedback channels. Stade *et al.* [14] show the importance of making the feedback interface prominently visible to the users; by presenting a link whenever the users accessed the system, they collected over 80 feedback entries in two weeks. Kolpondinos and Glinz [4] managed to involve stakeholders outside the organization's reach through the GARUSO gamified platform. By analyzing crowd feedback in two software ecosystems, Johnson *et al.* [15] discussed the challenges of managing the crowd and of analyzing the inputs when multiple, heterogeneous feedback channels exist. Open questions remain: what is the added value of these channels compared to more traditional elicitation techniques? In which way can we combine these channels? And how can we merge these inputs with the software product's roadmap?

IV. A PERSONAL JOURNEY ACROSS RESEARCH AND PRACTICE

Starting from research executed in collaboration with partners from industry, I will offer a personal, yet empirically grounded, perspective on the value of conducting CrowdRE research in practice by means of case study or action research.

The *REfine* platform [3] was our lab's first effort to integrate gamification elements into CrowdRE. Through a case study regarding a governance risk and compliance web platform created and used within the consultancy company KPMG, we successfully engaged a small crowd of users in a pull feedback platform. The study highlighted the importance of establishing a clear goal for the crowd (in our study, improving the usability on mobile devices), and how an active moderator in soliciting feedback and managing the platform is crucial.

In the context of the *Tournify* sports tournament planning app [16], the start-up company who developed the app deployed an elicitation platform for their users to express requirements via simplified user stories. This research showed that CrowdRE elicitation may lead to numerous inputs, but also confirmed the difficulty of sustaining user engagement over time: most contributors brought in a single idea.

In this 2021 edition of the RE conference [17], we presented our experience with applying CrowdRE in a governmental organization: the Royal Netherlands Marechaussee. We succeeded at scaling up CrowdRE: via two case studies, we mobilized 130 and 60 users from whom we gathered 78 and 32 ideas, and 531 and 316 votes, respectively. Many of those ideas turned out to be useful, several creative and novel. Our investigation showed that gamification elements had no effect, while it confirmed the importance of a human moderator who manages the crowd and keeps it active.

We have also investigated the practical effectiveness of automation applied to push feedback. Our *RE-SWOT* [18] platform adapts Strength-Weaknesses-Opportunities-Threats analysis to RE. Through an unsupervised approach, *RE-SWOT* extracts app features from the user reviews of multiple competing apps, and then classifies those features on the basis of the app rating given by the reviews. Through three case studies in the dating, travel, and games domains, *RE-SWOT* provided unknown information to the product owners, such as features that the competitors had recently implemented, but also showed that the inaccuracy of automation may introduce noise that seriously hampers the usability of the results.

V. CONCLUSION

In the keynote, I will follow the structure sketched in this extended abstract, and I will review the interplay between research value and value for practitioners. Based on the findings obtained in previous research (partially outlined in Sec. IV), I will put forward concrete directions for conducting CrowdRE research that has value for RE practice.

ACKNOWLEDGMENT

I thank all the members of the RE-Lab who conducted research on CrowdRE: in alphabetical order, Sjaak Brinkkemper,

Eduard C. Groen, Abel Menkveld, Micaela Parente, Remco Snijders, Martijn van Vliet, Boris Winter, and Jelle Wouters.

REFERENCES

- [1] M. Bano and D. Zowghi, "A systematic review on the relationship between user involvement and system success," *Information and Software Technology*, vol. 58, pp. 148–169, 2014.
- [2] E. Groen, N. Seyff, R. Ali, F. Dalpiaz, J. Doerr, E. Guzman, M. Hosseini, J. Marco, M. Oriol, A. Perini, and M. Stade, "The crowd in requirements engineering: The landscape and challenges," *IEEE Software*, vol. 34, no. 2, 2017.
- [3] R. Snijders, F. Dalpiaz, S. Brinkkemper, M. Hosseini, R. Ali, and A. Özüm, "REfine: A Gamified Platform for Participatory Requirements Engineering," in *International Workshop on Crowd-Based Requirements Engineering*, 2015.
- [4] M. Z. Kolpondinos and M. Glinz, "GARUSO: A gamification approach for involving stakeholders outside organizational reach in requirements engineering," *Requirements Engineering*, vol. 25, no. 2, pp. 185–212, 2020.
- [5] D. Pagano and W. Maalej, "User feedback in the Appstore: An empirical study," in *IEEE International Requirements Engineering Conference*, 2013, pp. 125–134.
- [6] W. Martin, F. Sarro, Y. Jia, Y. Zhang, and M. Harman, "A survey of app store analysis for software engineering," *IEEE Transactions on Software Engineering*, vol. 43, no. 9, pp. 817–847, 2016.
- [7] W. Maalej, Z. Kurtanović, H. Nabil, and C. Stanik, "On the automatic classification of app reviews," *Requirements Engineering*, vol. 21, no. 3, pp. 311–331, 2016.
- [8] V. T. Dhinakaran, R. Pulle, N. Ajmeri, and P. K. Murukannaiah, "App review analysis via active learning: Reducing supervision effort without compromising classification accuracy," in *International Requirements Engineering Conference*, 2018, pp. 170–181.
- [9] E. Guzman and W. Maalej, "How do users like this feature? a fine grained sentiment analysis of app reviews," in *IEEE International Requirements Engineering Conference*, 2014, pp. 153–162.
- [10] J. A. Khan, L. Liu, L. Wen, and R. Ali, "Crowd intelligence in requirements engineering: Current status and future directions," in *International Working Conference on Requirements Engineering: Foundation for Software Quality*, ser. LNCS, vol. 11412, 2019, pp. 245–261.
- [11] R. Santos, E. C. Groen, and K. Vilela, "An overview of user feedback classification approaches," in *International Workshop on Natural Language Processing for Requirements Engineering*, 2019.
- [12] M. van Vliet, E. C. Groen, F. Dalpiaz, and S. Brinkkemper, "Identifying and classifying user requirements in online feedback via crowdsourcing," in *International Working Conference on Requirements Engineering: Foundation for Software Quality*, ser. LNCS, vol. 12045, 2020, pp. 143–159.
- [13] R. R. Mekala, A. Irfan, E. C. Groen, A. Porter, and M. Lindvall, "Classifying user requirements from online feedback in small dataset environments using deep learning," in *IEEE International Requirements Engineering Conference*, 2021.
- [14] M. Stade, M. Oriol, O. Cabrera, F. Fotrousi, R. Schaniel, N. Seyff, and O. Schmidt, "Providing a user forum is not enough: First experiences of a software company with CrowdRE," in *International Workshop on Crowd-Based Requirements Engineering*, 2017.
- [15] D. Johnson, J. Tizard, D. Damian, K. Blincoe, and T. Clear, "Open CrowdRE challenges in software ecosystems," in *International Workshop on Crowd-Based Requirements Engineering*, 2020.
- [16] A. Menkveld, S. Brinkkemper, and F. Dalpiaz, "User story writing in crowd requirements engineering: The case of a web application for sports tournament planning," in *International Workshop on Crowd-Based Requirements Engineering*, 2019.
- [17] J. Wouters, R. Janssen, B. van Hulst, J. van Veenhuizen, F. Dalpiaz, and S. Brinkkemper, "CrowdRE in a Governmental Setting: Lessons from Two Case Studies," in *IEEE International Requirements Engineering Conference, Industrial Innovation Track*, 2021.
- [18] F. Dalpiaz and M. Parente, "RE-SWOT: from user feedback to requirements via competitor analysis," in *International Working Conference on Requirements Engineering: Foundation for Software Quality*, ser. LNCS, vol. 11412, 2019, pp. 55–70.