A Role for Provenance in Social Computation

Milan Markovic, Peter Edwards, and David Corsar

Computing Science & dot.rural Digital Economy Hub, University of Aberdeen,
Aberdeen, AB24 5UA
{m.markovic,p.edwards,dcorsar}@abdn.ac.uk

Abstract. We argue that existing systems to support social computation suffer from a lack of transparency and that this can be addressed by integrating provenance capture mechanisms into such systems. We discuss how Semantic Web technologies can be used to facilitate this, and how the provenance record could be used to support various forms of decision-making about tasks such as workforce selection.

1 Introduction

The widespread use of online interactive technologies has enabled new forms of computations based on the principles of collective intelligence [4, 2]. Robertson and Giunchiglia [4] define one such approach, social computation, as: *"a computation for which an executable specification exists, but the successful implementation of this specification depends upon computer-mediated social interaction between the human actors and its implementation"*. However, the use of humans in such computations introduces several issues including: reliability of workers, workforce selection, and quality of the generated results. To address these issues, current platforms such as Amazon's Mechanical Turk[1] provide basic reputation scores for workers based on acceptance of their product, tools for workforce selection based on worker's attributes (e.g. geolocation, qualifications) and means to assess results (e.g. by comparison with a gold standard).

We argue that recording the provenance of such activities and other aspects of social computation (e.g. formation of a group of participants) will increase the transparency of such systems, and so enable more sophisticated means of control. Such a provenance record would describe the activities performed throughout the computation, the entities (things) used and generated by those activities, and the agents associated with those activities [3]. This can then be used to enhance assessments of: workers reliability (e.g. forming beliefs about their trustworthiness based on their motives, past performance, capabilities, and relationships to trusted workers); results (e.g. by reconstructing and inspecting the events that lead to result generation); the process of the execution itself (e.g. how the group of workers necessary to complete the computation was formed).

The executable specification of a social computation can include *social properties* that define: "*the drivers for the adoption and spread of the computation*

[1] https://www.mturk.com/mturk/

through the social group with which it engages" [4]. For example, consider a system requiring a worker to provide a photograph of a current event, a social property could be: "to secure a reward, provide a photograph of the event or delegate the task to a trusted friend able to provide one". We argue that it is possible to use the provenance record generated by a social computation to infer worker compliance with those properties (i.e. to check if a worker's behaviour during the computation was consistent with such properties). Provenance information would also permit assessment of a worker's effect on the formation of the group of human participants performing the social computation (e.g. trusted worker Bob delegated the task to his friend Jack, whom he trusted and knew was at the event). Provenance can also be used to infer information about workers motive's (e.g. Bob was motivated to delegate the task in order to receive a reward). In addition, the provenance record can include information enabling the identification of worker's attributes such as their skills (e.g. Jack knew how to take a photograph) and capabilities (e.g. Jack was at the event and had a smartphone). We argue that using provenance to enable the kinds of reasoning highlighted here, would enhance the capabilities of decision-making processes such as trust assessment of workers and workforce selection.

2 Our Approach

We are investigating development of a provenance model for social computation that is aligned with Prov-DM[2], the current W3C provenance recommendation. An analysis of six platforms[3] identified aspects of social computation that a provenance model should describe: the task execution process; links to the social properties applicable for a task; how workers were motivated to participate; what skills and capabilities were associated with a worker when they performed a task; and constraints that were associated with the task description (e.g. requirement for photographs to be stamped by the device with its timestamp and geolocation). Prov-DM does not currently support explicit modelling of these aspects, and therefore one of our goals is to investigate and design a set of appropriate Prov-DM extensions to accommodate them.

Hendler and Berners-Lee [2] have previously argued that the fundamental role of Semantic Web technologies in social computation-like systems is to enable them to easily share data. For example, a process assessing the trustworthiness of Jack, based on the photograph he supplied, might consider Jack more trustworthy if it can determine that the picture was taken at the same time and place as the event. To do so, a system would compare the time and location associated with the photograph with those provided by a description of the event obtained from other data sources on the Web of Linked Data. In addition, such technologies provide a range of reasoning techniques that can be used to support

[2] http://www.w3.org/TR/prov-dm/

[3] These were: Amazon Mechanical Turk, CrowdFlower (http://crowdflower.com/), Zooniverse (http://zooniverse.org/), Passbrains (http://passbrains.com/); oDesk (http://odesk.com/); InnoCentive (http://innocentive.com/)

automated decision-making processes. For example, the fact that Bob delegated a task to Jack and did not provide the photograph, could result in in a naive system excluding Bob from future task assignments. However, Bob might be an important element contributing towards the formation of a group necessary to perform tasks (e.g. delegating trusted friends that provide results).

We argue that enhanced trust assessments of workers could lead to reductions in the number of workers required to perform additional result validation. Such validation steps are typical for current design patterns such as Find-Fix-Verify [1]. Furthermore, better understanding of the process of worker group formation and worker motivations could allow for the selection of smaller groups that perform computations resulting in the same or better results as larger groups. To evaluate our approach, we aim to develop a computational framework that utilises our extended provenance model, supported by semantic technologies. The framework should operate alongside existing platforms using an API to facilitate the capture and use of provenance.

3 Conclusions

In this paper we have argued that introduction of provenance capture mechanisms will not only increase transparency of social computations, but will also permit reasoning about aspects such as trustworthiness of workers and workforce recruitment. We suggest an approach to facilitate the capture and use of such provenance, with the support of semantic technologies and via extensions to Prov-DM. We are aware that there are a number of possible limitations of the proposed approach including scalability issues associated with processing of large provenance records; and difficulties in capturing certain aspects of provenance (e.g. worker's motivation). These remain interesting questions for our future work.

Acknowledgements The research described here is supported by the award made by the RCUK Digital Economy programme to the dot.rural Digital Economy Hub; award reference: EP/G066051/1

References

1. M.S. Bernstein, G. Little, R.C. Miller, B. Hartmann, M.S. Ackerman, D.R. Karger, D. Crowell, and K. Panovich. Soylent: a word processor with a crowd inside. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, pages 313–322. ACM, 2010.
2. J. Hendler and T. Berners-Lee. From the semantic web to social machines: A research challenge for AI on the world wide web. *Artificial Intelligence*, 174(2):156–161, 2009.
3. L. Moreau and P. Missier. Prov-dm: The prov data model. W3C Recommendation (April 2012), http://www.w3.org/TR/prov-dm/, 2012.
4. D. Robertson and F. Giunchiglia. Programming the social computer. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1987), 2013.