CSC2552: Review 6, Paper 1

Due on March 06

Ashton Anderson

499 words

Thomas Hollis

Paper 1

This 2016 paper by Benoit, Conway et. al. is a *crowdsourced human computation study* attempting to investigate the feasibility of a new process of political policy classification using crowdsourced text annotation. The main results suggest that the method not only helps repeatability, cost and speed but also generates results indistinguishable from expert approaches.

The most significant weakness of this paper is the significant time investment required from a research perspective to answer its *research question*. This time cost is a tradeoff with the robust, rigorous and ethical *methodology* chosen. However, using the CrowdFlower platform not only took more time but is also more risky as this platform has been less rigorously audited by academia compared to a more well known platform like MTurk. The authors nonetheless present strong justifications behind this compromise and do detailed quality control to mitigate this issue. Another limitation of this methodology is that any HIT online marketplace is vulnerable to the issue of *weak motivators* [1]. Indeed, since the main motivator of Crowdflower is money, there is no particular incentive for users to truly give their utmost attention to the task at hand. The Crowdflower turker's goal is rather directed toward completing the task as fast as possible while ensuring that they get paid. Fortunately, this limitation does not harm the results thanks to the *aggregation strategy* employed by the paper, which allows a large enough group of non-expert turkers to nonetheless achieve performance akin to expert labelling. An alternative approach could be designed such that the main *motivator* is replaced by fun or community spirit. The open sourcing of a 'politics game' or of a 'political review community' could allow such a labelling process to occur naturally, for free and in a self-sustainable manner. Indeed, such projects have indeed come into existence and been moderately successful. One notable example of such a self-sustaining community annotation for politics is the work done for the 'Global Vote' project since 2015 [2].

In a contrast, a major strength of this paper is the methodology's rigorous and structured design. Using the *split-apply-combine* approach documented in BitByBit [1], the developed strategy demonstrates not only flexibility to specific research needs but also maximal generality through aggregation. This is one of the great advantages of using *computer-assisted human computation* for a crowdsourced task. Another strength is of the paper, further improving generalisation, is the extra experimental step of applying this method to a technical environment and in different languages.

Implications of these results are particularly promising as they offer implementable frameworks for future crowdsourced data annotation, in a wide variety of fields and problem domains, at the quality level of experts. This framework's strong generalisation is also heavily supported both within the study and by other papers in this field [3]. I agree with the main conclusions of the paper as they are presented with cautious and precise language but nonetheless believe that a self-sustaining alternative is more desirable for continued future success.

[1] Salganik, M. J. (2017).Bit by bit: social research in the digital age. Princeton University Press.

[2] The Good Country (2019). The Global Vote. Available at: https://www.globalvote.org/ (Accessed: 28 Feb 2019).

[3] Alonso, O., and R. Baeza-Yates. (2011). Design and Implementation of Relevance Assessments Using Crowdsourcing. In Advances in Information Retrieval, Berlin: Springer.