Argument-Based Policy Consultation Through Crowdsourcing

Joe CRAWFORD, Katie ATKINSON, Trevor BENCH-CAPON
Department of Computer Science, The University of Liverpool, UK

Abstract.
This paper describes an on-going project investigating the use of crowdsourcing in policy consultation. We see this as particularly useful in the early consultation stages (e.g. White Paper) when the opinions of the public are sought to determine policy objectives. The project involves a number of discrete stages: thus far we have looked especially at question design, the generation of suitable test data and the suitability of various aggregation algorithms. On the basis of these results we can design software to collect opinions, and generate arguments. The test data will allow the arguments to be evaluated with respect to variously composed populations.

Keywords. e-participation, crowdsourcing, policy, argumentation

1. Introduction

In this short paper we describe an on-going project designed to support e-participation by using crowdsourcing at an early stage of consultation. The project involves a number of discrete stages: thus far we have considered question design and the generated suitable test data which we have used to determine the suitability of various aggregation algorithms through a number of experiments. On the basis of these results we will design software to collect opinions, and can use the test data to generate arguments for a variety of positions and differently composed populations.

Public consultation enters the policy process at two key stages: the Green Paper stage and the Green Paper stage. For the White Paper, a policy area for action has been selected and the purpose of the consultation is to ask some rather general questions to identify what the public see as the key problems in the area and to gather ideas about how they should be addressed and what any solution should attempt to achieve. On the basis of the responses to the Green Paper, a specific proposal for a policy is formulated and this, with its justification, is presented for consultation as a White Paper. Comments on and critiques of these proposals are received and the policy refined as a result.

In previous work [6] and [7] we have described systems to support e-participation at the White Paper stage, both based on a particular argumentation

scheme for presenting and justifying policy proposals. In [6] the Government proposal is advanced and public critique solicited to identify the key points of agreement and disagreement. In [6] alternative proposals are solicited from the public and then critiqued from the perspective of the Government's policy, to justify why they were not adopted. Both systems have been implemented, as described in [5]. All of this work, however, requires an underlying model of relevant aspects of the domain in the form of an Action-Based Alternating Transition System with Values (AATS+V) as described in [1]. The quality of the arguments and critiques depends crucially on the quality of this underlying model. Although the model is intended to be developed from responses to the White Paper, it is not always straightforward to determine the consensus and to capture it as an AATS+V. See [4] for a discussion.

Standard White Paper consultations usually comprise a general description of the area intended to provide basic information, and also a set of questions intended to solicit opinion and priorities that the policy needs to take into account. Responses usually take the form of various opinions on the questions together with some free text justification of these opinions. Currently the analyst must make sense of the responses, to identify a consensus as to what the current concerns are, and what needs to be done to address them. Our idea is to make use of current ideas regarding crowdsourcing (e.g. [3]) to construct the model. The project requires a number of pieces of work to be performed, and is on-going at the University of Liverpool. Thus far not all the stages have been completed, but we have completed the first few steps, and obtained substantial results on which we can build. In this short paper we provide an overview of the proposed stages, and the results produced thus far.

2. Question Design

Typically questions in a White Paper are open ones inviting a discursive response. To map to an AATS+V we need to present the questions in a form which can receive "yes" or "no" answers (perhaps also "do not know"). Of course, we can still solicit the more discursive answers to inform our policy making, but to build the AATS+V we do need "yes" or "no" answers. These answers will need to provide opinions on the elements of an AATS+V; in particular we need to establish: (1) *relevant current facts*, so that we can form the states of AATS+V and identify the current state; (2) *consequences of relevant actions*, so that we can identify the states that will be reached; (3)*values promoted and demoted* by these transitions.

In our current experiments we have looked at the question of what the policy on cannabis use should be. In recent years some countries and states have legalised or decrimiminalised personal use, but it remains an offence in the UK. Because we need definite questions some elements will need to be decided in advance. We will have to identify a set of relevant facts and the role of the consultation with respect to (1) will be to identify what is generally thought to be currently true with respect to these facts. Thus we will ask questions such as: *Is there excessive use of cannabis among the population?*. Where we wish to ask about mutually exclusive alternatives, we will pose the question using radio buttons, and map

to two questions in post-processing. Note that these are not things which are currently known with certainty: our aim is to obtain a consensus of public *opinion* with respect to them. Note also that we are looking to produce only a fragment of the AATS+V comprising the initial state and those states reachable from the initial state, rather than a complete AATS+V. This means that we can ask about more propositions, since we do not have the exponential complexity associated with including all possible states.

Turning to (2), the actor in which we are interested is the Government, although other actors, in particular the general public, will be relevant in that their response will determine the results of actions. We choose a range of possible Government actions and inquire about their effects: e.g *Suppose cannabis is legalised: will cannabis use increase?*. Again radio buttons may be used to obtain consistent responses to exclusive alternatives. Questions relating to (3) will again be conveniently presented using radio buttons, to be later mapped to several "yes/no" questions. We will pose such questions as: *With respect to civil liberties, will decriminalising cannabis be beneficial, detrimental or neutral?*. Finally we will need to ask what option the respondent favours, so that we can determine preferences, and relate them to the opinions.

3. Aggregating Responses

Having collected the responses, we need to aggregate them to decide on a consensus answer to each of them. Several methods have been proposed for aggregation (see [2]). We have generated test data and conducted a number of experiments to determine a suitable method, and to explore its appropriateness under a range of different conditions. We explored two methods in particular, both taken from [2], a *distance based* algorithm and a *greedy consensus* algorithm.

Standard crowdsourcing concerns what is true, and relies on the idea that although no one will be right on everything, people will, in general, be more right than wrong. Our task makes this assumption entirely reasonable, since our questions do not have a right answer: we are seeking a consensus of opinion, and so, while we would not expect any individual to hold all the "right" opinions, the majority will dictate what should be accepted "right" according to the consensus, and so people will be more often "right" than "wrong".

For our experiments we allocated "right" answers to a set of questions and then generated sets of answers with varying numbers of "right" answers. From this we can compose different populations with different characteristics. For example we can produce a population with an average of 70% "right" answers, with the "wrong" answers different for different individuals. We can then have different distributions: e.g. a uniform distribution in which everyone scores 70%, or populations in which error is normally distributed, with some doing significantly better and some significantly worse than the mean, or bimodal distributions in which one group does better and another group does worse than the mean. We also want to relate the opinions to the preferred option and so we developed three sets of "right"answers, one for each of the possible actions (prohibition, decriminalisation and legalisation), and then produced three sub-populations, each based

on varying one of these three standard answer sets. This enables us to test against populations with clear majorities, and those which are more equally split between options. Key results were:

- The greedy consensus algorithm clearly outperformed the distance based algorithm in identifying the majority position with respect to the questions;
- The above result was robust over all the various distributions (uniform, normal and bimodal)
- Although the greedy consensus identifies the majority position, it does not reflect the *size* of the majority. If the population is more or less evenly split over the three options, the distance based algorithm can be used to produce a compromise, rather than a winner takes all, position.

4. Using the Model

The application of the greedy consensus algorithm will thus allow us to populate the *state*, *action* and *transition* relations of a database representing an AATS+V as in [5]. This will enable us to use either of the two programs: we can both present arguments for a particular position, and critique proposals from the point of view of the consensus model (or the consensus of those advocating a particular option). This will typically produce arguments for (and against) several options.

5. Concluding Remarks

This paper has described an on-going project to provide support based on crowd-sourcing for the model building required to underpin the argumentation based policy consultation tools described in [6] and [7]. Thus far we have looked at question design, generated test data and thoroughly explored the properties of various possible aggregation algorithms. The next phases will be to build a tool to collect the opinions, and to experiment with the various sets of arguments produced from the model for various populations.

References

[1] K. Atkinson, T. Bench-Capon, D. Cartwright, and A. Wyner. Semantic models for policy deliberation. In *Thirteenth ICAIL*, pages 81–90, 2011.

[2] U. Endriss and R. Fernández. Collective annotation of linguistic resources: Basic principles and a formal model. In *51st Annual Meeting of the ACL*, pages 539–549, 2013.

[3] T. Milo. Making collective wisdom wiser. In *24th Database and Expert Systems Applications Conference*, pages 7–8, 2013.

[4] S. Pulfrey-Taylor, E. Henthorn, K. Atkinson, A. Wyner, and T. Bench-Capon. Populating an online consultation tool. In *The 24th JURIX Confernce*, pages 150–154, 2011.

[5] M. Wardeh, A. Wyner, K. Atkinson, and T. Bench-Capon. Argumentation based tools for policy-making. In *Fourteenth ICAIL*, pages 249–250, 2013.

[6] A. Wyner, K. Atkinson, and T. Bench-Capon. Towards a structured online consultation tool. In *3rd International Conference on Electronic Participation*, pages 286–297, 2011.

[7] A. Wyner, K. Atkinson, and T. Bench-Capon. Model based critique of policy proposals. In *4th International Conference on Electronic Participation*, pages 120–131, 2012.