Towards Effective Structure-Based Assessment of Proposals and Arguments in Online Deliberation

Sanja Tanasijevic, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
Klemens Böhm, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

ABSTRACT

Deliberation, i.e., discussing and ranking different proposals and making decisions, is an important issue for many communities, be they political, be they boards of experts for a scientific issue. Online deliberation however has issues, such as unorganized content, off-topic or repetition postings, or aggressive and conflicting behavior of participants. To address these issues, based on a relatively simple argumentation model and on feedback of different type, the authors propose to weight community members in an elaborate manner; this in turn is used to score arguments and proposals. Given such a scoring scheme, it is important to examine to which extent individuals have understood and accepted the approach, to identify characteristics of ‘good’ discussants and of strong arguments and proposals, and to study the robustness of the approach with regard to minor changes. To this end, the authors have carried out an experiment with a real-world community which had to make subjective decisions on issues relevant to them, and they have analyzed the data generated by it systematically, covering the different layers of their approach. The authors’ takeaway is that the approach proposed here is promising to improve deliberation in many settings.

KEYWORDS

Community, Forum, Online Deliberation, Social Data Analysis

1. INTRODUCTION

Deliberation is the act where communities identify possible solutions for a problem and the one(s) from this space that best meet their needs (Walton & Krabbe, 1995; v. Eemeren & Grootendorst, 2003). The spectrum of communities whose discussions rely on reasons and arguments is broad: It not only includes groups of citizens, from (small) municipalities to much larger administrative units. It also ranges from communities in science and technology, including the teams developing software, and communities of online gamers to groups of experts within large companies or organizations. Many communities are small, consisting of about, say, 100 or 200 individuals.

In practice, deliberation faces problems: Major flaws of group discussions are poorly organized content, repetitions, off-topic comments, bad wording and aggressive and conflicting behavior of participants. Some recent projects, e.g., Deliberatorium (Klein, 2011), have tried to apply a very formal argumentation model to bring structure to online discussions and to facilitate content evaluation. However, such rigid formalisms often undermine the natural discussion flow and require a lot of effort from participants. The question we want to investigate here is whether a simple, intuitive
argumentation model, but together with ratings by participants, possibly of different type, allows to identify useful points, arguments and convincing proposals.

Designing such a scoring scheme is not obvious. We for our part propose to weight participants based on the adherence to criteria which correspond to efficient discussion behavior, such as the absence of repetition or off-topic comments, clarity of argumentation etc. However, identifying arguments and deriving conclusions and decisions from a discussion still is difficult. Thus, a question we address is to what extent such a derivation can be based on the structure of the discussion. Further, in the discussions foreseen here, there is no objective truth criterion. Instead, criteria we target at include community satisfaction and consensus of opinions. This makes the assessment of approaches such as the one proposed here more difficult. Finally, the broad variety of communities relying on deliberation will make it necessary to accommodate small changes of the scoring scheme, targeting at specific communities. This means that our approach must be robust to such changes.

We have proposed a relatively simple argumentation model to categorize content and different rating types to assess its quality. The rationale has been to give a clear structure to the discussion and to nudge discussants towards deliberation. In more detail, participants discuss different proposals, each one in a separate thread (mainly by posting arguments in favor or against it). Participants also categorize their comments based on its content; examples of respective comment types are ‘pro argument’ or ‘contra argument’. They can also assess comments by other participants, by giving feedback regarding the argumentation presented, post comments that explicitly express agreement or disagreement etc. The assessment can also refer to the clarity of writing, to the tonality of comments, or to the types of the comments. Based on all this information, our approach assesses potential solutions to discussion subjects which participants have proposed in the course of the discussion. With our approach, collecting ideas for solutions is as important as their evaluation. This is in slight contrast to other recent deliberation projects such as ConsiderIt (Kriplean, Morgan, Freelon, Borning, & Bennett, 2012), which focuses on the collection of pro and contra arguments.

The contributions of this paper are as follows: We motivate and describe our criteria for efficiency of deliberation, e.g., originality of posts, comments focused on the topic etc. Adherence of participants to these criteria results in different participant weights. To incentivize such favorable behavior, an important design decision of ours has been to give participants different degrees of influence on the evaluation of the argumentation, contingent on their weights. We then describe our argumentation model and its expected effects on the discussion structure. Next, we explain the different rating types, as mentioned in the previous paragraph. In contrast to, say, Facebook’s like option or up and down votes on Reddit (Gilbert, 2013) whose aim is to identify popular content, our rating system serves the deliberative nature of the discussion, tries to keep the discussion streamlined without repetitions, off-topic postings, or conflicting behavior and to support the evaluation of comments and proposals. Next, we then present our scoring scheme for evaluating comments and proposals. Comment scores rely on community consensus, on agreements and disagreements received, and on weights of author and raters. Subsequently, proposals are scored based on the scores of comments referring to them.

We then present results gained in a comprehensive analysis of the data gathered in an experimental study with around 200 participants. In a four-week discussion, the participants have generated 954 comments and 3849 ratings. We have found forum participation to be satisfying. Out of 200 participants in total, 164 have posted at least one comment and 175 have posted at least one rating. To illustrate further, we have been able to conclude the following points from the data: Participants have adopted our argumentation model quite easily and without any serious flaws. Few comments were uncategorized, and comment types reported by the authors and other community members have rarely been mismatched. There have been very few off-topic or repetition comments.

We readily admit that we have deviated from the ideal scientific procedure in that we have not carried out several studies with different settings or another study without a scoring scheme, in order to do meaningful comparisons. This mainly is because this current study in isolation, from compiling the issues to be discussed and testing them with a smaller group of individuals to implementing the
Towards an Integrated Personal Software Process and Team Software Process Supporting Tool
www.igi-global.com/chapter/towards-integrated-personal-software-process/61216?camid=4v1a

Model-Based Analysis and Engineering of Automotive Architectures with EAST-ADL
www.igi-global.com/chapter/model-based-analysis-and-engineering-of-automotive-architectures-with-east-adl/116112?camid=4v1a