Likes as Argument Strength for Online Debates

Anthony P. Young[0000−0003−0747−3866]

Department of Informatics, King's College London,
Bush House, Strand Campus, 30 Aldwych, WC2B 4BG
peter.young@kcl.ac.uk

Abstract. We investigate how "likes" in online debates can be incorporated as argument strength when determining the winners of the debate using labelled bipolar argumentation frameworks (BAFs); this builds on previous work where unlabelled BAFs were considered. We ask whether an existing result that "reading by likes" *on average* exposes more winning arguments to the reader than "reading by time" is preserved.

Keywords: Online debates, bipolar argumentation frameworks, Kialo.

1 Introduction

Online debates have grown so large and acrimonious that no one realistically has the time to read everything and hence get a sense of the *actually* winning arguments (*winners*) after all points have been considered. Argumentation theory has been applied to the analysis of online debates at scale, and the literature has addressed problems such as how to formally represent the debate, identify valid arguments and how to efficiently calculate the winning arguments (e.g. [1,2,6]). Many platforms that host these debates have incorporated *comment sorting policies* (CSPs) in their user interface to guide interested readers. Typically, such policies sort and display the arguments in the debate based on a numerical attribute, e.g. from most liked to least liked, or from oldest to newest by time of posting, or from the most replied to the least replied. Previous work has articulated a pipeline that measures the proportion of actual winners a reader is shown given a CSP and the number of arguments read, starting from the opening argument, and claims that for Kialo debates,[1] sorting (and hence reading) from most to least liked (*reading by likes*), *on average*, shows more actually winning arguments than sorting from oldest to newest (*reading by time*), for readers who may not have read the entire debate [8,9]. However, the pipeline only uses the number of likes in each argument as a means of simulating the CSP; the likes attribute has not been incorporated as an obvious form of argument strength when determining the winners. We therefore ask: is the claim of [8] preserved when we include likes as a form of argument strength when determining the winners?

[1] https://www.kialo.com/, last accessed 17/7/2021.

2 Technical Background and Problem Statement

To evaluate how effective a CSP is in displaying winning arguments to a reader who may not have read the entire debate, we have devised the following pipeline shown in Figure 1 [8,9]. Its steps are as follows:

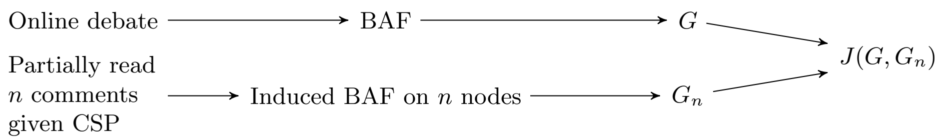

Fig. 1. A schematic of the pipeline from [8,9] used to evaluate comment sorting policies

1. We start with an online debate, which has been mined and cleaned into a bipolar argumentation framework (BAF) [3]. This is a tree whose vertices are arguments - each argument has some text and associated numerical attributes such as likes and time of posting - and whose directed edges are replies that are either attacking or supporting; for arguments a and b, $a \to b$ denotes that a replies to b, and b is prior to a in time. The opening argument is called the *root*, arguments that have not yet received any replies are *leaves*, and every non-root argument replies to exactly one other argument. Assume the tree has $N \in \mathbb{N}^+$ arguments.
2. We calculate the set of winning arguments in the usual way (e.g. [3,4,5]). The set of winning arguments exists and is unique, because a tree BAF gives a directed acyclic argumentation framework (e.g. [7, Corollary 8.13]). We call this set the *actual winners*, denoted G as it is the grounded extension [4].
3. To simulate a CSP, we sort the BAF into a line starting from the root of the tree (the opening argument), and reading "against" the directed edges until we reach a leaf (depth-first search, DFS), such that if there is a choice of which arguments to reach next,[2] we choose the argument with the highest or lowest attribute (e.g. likes or time or number of replies). We give an example:

   Example 1. Consider an online debate of five arguments a_0 to a_4, where a_1 replies to a_0, a_2 and a_4 reply to a_1, and a_3 replies to a_2. This is visualised in Figure 2. Suppose that a_2 is more liked than a_4, but a_4 is earlier than a_2. If we sort from most to least liked following DFS, the sorting order will be a_0, a_1, a_2, a_3, a_4; due to DFS, argument a_3 is read prior to a_4 even if a_3 is less liked than a_4. If we sort from earliest to latest, the sorting order is a_0, a_1, a_4, a_2, a_3. Due to the direction of replies, a_3 is necessarily later than a_2, which in turn is later than a_4.

4. Given this sorting of the BAF tree, we choose an initial segment of $0 \leq n \leq N$ arguments, these arguments induce a sub-BAF, whose winning arguments

[2] i.e. we are at an argument with in-degree > 1.

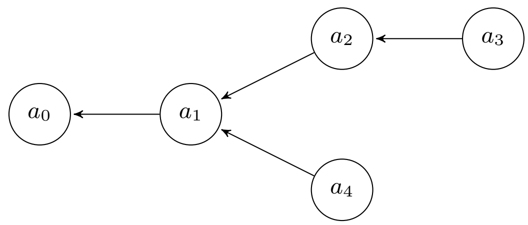

Fig. 2. The online debate from Example 1 to illustrating comment sorting

can also be calculated. Just like in Step 2, this set of winners exists and is unique, and we denote it G_n; we call G_n the *provisional winners*. The intuition is that if our reader cannot read the entire debate, he or she will apply the CSP and sort the comments from most to least liked, and read the top n such comments. Assuming this reader is rational in the sense of argumentation theory, he or she will think that the arguments in G_n have won the debate. Of course, when $n \neq N$, $G_n \neq G$ in general.

5. We then quantitatively compare G and G_n via their *Jaccard coefficient* [8,9]: for finite sets A, B,

$$J(A,B) := \frac{|A \cap B|}{|A \cup B|} \in [0,1] \text{ and } J(\varnothing,\varnothing) := 1. \tag{1}$$

We then calculate an averaged Jaccard coefficient, $\frac{1}{N}\sum_{n=1}^{N} J\left(G, G_n\right)$ for each debate in our dataset, and examine the distribution of such values for the entire dataset of debates for each CSP.

It is claimed [8] that for Kialo debates, reading by likes is *on average* better (i.e. a statistically significantly larger median averaged Jaccard across debates in our dataset) than reading by time.

The pipeline is quite modular and adaptable. For example, in Stage 1, we can incorporate appropriate argument mining or natural language processing techniques depending on how structured the online debate dataset is [8]. However, the likes attribute is only used to sort according to the CSP at present. It is tempting to interpret likes as some form of argument strength - whether that is to do with whether the argument makes a "good point", or that the argument is somehow more plausible, or whether the people who up-vote arguments by increasing their likes want the argument to be accepted regardless of its truth. We therefore motivate two tasks:

1. Conceptually clarify what "likes" means in the dataset of online debates to be analysed. For example, in Kialo, "likes" is a measure of how plausible an argument is, and is calculated by people voting for various arguments after the debate has been completed [8]. This interpretation may not carry over to less clean debates such as those in Reddit.[3]

[3] `https://www.reddit.com/`, last accessed 21/9/2021.

2. Formally incorporate "likes" as a suitable notion of argument strength which will then help determine the actual and provisional winners, respectively in Steps 2 and 4 of the pipeline. The model we have chosen is labelled BAFs [5], as it is recent, general and motivated by social media analytics. Specific considerations include - how should we define the algebra of likes? How can we implement this method of finding winning arguments?

By completing these tasks, we seek to answer the question: does the conclusion of [8] still hold when one incorporates "likes" as argument strength?

3 Conclusion and Next Steps

Online debates can be analysed with argumentation theory, and existing work has claimed that for Kialo debates, reading by likes displays more actual winners than reading by time. However, likes has not been used as argument strength, which may present a different set of winners. We propose to fill this gap by modelling likes using labelled BAFs. We will investigate whether reading by likes is still better than reading by time for Kialo. If this does hold, then we will seek to explore whether this also holds for less clean debates and why this is so.

References

1. Tom Bosc, Elena Cabrio, and Serena Villata. Tweeties Squabbling: Positive and Negative Results in Applying Argument Mining on Social Media. *COMMA*, 2016:21–32, 2016.
2. Elena Cabrio and Serena Villata. A natural language bipolar argumentation approach to support users in online debate interactions. *Argument & Computation*, 4(3):209–230, 2013.
3. Claudette Cayrol and Marie-Christine Lagasquie-Schiex. On the Acceptability of Arguments in Bipolar Argumentation Frameworks. In *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, pages 378–389. Springer, 2005.
4. Phan Minh Dung. On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n-Person Games. *AIJ*, 77(2):321–357, 1995.
5. Melisa Gisselle Escañuela Gonzalez, Maximiliano Celmo David Budán, Gerardo Simari, and Guillermo Ricardo Simari. Labeled Bipolar Argumentation Frameworks. *Journal of Artificial Intelligence Research*, 70:1557–1636, 2021.
6. John Lawrence and Chris Reed. Argument mining: A survey. *Computational Linguistics*, 45(4):765–818, 2020.
7. Anthony P. Young. Notes on Abstract Argumentation Theory. *arXiv preprint arXiv:1806.07709*, 2018.
8. Anthony P. Young, Sagar Joglekar, Gioia Boschi, and Nishanth Sastry. Ranking comment sorting policies in online debates. *Argument & Computation*, 12(2):265–285, 2021.
9. Anthony P. Young, Sagar Joglekar, Kiran Garimella, and Nishanth Sastry. Approximations to truth in online comment networks. In *Proceedings of the Workshop on Argumentation and Society at the 7th International Conference on Computational Models of Argument*, 2018.