ArgVote: Which Party Argues Like Me? Exploring an Argument-Based Voting Advice Application

Markus Brenneis and Martin Mauve

Heinrich-Heine-Universität, Universitätsstraße 1, 40225 Düsseldorf, Germany,
Markus.Brenneis@uni-duesseldorf.de,
https://cs.hhu.de/

Abstract. A lof of people use Voting Advice Applications (VAAs) as a decision-making tool to assist them in deciding which political party to vote for in an election. We think that arguments for/against political positions also play an important role in this decision process, but they are not considered in classical VAAs. Therefore, we introduce a new kind of VAA, *ArgVote*, which considers opinions on arguments when calculating voter–party similarity. We present the results of an empirical study comprising two groups who used ArgVote with and without arguments. Our results indicate that arguments improve the understanding of political issues and different opinions, and that people enjoy the interaction with arguments. On the other hand, the matching algorithm which considers arguments was not better, and user interface improvements are needed. The user profiles we collected are provided to assist further research.

Keywords: Argumentation, Data Set, Voting Advice Applications

1 Introduction

Many people [1, 2] around the world use voting advice applications (VAA) like *Vote Compass* or the German *Wahl-O-Mat*. They inform themselves about positions of different parties concerning current political issues before general elections to receive help in deciding for whom to vote. In many applications, the similarities between voters and parties are calculated with a high-dimensional proximity model [3], based on proximity voting logic [4], where parties are matched with voters based on their opinions concerning a number of political positions.

Classical VAAs, however, do not consider *why* parties and voters maintain certain views. Consider, for instance, Party A being against nuclear power because it thinks nuclear power plants are dangerous, and Party B is against nuclear power because nuclear waste cannot be stored safely. If a voter thinks that nuclear power plants are safe, they are certainly closer to Party B than to Party A. But a classical VAA, which only asks whether the voter is for or against nuclear power, would not capture this information. Therefore, we assume that not only the opinions concerning political positions, but also the arguments used to sustain these positions are relevant for the personal party preference.

The final authenticated version is available online at https://doi.org/10.1007/978-981-16-2765-1_1.

Hence, we have developed *ArgVote*, a new kind of VAA, which does not only consider the political positions, but also the arguments used to arrive at the given position. In an online survey comprising two groups, we tested the acceptance of our new application and whether its new matching algorithm performs better than that of a classical VAA. We also questioned whether people are more informed when arguments are presented, and if they can indicate their own political opinion more easily.

In the next section, we explain why and how we developed an argument-based VAA. Then we present our methods and hypotheses. In Section 4, we show our results and subsequently discuss their consequences. Finally, we have a look at related work and summarize our findings.

2 Designing an Argument-Based VAA

We now sum up the key motivations for developing an argument-based VAA, and then present how our new application ArgVote looks like.

2.1 Limitations of Classical VAAs

As described in the introduction, we think that the reasons why a party has certain attitudes are also important for providing sensible support for a voting decision. If, in our example, the problem with nuclear waste was solved, then Party B would be likely to change its attitude towards nuclear power, as would a voter who was against nuclear power for the same reason. This reinforces our stance that arguments are relevant.

What is more, voters might not be familiar with an issues raised within a VAA, and they tend not to "look up additional information on the web and oftentimes 'just' provide a neutral no opinion answer" [5]. We conjecture that providing arguments for and against a position right within the VAA increases the informedness of voters, who can then better express their opinion and get more meaningful results, i.e. a more suitable voting advice.

Another advantage of arguments is making it harder for parties to "cheat" when the parties provide the answers to the questions in the VAA themselves. Sometimes, parties indicate to be neutral instead of taking an unpopular position to improve their results [2], which leads to inconsistencies between the official stance of a party and its reasons.

2.2 How ArgVote Works

Based on the design of the German VAA Wahl-O-Mat, we have developed ArgVote which additionally displays arguments for and against agreeing with a position (see Figure 1). The arguments can be displayed before the voter indicates their opinion, but ArgVote also explicitly asks the voter to (optionally) choose their arguments after opinion input. If available, (counter)arguments for/again the arguments displayed can be navigated through. The arguments

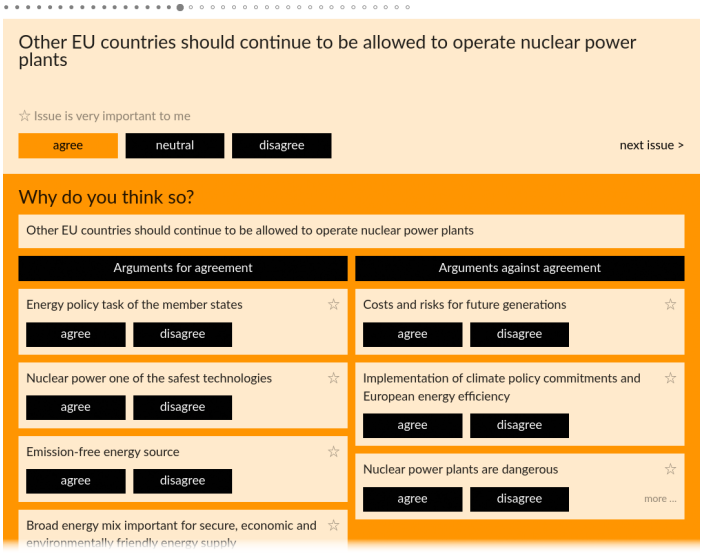

Fig. 1. Main user interface of ArgVote: The user is asked for their arguments after indicating their opinion on an issue, but they can also display the arguments beforehand. "More ..." can be used to see (counter)arguments to arguments.

presented in ArgVote were provided by the parties beforehand. As in the Wahl-O-Mat, political issues, but also arguments, can be marked as important, giving them a higher weight in the matching algorithm. After the last question, a user can compare their arguments with the parties' arguments and sees a bar chart indicating how much they agree with the attitudes of the individual parties.

In the classical matching algorithm used by the Wahl-O-Mat [6], party and voter have a distance 0 for an issue if they have the same opinion, 0.5 if they are different and one is neutral, and 1 otherwise; the value is doubled for issues marked as *important*. ArgVote's matching algorithm is based on our pseudometric for weighted argumentation graphs [7], which also considers the opinions on arguments for/against the positions. *agree*, *neutral* and *disagree* are translated to opinion values 0.5, 0, or -0.5, respectively, in this model. When a position or argument is marked as important, the corresponding edge in the argumentation tree gets a doubled weight.

The relative importance of opinions on arguments and positions can be balanced with a parameter α of the used pseudometric (similar to PageRank's [8] damping factor). ArgVote uses $\alpha = 0.3$, giving the opinions on positions a slighter higher influence than the arguments used. This choice is motivated by the results of an earlier empirical study of ours [9], which indicated that opinions on positions are considered more important by most people than opinions on arguments. From the same study, we also learned that the results of the chosen pseudometric matches human intuition well, and thus, are understandable, in many argumentative contexts.

3 Hypotheses and Methods

With ArgVote, we want to identify differences after using an argument-based VAA and a classical VAA. For our experiment, we recruited German participants from within our personal contacts[1] and let them use ArgVote in two different modes: Group 1 used ArgVote as described above, Group 2 (control group) used ArgVote without arguments displayed under the theses, i.e. it basically behaved like the Wahl-O-Mat. Before the participants used ArgVote, we asked them for their sympathy with the biggest German parties (in alphabetical order: AfD, CDU, Die Linke, FDP, SPD, and Grüne), which were included in ArgVote.

The content for ArgVote was copied from the Wahl-O-Mat of the European Parliament Election 2019, which had been the last election where all Germans were allowed to vote, and comprised 38 positions. We only used the first 15 positions in both groups to reduce the time needed for participation. The complete argumentation corpus contains 294 arguments for all political theses and 147 arguments for the first 15 issues. It was created by three annotators based on the justification statements the parties provided in the Wahl-O-Mat. All annotators independently annotated for each argument whether it is used by a party. The annotator agreement in terms of Krippendorff's alpha [10, p. 211 ff.] is 78%.

In our experiment, we want to research the differences between both groups regarding subjective informedness, ease of indicating an opinion on a thesis, better matching results compared to own party preferences, and usability assessment of ArgVote. After using ArgVote, we asked participants what features of ArgVote they used, how hard they were to use, and how well-informed they feel about policies. Moreover, we count how often user indicate no opinion.

ArgVote also asks different questions about how much participants like their matching results (in overall and concerning the top position) to get a subjective rating of how good the result is. We also checked how close the calculated matching matches a participant's party sympathy rating using the rank-biased overlap (RBO) [11]; RBO compares two sorted lists, where difference in the top-positions are punished more than differences in bottom-positions. We also compare the average rank of a user's party, as also done before for other VAAs by [3].

To wrap up, we have the following hypotheses:

1. Group 1 feels more informed after using ArgVote with arguments than Group 2 (control group without arguments).
2. It is easier for Group 1 to indicate an opinion for a political thesis.
3. Group 1 does not consider ArgVote harder to use.
4. Matching results of Group 1 better match participants' party preferences.

We want to clarify that we mainly focus on checking whether our general idea works well. If it works well, a bigger study can be considered, where improvements on the user interface, the selection and formulation of the arguments, and a more representative sample can be considered.

[1] We first planned to do on-campus recruiting of participants, but this was not possible due to the lockdown at that time.

4 Results

We now present our key findings, starting with the comparison of the experimental groups, and then checking our hypotheses presented in the previous section. The dataset containing the VAA questions and argumentation corpus, as well as the collected user profiles are provided online[2].

4.1 General Information on Participants and Groups

60 participants successfully completed our survey (including two attention check questions). 30 were in Group 1 (with arguments), 30 in Group 2 (control group without arguments). 63% of the participants were male (German population: 49% [12]), the average age was 27 (German population average: 45 [12]), and more than 96% had at least a higher education entrance qualification (Hochschulreife; German average: 34% [12]).

4.2 Hypothesis 1: Informedness

Looking at the subjective answers about informdness, which had been asked after using ArgVote and are presented in Table 1, we could not deduce that Group 1 got a higher awareness of political topics, nor the differences of parties became clearer. But we saw that Group 1 got a clearer picture why there were different opinions, and they understood political issues significantly better.

Table 1. Subjective level of informdness on a Likert scale from *do not agree at all* (1) to *fully agree* (5), p-values according to a Mann–Whitney rank test (MW) [13].

Question	Group 1	Group 2	p (MW)
By using ArgVote I became aware of political issues.	**2.87**	2.60	.20
After using ArgVote, the difference between the parties is clearer to me.	2.70	**2.87**	.77
Using ArgVote helped me understand some political issues better.	**3.40**	2.33	$< .001$
After using ArgVote, it is clearer to me why there are different opinions on certain theses.	**3.30**	2.77	.054

4.3 Hypothesis 2: Ease of Indicating Opinion

There was no big difference between both groups regarding the number of neutral answers and skipped questions. On average, 28% of participants in Group 1 chose a neutral answer, whereas 30% of participants in the control group did so (no significant difference, $p = 1$ with a χ^2 test). The average skip rate (i.e. providing no opinion on an issue) was 1.1% in Group 1, and 1.3% in Group 2 ($p = .58$).

[2] https://github.com/hhucn/argvote-dataset

Table 2. Assessments of task difficulty on a Likert scale from *very hard* (1) to *very easy* (5).

Question	Group 1	Group 2	p (MW)
give my opinion on the theses	**3.67**	3.47	.25
mark a thesis as important	**3.32**	3.24	.36
agree/disagree with arguments	3.78	n/a	n/a
mark an argument as important	3.48	n/a	n/a

On the other hand, the subjects in Group 1 considered indicating an opinion on a thesis slightly easier than those in Group 2 (cf. Table 2). We can also see that (dis)agreeing with arguments was not considered much more difficult than giving an opinion on a thesis, which means that this additional task was not too hard for VAA users. Group 1 also strongly agreed that seeing arguments next to the theses is useful (4.40 on a Likert scale from 1 to 5).

4.4 Hypothesis 3: Ease of Use

As depicted in Table 3, subjects in both groups understood ArgVote, had no problems with navigating, and tended to use the tool again. Group 1 considered the user interface more cluttered and less self-explanatory, which makes sense because of the additional features available. Surprisingly, Group 1 enjoyed using ArgVote more, maybe because it offered a new kind of interaction.

Table 3. Assessments of usability on a Likert scale from *do not agree at all* (1) to *fully agree* (5).

Question	Group 1	Group 2	p (MW)
ArgVote appeared cluttered to me.	2.23	**1.87**	.90
ArgVote was self-explanatory.	3.83	**4.26**	.94
I did not understand how ArgVote works.	1.47	**1.30**	.75
I had no problems navigating ArgVote.	4.16	**4.33**	.69
I would use ArgVote again.	4.20	**4.23**	.63
I enjoyed using ArgVote.	**3.93**	3.60	.052

On the objective side, the time participants stayed in ArgVote and its introduction page was significantly longer in Group 1 (median 17.9 minutes) than in Group 2 (6.08 minutes). This increase was expected because interacting with the arguments needs more time, but it also shows that participants actually did spend time with arguments and did not ignore them.

4.5 Hypothesis 4: Better Matching

We anticipated that taking into account the opinion on arguments (Group 1) yields results better matching individuals' party preferences. In fact, the RBO

(with its parameter $p = 0.7$) in Group 1 (0.67) was worse than the RBO for Group 2 (0.71). Looking at how often the calculated top-1 position matches the party preference, we see something similar (Group 1: 37%, Group 2: 50%). We also got better results for Group 2 when considering the average position at which the user's preferred party is put (Group 1: 1.97, Group 2: 1.60).

Table 4. Assessments of the party matching after using ArgVote on a Likert scale from *do not agree at all* (1) to *fully agree* (5).

Question	Group 1	Group 2	p (MW)
I am confused about the result.	**2.07**	2.17	.35
I am happy with which party is displayed at position 1.	3.80	**3.87**	.62
I can understand which party is displayed at position 1.	**4.23**	4.10	.41
I can understand the displayed percentage of agreement with the party at position 1.	4.00	**4.07**	.75
I consider the overall order of the parties as a whole to be reasonable.	3.90	**3.97**	.63
I consider the percentage of agreement of the parties as a whole to be reasonable.	**3.97**	3.60	.12

The subjective satisfaction with the matching result was basically the same in both groups (cf. Table 4). Participants in Group 1 understood the percentages presented in the matching slightly better.

5 Discussion

Our results give a first hint that incorporating arguments in a VAA makes sense since people tended to be more informed, to give their own opinion more easily and enjoyed the new kind of interaction. But there are some limitations in our current approach, especially when considering using ArgVote for a real election.

We are well aware that our participants were not representative for the German population, which was due to our recruiting process which mainly targeted young students at a university. But our results still give first important hints on whether our approach of incorporating arguments into a VAA is sensible. A bigger study with older and less educated people would be needed, though, to see if they perceive ArgVote as positively as our highly academic, young sample.

The user interface (UI) was considered more cluttered, hence reducing the pieces of information shown at once should be considered, e.g. by pre-filtering the arguments presented in a sensible way. This could, however, lead to the feeling of being manipulated. Related to this, a mobile-friendly UI is not yet available, but is important in a time in which most site views on the Internet come from mobile devices.

From free text comments, we could also learn that the UI regarding the presentation of arguments should be improved, e.g. it was not always clear what "agreeing" with an argument means ("the argument makes sense to me in this

context" vs. "in my view, this statement is correct (but is possibly no good argument)"). Some users also wished to "partially agree" with theses or arguments, as possible e.g. in the VAA ParteiNavi; this could be handled by the underlying pseudometric, but was not possible through our UI.

An important question is where the argument come from. For simplicity, we used the arguments provided in the parties' statements in the Wahl-O-Mat for each issue in our experiment. It can be assumed, though, that a party does not mention every argument it (dis-)agrees with in its statement, which means that the dataset created that way is incomplete. Furthermore, the general party sympathy might also not be in line with the stance on the 15 European topics presented to the participants. Those aspects could also explain why the argumentation-based matching algorithm performed worse when compared to party sympathy, but had a better subjective rating.

A better approach would be asking parties to provide all their arguments and also providing opinions on other parties' arguments, possibly through argumentation platforms like kialo. It has to be considered, though, that participation in such a platform would be hard for small parties with few resources. A related question is whether voters should be able to provide arguments for the corpus, too. Furthermore, some parties mention compromise proposals in their reasons for positions, but those cannot be mapped to arguments, and hence, cannot currently be presented in ArgVote.

Another aspect is whether reaching a good agreement with a user's party preference and user satisfaction are actually the goal of a VAA. The personally preferred party might actually not match the party which would represent one's interests best, but lacking a sensible ground truth, we think that party sympathy is the best approximation we can get.

6 Related Work

We are not aware of other VAAs which incorporate opinions on arguments in their matching algorithms. But there are other kinds of VAAs which also use other approaches than pre-defined distance functions to determine party–voter similarity of classical VAAs, or provide arguments in their interface.

So-called Social VAAs (SVAAs, e.g. *Choose4Greece*) use collaborative filtering, where recommendations are made based on the voting intention of similar users.[14] As shown in [15], the results of SVAAs can be better than those of traditional VAAs. For evaluation purposes, the voting intention given by the users was used, which has the limitations we have already discussed in Section 5. A problem of model-based SVAAs is that their results are not easy to explain[4], whereas understandable results were a design-goal of *ArgVote*, which influenced the choice and design of the underlying pseudometric.

The Learning VAA by [4] took another approach by learning individual distance matrices for each issue instead of using one global, fixed distance function.

Finding political parties closest to a user was also studied in [16]. A user's party could be predicted based on their opinions on ideological positions with an

accuracy of 80%. They used a dataset from debate.org, and applied collaborative filtering to make clustering with sparse information easier. It has to be noted, though, that there have only been two parties (Republicans and Democrats) in that experiment, but we considered six parties.

Some VAAs, like the Greek *Votematch* or *VoteSwiper*, can show additional information on a position, including arguments, but they do not consider argument agreements in their matching. Similarly, the Dutch VAA *Young Voice* provided short videos with pro and contra arguments for each thesis. In the study presented in [17], showing additional information like arguments did increase the number of issues for which an opinion was given, but it did not improve the comprehension of issues. We could not confirm this result in our study.

Our argument annotation process was similar to the method by [18]. They first created a corpus of all possible arguments, and then multiple annotators decided for each text–argument pair if the argument is present in the text.

7 Conclusion and Future Work

We have introduced ArgVote, a new kind of Voting Advice Application which can display arguments next to theses and considers opinions on arguments when calculating the user–party similarity. In an empirical study, we compared ArgVote with and without arguments. We got first hints that the arguments help with forming an opinion on a thesis, understanding different political positions, and make users enjoy the application. The matching results matched the subjective party sympathy worse, though, and our sample was not representative.

The dataset with arguments and user profiles is provided to the community, e.g. for improving the matching algorithm. Using this dataset, the performance of other, possibly more intelligent matching algorithms can be evaluated.

For future work, the user interface of ArgVote should be streamlined to feel less cluttered and reduce the time needed to use the VAA. One possibility would be considering completely different user interactions, e.g. an interactive chat bot, which could reduce the perceived time needed for dealing with the VAA by asking questions on different days. A more representative study should check the influence of the arguments on people who are older and less educated. Another major open question for a real-world application is how the argument corpus should be created.

Acknowledgements We thank the students Marc Feger and Jan Steimann for helping with the annotation of the argumentation dataset, and Henrik Domansky, Lucas Constantin Wurthmann, and Stefan Marschall for providing valuable feedback on the study design.

References

1. Van Camp, K., Lefevere, J., Walgrave, S.: The content and formulation of statements in voting advice applications. Matching Voters With Parties and Candidates. Voting Advice Applications in Comparative Perspective pp. 11–32 (2014)

2. Wagner, M., Ruusuvirta, O.: Matching voters to parties: Voting advice applications and models of party choice. Acta politica **47**(4), 400–422 (2012)
3. Mendez, F.: What's behind a matching algorithm? a critical assessment of how voting advice applications produce voting recommendations. Matching Voters with Parties and Candidates (2014)
4. Romero Moreno, G., Padilla, J., Chueca, E.: Learning VAA: A new method for matching users to parties in voting advice applications. Journal of Elections, Public Opinion and Parties pp. 1–19 (2020)
5. Kamoen, N., Holleman, B.: I don't get it: Response difficulties in answering political attitude statements in voting advice applications. In: 7th International European Survey Research Association Conference, vol. 11, pp. 125–140. European Survey Research Association (2017)
6. Marschall, S.: The online making of citizens: Wahl-O-Mat. The making of citizens in Europe: New perspectives on citizenship education pp. 137–141 (2008)
7. Brenneis, M., Behrendt, M., Harmeling, S., Mauve, M.: How Much Do I Argue Like You? Towards a Metric on Weighted Argumentation Graphs. In: Proceedings of the Third International Workshop on Systems and Algorithms for Formal Argumentation (SAFA 2020), no. 2672 in CEUR Workshop Proceedings, pp. 2–13. Aachen (2020)
8. Page, L., Brin, S., Motwani, R., Winograd, T.: The PageRank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab (1999). Previous number = SIDL-WP-1999-0120
9. Brenneis, M., Mauve, M.: Do I Argue Like Them? A Human Baseline for Comparing Attitudes in Argumentations. In: B. Fazzinga, F. Furfaro, F. Parisi (eds.) Proceedings of the Workshop on Advances In Argumentation In Artificial Intelligence 2020, no. 2777 in CEUR Workshop Proceedings, pp. 1–15. Aachen (2020). URL http://ceur-ws.org/Vol-2777/paper21.pdf
10. Krippendorff, K.: Content Analysis: An Introduction to Its Methodology, 2 edn. Sage Publications (2004)
11. Webber, W., Moffat, A., Zobel, J.: A similarity measure for indefinite rankings. ACM Transactions on Information Systems (TOIS) **28**(4), 1–38 (2010)
12. Statistisches Bundesamt (Destatis): Gesellschaft und Umwelt. URL https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/_inhalt.html
13. Mann, H.B., Whitney, D.R.: On a test of whether one of two random variables is stochastically larger than the other. The annals of mathematical statistics pp. 50–60 (1947)
14. Tsapatsoulis, N., Agathokleous, M., Djouvas, C., Mendez, F.: On the design of social voting recommendation applications. International Journal on Artificial Intelligence Tools **24**(03), 1550,009 (2015)
15. Katakis, I., Tsapatsoulis, N., Mendez, F., Triga, V., Djouvas, C.: Social voting advice applications - definitions, challenges, datasets and evaluation. IEEE Transactions on Cybernetics **44**(7), 1039–1052 (2013)
16. Gottipati, S., Qiu, M., Yang, L., Zhu, F., Jiang, J.: Predicting user's political party using ideological stances. In: International Conference on Social Informatics, pp. 177–191. Springer (2013)
17. Tziraki, C.M.: Towards accurate voting advice: How does Additional Information and Party Identification Influence the users' Comprehension of, and Answers to VAA Statements? Master's thesis, Tilburg University (2020)
18. Boltužić, F., Šnajder, J.: Back up your stance: Recognizing arguments in online discussions. In: Proceedings of the First Workshop on Argumentation Mining, pp. 49–58 (2014)