Value-Sensitive Disagreement Analysis for Online Deliberation

Van Der Meer, Michiel; Vossen, Piek; Jonker, Catholijn M.; Murukannaiah, Pradeep K.

DOI
10.3233/FAIA240231

Publication date
2024

Document Version
Final published version

Published in
HHAI 2024: Hybrid AI Systems for the Social Good

HHAI 2024: Hybrid Human AI Systems for the Social Good
F. Lorig et al. (Eds.)

doi:10.3233/FAIA240231

Value-Sensitive Disagreement Analysis for Online Deliberation

Michiel VAN DER MEER [a,1], Piek VOSSEN [b], Catholijn M. JONKER [c], and Pradeep K. MURUKANNAIAH [c]
[a] *LIACS, Leiden University*
[b] *CLTL, Vrije Universiteit Amsterdam*
[c] *II, TU Delft*

Abstract. Disagreements are common in online societal deliberation and may be crucial for effective collaboration, for instance in helping users understand opposing viewpoints. Although there exist automated methods for recognizing disagreement, a deeper understanding of factors that influence disagreement is currently missing. We investigate a hypothesis that differences in *personal values* influence disagreement in online discussions. Using Large Language Models (LLMs) for estimating both profiles of personal values and disagreement, we conduct a large-scale experiment involving 11.4M user comments. We find that the dissimilarity of value profiles correlates with disagreement only in specific cases, but that incorporating self-reported value profiles changes these results to be more undecided.

Keywords. perspectives, values, natural language processing, hybrid intelligence

1. Introduction

A large number of users participate in online deliberations on societal issues such as climate change [1] and vaccination hesitancy [2]. Disagreement is important to deliberation [3,4], since it helps users understand the opposing viewpoints [5,6]. Measuring disagreements may therefore be useful for measuring deliberation quality [7]. However, recognizing the level of disagreement in a discussion is hard because a deeper understanding of factors that influence disagreement is currently missing [8].

To a degree, automated methods can be used to analyze disagreement in opinions expressed through text. Contemporary approaches are usually based on LLMs, and encompass a variety of tasks like Sentiment Analysis [9], Stance Detection [10], and Argument Mining [11]. However, these methods are shown to disregard diverse opinions [12,13,14], and only paint a partial image of the opinions at play. Analyses of deliberations require a more accurate description of the diverse opinions to be effective [15,16].

We propose to instead use a proxy for disagreement by focusing on *value conflicts*, a potential root cause for disagreement [17]. Preferences among values are used for representing the motivations underlying opinions and actions. In this abstract, we briefly describe how we tested the hypothesis that when users with conflicting values engage in a discussion, diverging views come up [18].

[1] Corresponding author: Michiel van der Meer, m.t.van.der.meer@liacs.leidenuniv.nl.

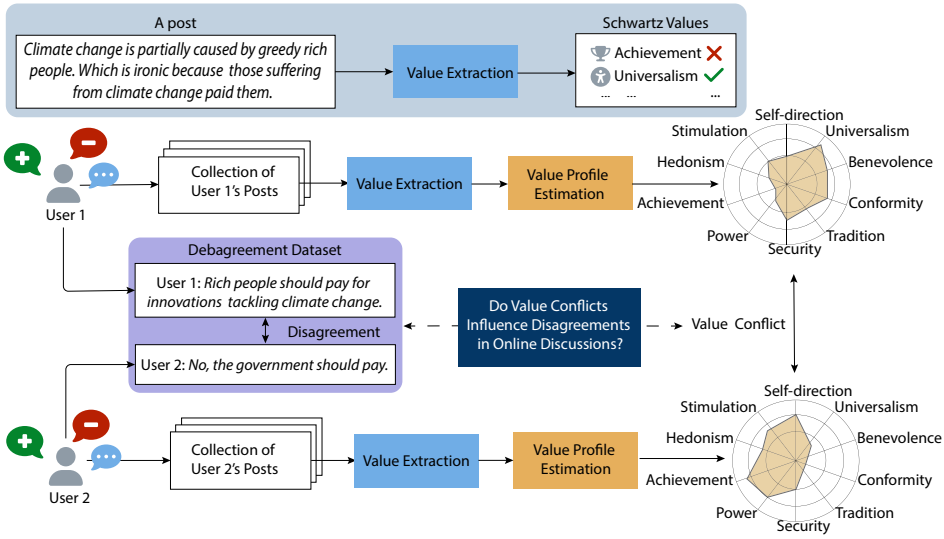

Figure 1.: Analyzing disagreement & value conflicts via Value Profile Estimation (VPE).

2. Value-Sensitive Disagreement Analysis

To evaluate our hypothesis, we estimate individual value profiles based on user comments on Reddit, a social media platform. A value profile captures the relative importance a user ascribes to values from the well-known Schwartz theory of basic values [19]. We compare the similarities among profiles to cases of (dis-)agreement among users. This allows us to investigate the association between value conflict (low similarity) and disagreement. Figure 1 shows a detailed overview of our approach. We train a classifier to perform Value Extraction on individual user comments. We aggregate predictions for a single user and construct the value profile by scoring the relative frequency of value mentions. We test if there is a significant difference among the profiles of user pairs that have been shown to agree with each other versus those that disagree with each other.

We gather 11.4M comments from 19K users on Reddit to construct value profiles. We perform up to 200 tests with different settings to investigate our hypothesis. We further experiment with replacing *estimated* value profiles with *self-reported* ones. To do so, we collect 572 judgments from 26 annotators in combination with self-reported value profiles. We mostly observe negative evidence of a correlation between profile similarity and disagreeing users when using VPE. This corroborates the difficulty of solely automated discussion analysis with LLMs. When including self-reports, the results change for a majority of the cases, alluding to differences in how profiles are obtained. The specific cases in which value conflicts correlate with disagreements are those where values are likely to be relevant and diverse. Therefore, while value conflicts may not be directly related to disagreement, they signal the diversity of the underlying motivations of participants. How context influences the meaning and relevance of values is usually unknown, though Hybrid Intelligence (HI) approaches can be used to inform decision-making [20]. This opens up human-machine collaboration opportunities for a more constructive conversation [21,22,23].

References

[1] Beel J, Xiang T, Soni S, Yang D. Linguistic Characterization of Divisive Topics Online: Case Studies on Contentiousness in Abortion, Climate Change, and Gun Control. Proceedings of the International AAAI Conference on Web and Social Media. 2022 May;16(1):32-42. Available from: https://ojs.aaai.org/index.php/ICWSM/article/view/19270.

[2] Weinzierl MA, Harabagiu SM. From Hesitancy Framings to Vaccine Hesitancy Profiles: A Journey of Stance, Ontological Commitments and Moral Foundations. In: Proceedings of the International AAAI Conference on Web and Social Media. vol. 16; 2022. p. 1087-97. Available from: https://ojs.aaai.org/index.php/ICWSM/article/view/19360/19132.

[3] Polletta F, Gardner B. The Forms of Deliberative Communication. In: The Oxford Handbook of Deliberative Democracy. Oxford University Press; 2018. Available from: https://doi.org/10.1093/oxfordhb/9780198747369.013.45.

[4] Klein M. Enabling large-scale deliberation using attention-mediation metrics. Computer Supported Cooperative Work (CSCW). 2012;21(4-5):449-73. Available from: :https\protect\protect\leavevmode@ifvmode\kern+.2222em\relax//link.springer.com/content/pdf/10.1007/s10606-012-9156-4.pdf:URL.

[5] Lin H, Kim Y. Learning from disagreement on social media: The mediating role of like-minded and cross-cutting discussion and the moderating role of fact-checking. Computers in Human Behavior. 2023;139:107558. Available from: https://www.sciencedirect.com/science/article/pii/S0747563222003788.

[6] Saveski M, Gillani N, Yuan A, Vijayaraghavan P, Roy D. Perspective-taking to reduce affective polarization on social media. In: Proceedings of the International AAAI Conference on Web and Social Media. vol. 16; 2022. p. 885-95. Available from: https://ojs.aaai.org/index.php/ICWSM/article/view/19343/19115.

[7] Esterling KM, Fung A, Lee T. How much disagreement is good for democratic deliberation? Political Communication. 2015;32(4):529-51. Available from: https://www.tandfonline.com/doi/pdf/10.1080/10584609.2014.969466.

[8] Stromer-Galley J, Bryant L, Bimber B. Context and medium matter: Expressing disagreements online and face-to-face in political deliberations. Journal of Deliberative Democracy. 2020;11(1).

[9] Liu B. Sentiment Analysis and Opinion Mining. Springer International Publishing; 2012. Available from: http://dx.doi.org/10.1007/978-3-031-02145-9.

[10] Küçük D, Can F. Stance detection: A survey. ACM Computing Surveys (CSUR). 2020;53(1):1-37.

[11] Lawrence J, Reed C. Argument mining: A survey. Computational Linguistics. 2020;45(4):765-818.

[12] Ng LHX, Carley KM. Is my stance the same as your stance? A cross validation study of stance detection datasets. Information Processing & Management. 2022;59(6):103070.

[13] Cabitza F, Campagner A, Basile V. Toward a Perspectivist Turn in Ground Truthing for Predictive Computing. Proceedings of the AAAI Conference on Artificial Intelligence. 2023 Jun;37(6):6860-8. Available from: https://ojs.aaai.org/index.php/AAAI/article/view/25840.

[14] Van Der Meer M, Vossen P, Jonker C, Murukannaiah P. An Empirical Analysis of Diversity in Argument Summarization. In: Graham Y, Purver M, editors. Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers). St. Julian's, Malta: Association for Computational Linguistics; 2024. p. 2028-45. Available from: https://aclanthology.org/2024.eacl-long.123.

[15] Vecchi EM, Falk N, Jundi I, Lapesa G. Towards Argument Mining for Social Good: A Survey. In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers); 2021. p. 1338-52. Available from: https://aclanthology.org/2021.acl-long.107.pdf.

[16] Shortall R, Itten A, van der Meer M, Murukannaiah P, Jonker C. Reason against the machine? Future directions for mass online deliberation. Frontiers in Political Science. 2022. Available from: https://www.frontiersin.org/articles/10.3389/fpos.2022.946589/full.

[17] Stromer-Galley J, Muhlberger P. Agreement and Disagreement in Group Deliberation: Effects on Deliberation Satisfaction, Future Engagement, and Decision Legitimacy. Political Communication. 2009;26(2):173-92. Available from: https://doi.org/10.1080/10584600902850775.

[18] van der Meer M, Vossen P, Jonker C, Murukannaiah P. Do Differences in Values Influence Disagreements in Online Discussions? In: Bouamor H, Pino J, Bali K, editors. Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. Singapore: Association for Com-

putational Linguistics; 2023. p. 15986-6008. Available from: https://aclanthology.org/2023.emnlp-main.992.

[19] Schwartz SH. An overview of the Schwartz theory of basic values. Online readings in Psychology and Culture. 2012;2(1):2307-0919. Available from: :https\protect\protect\leavevmode@ifvmode\kern+.2222em\relax//core.ac.uk/download/pdf/10687025.pdf:URL.

[20] Liscio E, Lera-Leri R, Bistaffa F, Dobbe RI, Jonker CM, Lopez-Sanchez M, et al. Inferring Values via Hybrid Intelligence. In: HHAI 2023: Augmenting Human Intellect. IOS Press; 2023. p. 373-8.

[21] Hadfi R, Ito T. Augmented Democratic Deliberation: Can Conversational Agents Boost Deliberation in Social Media? In: Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems; 2022. p. 1794-8. Available from: https://ifaamas.org/Proceedings/aamas2022/pdfs/p1794.pdf.

[22] van der Meer M, Liscio E, Jonker CM, Plaat A, Vossen P, Murukannaiah PK. HyEnA: A Hybrid Method for Extracting Arguments from Opinions. In: Proceedings of the first International Conference on Hybrid Human-Artificial Intelligence (HHAI 2022). Amsterdam, the Netherlands: IOS Press; 2022. p. 1-15. Available from: https://liacs.leidenuniv.nl/~meermtvander/publications/hyena/.

[23] Alshomary M, El Baff R, Gurcke T, Wachsmuth H. The Moral Debater: A Study on the Computational Generation of Morally Framed Arguments. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Dublin, Ireland: Association for Computational Linguistics; 2022. p. 8782-97. Available from: https://aclanthology.org/2022.acl-long.601.