Analysing Pathos in User-Generated Argumentative Text

Natalia Evgrafova, Véronique Hoste, Els Lefever
LT3, Ghent University, Belgium
Groot-Brittanniëlaan 45, 9000 Ghent
{natalia.evgrafova, veronique.hoste, els.lefever}@ugent.be

Abstract

While persuasion has been extensively examined in the context of politicians' speeches, there exists a notable gap in the understanding of the pathos role in user-generated argumentation. This paper presents an exploratory study into the pathos dimension of user-generated arguments and formulates ideas on how pathos could be incorporated in argument mining. Using existing sentiment and emotion detection tools, this research aims to obtain insights into the role of emotion in argumentative public discussion on controversial topics, explores the connection between sentiment and stance, and detects frequent emotion-related words for a given topic.

Keywords: argument mining, sentiment analysis, emotion detection, social media, political discourse

1. Pathos in Political Argument Mining

An essential aspect of political activity is persuasive communication (Windisch, 2008). According to Wolton (1989), political communication is a platform where politicians, journalists, and the public (through opinion polls) openly express their views on politics. Windisch (2008) rightly argues that public opinion today is expressed through a variety of communication channels available. This trend has led to the development of both automated and non-automated methods for public opinion collection and analysis. With the advances in big data, tools such as sentiment analysis, opinion and argument mining have been developed to understand and predict public attitudes towards various entities, from products or films to governmental initiatives and salient social problems.

Persuasive communication in politics implies effective argumentation, achieved within logos, pathos, and ethos dimensions (Cardoso et al., 2023). The logos dimension is associated with the logical structure of arguments, the pathos dimension is related to appeals to emotion, and ethos is concerned with credibility and appeals to authority (Cardoso et al., 2023; Habernal and Gurevych, 2017).

In natural language processing, it is the field of argument mining that aims to automatically extract, analyse and understand arguments from natural language text. Within argument mining, researchers have been developing methods to classify argumentative and non-argumentative spans of text, detect topics, aspects, stances, and other argument components, and generate high-quality arguments (Cabrio and Villata, 2018). While their work has predominantly focused on the logos dimension, with a particular success for argumentative essays and other well-structured text, analysing the pathos dimension has not been as thorough despite the fact that it plays an important role in the social media argumentative discourse, especially political discussions. Such discussions often include emotional and metaphoric language that cannot be properly analysed within the logos dimension (Habernal and Gurevych, 2017). Moreover, attempts to analyse these arguments are associated with challenges related to overlaps of sentiment analysis and argument mining (Cabrio and Villata, 2018), and low inter-annotator agreement for emotional components of arguments (Habernal and Gurevych, 2017).

To the best of our knowledge, the exploration of the pathos dimension of argumentation, including the appeal to emotion, have been primarily dealt with within fallacy detection (Goffredo et al., 2023; Vijayaraghavan and Vosoughi, 2022; Sahai et al., 2021). In argumentation, logical reasoning is considered to be more legitimate than emotional language (Duckett, 2020). However, an appeal to emotion does not necessarily mean fallacious argumentation (Walton, 2005; Duckett, 2020), and its usage could be justified when it comes to value-contested debates such as assisted dying, abortion, war, independence (Duckett, 2020).

In Walton (1992), we read that certain types of emotional appeals "are very powerful as arguments in themselves", though there is always a chance they could be fallacious, namely, irrelevant or logically weak, but that is not always the case and, more importantly, does not always limit their effect. It is often through emotional language that users express their beliefs, values, and moral motivations. This is why we argue that confining argument mining solely to the logos dimension and reserving the

Proceedings of the Second Workshop on Natural Language Processing for Political Sciences (PoliticalNLP 2024), pages 39–44
21 May, 2024. © 2024 ELRA Language Resource Association: CC BY-NC 4.0

pathos for fallacy detection may prove overly restrictive. Simply marking argumentation as fallacious might not substantially improve our understanding of the prevalent reasons for taking one stance or the other.

An effort to include the pathos dimension in annotations was made by Habernal and Gurevych (2017) in their study on argument mining in user-generated web discourse. In this study, the corpus included documents that were retrieved from heterogeneous web resources (comments on articles, forum and blog posts). 6% of documents were purely emotional without logical backing and could not be analysed in terms of their logical structure. Though in some cases claims and premises could be identified, persuasiveness was achieved through emotion. Following the given annotation guidelines, annotators had to classify arguments as "appeal to emotion" in case the argumentation relied on figurative language or obvious exaggerations. The task posed a significant difficulty reflected in Krippendorf's agreement of only $\alpha U = 0.30$. Consequently, the authors chose to focus on the logos dimension. The Internet Argument Corpus (IAC) (Walker et al., 2012) included annotation for Fact/Emotion-based arguments with relatively low agreement results $\alpha U = 0.32$. The low inter-annotator agreement proves that new approaches should be developed for incorporating the pathos dimension into argument mining. Logical and emotional components in arguments are intertwined (van Eemeren and van Haaften, 2023), expressed in different degrees and supplement each other for the purpose of persuasiveness, making emotion an integral part of public reasoning (Stucki and Sager, 2018) that should not be disregarded.

In this work we focused on an exploratory pathos analysis of argumentative text for the task of argument mining. The contributions of the paper are the following: (1) analysing the relationship between sentiment and stance, (2) comparing the results of a manual analysis of the emotional components of arguments on a given topic with an automated extraction of emotion-related words, and (3) suggesting ways to incorporate the pathos dimension for the argument mining task.

2. The Datasets

For the analysis of sentiment, emotional words, and the connection between sentiment and stance, we selected two datasets consisting of user-generated arguments on contentious topics with stance annotations. We prioritised stance annotations over sentiment annotations as stance is more difficult to detect automatically, but we could automatically annotate the argument sentiment with sufficient accuracy. We chose the Webis *args.me* corpus (Ajjour et al., 2019) containing 387,606 arguments from debate portals and the IBM ArgKP-2021 dataset (Friedman et al., 2021) of crowdsourced arguments — based on the ArgKP dataset (Bar-Haim et al., 2020) and Gretz et al. (2019). As some of the topics in the Webis corpus contained very few comments, we decided on the 30 most commented topics from the corpus and deleted very short comments (up to 10 words) such as "I win" that were part of users' communication on the platform, which resulted in 8902 comments. From the IBM ArgKP-2021 dataset we kept 7238 unique arguments on 31 topics.

3. Corpora Analysis

3.1. Sentiment and Stance

The relationship between sentiment and stance is complex. One of the hypotheses could be that the sentiment is more positive in PRO stances and more negative in CON stances. This could be true for certain datasets and explain why BERT-based models tend to rely on sentiment words for stance prediction (Trautmann, 2020). However, the same arguments might be attacking a certain topic or aspect and support the other, regardless of their sentiment; for example, an argument attacking coal energy might be supporting wind energy (Daxenberger et al., 2020). Understanding how sentiment is connected with stance might not only provide insights into the pathos dimension, but also help design corpora in such a way as to minimise errors in machine-learning.

The first task that we addressed was the analysis of the sentiment and stance distribution in the chosen corpora. To compensate for the lack of sentiment annotations, the arguments in both corpora were automatically annotated for positive, neutral, or negative overall sentiment using two existing transformer-based models for sentiment analysis with reasonable recall scores. When selecting a model, our aim was to ensure that it was trained and fine-tuned on internet user-generated texts. We prioritised models based on tweets due to their closer resemblance to our dataset: tweets are user-generated, vary in length, and incorporate colloquial and emotional language. The first model we used was the recent version of the fine-tuned twitter-roberta-base-sentiment-latest (Camacho-Collados et al., 2022). The second model we applied was the fined-tuned pysentimiento bertweet-base-sentiment-analysis model (Pérez et al., 2023). The comparison of the resulting sentiment labels from the two models showed an overlap of about 78% in both corpora, which meant that the majority of arguments were correctly labelled in terms of their sentiment. The further exploration was based on the labels from the twitter-roberta-base-sentiment-latest as it allowed for longer texts (512 tokens compared to 128 in pysentimiento).

The preliminary analysis of the Webis dataset revealed a larger proportion of neutral and positive texts in PRO arguments (61,3%) compared to CON (50,0%).

In the IBM corpus, the proportion of negative arguments for PRO was higher (55.1% compared to 47.2% in CON), as was the proportion of positive arguments (14.4% compared to 11.5% in CON). Conversely, there were more arguments with a neutral sentiment in CON (41.3% compared to 30.5% in PRO), see Fig. 1. A qualitative analysis showed that for this corpus, the larger proportion of negative sentiment in PRO could be explained by the fact that many topics, which are major claims in this corpus, are formulated with a negative framing, for instance, “Assisted suicide should be criminal offence”, “We should ban human cloning”, “Home schooling should be banned”. These claims imply a negative stance towards the main topic (assisted suicide, home schooling, human cloning) making arguments that are against main topics actually belonging to the PRO category if the topic is positively framed.

To check if swapping the stance labelling for such topics results in major changes in the sentiment distribution, we studied 31 topics of the IBM dataset and manually changed the stances for 18 topics that were framed negatively.

The results for the modified IBM corpus (see Fig. 2) revealed a much bigger proportion of positive and neutral sentiment in PRO and a substantial increase in the negative sentiment in CON. This indicated that topic framing influenced the distribution of sentiment across stances.

In the 30 topics of the Webis dataset there were only two negatively framed topics “Abortion should be illegal” and "Gun Control" with 323 comments, which could not result in much change. Among other topics, there were "Gun rights", "Abortion", "Gay marriage", "Euthanasia" that implied a positive claim even though some of them were expressed in one word only, for example, "Abortion" could be extended to "Abortion should be allowed" without causing changes in the stances of the arguments. However, it should be noted that arguments in this corpus included quotations of the opposing position, and there were some non-argumentative user-interaction comments, which could also influence the sentiment distribution.

To conclude this section, from certain datasets a model can learn to rely too much on sentiment for stance prediction, but datasets can be modified to reduce errors. One of the ways to decrease the impact of sentiment on stance is to check the sentiment and stance distribution in the training corpus and ensure the balance. Another way would be to conduct training that involves positively and negatively framed similar topics (ex: "Abortion should be allowed", "Abortion should be banned", or "Abortion rights", "Abortion ban") for the same arguments, which could yield more robust results. On the whole, sentiment is intricately connected with stance and is highly influenced by a topic and its framing, certain checks and dataset modifications could be used to lower the chances of short-cut machine learning.

3.2. Emotion Words in Arguments

To get the first insights into the emotional dimension of arguments, we chose the "Death penalty" ("Capital punishment") topic to analyse in both corpora. First, we conducted manual analysis on a sample of arguments to identify emotional components. Next, we automatically annotated arguments with the NRClex 4.0 affect generator[1] based on the NLTK library’s WordNet synonym sets (Bird et al., 2009) and the NRC lexicon (Mohammad and Turney, 2013). The final labels for emotion included emotion-related words from each argument and a list of emotions associated with them. This enabled comparison with the results of the manual analysis.

The qualitative analysis of 171 arguments (90 PRO, 81 CON) for the topic of "death penalty" in the Webis corpus showed that in PRO arguments the prevalent emotional language was used to describe criminals that were “dangerous”, “heartless”, “cold-blooded”, often mentioning paedophiles, that “deserved” this punishment for the “heinous”, “violent” criminal act and “awful", "horrible" things they committed. This was contrasted with “innocent” victims “condemned to a terrible life” who “deserve” “true”, “proportional” justice. The Old Testament quotations and especially the "eye for eye" principle were referred to in the context of the punishment being "justified" and serving as an "efficient deterrent" that "inspires" "fear" into criminals. The concept of capital punishment was described as a means of "protection" of the "innocent" that brings "peace", "solace" and "closure", and "saves" other people.

The CON arguments described capital punishment as "a murder", "cruel", "outdated", "barbaric", "racist", "sexist, "unnecessary", "expensive", "hypocritical", they included appeals to "forgiveness", a chance for criminals to "repent" and used "innocent" to refer to the wrongly executed people.

In the CON arguments, there were also frequent mentions of "better", "more efficient", "other" ways to punish criminals, as well as references to "equality", "human rights",and questioning if death penalty is a "good" deterrent, describing life in prison as a "greater punishment".

[1](C) 2019 Mark M. Bailey, PhD

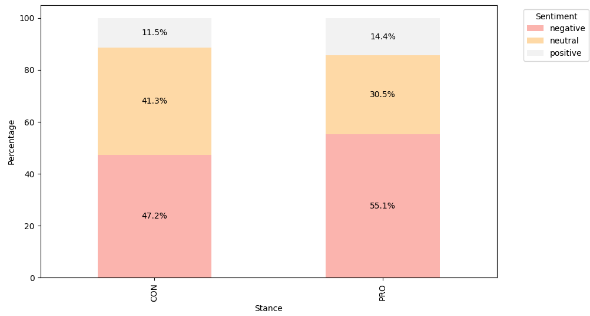

Figure 1: IBM sentiment distribution by stance

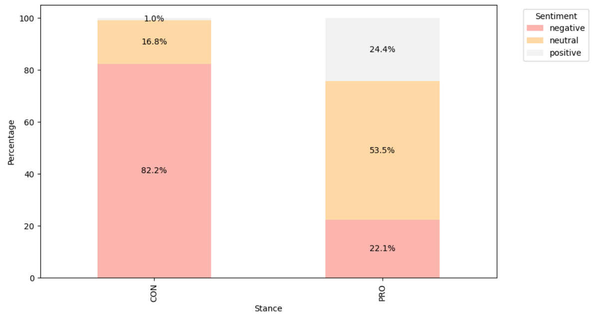

Figure 2: inverted IBM sentiment distribution by stance

As observed by Walton (2005), some argumentative discussions include an argument about how to define key concepts. This was also seen in our data, e.g. "Death penalty is a murder", as opposed to "Death penalty is a deterrent". Such *definitions* are highly important in ethically controversial debates, and they tend to differ in terms of the emotional spin for opposing parties (Walton, 2005).

The next step was to compare these results with what a simple emotional words detection could yield depending on arguments' stance. For this purpose, we relied on the NRClex 4.0 affect generator to extract only emotion-related words from all the comments. Subsequently, we segregated these words based on the stance of the arguments within the selected topic, creating unique sets for both 'PRO' and 'CON' stances. During this process, words exclusive to each stance were identified, with any overlapping words removed. After that, we counted the frequencies of these unique words and used these counts to generate word clouds that represented the most common emotion-related words for each stance.

The word clouds for the 171 arguments from the Webis corpus that had been previously manually analysed showed the prevalence of "protecting" people, the risk of prisoners' "escape" and references to "brutal", "horrible" things for PRO and a higher frequency of religious references to "hell", "repent", and "spirit" in the CON category (see Fig. 3).

For comparison, the world clouds for the "capital punishment" topic from the IBM corpus featured "violence" and "ineffective" as the most frequent emotional words for CON and the high frequency of "heinous" and "deserve" for PRO (see Fig. 4).

Overall, this exploration provided general insights into emotional language associated with stances. Nevertheless, some results were not easy to interpret without knowing what they referred to or a deeper knowledge of the context.

Based on these observations, we suggest considering the following in order to detect pathos in argumentative text: (1) the emotional components of arguments are defining particular aspects ("heinous crime", "innocent victim"). A more fine-grained comparison of emotional words by aspect could bring about more insightful results. Aspects can also be emotionally loaded and expressed by a variety of lexical means (e.g. "Solace"/"Peace"/"Closure" for crime victims); (2) apart from retrieving emotional words by aspect, pathos could be further explored through extraction of key concepts persuasive definitions (e.g. "Death penalty is a murder") as they often contain emotional words that differ for PRO and CON stances (Walton, 2005); (3) for comparison of PRO and CON, the topic should be clearly defined and be controversial, its framing, negative or positive, should be taken into account.

Figure 3: "Death penalty", 171 comments, Webis

Figure 4: "Capital punishment", 236 comments, IBM

4. Conclusion and Future Work

This paper explored the possibility of incorporating the pathos dimension of argumentation for the task of argument mining. The IBM ArgKP-2021 and the Webis args.me corpora were used for the analysis of the relation between sentiment and stance, and emotional words detection. A part of the Webis corpus was manually analysed to compare the results and develop ideas for automation of the pathos dimension analysis.

The automatic detection of emotional words based on a lexicon-based approach provided the first insights into the pathos dimension of arguments based on their stance, however, certain results were difficult to interpret without contextual information and understanding which aspects those words referred to.

A first avenue for future work is the extraction of definitions and emotional words associated with specific aspects in the argumentative text across diverse topics. Another important direction consists in developing effective annotation guidelines and creating pathos-annotated corpora of the argumentative texts. Finally, we aim to explore various pipelines to automate stance-dependent pathos analysis in argumentative texts on contentious topics.

Gaining deeper insights into the pathos dimension of arguments in political social media text can shed light on the role of emotion across controversial topics and in forming public opinions. Further developments could deepen our comprehension of what convinces the public, which stories and interpretations get spread in different languages, and how the public responds to these stories, what is reproduced and challenged.

Acknowledgements

This work was supported by the Research Foundation – Flanders (FWO) under grant FWO.OPR.2023.0004.01 (G019823N).

5. Bibliographical References

Roy Bar-Haim, Lilach Eden, Roni Friedman, Yoav Kantor, Dan Lahav, and Noam Slonim. 2020. From arguments to key points: Towards automatic argument summarization. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4029–4039. Association for Computational Linguistics.

Elena Cabrio and Serena Villata. 2018. Five years of argument mining: a data-driven analysis. In *27th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 5427–5433. International Joint Conferences on Artificial Intelligence Organization.

Jose Camacho-Collados, Kiamehr Rezaee, Talayeh Riahi, Asahi Ushio, Daniel Loureiro, Dimosthenis Antypas, Joanne Boisson, Luis Espinosa Anke, Fangyu Liu, and Eugenio Martínez Cámara. 2022. TweetNLP: Cutting-edge natural language processing for social media. pages 38–49. Association for Computational Linguistics.

Henrique Lopes Cardoso, Rui Sousa-Silva, Paula Carvalho, and Bruno Martins. 2023. Argumentation models and their use in corpus annotation: Practice, prospects, and challenges. *Natural Language Engineering*, 29(4):1150–1187.

Johannes Daxenberger, Benjamin Schiller, Chris Stahlhut, Erik Kaiser, and Iryna Gurevych. 2020. ArgumenText: Argument classification and clustering in a generalized search scenario. *Datenbank-Spektrum*, 20(2):115–121.

Stephen Duckett. 2020. Pathos, death talk and palliative care in the assisted dying debate in Victoria, Australia. *Mortality*, 25(2):151–166.

Pierpaolo Goffredo, Mariana Chaves, Serena Villata, and Elena Cabrio. 2023. Argument-based detection and classification of fallacies in political debates. pages 11101–11112. Association for Computational Linguistics.

Shai Gretz, Roni Friedman, Edo Cohen-Karlik, Assaf Toledo, Dan Lahav, Ranit Aharonov, and Noam Slonim. 2019. A large-scale dataset for argument quality ranking: Construction and analysis.

Ivan Habernal and Iryna Gurevych. 2017. Argumentation mining in user-generated web discourse. *Computational Linguistics*, 43(1):125–179.

Juan Manuel Pérez, Mariela Rajngewerc, Juan Carlos Giudici, Damián A. Furman, Franco Luque, Laura Alonso Alemany, and María Vanina Martínez. 2023. pysentimiento: A Python toolkit for opinion mining and social NLP tasks.

Saumya Sahai, Oana Balalau, and Roxana Horincar. 2021. Breaking down the invisible wall of informal fallacies in online discussions. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, pages 644–657. Association for Computational Linguistics.

Iris Stucki and Fritz Sager. 2018. Aristotelian framing: logos, ethos, pathos and the use of evidence in policy frames. *Policy Sciences*, 51(3):373–385.

Dietrich Trautmann. 2020. Aspect-based argument mining. In *Proceedings of the 7th Workshop on Argument Mining*, pages 41–52. Association for Computational Linguistics.

Frans H. van Eemeren and Ton van Haaften. 2023. The making of argumentation theory: A pragma-dialectical view. *Argumentation 2023*, 37(3):341–376.

Prashanth Vijayaraghavan and Soroush Vosoughi. 2022. TWEETSPIN: Fine-grained propaganda detection in social media using multi-view representations. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3433–3448. Association for Computational Linguistics.

Marilyn Walker, Jean Fox Tree, Pranav Anand, Rob Abbott, and Joseph King. 2012. A corpus for research on deliberation and debate. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 812–817. European Language Resources Association.

Douglas Walton. 1992. *The Place of Emotion in Argument*. Penn State University Press.

Douglas Walton. 2005. *Fundamentals of Critical Argumentation*. Cambridge University Press.

Uli Windisch. 2008. Daily political communication and argumentation in direct democracy: advocates and opponents of nuclear energy. *Discourse Society*, 19(1):85–98.

Dominique Wolton. 1989. La communication politique: construction d'un modele. *Hermès, La Revue*, 4(1):27–42.

6. Language Resource References

Yamen Ajjour and Henning Wachsmuth and Johannes Kiesel and Martin Potthast and Matthias Hagen and Benno Stein. 2019. *Data Acquisition for Argument Search: The args.me corpus*. Springer. PID https://zenodo.org/records/3274636.

Bird, Steven and Klein, Ewan and Loper, Edward. 2009. O'Reilly. PID https://www.nltk.org/nltk_data/.

Friedman, Roni and Dankin, Lena and Hou, Yufang and Aharonov, Ranit and Katz, Yoav and Slonim, Noam. 2021. *Overview of the 2021 Key Point Analysis Shared Task*. Association for Computational Linguistics. PID https://paperswithcode.com/dataset/argkp-2021.

Mohammad, Saif M. and Turney, Peter D. 2013. *Crowdsourcing a Word–Emotion Association Lexicon*. PID http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm.