Computational Assessment of Pathotic Argumentation Quality

Dexter Williams

University of Wisconsin-Madison

Abstract

What is a good argument? What is a good *pathotic* (emotional) argument? In studies on the computational assessment of argument quality, this second question has been subsumed by the first. Pathos has been treated as just a dimension of argument quality that should be considered when computationally assessing overall quality. In my research, I aim to provide an individualized treatment of pathotic argument quality by proposing a taxonomy for its study and outlining models for its assessment. The final goal of my PhD research is to introduce interactional pathotic argument quality assessment into dialogue systems. I am motivated by theoretical works that raise the status of pathos from its current treatment in studies of argument quality.

1 Introduction

The foundational work on the computational assessment of argument quality analyzed argument quality as a taxonomy (hereon referred to as the AQ taxonomy), with rhetorical, logical, and dialectical dimensions [Wachsmuth et al., 2017]. The AQ taxonomy unifies previously separate approaches to assessing argument quality, providing a common ground for future research.

The arousal of emotions as a persuasive technique (*pathos*) that can improve the strength and quality of an argument has been acknowledged since the works of pre-Socratic philosophers [Kennedy, 1994]. The AQ taxonomy incorporates pathos in 2 of its 13 dimensions: appropriateness and emotional appeal. Assessing appropriateness involves identifying whether the language of an argument has an acceptable style and (vaguely) the right amount of emotional content. Assessing emotional appeal involves identifying cases where the emotional appeal present in an argument makes a hearer more open to the speaker's claims.

From reviewing the literature, I have not observed any studies that solely examine pathotic argument quality assessment. However, pathotic argument quality is partially present in discussions on argument quality in general (see Figure 1). Some have moved away from the AQ taxonomy; [Vecchi et al., 2021] also considers the role of deliberative quality, and analyzes argument quality using a taxonomy of deliberative quality, although they acknowledge that the emotional dimensions of the deliberative taxonomy are equivalent to those in the AQ taxonomy. [Ziegenbein et al., 2023] create a fine-grained taxonomy just for the appropriateness sub-dimension of the AQ taxonomy, where inappropriateness can be due to toxic emotions, namely excessive intensity of emotions, or emotional deception. However, the taxonomy proposed by [Ziegenbein et al., 2023] contains mostly non-emotional dimensions, and is only concerned with how argument quality can be minimized by the use of pathos and not maximized [Wachsmuth et al., 2024].

Why should pathotic argument quality be examined separately from other aspects of argument quality, when presumably it is captured by more general taxonomies of argument quality? In studies on the computational assessment of argument quality, pathos has been treated as something that affects ar-
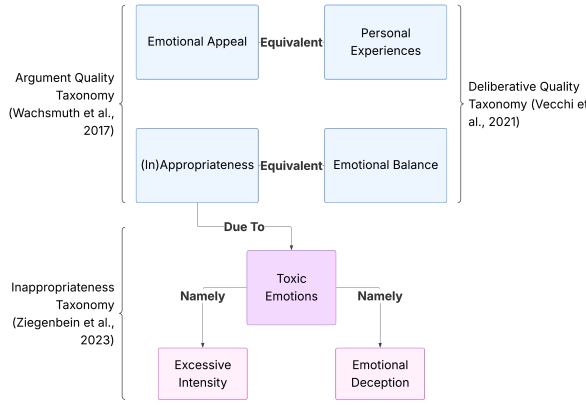
Figure 1: Links between the pathotic dimensions of previous argument quality taxonomies.

Figure 2: My proposed taxonomy of pathotic argument quality.

gumentation in general, but never as the object of argumentation itself. To elaborate, computational assessment of pathotic argument quality can be defined in one of two ways:

Weak Assessment: Assessing pathotic argument quality involves assessing the *impact of pathos* on the quality of an argument.

Strong Assessment: Assessing pathotic argument quality involves assessing the *quality of a pathotic argument qua pathotic argument.*

Studies in the computational assessment of argument quality currently adopt the weak assessment view. Strong assessment can be traced to several theoretical works that raise the cognitive status of emotions in arguments [Braet, 1992, Brinton, 1988, Micheli, 2010]. This raising of the cognitive status of emotions allows for a different kind of argumentation that is uniquely emotional. If the strong assessment view is adopted, it would be useful to have a taxonomy that aids in the computational assessment of this second, uniquely emotional kind of argumentation.
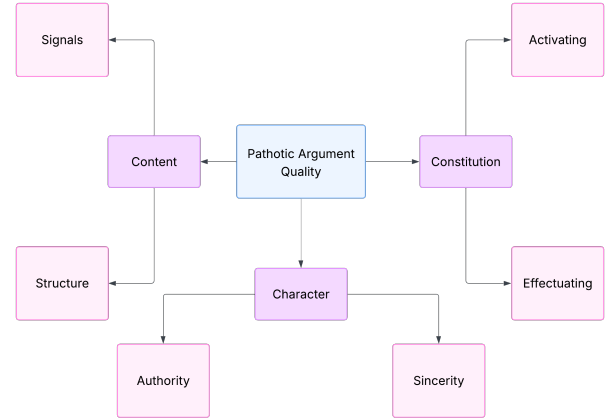
2 Taxonomy of Pathotic Argumentation Quality

I will briefly summarize my initial proposed taxonomy of pathotic argument quality from [Williams, 2025], shown in Figure 2:

Pathotic Content: Any content-related factors that affect the pathotic quality of an argument. I subdivide these into two broad kinds: *signals* and *structure.* The former relates to certain words, phrases, and non-verbal or para-verbal cues that induce or elicit emotion, such as gestures. The latter relates to how the argument structure (e.g., its scheme) can be said to elicit emotion.

Pathotic Constitution: Captures how the state of mind (mental constitution) of the hearer can affect pathotic argument quality. *Activating* pathotic constitution involves activating preexisting emotional goals in the hearer (e.g., "to feel happy") to induce a change in belief. *Effectuating* pathotic constitution involves inducing a new emotion in the hearer for them to generate a new goal or belief.

Pathotic Character: The emotional reaction of the hearer to the speaker's character can affect the pathotic quality of arguments given by the speaker. I broadly distinguish between *authority* and *sincerity*, where the former relates to the hearers believing the speaker has good personal grounds for giving their emotional arguments, and the latter relates to the hearers believing the speaker genuinely feels the emotions they are arguing about.

This initial taxonomy is grounded in, and corresponds to, the location of the three modes of persuasion (*logos*, *ethos*, *pathos*) in the Aristotelian tripartite view of rhetoric [Aristotle, 2007]. To elaborate, *logos* involves persuasion through the argument itself, *ethos* involves persuasion through the speaker, and *pathos* involves persuasion through the hearers. These locations of argument, speaker, and hearer are reused in my initial taxonomy as places in which emotion affects the quality of an argument. The hearer feels emotions towards the *content* of the arguments presented, the *character* of the speaker, or the existing emotional *constitution* of the hearer.

The AQ taxonomy was built on a survey of existing works on argument quality assessment, as well as theoretical considerations [Wachsmuth et al., 2017]. However, only a small number of studies are concerned with pathotic argument quality assessment; [Wachsmuth et al., 2024] only cites the philosophical works of Govier and Aristotle as their reason for including emotional appeal. My initial taxonomy is grounded in a broader set of theoretical works than found in [Wachsmuth et al., 2024], as well as the small number of existing studies that tackle the computational assessment of pathotic argument quality. I aim to make my taxonomy compatible with the weak and strong assessment views of pathotic argument quality assessment. The first contribution of my PhD research will be a taxonomy of a pathotic argument quality, and the evaluation of my initial taxonomy will be based on feedback from the argumentation community, focusing on the utility of the taxonomy for grounding research on computational assessment of pathotic argument quality.

3 Methods for Assessment

Broadly, argument quality is assessed with either *intrinsic* or *interactional* methods. In intrinsic assessment, the quality of an argument is determined by evaluating it without seeing it in practice. For example, [Wachsmuth et al., 2017] annotated 304 individual arguments with scores of 1 (low quality) to 3 (high quality) along each of the 13 dimensions of the AQ taxonomy. In their scoring of the AQ dimensions [Wachsmuth et al., 2017] observe moderate annotator agreement overall at .51 (Krippendorf's α); however, emotional appeal had the lowest annotator agreement of any AQ dimension at .26.

Intrinsic assessment is not ideal for pathotic argument quality assessment as it is challenging to account for the more hearer-dependent parts of my pathotic argument quality taxonomy: character and constitution. An alternative is to assess pathotic argument quality *interactionally* by evaluating arguments based on the reactions or responses of the hearers. For example, [Cano-Basave and He, 2016] studies the impact of persuasive argumentation in political debates. As features, they use a set of generated emotion-related semantic frames on the part of the speaker (pathotic character), as well as non-linguistic pathotic signals from the audience (e.g., booing and cheering). They then observe how well these features predict whether a speaker's standing in the polls improves after a debate. There are other examples of interactional pathotic argument quality assessment [Konat et al., 2024, Wu et al., 2024], with some considering more than just pathotic content [Lukin et al., 2017, El Baff et al., 2020].

Beyond just observing the pathotic signals of a hearer in response to an argument, the emotional dynamics [Hipson and Mohammad, 2021] of each hearer can be tracked, allowing for the emotional baselines of the audience to be computed. From these baselines, displacements in emotion indicate emotional reactions to arguments or claims. This method assumes, like many examples of interactional assessment, that temporally immediate reactions to arguments should contain the emotional responses to those arguments. A more rigorous approach would be to predict the exact emotional causes of an audience's

responses. Emotion-cause pair extraction in conversations (ECPEC) is a computational task that aims to extract emotional utterances from a conversation and the associated utterances that caused those emotions [Li et al., 2023]. ECPEC and emotion dynamics methods have not yet been applied to pathotic argument quality assessment. A barrier to interactional assessment of pathotic argument quality is a lack of suitable datasets; while a small number of ECPEC datasets exist, the corpora used are not strictly argumentative. The second contribution of my PhD research will be an ECPEC dataset focusing on argumentative text, incorporating the dimensions of my proposed taxonomy.

4   Applications

The applications of automatic pathotic argumentation quality assessment are unique and interesting, and can be grouped into those which treat pathotic argument quality minimally or maximally. The conceptual notions of minimal and maximal argument quality refer to, for a given task/context, whether argument quality is seen as something that is minimized (there is a low bar of acceptability that needs to be met) or maximized [Wachsmuth et al., 2024]. Pathotic argument quality has mostly been explored from a minimal perspective, such as in the inappropriateness taxonomy of [Ziegenbein et al., 2023]. For real-world applications, several tasks view uses of emotion as harming argument quality, for example, in fake news detection [Scheibenzuber et al., 2023] and polarization analysis [Gajewska et al., 2024].

In my research, I focus on applications that view pathotic argument quality as something to be maximized. This is in part to account for both the weak and strong assessment views of pathotic argument quality raised in section (1). The weak assessment view naturally lends itself to minimal assessment. Both polarization analysis and fake news detection view emotion as something that potentially impacts the quality of overall argumentation, or at least indicates the presence of low-quality argumentation. However, there are examples of compelling, uniquely emotional arguments that correspond with the strong assessment view whose quality is maximized by their emotional nature [Micheli, 2010].

Applications adjacent to argumentation that view the use of emotion in dialogue as a positive exist, particularly in the rapid development of affective computing. For example, [Feng et al., 2024] found that infusing emotions into the natural language understanding, dialogue management, and natural language generation of a task-oriented dialogue system improves both the user's emotional experience as well as the task success. I propose that the most promising applications of my pathotic argument quality taxonomy and interactional methods of assessment are the development of emotion-aware argument-based dialogue systems. This application is promising because dialogue systems are inherently interactional. An emotion-aware argument-based dialogue system would naturally incorporate interactional pathotic argument quality assessment. The final contribution of my research will be to incorporate interactional methods of pathotic argument quality assessment into dialogue systems, either by enriching existing dialogue systems or implementing a new dialogue system solely concerned with emotional argumentation.

5   Conclusion

This paper summarizes the preliminary work of my PhD research, in which I explore computational pathotic argument quality assessment. I have outlined a theoretical taxonomy of pathotic argument quality, as well as methodological considerations, which push pathotic argument quality assessment in a more context/hearer-dependent and interactional direction. Finally, I have outlined my proposed application of this theory and methodology: interactional argument-based emotion-aware dialogue systems.

6   Acknowledgements

References

[Aristotle, 2007] Aristotle (2007). *On rhetoric: A theory of civic discourse.* Clarendon Aristotle Series. Oxford University Press, Oxford.

[Braet, 1992] Braet, A. C. (1992). Ethos, pathos and logos in Aristotle's Rhetoric: A re-examination. *Argumentation*, 6(3):307–320.

[Brinton, 1988] Brinton, A. (1988). Pathos and the "Appeal to Emotion": An Aristotelian analysis. *History of Philosophy Quarterly*, 5(3):207–219.

[Cano-Basave and He, 2016] Cano-Basave, A. E. and He, Y. (2016). A study of the impact of persuasive argumentation in political debates. In Knight, K., Nenkova, A., and Rambow, O., editors, *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1405–1413, San Diego, California. Association for Computational Linguistics.

[El Baff et al., 2020] El Baff, R., Al Khatib, K., Stein, B., and Wachsmuth, H. (2020). Persuasiveness of news editorials depending on ideology and personality. In Nissim, M., Patti, V., Plank, B., and Durmus, E., editors, *Proceedings of the Third Workshop on Computational Modeling of People's Opinions, Personality, and Emotion's in Social Media*, pages 29–40, Barcelona, Spain (Online). Association for Computational Linguistics.

[Feng et al., 2024] Feng, S., Lin, H., Geishauser, C., Lubis, N., van Niekerk, C., Heck, M., Ruppik, B. M., Vukovic, R., and Gasic, M. (2024). Infusing emotions into task-oriented dialogue systems: Understanding, management, and generation. In Kawahara, T., Demberg, V., Ultes, S., Inoue, K., Mehri, S., Howcroft, D., and Komatani, K., editors, *Proceedings of the 25th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 699–717, Kyoto, Japan. Association for Computational Linguistics.

[Gajewska et al., 2024] Gajewska, E., Budzynska, K., Konat, B., Koszowy, M., Kiljan, K., Uberna, M., and Zhang, H. (2024). Ethos and pathos in online group discussions: corpora for polarisation issues in social media. arXiv:2404.04889 [cs.CL].

[Hipson and Mohammad, 2021] Hipson, W. E. and Mohammad, S. M. (2021). Emotion dynamics in movie dialogues. *PLOS ONE*, 16(9):e0256153. Publisher: Public Library of Science.

[Kennedy, 1994] Kennedy, G. A. (1994). Greek rhetorical theory from Corax to Aristotle. In *A New History of Classical Rhetoric*, pages 30–63. Princeton University Press.

[Konat et al., 2024] Konat, B., Gajewska, E., and Rossa, W. (2024). Pathos in natural language argumentation: emotional appeals and reactions. *Argumentation*, 38(3):369–403.

[Li et al., 2023] Li, W., Li, Y., Pandelea, V., Ge, M., Zhu, L., and Cambria, E. (2023). ECPEC: Emotion-cause pair extraction in conversations. *IEEE Transactions on Affective Computing*, 14(3):1754–1765.

[Lukin et al., 2017] Lukin, S., Anand, P., Walker, M., and Whittaker, S. (2017). Argument strength is in the eye of the beholder: audience effects in persuasion. In Lapata, M., Blunsom, P., and Koller, A., editors, *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 742–753, Valencia, Spain. Association for Computational Linguistics.

[Micheli, 2010] Micheli, R. (2010). Emotions as objects of argumentative constructions. *Argumentation*, 24(1):1–17.

[Scheibenzuber et al., 2023] Scheibenzuber, C., Neagu, L.-M., Ruseti, S., Artmann, B., Bartsch, C., Kubik, M., Dascalu, M., Trausan-Matu, S., and Nistor, N. (2023). Dialog in the echo chamber: Fake news framing predicts emotion, argumentation and dialogic social knowledge building in subsequent online discussions. *Computers in Human Behavior*, 140:107587.

[Vecchi et al., 2021] Vecchi, E. M., Falk, N., Jundi, I., and Lapesa, G. (2021). Towards argument mining for social good: A survey. In Zong, C., Xia, F., Li, W., and Navigli, R., editors, *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)*, pages 1338–1352, Online. Association for Computational Linguistics.

[Wachsmuth et al., 2024] Wachsmuth, H., Lapesa, G., Cabrio, E., Lauscher, A., Park, J., Vecchi, E. M., Villata, S., and Ziegenbein, T. (2024). Argument quality assessment in the age of instruction-following large language models. In Calzolari, N., Kan, M.-Y., Hoste, V., Lenci, A., Sakti, S., and Xue, N., editors, *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 1519–1538, Torino, Italia. ELRA and ICCL.

[Wachsmuth et al., 2017] Wachsmuth, H., Naderi, N., Hou, Y., Bilu, Y., Prabhakaran, V., Thijm, T. A., Hirst, G., and Stein, B. (2017). Computational argumentation quality assessment in natural language. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 176–187, Valencia, Spain. Association for Computational Linguistics.

[Williams, 2025] Williams, D. (2025). That feels convincing: A taxonomy of pathotic argumentation quality. In *Proceedings of the 5th European Conference on Argumentation*. Forthcoming.

[Wu et al., 2024] Wu, Y., Zhang, M., Shen, Y., Liu, N., and Li, Y. (2024). How do project updates influence fundraising on online medical crowdfunding platforms? Examining the dynamics of content updates. *IEEE Transactions on Engineering Management*, 71:9135–9149.

[Ziegenbein et al., 2023] Ziegenbein, T., Syed, S., Lange, F., Potthast, M., and Wachsmuth, H. (2023). Modeling appropriate language in argumentation. In Rogers, A., Boyd-Graber, J., and Okazaki, N., editors, *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4344–4363, Toronto, Canada. Association for Computational Linguistics.