HyEnA: Hybrid Intelligence for Argument Mining

Michiel van der Meer[1,2][0000−0003−1877−6002], Catholijn M. Jonker[1,2], and Pradeep K. Murukannaiah[2]

[1] Leiden Institute for Advanced Computer Science (LIACS), Leiden University
[2] Interactive Intelligence (II), TU Delft

1 Introduction

Online deliberations are an essential means of scaling stakeholder participation in decision-making processes. Such deliberations, especially when unmoderated, contain large amounts of noise. Argument Mining (AM) methods use Natural Language Processing (NLP) techniques to extract strongly argumentative content from this noise [8]. When conducting time-sensitive analyses, like assessing citizens' opinions on governmental decisions during the COVID-19 pandemic [12], such automated analyses can provide crucial insights swiftly.

However, AM methods (1) suffer from data hunger and poor cross-domain generalization, (2) do not compress information, and (3) adapt poorly to low-frequency arguments, even in the same domain. Capturing arguments in new domains from a diverse set of perspectives is important for both supporting decision-making and informing better policy [7]. Key Point Analysis (KPA) is an alternative avenue for extracting arguments from opinions [3]. However, KPA models are trained on expert argument annotations instead of user-submitted content, misrepresenting the inherent subjectivity of the task [2]. Further, employing only experts defeats the purpose of inclusive participation in decision-making.

We propose HyEnA (Hybrid Extraction of Arguments) [10], a hybrid (human + AI) method [1] that combines the strengths of computational models with human understanding. HyEnA extracts a diverse set of arguments from a textual opinion corpus by guiding humans on annotation tasks, aided by NLP methods.

2 HyEnA Method for Extracting Arguments

An overview of HyEnA is given in Figure 1. HyEnA analyzes an opinion corpus containing comments gathered from citizens on COVID-19-related policy decisions for key arguments. Humans work together with NLP models to come to a set of diverse arguments. The method is split into two phases: (1) an individual **annotation** phase, and (2) a collaborative **consolidation** phase.

Initially, annotators independently collect a set of arguments. Multiple annotators work in parallel aided by a Farthest-First opinion sampling approach to preserve a large diversity of perspectives. During consolidation, individual sets of arguments are merged into a single set of key arguments. To reduce the number of manual comparisons required we use `Power` [4], an entity resolution algorithm that mixes crowd annotation with similarity metrics.

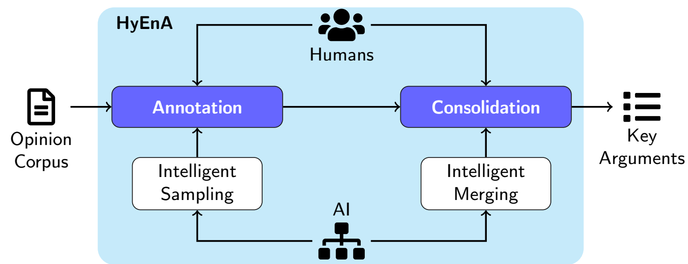

Fig. 1. Overview of the HyEnA method.

We compare our hybrid approach with a fully manual analysis of the same opinion corpus [12] and an automated approach [3]. We ask for a new set of annotators to evaluate the match between opinions and arguments and check which arguments stemming from the manual analysis were rediscovered by HyEnA.

3 Results and Discussion

We applied HyEnA to three opinion corpora on COVID-19 measures. A total of 348 Prolific crowd workers were employed among the three tasks: (1) annotation, (2) consolidation, and (3) evaluation. Workers performing the annotation task extracted arguments from noisy opinions leaving only a limited number of opinions marked as containing repeated arguments (about 15%) on average. In the consolidation phase, `Power` reduced the work required to be performed manually by 60%, and a subsequent clustering yielded groups of semantically similar argument clusters, from which single representative arguments are selected.

Compared to the automated baseline, HyEnA shows an increment in precision, indicating that our evaluation crowd workers more often judged the extracted argument as correct. In terms of coverage, the number of opinions mapped to an argument, both methods score on par with each other when considering coverage over the entire opinion corpus. However, when restricted to a limited diverse set of opinions, the coverage for the automated method drops considerably, indicating that popular arguments (i.e. frequently restated arguments) make up the majority of the baseline's coverage score. Compared to manual analysis, HyEnA analyzed fewer opinions but discovered a comparable number of arguments. Further, the diverse opinions sampled in HyEnA resulted in novel arguments not included in the manual analysis.

HyEnA shows promising results on a real-world argument extraction task and the inclusion of human effort. Several aspects of the hybrid method remain to be scrutinized. The majority of the annotation is centered on resolving argument identity, whereas more argument-informed similarity metrics [13] can help to improve the automated labeling. Additionally, arguments are not the only aspects that constitute citizens' perspectives. Merging argument extraction with the identification of personal values [9, 6], sentiment [5] or attribution [11] may further help to structure the opinionated feedback.

References

1. Akata, Z., Balliet, D., de Rijke, M., Dignum, F., Dignum, V., Eiben, G., Fokkens, A., Grossi, D., Hindriks, K., Hoos, H., Hung, H., Jonker, C., Monz, C., Neerincx, M., Oliehoek, F., Prakken, H., Schlobach, S., van der Gaag, L., van Harmelen, F., van Hoof, H., van Riemsdijk, B., van Wynsberghe, A., Verbrugge, R., Verheij, B., Vossen, P., Welling, M.: A research agenda for hybrid intelligence: Augmenting human intellect with collaborative, adaptive, responsible, and explainable artificial intelligence. Computer **53**(8), 18–28 (2020). https://doi.org/10.1109/MC.2020.2996587
2. Aroyo, L., Welty, C.: Truth is a lie: Crowd truth and the seven myths of human annotation. AI Magazine **36**(1), 15–24 (2015)
3. Bar-Haim, R., Kantor, Y., Eden, L., Friedman, R., Lahav, D., Slonim, N.: Quantitative argument summarization and beyond: Cross-domain key point analysis. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). pp. 39–49. Association for Computational Linguistics, Online (Nov 2020). https://doi.org/10.18653/v1/2020.emnlp-main.3, https://aclanthology.org/2020.emnlp-main.3
4. Chai, C., Li, G., Li, J., Deng, D., Feng, J.: Cost-effective crowdsourced entity resolution: A partial-order approach. In: Proceedings of the 2016 International Conference on Management of Data. p. 969–984. SIGMOD '16, Association for Computing Machinery, New York, NY, USA (2016). https://doi.org/10.1145/2882903.2915252, https://doi.org/10.1145/2882903.2915252
5. Ibeke, E., Lin, C., Wyner, A., Barawi, M.H.: Extracting and understanding contrastive opinion through topic relevant sentences. In: Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers). pp. 395–400 (2017)
6. Kiesel, J., Alshomary, M., Handke, N., Cai, X., Wachsmuth, H., Stein, B.: Identifying the human values behind arguments. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 4459–4471 (2022)
7. Klein, M.: Enabling large-scale deliberation using attention-mediation metrics. Computer Supported Cooperative Work (CSCW) **21**(4-5), 449–473 (2012). https://doi.org/10.1007/s10606-012-9156-4
8. Lawrence, J., Reed, C.: Argument Mining: A Survey. Computational Linguistics **45**(4), 765–818 (01 2020). https://doi.org/10.1162/coli_a_00364, https://doi.org/10.1162/coli_a_00364
9. Liscio, E., van der Meer, M., Siebert, L.C., Jonker, C.M., Mouter, N., Murukannaiah, P.K.: Axies: Identifying and evaluating context-specific values. In: Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems. p. 799–808. AAMAS '21, International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC (2021)
10. van der Meer, M., Liscio, E., Jonker, C.M., Plaat, A., Vossen, P., Murukannaiah, P.K.: Hyena: A hybrid method for extracting arguments from opinions. In: Proceedings of the first International Conference on Hybrid Human-Artificial Intelligence (HHAI 2022). pp. 1–15. IOS Press, Amsterdam, the Netherlands (2022)
11. Morante, R., Van Son, C., Maks, I., Vossen, P.: Annotating perspectives on vaccination. In: Proceedings of The 12th Language Resources and Evaluation Conference. pp. 4964–4973 (2020)

12. Mouter, N., Hernandez, J.I., Itten, A.V.: Public participation in crisis policymaking. how 30,000 dutch citizens advised their government on relaxing covid-19 lockdown measures. PLOS ONE **16**(5), 1–42 (05 2021). https://doi.org/10.1371/journal.pone.0250614, https://doi.org/10.1371/journal.pone.0250614
13. Opitz, J., Heinisch, P., Wiesenbach, P., Cimiano, P., Frank, A.: Explainable unsupervised argument similarity rating with abstract meaning representation and conclusion generation. In: Proceedings of the 8th Workshop on Argument Mining. pp. 24–35 (2021)