Fine-Grained Emotion Knowledge Extraction in Human Values: An Interdisciplinary Analysis

Amir Reza Jafari[1], Praboda Rajapaksha[1,2], Reza Farahbakhsh[1], Guanlin Li[1], Noel Crespi[1]

[1] Samovar, Telecom SudParis, Institut Polytechnique de Paris, 91120 Palaiseau, France,
[2] Aberystwyth University, Aberystwyth, SY23 3DB Ceredigion, UK
amir-reza.jafari_tehrani@telecom-sudparis.eu

1 Introduction

Human arguments serve as a means to express and substantiate ideas, and the way they are presented can vary in terms of style, language, and objectives. In recent years, detecting and understanding 'human values' behind arguments become important with the proliferation of digital communication [1,2]. Specifically, this task gained more attention as it was introduced in the SemEval 2023 competition. This task focuses on detecting human values using a multi-level taxonomy comprising 54 values [2]. Detecting these values using machine learning presents considerable challenges due to the multitude of values and their implicit usage in arguments, and several existing research employed various approaches to improve the performance [3,4,5,6][3].

The process of extracting information from human values involves interpreting the ethical, cultural, and emotional foundations that drive individuals and groups to express their opinions to support their discussion. Fine-grained emotion analysis is a valuable disciplinary research area to reinforce detecting and understanding human values because it provides a deeper, more nuanced insight into the emotional support of value expressions. We investigated this subject to improve implicit hate speech detection earlier [7], and in this paper, we aim to utilize emotion analysis to enhance the process of human value detection by providing an interdisciplinary perspective on how values encompass emotional knowledge [8], [9].

In this study, we analyze fine-grained emotions using GoEmotion [10] in human values dataset [11], which is an annotated dataset for the SemEval 2023 Task 4. We also conducted an experiment by using emotion features for the human value detection task. In order to extract information from the discussions and construct a large-scale knowledge graph, unravelling these emotions becomes vital for understanding the characteristics behind different values through data-driven analyses.

2 Method

In our study, we investigated the 'fine-grained emotions' analysis applied to the human values dataset, which consists of religious texts, political discussions, free-text arguments, newspaper editorials, and online democracy platforms. To carry out our analysis, we used Huggingface implementation of the original taxonomy of the GoEmotions

[3] https://touche.webis.de/publications.html

model [4], which consists of 27 emotions + neutral which is inherently structured in a finer-grained Hierarchical Grouping (positive, negative, ambiguous). This model is based on Transformers and computes a probabilistic score for each emotion within the range of 0 to 1, thereby quantifying the likelihood of its occurrence. We set the threshold at 0.1 for the purposes of our experiments to include a wide range of emotions for each value.

Furthermore, the human value dataset contains 9324 arguments consolidated multi-level taxonomy [11]. At the first level, this taxonomy comprises a total of 54 fundamental human values. Subsequently, at the second level, these values are further categorized into distinct value categories. Although the dataset extends to encompass additional levels, such as Higher-Order Values (Level 3), Personal/Social Focus (Level 4a), and Motivation (Level 4b), our study focused solely on the initial two levels for our analytical experiments.

3 Results and Discussion

Figure 1 depicts the detailed distribution of fine-grained emotions within Level 2 of the human value dataset [11]. In Positive polarity, most of the human values are under the "*Approval*" category, while "*Disapproval*" and "*Annoyance*" emerge as the dominant emotion categories of Negative sentiments. This indicates that when individuals argue about negative opinions, they often do so to convey disapproval of certain values or experiences that provoke annoyance, and positive opinions convey approval of certain values. Additionally, "*Realization*" is prominent among the ambiguous emotions, which indicates that these values do not strongly express approval or disapproval. Notably, these emotions hold significant potential as additional features for enhancing the classification process. Moreover, as the dataset comprises arguments characterized by a premise, a conclusion, and a stance indicator that specifies whether the premise is in favor or against the conclusion, it is noticeable that the emotions extracted from this data mostly contain opposing emotions, such as "Approval" and "Disapproval".

For human values detection at Level 2, we leveraged the BERT model [12] and concatenated the extracted emotion features with textual features, yielding the results presented in Table 1. These results emphasise the promising possibility of enhancing the performance of human value detection through the incorporation of emotion-related information and the utilization of emotion features within the classification model.

In pursuing future research directions, we plan to extend our fine-grained emotion analysis and employ emotion features to encompass other levels of human values. This involves assessing the influence of emotional features on each distinct level to enhance human value detection performance. Moreover, we intend to construct a comprehensive knowledge graph founded on the insights gleaned from emotional features and investigate other aspects of information extraction for understanding human values. Furthermore, we will utilize Large Language Models (LLMs) and techniques such as chain-of-thought (CoT) prompting [13] as they enable complex reasoning capabilities. This integrated approach, encompassing emotion-driven analysis, knowledge graph development, and LLM-based reasoning, represents our strategic vision for pushing the frontiers of human value detection tasks as an interdisciplinary analysis.

[4] https://github.com/monologg/GoEmotions-pytorch

Table 1. Experimental Result: Human value classification (level 2) using emotion features.

Value Category	Self-direction: thought	Self-direction: action	Stimulation	Hedonism	Achievement	Power: dominance	Power: resources	Face	Security: personal	Security: societal	Tradition	Conformity: rules	Conformity: interpersonal	Humility	Benevolence: caring	Benevolence: dependability	Universalism: concern	Universalism: nature	Universalism: tolerance	Universalism: objectivity	Macro avg
Precision	0.33	0.42	0.16	0.21	0.47	0.20	0.39	0.19	0.62	0.45	0.35	0.38	0.19	0.20	0.41	0.33	0.49	0.49	0.17	0.39	0.32
Recall	0.64	0.75	0.52	0.84	0.73	0.57	0.77	0.55	0.89	0.88	0.45	0.77	0.30	0.44	0.88	0.66	0.93	0.86	0.44	0.65	0.68
F1	0.44	0.54	0.24	0.34	0.57	0.30	0.52	0.28	0.73	0.60	0.39	0.51	0.23	0.28	0.56	0.44	0.64	0.62	0.25	0.49	0.44

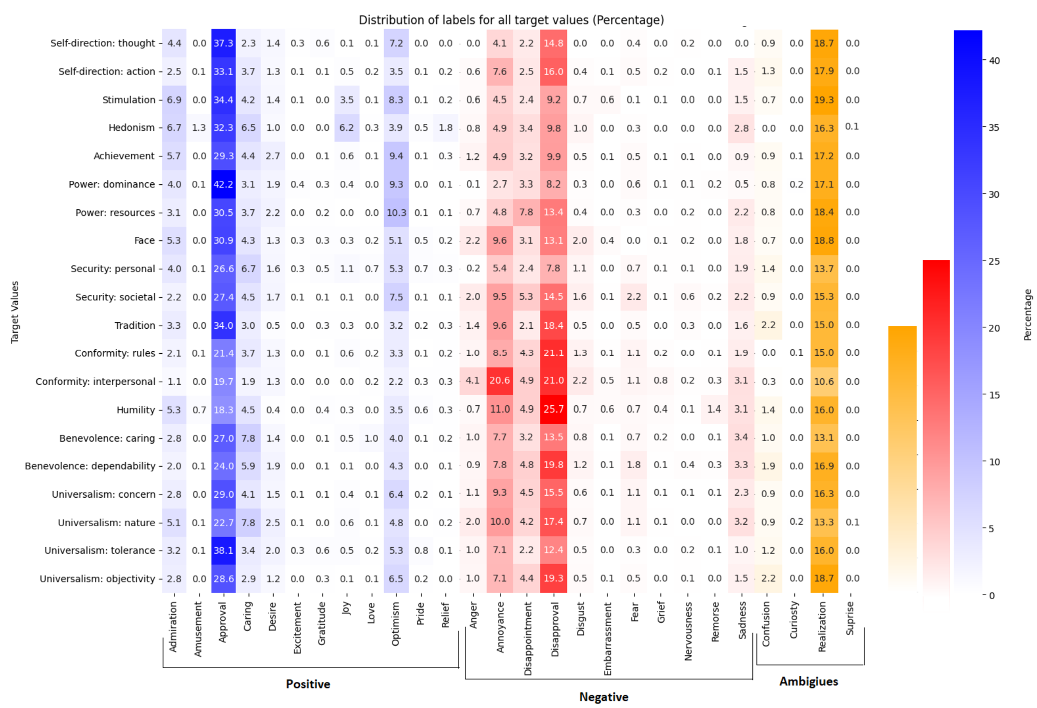

Fig. 1. Fine-grained emotion distribution of human values. This distribution indicates the importance of specific emotions with different polarities in human value arguments that can be a good candidate to consider as a feature for improving the performance of human value detection. The y-axis represents the Value category of the dataset at level 2, and the x-axis shows the fine-grained emotions from the original taxonomy of GoEmotions. The blue color in this heatmap is assigned for emotions under positive sentiment categories, and red and orange indicate emotions of negative and ambiguous categories, respectively.

References

1. T. Maheshwari, A. N. Reganti, S. Gupta, A. Jamatia, U. Kumar, B. Gambäck, and A. Das, "A societal sentiment analysis: Predicting the values and ethics of individuals by analysing social media content," in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, 2017, pp. 731–741.
2. J. Kiesel, M. Alshomary, N. Handke, X. Cai, H. Wachsmuth, and B. Stein, "Identifying the human values behind arguments," in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2022, pp. 4459–4471.
3. D. Schroter, D. Dementieva, and G. Groh, "Adam-smith at SemEval-2023 task 4: Discovering human values in arguments with ensembles of transformer-based models," in *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*. Toronto, Canada: Association for Computational Linguistics, Jul. 2023, pp. 532–541. [Online]. Available: https://aclanthology.org/2023.semeval-1.74
4. G. Papadopoulos, M. Kokol, M. Dagioglou, and G. Petasis, "Andronicus of rhodes at SemEval-2023 task 4: Transformer-based human value detection using four different neural network architectures," in *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*. Toronto, Canada: Association for Computational Linguistics, Jul. 2023, pp. 542–548. [Online]. Available: https://aclanthology.org/2023.semeval-1.75
5. S. Honda and S. Wilharm, "Noam Chomsky at SemEval-2023 task 4: Hierarchical similarity-aware model for human value detection," in *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*. Toronto, Canada: Association for Computational Linguistics, Jul. 2023, pp. 1359–1364. [Online]. Available: https://aclanthology.org/2023.semeval-1.188
6. H. H. Hemati, S. H. Alavian, H. Sameti, and H. Beigy, "Sutnlp at semeval-2023 task 4: Lg-transformer for human value detection," in *Proceedings of the The 17th International Workshop on Semantic Evaluation (SemEval-2023)*, 2023, pp. 340–346.
7. A. R. Jafari, G. Li, P. Rajapaksha, R. Farahbakhsh, and N. Crespi, "Fine-grained emotions influence on implicit hate speech detection," *IEEE Access*, vol. 11, pp. 105 330–105 343, 2023.
8. N. H. Ibrahim, A. Azmi, and A. M. Lokman, "Towards human values, ethics and positive emotion in good governance: A reliability test of good governance index model," in *Proceedings of the 8th International Conference on Kansei Engineering and Emotion Research: KEER 2020, 7-9 September 2020, Tokyo, Japan*. Springer, 2020, pp. 421–429.
9. J. J. Deng, C. H. C. Leung, P. Mengoni, and Y. Li, "Emotion recognition from human behaviors using attention model," in *2018 IEEE First International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*, 2018, pp. 249–253.
10. D. Demszky, D. Movshovitz-Attias, J. Ko, A. Cowen, G. Nemade, and S. Ravi, "Goemotions: A dataset of fine-grained emotions," *arXiv preprint arXiv:2005.00547*, 2020.
11. N. Mirzakhmedova, J. Kiesel, M. Alshomary, M. Heinrich, N. Handke, X. Cai, B. Valentin, D. Dastgheib, O. Ghahroodi, M. A. Sadraei *et al.*, "The touch\'e23-valueeval dataset for identifying human values behind arguments," *arXiv preprint arXiv:2301.13771*, 2023.
12. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.
13. J. Wei, X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, Q. V. Le, D. Zhou *et al.*, "Chain-of-thought prompting elicits reasoning in large language models," *Advances in Neural Information Processing Systems*, vol. 35, pp. 24 824–24 837, 2022.