

Advanced Multi-Class Emotion Detection in Text: Leveraging Ensemble Techniques and Addressing Class Imbalance

Author: Courage Adzewoda

ID: 10962382

University of Ghana, Computer Science

1. Introduction

The ability to understand human emotions from text is crucial for various applications such as sentiment analysis in social media, customer service chatbots, and mental health monitoring. Traditional sentiment analysis often focuses on binary classification (positive or negative); however, human emotions are far more nuanced. This project proposes an advanced approach for multi-class emotion detection in text, leveraging ensemble techniques and addressing the challenge of class imbalance.

2. Problem Statement

The complete range of human emotions seen in written texts is difficult for most text-analysis techniques to recognize. A major problem is categorizing emotions into different groups such as fear, anger, sadness, and joy. Additionally, emotional expressions in the text can be imbalanced, with some emotions appearing more frequently than others. This class imbalance can significantly hinder the accuracy of machine-learning models.

3. Proposed Approach

This project will address these challenges by developing a novel multiclass emotion detection system in text and addressing class imbalances. Here is the breakdown:

- **Natural Language Toolkit (NLTK):** We will utilize NLTK, a powerful Python library for natural language processing, to perform tasks such as text preprocessing (tokenization, regular expressions for stemming, and stop word removal).
- **Synthetic Minority Oversampling Technique (SMOTE):** We used SMOTE, a method that creates synthetic examples for underrepresented emotions in the dataset, to solve the class imbalance. This ensured a more balanced training set for the model.
- **Ensemble Learning:** We leverage ensemble methods, in which multiple machine learning models are combined to create a more robust and accurate classifier. This approach can improve the overall performance of emotion detection compared with a single model.

4. Methodology

The project followed the following key steps:

1. **Data Collection:** A dataset from Kaggle labeled with various emotions in text form was gathered.
2. **Data preprocessing:** The text data are cleaned and prepared using NLTK, including tokenization, stop word removal, and feature extraction.
3. **Class Imbalance Handling:** Apply SMOTE to generate synthetic samples for underrepresented emotions and balance the dataset for a balanced training set for the model.
4. **Model Training:** Machine-learning models (e.g., Naive Bayes, Random Forest, and Logistic Regression) will be trained to classify emotions.
5. **Ensemble Creation:** Combine trained models into an ensemble using techniques such as bagging, boosting, and voting to determine the one that performs best.
6. **Evaluation:** A held-out test set was used to assess the ensemble model's performance using measures including accuracy, recall, precision, and F1 score for every emotion class.

5. Expected Outcome

This project aims to develop a highly accurate and robust system for multiclass emotion detection in text. By leveraging ensemble methods and addressing class imbalance, we expect to achieve significant improvements over traditional sentiment analysis approaches.

7. Conclusion

This project can significantly advance the field of multi-class emotion detection in text. By leveraging ensemble learning techniques and addressing class imbalance, the proposed system can offer a powerful tool for various applications that rely on understanding human emotions through written communication.

8. References

- Balakrishnan, V., Cynthia Martin, M., Kaur, W., & Javed, A. (2019). A comparative analysis of detection mechanisms for emotion detection. *Journal of Physics: Conference Series*, 1339(1). <https://doi.org/10.1088/1742-6596/1339/1/012016>
- Bandhakavi, A., Wiratunga, N., & Massie, S. (1541b). *AFFECTIVE COMPUTING AND SENTIMENT ANALYSIS Lexicon Generation for Emotion Detection from Text*. <http://news.sina.com.cn/society>
- Chatterjee, A., Narahari, N., Joshi, M., & Agrawal, P. (n.d.). *SemEval-2019 Task 3: EmoContext Contextual Emotion Detection in Text*. www.pandorabots.com/mitsuku
- Dietterich, T. G. (n.d.). *Ensemble Methods in Machine Learning*. <http://www.cs.orst.edu/~tgd>
- Khatri, S. K., Amity University, Amity University. Amity Institute of Information Technology, Institute of Electrical and Electronics Engineers. Uttar Pradesh Section, & Institute of Electrical and Electronics Engineers. (n.d.). *2018 7th International Conference on Reliability*,

- Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions) : August 29-31, 2018, venue, Amity University Uttar Pradesh, Noida, India.*
- Khatri, S. Kumar. (2018). *2018 7th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions) : August 29-31, 2018, venue, Amity University Uttar Pradesh, Noida, India.* IEEE.
- Kodem, S., Sudheer Reddy, D., Siva Rama Krishna, B. N., Suman, N., Sathish, V., & Banala, M. (2024). Emotion detection from text using recurrent neural network. *AIP Conference Proceedings*, 2971(1). <https://doi.org/10.1063/5.0196068>
- Nandwani, P., & Verma, R. (2021). A review on sentiment analysis and emotion detection from text. In *Social Network Analysis and Mining* (Vol. 11, Issue 1). Springer. <https://doi.org/10.1007/s13278-021-00776-6>
- Nie, J.-Y., Institute of Electrical and Electronics Engineers, & IEEE Computer Society. (n.d.). *2017 IEEE International Conference on Big Data : proceedings : Dec 11- 14, 2017, Boston, MA, USA.*
- Poria, S., Cambria, E., Bajpai, R., & Hussain, A. (2017). A review of affective computing: From unimodal analysis to multimodal fusion. *Information Fusion*, 37, 98–125. <https://doi.org/10.1016/j.inffus.2017.02.003>
- Rahman, M. M., & Shova, S. (2023). *Emotion Detection From Social Media Posts.* <http://arxiv.org/abs/2302.05610>
- Rechowicz, K. J., & Elzie, C. A. (2023). The use of artificial intelligence to detect students' sentiments and emotions in gross anatomy reflections. *Anatomical Sciences Education*. <https://doi.org/10.1002/ase.2273>