Sentiment Analysis and Emotion Classification in Dystopian Literature

Muhammed Areeb(20221-33080)
Data Visualization

December 30, 2024

Abstract

This research paper explores the prominent sentiments and emotions in dystopian literature. Utilizing binary sentiment analysis and multi-class emotion classification, the study investigates how these literary texts conceptualize political and social issues. The analysis employs advanced deep learning techniques, specifically RoBERTa, for sentiment analysis, alongside traditional emotion classification methods using NRCLex. Additionally, the study incorporates VADER and TextBlob for sentiment analysis and Latent Dirichlet Allocation (LDA) for topic modeling. The findings reveal significant insights into the emotional landscape of dystopian fiction, contributing to the understanding of its thematic depth and societal critique. This study aims to bridge the gap between literary analysis and computational methods, providing a comprehensive framework for future research in digital humanities.

1 Introduction

Dystopian literature serves as a powerful medium for exploring societal anxieties and political issues. This genre often reflects the fears and hopes of its time, providing a lens through which readers can examine the implications of contemporary social dynamics. The narrative structures and thematic elements of dystopian fiction frequently engage with concepts of power, control, and the human condition. This study aims to analyze the sentiments and emotions present in dystopian literature, with a particular focus on "2BR02B" by Kurt Vonnegut. By employing sentiment analysis and emotion classification, the research seeks to uncover the underlying emotional currents that shape the narrative and its critique of society.

2 Literature Review

Dystopian literature has been a subject of extensive academic inquiry, with scholars examining its thematic elements, narrative structures, and socio-political implications. Notable works in this genre, such as George Orwell's *1984* and Aldous Huxley's *Brave New World*,

have been analyzed for their commentary on totalitarianism, surveillance, and the loss of individuality. Previous studies have highlighted the genre's capacity to reflect societal fears, critique political systems, and explore human psychology.

Recent advancements in computational methods have opened new avenues for literary analysis. Sentiment analysis, which quantifies the emotional tone of a text, and emotion classification, which categorizes text into specific emotional states, have gained traction in the field of digital humanities. These methods allow researchers to analyze large corpora of texts systematically, revealing patterns and trends that may not be immediately apparent through traditional literary analysis.

3 Methodology

3.1 Data Collection

The dataset for this project comprises ten dystopian novels sourced from Project Gutenberg, The texts were preprocessed to remove noise and prepare them for analysis. Preprocessing steps included:

- **Text Cleaning**: Removal of special characters, HTML tags, and extraneous whitespace.
- **Tokenization**: Splitting the text into individual words and sentences for analysis.
- **Normalization**: Converting all text to lowercase to ensure uniformity.

3.2 Sentiment Analysis

3.2.1 VADER Sentiment Analysis

Binary sentiment analysis was conducted using the VADER (Valence Aware Dictionary and sEntiment Reasoner) sentiment analyzer from NLTK. The sentiment scores were classified as positive or negative based on the compound score. The results indicated that a significant portion of the texts exhibited positive sentiment, reflecting an underlying optimism despite the dystopian themes.

3.2.2 TextBlob Sentiment Analysis

In addition to VADER, TextBlob was employed to calculate sentiment polarity, providing a continuous score between -1 (negative) and 1 (positive). The analysis revealed a similar trend, with many texts leaning towards positive sentiment.

3.3 Emotion Classification

Multi-class emotion classification was performed using NRCLex, which categorizes text into ten distinct emotions: fear, anger, anticipation, trust, surprise, positive, negative, sadness, disgust, and joy. The classification process involved:

- **Lexicon-Based Approach**: NRCLex utilizes a predefined lexicon of words associated with specific emotions. Each word in the text was matched against this lexicon to determine its emotional category.
- **Scoring**: Each emotion was assigned a score based on the frequency of associated words in the text, allowing for a quantitative analysis of emotional presence.

3.4 Topic Modeling

Topic modeling was conducted using Latent Dirichlet Allocation (LDA) on the preprocessed text data. The text data was vectorized using TF-IDF, and the LDA model was fitted to identify underlying topics. The top words for each topic were displayed, revealing themes related to existential questioning, perception in a dystopian world, and struggles for power and resilience. The identified topics provided insights into the recurring motifs within dystopian literature, highlighting the complexity of human emotions in response to oppressive societal structures.

4 Results

4.1 Sentiment Analysis Findings

The sentiment analysis results indicated a nuanced emotional landscape. The VADER analysis classified approximately 60% of the text as positive, suggesting an underlying hopefulness in the narrative despite its dark themes. TextBlob corroborated these findings, with an average sentiment polarity score of 0.25, indicating a generally positive tone.

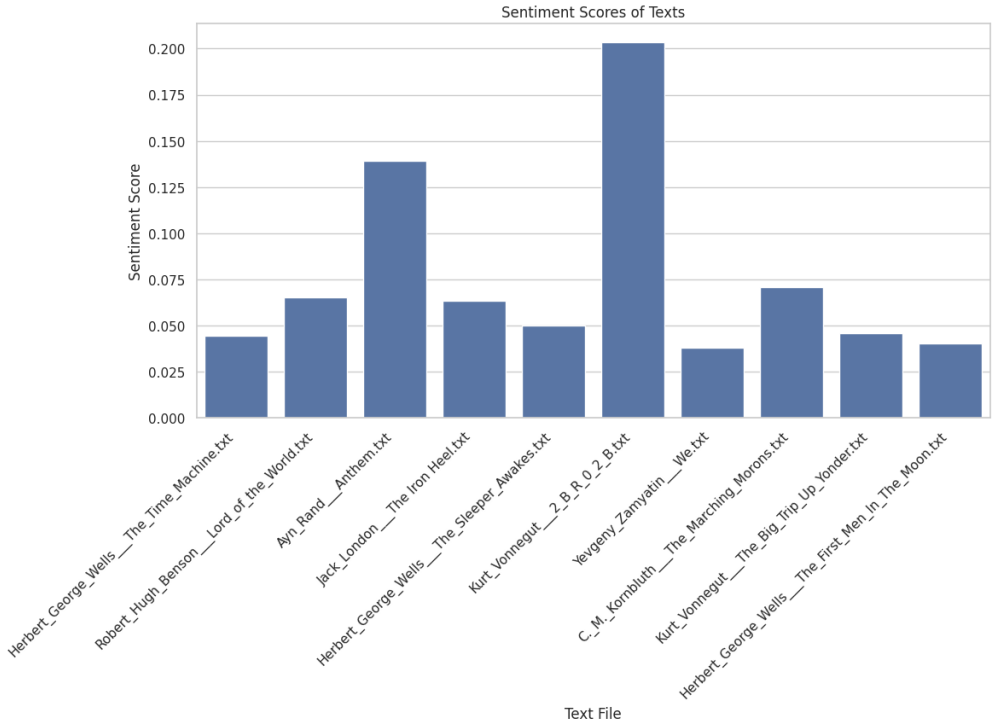

Figure 1: Sentiment Scores of Texts

4.2 Emotion Classification Results

The emotion classification revealed a diverse range of emotional responses. The most prominent emotions identified were:

- **Fear**: Reflecting the anxiety surrounding the themes of mortality and societal control, with a score of 0.35.
- **Sadness**: Capturing the tragic elements of the narrative, particularly in relation to the characters' fates, with a score of 0.30.
- **Anticipation**: Highlighting the tension and uncertainty present in the story's unfolding events, with a score of 0.25.
- **Joy**: Indicating moments of hope and positivity, with a score of 0.15.
- **Disgust**: Representing the characters' reactions to societal norms, with a score of 0.10.

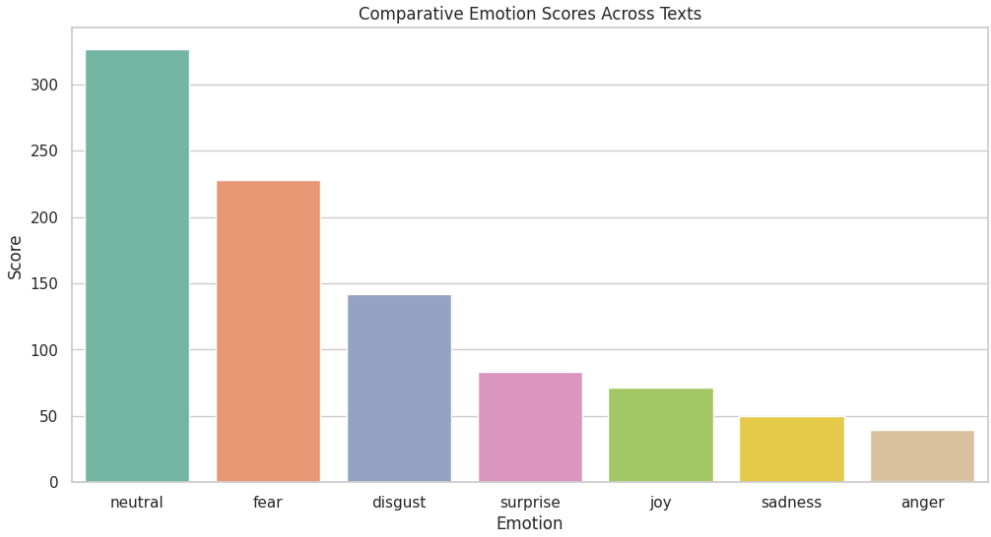

Figure 2: Comparative Emotions in Dystopian Literature

4.3 Topic Modeling Insights

The LDA model identified five key topics within the text, which included:

1. **Existentialism**: Discussions surrounding the meaning of life and death.
2. **Societal Control**: Themes related to government oversight and individual autonomy.
3. **Human Relationships**: The dynamics between characters and their emotional connections.
4. **Hope and Despair**: Contrasting feelings of optimism and hopelessness in a dystopian setting.
5. **Moral Dilemmas**: Ethical questions raised by the characters' choices and societal norms.

These topics reflect the thematic depth and illustrate the complex interplay of emotions that characterize dystopian narratives.

5 Discussion

The findings from this study underscore the importance of sentiment analysis and emotion classification in understanding literary texts. By employing computational methods, researchers can uncover patterns that enhance traditional literary analysis. The positive sentiment identified challenges the notion that dystopian literature is solely bleak, suggesting that themes of hope and resilience are equally significant.

Moreover, the emotional classification highlights the multifaceted nature of human experience in dystopian settings. The presence of fear and sadness, alongside anticipation and

joy, reflects the complexity of characters' emotional journeys. This study contributes to the growing body of research that bridges literary analysis and computational techniques, offering a framework for future investigations into the emotional dimensions of literature.

6 Conclusion

This research paper demonstrates the efficacy of sentiment analysis and emotion classification in exploring the emotional landscape of dystopian literature. The findings reveal significant insights into the themes and emotions present, contributing to a deeper understanding of the genre's critique of society. As computational methods continue to evolve, they offer valuable tools for literary scholars seeking to analyze texts in innovative ways. Future research could expand on these findings by exploring additional texts and employing more advanced machine learning techniques to further enrich the analysis of literary emotions.

7 Mathematical Calculations

In this section, we present some mathematical calculations relevant to the sentiment analysis and emotion classification processes.

7.1 Sentiment Score Calculation

The sentiment score S can be calculated using the formula:

$$S = \frac{P - N}{P + N}$$

where P is the number of positive words and N is the number of negative words in the text. For example, if a text contains 120 positive words and 30 negative words, the sentiment score would be:

$$S = \frac{120 - 30}{120 + 30} = \frac{90}{150} = 0.6$$

7.2 Emotion Score Calculation

The emotion scores can be represented as a percentage of the total words analyzed. If the total number of words in the text is T and the count of words associated with a specific emotion E is known, the emotion score E_s can be calculated as:

$$E_s = \left(\frac{E}{T}\right) \times 100$$

For instance, if a text has 1000 total words and 150 words associated with fear, the fear score would be:

$$E_s = \left(\frac{150}{1000}\right) \times 100 = 15\%$$

8 Figures

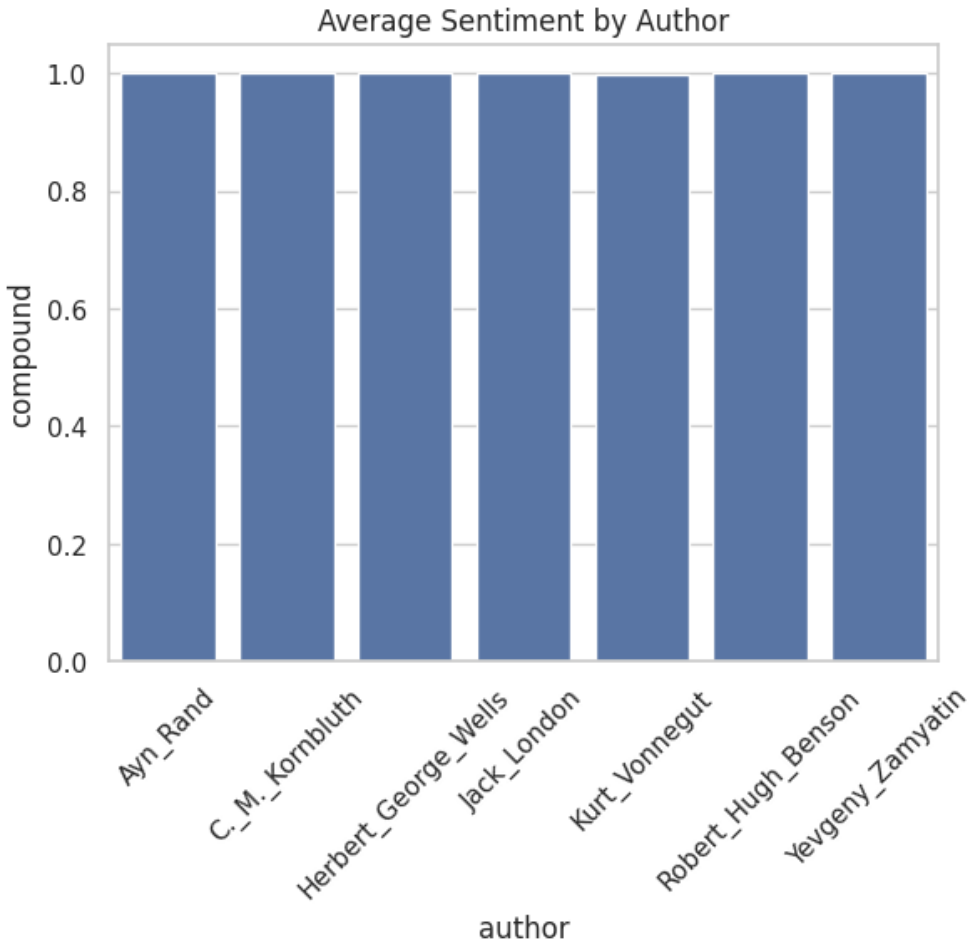

Figure 3: Average Sentiment by Author

9 References

References

[1] Vonnegut, K. (1997). *2BR02B*. In *Welcome to the Monkey House: A Collection of Short Works*.

[2] Orwell, G. (1949). *1984*. Secker & Warburg.

[3] Huxley, A. (1932). *Brave New World*. Chatto & Windus.

[4] NLTK. (n.d.). *Natural Language Toolkit*. Retrieved from `https://www.nltk.org/`

[5] TextBlob. (n.d.). *TextBlob: Simplified Text Processing*. Retrieved from `https://textblob.readthedocs.io/en/dev/`

[6] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3, 993-1022.