Ssnites at Touché: Ideology and Power Identification in Parliamentary Debates using BERT Model

Notebook for the Touché Lab at CLEF 2024

Keerthick V, Kaushik Ananth Kumar S, Kathir Kaman A, Mirunalini P and Sripriya N

SSN College of Engineering, Chennai, Tamil Nadu, India

Abstract

Debates in national parliaments do not only affect the fundamental aspects of citizens' life, but often a broader area, or even the whole world. As a form of political debate, however, parliamentary speeches are often indirect and present a number of challenges to computational analyses. In this task, we focus on identifying two variables associated with speakers in a parliamentary debate: their political ideology and whether they belong to a governing party or a party in opposition. Both subtasks are formulated as binary classification tasks.

Keywords

Ideology, Speech, Power

1. Introduction

This study presents the development and evaluation of machine learning models for the shared task "Ideology and Power Identification in Parliamentary Debates" as part of CLEF 2024. Utilizing a dataset derived from the ParliaMent corpora , we aim to classify parliamentary speeches based on two distinct variables: the political ideology of the speaker (left or right) and their party affiliation (governing party or opposition). The dataset comprises speeches from multiple national and regional parliaments, providing a diverse linguistic and political landscape. Each speech is represented by a unique identifier, the speaker's identifier, the speaker's sex, the transcribed text of the speech, an English translation of the text (where applicable), and binary labels for political ideology and party affiliation. Our methodology involves preprocessing the text data, translating non-English speeches where necessary, and extracting relevant linguistic features. We then apply various supervised machine learning algorithms to build classification models. These models are trained on the provided dataset and evaluated using metrics such as accuracy, precision, recall, and F1 score to determine their effectiveness. The results highlight the challenges and opportunities in automatically identifying political ideology and power dynamics in parliamentary debates. This research contributes to the field of political text analysis by providing robust methodologies for understanding the underlying political contexts of parliamentary discourse, ultimately aiding in the development of tools for more transparent and informed governance.

2. Background

Parliamentary debates are a cornerstone of democratic governance, providing a forum where elected representatives discuss, deliberate, and decide on issues that affect national and international policies. These debates not only influence legislation but also reflect the ideological stances and political strategies of different parties. Understanding the nuances of these debates is crucial for multiple reasons: Policy Analysis: Stakeholders, including policymakers, researchers, and the public, need to understand the positions and arguments presented by different political actors to evaluate the potential impact of proposed legislation. Political Transparency: Analyzing parliamentary debates enhances transparency by making it easier to track how representatives align with their stated policies and party platforms.

CLEF 2024: Conference and Labs of the Evaluation Forum, September 09–12, 2024, Grenoble, France
keerthick2210372@ssn.edu.in (K. V); kaushikananthkumar2210199@ssn.edu.in (K. A. K. S); kathirkaman2210947@ssn.edu.in (K. K. A); miruna@ssn.edu.in (M. P); sripriyan@ssn.edu.in (S. N)

Public Engagement: Improved understanding of parliamentary discourse can foster greater public engagement and awareness of political processes. Despite their importance, parliamentary speeches are inherently complex. They often contain indirect language, rhetorical devices, and contextual references that can obscure the speaker's true intent. Traditional methods of analyzing political texts rely heavily on human interpretation, which is time-consuming and subjective. Consequently, there is a growing interest in applying computational techniques to automate and enhance the analysis of political speeches.

3. System Overview

System Overview The system developed for this study integrates multiple components designed to process, analyze, and classify parliamentary speeches. The system architecture can be divided into several key modules:

3.1. Data Ingestion and Preprocessing

Data Collection: Acquires parliamentary debate transcripts from given sources. Text Cleaning: Removes non-essential elements such as procedural annotations and speaker identifiers. Tokenization: Splits text into individual words or tokens. Annotation: Labels each speech segment with political ideology and party status using expert annotations.

3.2. Feature Extraction

Linguistic Features: Extracts n-grams, part-of-speech tags, and syntactic structures. Semantic Features: Utilizes BERT [1] to capture contextual meanings and dependencies within the text. Domain-Specific Features: Identifies references to political entities, policy issues, and rhetorical devices.

3.3. Model Training

Model Selection: Implements a machine learning transformer BERT-based model[2]. Training Process: Uses stratified train-test splits and cross-validation to ensure balanced and robust training. Hyperparameter Tuning: Adjusts model parameters to optimize performance. CUDA Acceleration: Employs CUDA [3] to accelerate the training of BERT-based models [4], improving efficiency and scalability .

3.4. Classification

Ideology Classification: Classifies speeches into left or right political ideology using trained models. Party Status Classification: Determines whether the speaker belongs to a governing or opposition party.

3.5. Evaluation and Analysis

Performance Metrics: Assesses model performance using accuracy, precision, recall, and F1-score. Error Analysis: Identifies common misclassification errors and their potential causes. Comparative Analysis: Compares the performance of different models and feature sets.

3.6. Output and Visualization

Results Presentation: Displays classification results in a user-friendly format. Visualization Tools: Provides visualizations such as confusion matrices and performance charts to aid in interpreting the results.

4. Results

In our study, we applied a BERT model [5] to two tasks: identifying the ideology of a speaker's party and determining whether the speaker's party is currently governing or in opposition, using parliamentary speeches in multiple languages. The model achieved an average F1 score of 0.5894 for the orientation task and 0.6026 for the power task. These results indicate moderate performance, suggesting that while the model performs better than random guessing, there is substantial room for improvement. Factors such as the variability and complexity of political discourse across languages, potential class imbalances in the dataset, and the nuanced nature of political ideologies likely contributed to these outcomes. Future work should focus on enhancing dataset quality, incorporating additional contextual features, and exploring advanced modeling techniques to improve accuracy.

5. Conclusion

This study demonstrates the viability of binary classification for political ideology and party status in parliamentary debates, contributing valuable insights into political discourse analysis. Our findings underscore the potential of computational techniques in political science, paving the way for more sophisticated and scalable analysis methods.

References

[1] S.-H. Chiu, B. Chen, Innovative bert-based reranking language models for speech recognition, in: 2021 IEEE Spoken Language Technology Workshop (SLT), IEEE, 2021, pp. 266–271.

[2] E. Demir, M. Bilgin, Sentiment analysis from turkish news texts with bert-based language models and machine learning algorithms, in: 2023 8th International Conference on Computer Science and Engineering (UBMK), IEEE, 2023, pp. 01–04.

[3] K. Ho, H. Zhao, A. Jog, S. Mohanty, Improving gpu throughput through parallel execution using tensor cores and cuda cores, in: 2022 IEEE Computer Society Annual Symposium on VLSI (ISVLSI), IEEE, 2022, pp. 223–228.

[4] M. Koroteev, Bert: a review of applications in natural language processing and understanding, arXiv preprint arXiv:2103.11943 (2021).

[5] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805 (2018).

[6] X. Zhao, J. Greenberg, Y. An, X. T. Hu, Fine-tuning bert model for materials named entity recognition, in: 2021 IEEE International Conference on Big Data (Big Data), IEEE, 2021, pp. 3717–3720.

[7] J. D. M.-W. C. Kenton, L. K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, in: Proceedings of naacL-HLT, volume 1, 2019, p. 2.

[8] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, Roberta: A robustly optimized bert pretraining approach, arXiv preprint arXiv:1907.11692 (2019).