A SAMPLE OPINION ANALYSIS SYSTEM: SENTIMENT ANALYSIS APPROACH

Murat Küçük & Tuğberk Varoğlu & Yusuf Atmaca
Department of Computer Engineering
İzmir Institute of Technology
Urla, İzmir 35433, TR
{muratkucuk,tugberkvaroglu,yusufatmaca}@std.iyte.edu.tr

ABSTRACT

This study focuses on analyzing the sentiment of online news articles across multiple languages. The objective was to classify news summaries as positive or negative while addressing challenges related to multilingual data processing and ensuring the accuracy of results. Summaries were collected from diverse sources and refined to improve data quality. Translation into a common language was performed to streamline analysis, followed by the identification of relevant entities and sentiment classification.

The experimental results demonstrated variations in performance, emphasizing the importance of data quality and additional training for improving outcomes. The study highlights the effectiveness of creating a structured pipeline for multilingual sentiment analysis, providing valuable insights into handling diverse datasets and improving classification accuracy.

1 INTRODUCTION

The rapid growth of digital content and online news has created an abundance of multilingual textual data, posing significant challenges for natural language processing (NLP) applications. Among these challenges is the need for efficient and accurate opinion analysis systems capable of handling texts in various languages. *These systems play a crucial role in understanding public sentiment*, enabling decision-making in fields such as media analysis, business strategy, and public policy.

This project aims *to design and implement a sample opinion analysis system to predict the sentiment polarity—positive or negative*—of online news articles. By focusing on news data from different sources and leveraging the advancements in *neural machine translation* (NMT) and *sentiment analysis*, we address key challenges in handling diverse languages and extracting meaningful insights from textual content.

The system comprises four main phases.

- **Online News Data Collection**: A diverse dataset of news articles is collected from different sources in *English*, *French*, and *German*, to reflect the multilingual nature of online media.
- **Neural Machine Translation**: The collected articles are *translated into Turkish*, ensuring the preservation of semantic and syntactic nuances, while accurately identifying entities and words.
- **Sentiment Analysis**: The translated Turkish texts are analyzed *to determine their sentiment polarity*, classifying them as *positive* or *negative*. Scalability is emphasized during this phase to enable efficient processing of large datasets.
- **Method**: *Five state-of-the-art models* from the *HuggingFace (2024)* library are explored to implement both *zero-shot* and *few-shot* learning techniques, providing insights into their effectiveness and adaptability.

This project establishes a robust and scalable multilingual opinion analysis pipeline while evaluating the performance of advanced neural models in real-world applications. By integrating multilingual data processing, machine translation, and sentiment analysis, the system demonstrates the practical implementation of state-of-the-art NLP techniques for sentiment prediction. Through comprehensive experimental analysis, the project emphasizes the impact of zero-shot and few-shot learning techniques in handling low-resource languages like Turkish. The findings contribute to the growing field of sentiment analysis in a multilingual context, offering actionable insights for future research and development in global opinion analysis systems.

The *methodology* section describes the models, datasets, and techniques used for translation and sentiment analysis. The *experimental work* section details the setup, results, and error analysis. Finally, the *conclusion* highlights the findings and provides insights for future work, particularly in applying NLP techniques to low-resource languages.

2 METHOD

Our project utilizes a pipeline-based approach to analyze the sentiment polarity of online news summaries in Turkish. The process involves translation, named entity recognition (NER) tagging, and sentiment analysis. Below, we describe the selected language models, their components, and the implementation specifics for each phase.

The first step in our pipeline was to collect news summaries from various open-source RSS sources in English, French, and German. We extracted the summaries from the prepared RSS feeds using the `feedparser` library. We chose summaries instead of titles because they provided more descriptive and detailed sentences, making them more suitable for sentiment analysis. Once collected, named entity recognition (NER) was applied to the summaries in their original languages to improve dataset quality by identifying and retaining entries with at least one named entity, such as persons, organizations, or locations. By filtering out entries without NER tags, we ensured that low-quality data was excluded before the translation phase, thereby reducing unnecessary translation costs. For NER tagging, we used `spaCy` for German and English. However, for French, instead of relying on `spaCy` for sentence separation, we employed a manual regex-based approach. This decision was made because French uses a wide variety of punctuation marks, which often led to incorrect sentence splits. The custom regex was designed to split sentences at typical sentence-ending characters, such as periods, exclamation marks, and question marks, ensuring better sentence segmentation.

Once the dataset was refined, we used the `facebook/nllb-200-distilled-600M` model to translate the remaining summaries into Turkish. This model, a distilled version of NLLB-200, is specifically designed for multilingual translation tasks across 200 languages. Its architecture leverages an encoder-decoder transformer, making it well-suited for low-resource languages like Turkish. The model employs `SentencePiece` tokenization, which handles subword-based tokenization to manage out-of-vocabulary words effectively. During implementation, translation quality was evaluated using the BLEU metric, comparing outputs from the model with Google Translate outputs.

After translation, we performed NER tagging on the Turkish dataset using the `akdeniz27/bert-base-turkish-cased-ner` model. This model is a fine-tuned version of `dbmdz/bert-base-turkish-cased`, trained on a curated version of a well-known Turkish NER dataset. Its evaluation was conducted using test sets proposed in the paper "Küçük, D., Küçük, D., Arıcı, N. 2016. A Named Entity Recognition Dataset for Turkish." The model demonstrated impressive performance metrics across various test sets, achieving an overall accuracy of 99.61%, precision of 97.20%, recall of 95.16%, and F1-score of 96.17%. This high level of accuracy ensured the reliable identification of entities such as persons, organizations, and locations in the translated Turkish dataset, maintaining the integrity of the information during the NER process.

For sentiment analysis, we chose five pre-trained Turkish sentiment analysis models:

- `savasy/bert-base-turkish-sentiment-cased`,
- `Gorengoz/bert-turkish-sentiment-analysis-cased`,
- `saribasmetehan/bert-base-turkish-sentiment-analysis`,
- `emre/turkish-sentiment-analysis`, and
- `akoksal/bounti`

Each model utilizes a `BERT`-based architecture and has been fine-tuned on specific Turkish datasets. For example, `savasy/bert-base-turkish-sentiment-cased` was trained on a merged dataset of *Turkish movie reviews*, *product reviews*, and *tweets*, comprising 48,290 entries. Similarly, `saribasmetehan/bert-base-turkish-sentiment-analysis` was fine-tuned on the `winvoker/turkish-sentiment-analysis-dataset` with a training set of 10,000 entries and a test set of 2,000 entries, achieving an accuracy of 96.2% on the evaluation set.

3 Experimental Work

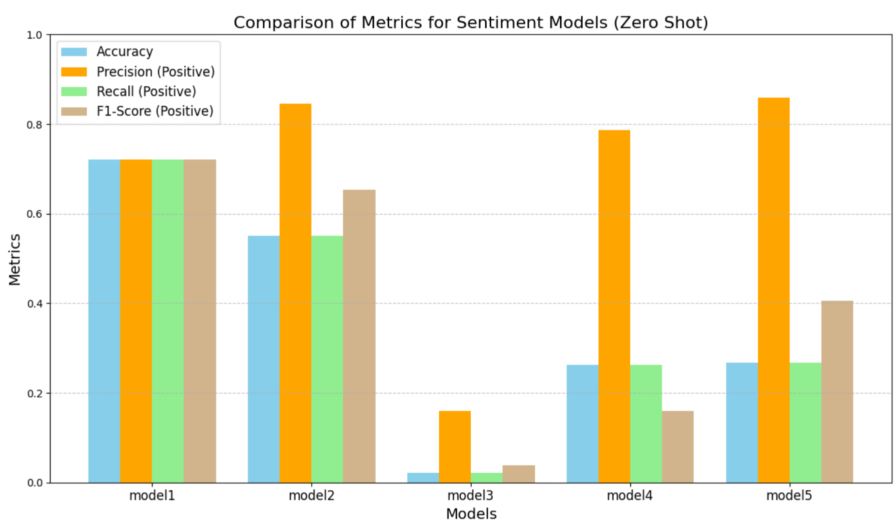
To conduct sentiment analysis on multilingual news articles, we structured the project into several key phases. We began by collecting *5,031* news articles from open-source RSS feeds in English, French, and German, ensuring diversity in content. These articles underwent a filtering process to improve dataset quality. Using Named Entity Recognition (NER) tagging, we selected *900* entries from each language that contained at least one recognized entity. This step ensured the relevance of the data, particularly for downstream sentiment analysis tasks.

After filtering, we manually labeled *225* entries across the three languages—*78* from German, *76* from French, and *71* from English—as positive or negative for evaluation purposes. During this process, entries lacking meaningful textual content, such as "Von Hans-Joachim Vieweger," were manually removed to maintain dataset integrity. Post-filtering, the dataset comprised *776* entries in English (*49* negative, *22* positive), *797* in German (*46* negative, *32* positive), and *799* in French (*58* negative, *18* positive).

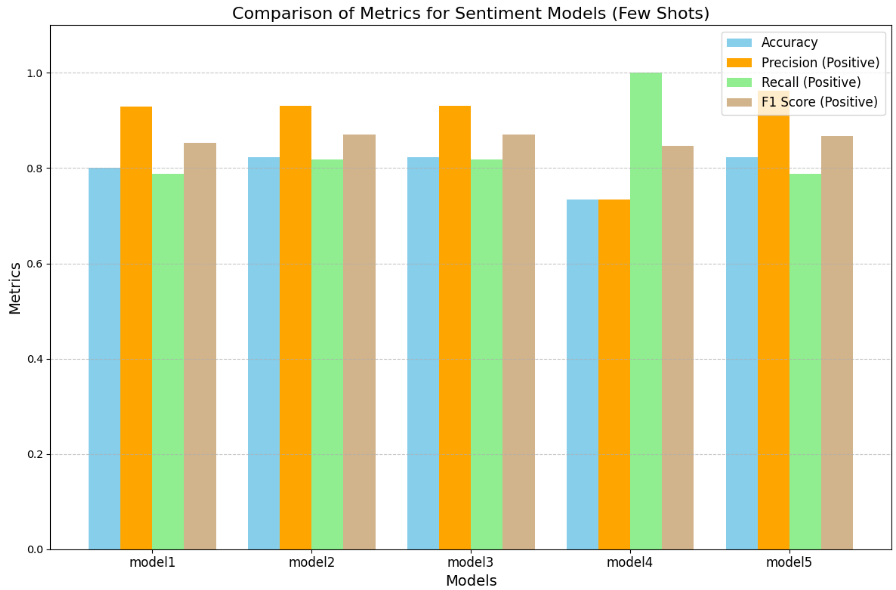
The translation quality was evaluated using the BLEU metric, using *Google Translate outputs* as a reference. The BLEU scores achieved were *11.81 for German*, *18.31 for English*, and *18.60 for French*, with a combined BLEU score of *27.50*. These scores indicate reasonable translation accuracy for multilingual data.

Key metrics such as *accuracy*, *precision (positive)*, *recall (positive)*, and *F1-score (positive)* were calculated for each model. Model 1 demonstrated the highest overall accuracy of 0.72, followed by Model 2 with 0.55. However, Model 3 exhibited very low performance across all metrics, suggesting its ineffectiveness for zero-shot sentiment analysis in this context.

Model 1 also outperformed others in terms of positive precision and F1-score, while Models 4 and 5 showed moderate performance. These results highlight the variability in performance among pre-trained models and the importance of selecting the right model for specific tasks.



To evaluate the effect of few-shot learning, we fine-tuned the five models using 45 selected entries from the manually labeled dataset. The fine-tuned models were then tested on the remaining 210 entries. The few-shot results, visualized in the attached bar chart, show significant improvements in performance across most metrics compared to the zero-shot setup.



Model 1 retained its superior performance, achieving the highest accuracy and F1-scores for both positive and negative sentiments. Model 2 also demonstrated a substantial increase in accuracy and recall, indicating its adaptability to few-shot learning. The remaining models exhibited varying levels of improvement, with Model 3 showing the most notable gains relative to its zero-shot performance, though it remained the weakest overall.

The error rates for the zero-shot setup revealed that Model 1 had the lowest error rate (0.28), while Model 3 had the highest (0.98). The higher error rates in Models 3, 4, and 5 suggest that these models may require more extensive fine-tuning or are less compatible with the dataset's characteristics. In the few-shot setup, error rates decreased significantly for all models, demonstrating the effectiveness of fine-tuning in improving performance.

4 Conclusion and Discussion

The results indicate that pre-trained sentiment analysis models can achieve varying degrees of success depending on their architecture and training data. Few-shot learning proved highly effective in enhancing model performance, particularly for less effective zero-shot models. The study highlights the importance of dataset quality, model selection, and the use of supplementary training data for improving sentiment analysis tasks.

In future work, we plan to expand the dataset by incorporating additional languages and explore the impact of larger fine-tuning datasets on model performance. The integration of domain-specific data may further improve the models' applicability to real-world sentiment analysis tasks.

Overall, this project offers a clear example of how to combine NMT, NER, and sentiment analysis into a cohesive pipeline capable of handling multilingual data. By identifying key challenges and providing targeted solutions—such as subword merging for NER—it sets the stage for more advanced multilingual NLP applications aimed at extracting sentiment and other valuable signals from ever-growing global media content.

References

HuggingFace. Hugging face: State-of-the-art machine learning for nlp, vision, and more. https://huggingface.co, 2024. Accessed: 2024-12-30.