COSE474-2024F: Final Project Proposal
"Sentiment analysis using text"

2022320149 Subeen Ho

1. Introduction

In recent years, the understanding and management of emotional states have become critical concerns, especially in the context of mental health. Many individuals struggle to accurately recognize and interpret their own emotions, which can lead to emotional distress and contribute to mental health challenges. Sentiment analysis, specifically focused on detecting and analyzing emotions in textual data, provides a valuable tool for improving emotional awareness. By leveraging sentiment analysis models, it is possible to assess an individual's emotional state and detect signs of prolonged negative sentiment, which may indicate underlying psychological issues.

This project aims to utilize sentiment analysis to monitor and evaluate emotional states, particularly focusing on the persistence of negative emotions in textual data. By identifying sustained negative sentiments, we seek to better understand emotional trajectories and offer insights into possible interventions for emotional well-being.

2. Motivation

The motivation for this project arises from the growing global concern about mental health issues. According to the World Health Organization (WHO), mental health conditions have risen sharply worldwide, with one of the key factors being the inability of individuals to recognize and understand their own emotions. Studies have shown that self-awareness of emotional states is directly linked to better mental well-being and that individuals who can identify and regulate their emotions are less prone to emotional distress.

Despite the increasing need for tools that help individuals monitor and manage their emotional states, many existing methods focus primarily on providing support for people already experiencing mental health issues. By detecting persistent negative sentiments in everyday text, we hope to provide a tool for early identification of potential emotional struggles. This tool could enable individuals to become more aware of their emotions and intervene before negative feelings escalate into more serious psychological issues.

3. Problem definition & Challenges

The goal of this sentiment analysis project is to detect and classify sentiments expressed in text as either positive or negative. However, unlike traditional sentiment analysis, this project aims to assess the persistence of negative emotions. If an individual expresses sustained negative sentiments in their text over time, it may indicate a more profound emotional or psychological concern, such as depression.

The challenges associated with this project include:

1. Contextual Sentiment Analysis

   Words or phrases can have varying meanings depending on the context in which they are used. A critical challenge is to accurately interpret the sentiment expressed by an individual, especially in cases where negations or sarcasm may alter the perceived sentiment. For example, phrases like "I'm not unhappy" could be misclassified without careful consideration of the context.

2. Handling Persistent Negative Sentiments

   Analyzing the persistence of negative sentiment is crucial to understanding the emotional trajectory of individuals. The challenge lies in identifying whether negative sentiments are transient or sustained over time. This requires an analysis that not only identifies individual sentiments but also tracks sentiment shifts across multiple interactions or over extended periods.

I aim to use the KoBERT model to infer the meaning of words within their contextual usage.

4. Concise description of contribution

The main contribution of this project lies in extending the domain of sentiment analysis, which has primarily been applied to negative sentiment analysis in domains such as social media posts, to the movie review dataset.

While existing sentiment analysis models have mostly focused on monitoring and evaluating negative emotions in text or posts, this study applies it to the NSMC dataset (movie reviews), tracking persistent negative sentiments within reviews and utilizing this information to identify potential signs of a user's psychological state or mental health issues. This approach opens up the possibility of expanding sentiment analysis techniques beyond movie reviews to various other text analysis domains. Additionally, by identifying the persistence of negative emotions, this research provides an opportunity for early detection of users who may need mental health support or preventative intervention.

5. Methods

1. pseudo code

   - Load Dataset
     dataset = loaddataset(nsmc)
     tokenizer = BertTokenizer.frompretrained(”bert-base-multilingual-cased”)
   - Preprocess Data

     function preprocessfunction(examples):
     Tokenize text using tokenizer
     Generate time-series data
     Return tokenized inputs, time-series, and labels

   - Define Models
     Baseline Model: BertForSequenceClassification
     Time-Series Model: Custom BertWithTimeSeries model
     Time-Series with EarlyStopping Model: Same as time-series, with EarlyStoppingCallback

   - Define Training
     baselinetrainingargs = TrainingArguments(...)
     timeseriestrainingargs = TrainingArguments(...)
     earlystoptrainingargs = TrainingArguments(...)

   - Train Models
     baselinetrainer.train()
     timeseriestrainer.train()
     earlystoptrainer.train()

   - Evaluate Models
     Evaluate each model on the test dataset
     Compute metrics: accuracy, precision, recall, F1

   - Visualize Results
     Compare metrics across models (baseline, time-series, early-stopping) using a bar chart

2. Model Architecture

   We use KoBERT, a BERT-based transformer model pre-trained on Korean text, to leverage its language understanding capabilities for sentiment analysis. KoBERT is fine-tuned on the NSMC dataset, which consists of movie reviews labeled with either positive or negative sentiments.

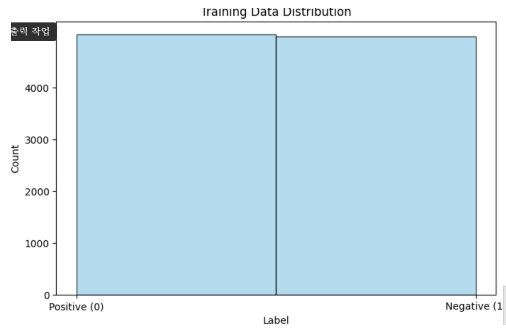

Figure 1. data distribution

3. Data Preprocessing

   (a) Tokenization
   Text is tokenized using BertTokenizer, converting each movie review into a sequence of tokens that the model can process.

   (b) Time-series Data
   For each review, we simulate time-series data (representing user behavior over time) and include it as an additional input feature. This time-series data helps track whether negative sentiment persists over multiple reviews from the same user.

   The figure illustrates the distribution of positive and negative reviews within the dataset. As shown, the dataset is well-balanced, with an almost equal number of positive and negative samples, indicating that the data is not biased towards either sentiment. This balance ensures a fair representation of both sentiment classes, contributing to the robustness of the sentiment analysis model.

4. Sentiment Classification

   The core of the model is a sentiment classification task, where the goal is to predict whether a given review is positive or negative based on the text content. The sentiment is predicted using cross-entropy loss.

5. Evaluation Metrics

   The model is evaluated using traditional metrics like accuracy, precision, recall, and F1 score. Additionally, we assess the sustainability of negative emotions by calculating how often users' reviews are negative over time, considering this as a potential signal for emotional distress.

6. Experiments

1. dataset

   The NSMC dataset contains movie reviews labeled as either positive or negative, providing a foundation for training sentiment analysis models. While movie reviews might seem a limited scope for sentiment analysis, they offer rich, emotionally charged language that is useful for training models designed to detect emotions.

   The choice of the NSMC dataset was driven by several factors:

   (a) Large-scale data
   With a substantial amount of labeled data, the dataset offers sufficient examples to train robust sentiment classification models.

   (b) Relevance to emotional analysis
   Movie reviews, like other forms of textual content, often reflect deep emotional engagement, which is valuable for detecting sustained emotional patterns.

2. computer resource

   The model was trained using Google Colab with an L4 GPU, which allowed for efficient computation during the training of the sentiment analysis model. The training was carried out using the Hugging Face Transformers library, which integrates seamlessly with PyTorch for fast GPU-accelerated training.

3. experimental design

   The project utilized KoBERT, pre-trained for the Korean language, to perform sentiment classification. The model was fine-tuned using the NSMC dataset. We experimented with two model configurations:

   (a) baseline model
   A basic binary sentiment classification model trained without incorporating any time-series information or advanced techniques like early stopping.

   (b) Time-series model
   A sentiment analysis model that incorporates time-series data, aiming to capture sentiment trends over multiple data points.

   (c) This configuration builds on the previous one, adding the early stopping technique to prevent overfitting and improve generalization by halting training when validation loss stops improving.

4. Quantitative results

   (a) Baseline model
   The performance of the baseline model without any modifications is presented in Figure 1.

Epoch	Training Loss	Validation Loss	Accuracy	Precision	Recall	F1
1	0.371600	0.342376	0.850140	0.850298	0.850140	0.850109
2	0.300600	0.322350	0.861740	0.863556	0.861740	0.861607
3	0.224700	0.336594	0.868800	0.868800	0.868800	0.868800

Figure 2. Baseline model

   (b) Time-series model
   The time-series model includes temporal features derived from review timestamps, enabling the model to capture sentiment trends over time. The performance of the model trained with these additional time-series features is shown in Figure 2.

Epoch	Training Loss	Validation Loss	Accuracy	Precision	Recall	F1
1	0.354200	0.329521	0.853640	0.853706	0.853640	0.853624
2	0.286700	0.311739	0.869460	0.869671	0.869460	0.869453
3	0.207800	0.332441	0.872680	0.872745	0.872680	0.872667

Figure 3. Time-series model

   (c) Time-series with early stopping
   The time-series model with early stopping incorporates temporal features and halts training when performance on the validation set stops improving, preventing overfitting. The results of this model are shown in Figure 3.

Epoch	Training Loss	Validation Loss	Accuracy	Precision	Recall	F1
1	0.353100	0.330382	0.854700	0.854708	0.854700	0.854701
2	0.292400	0.311581	0.866360	0.866476	0.866360	0.866358
3	0.205000	0.347778	0.869560	0.869720	0.869560	0.869534

Figure 4. Time-series with early stopping

   (d) Final model
   The model was trained on a dataset of 10,000 samples and evaluated using 5,000 samples. Each review was randomly generated by one of 500 users. This model focuses on identifying persistent negative sentiment, enabling an analysis of how user-generated content can reflect ongoing emotional states.

Epoch	Training Loss	Validation Loss	Accuracy	Precision	Recall	F1
1	0.495000	0.499970	0.774400	0.775655	0.774400	0.774232
2	0.487900	0.475220	0.786200	0.790763	0.786200	0.785193
3	0.337900	0.479863	0.806200	0.806254	0.806200	0.806204

Figure 5. Fianl model anlaysis

5. Qualititative results

   The figure 6 compares the performance of the baseline model, time-series model, and time-series model with early stopping, evaluated across four metrics:

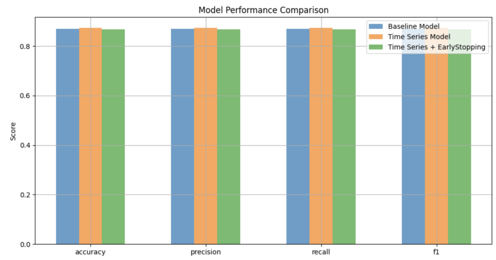

Figure 6. Performance comparison of the baseline model, time-series model, and time-series model with early stopping across accuracy, precision, recall, and F1 score.

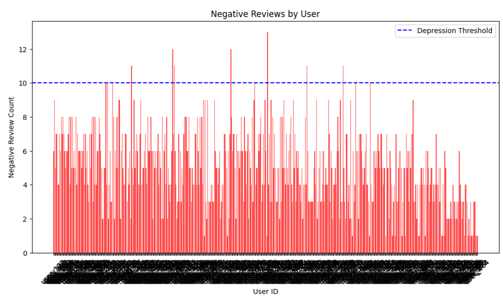

Figure 7. Figure X: Number of negative reviews per user. The blue dashed line represents the depression threshold

accuracy, precision, recall, and F1 score. As shown, the time-series models consistently outperform the baseline model, indicating the benefits of incorporating time-series data in sentiment analysis tasks.

The Figure 7 above illustrates the number of negative reviews written by each user, with the x-axis representing user IDs and the y-axis showing the count of negative reviews. The blue dashed line indicates the depression threshold, identifying users with consistently high numbers of negative reviews.

By applying the trained sentiment analysis model to classify reviews as positive or negative, this analysis highlights users whose negative review counts exceed the threshold. These users may exhibit patterns of sustained negative sentiment, offering potential insights into their emotional states. This result demonstrates the model's capability to detect sustained negative sentiment across user activity effectively.

6. Discussion

While the model demonstrated promising results in sentiment classification, there are several factors that may have limited its performance. First, the choice of dataset, NSMC, was suitable for binary sentiment analysis, but alternative datasets could potentially yield better results. Given that the dataset predominantly consists of movie reviews, it may not fully capture the nuances of other forms of textual data, such as social media posts, which could provide more varied sentiment expressions.

Additionally, the difference in performance between the baseline, time-series, and time-series with early stopping models was relatively minimal. While the time-series models showed slightly better results, the improvement was not as significant as anticipated. This suggests that the inclusion of time-series data may not have had a major impact on sentiment classification for this specific dataset. Moreover, the computational resources available were limited, preventing further experimentation with increased epoch numbers. An extended training duration could have potentially led to better model convergence, thus enhancing overall performance.

7. Future direction

1. Extended Dataset

   Beyond the current dataset (NSMC), future work could incorporate a wider variety of emotion analysis datasets. For instance, data from social media, news articles, or user reviews could be used to further enhance the model's performance. This would help improve the model's ability to generalize to different types of text and user demographics.

2. Multimodal Analysis

   Currently, the model analyzes text alone, but future research could explore multimodal data, combining text with images or speech to enhance emotion detection. For example, analyzing a user's profile picture or voice feedback could contribute additional information for more accurate emotion classification.

3. Real-time Sentiment Analysis

   Future work could also focus on real-time sentiment analysis, enabling the model to analyze user reviews or social media data as it is generated. This would allow for the tracking of emotional changes or the persistence of negative emotions, providing faster insights for emotional health management.

8. references

Sun, Chi, et al. "BERT Post-Training for Review Reading Comprehension and Aspect-based Sentiment Analysis." Proceedings of NAACL-HLT 2019.

Vaswani, Ashish, et al. ”Attention Is All You Need.” Advances in Neural Information Processing Systems (NeurIPS), 2017.