Advanced Machine Learning course Project: Tone Tag text Classifier

Aleksandr Vashchenko
a.vashchenko@innopolis.university
Grigoriy Nesterov
g.nesterov@innopolis.university
Danil Shalagin
d.shalagin@innopolis.university

1 Introduction

In the realm of digital communication, the ability to accurately interpret the emotional nuances and underlying subtext of text messages is of paramount importance. This is particularly true for neurodiverse individuals, who may encounter difficulties in discerning these subtleties. The current project is dedicated to the development of a tool that can assist in this process, thereby significantly enhancing the communication experience by providing a deeper understanding of the emotional context of messages.

The project explores the application of tone tags for text classification, leveraging advanced neural network architectures that are capable of learning in a context-aware manner. The goal extends beyond merely classifying the text into broad categories such as positive or negative. Instead, it aims to identify the specific tone or emotion conveyed in the text, thereby providing a more nuanced understanding of the message. This report provides a comprehensive overview of the motivation behind the project, the methodology employed, the experiments conducted, and the results obtained, thereby offering a detailed insight into this innovative endeavor.

2 Motivation

The motivation for this project stems from a noticeable gap in the field of text classification. While there are numerous methods available for classifying text, there is a distinct lack of solutions specifically designed to assist neurodiverse individuals in understanding the emotional nuances and subtext of messages. This project aims to address this gap by developing a tool that can classify texts based on tone tags, thereby providing a more nuanced understanding of the emotional context of messages.

Existing solutions, such as binary text classification into negative and positive classes or multi-class classification into a range of emotions, do not adequately cater to the unique challenges faced by neurodiverse individuals. These methods often fail to capture the subtleties of emotional nuances in text, leading to a lack of precision in their classifications. By developing a tool that can accurately discern these nuances, we aim to enhance the communication experience for neurodiverse individuals, thereby addressing a significant need in the field of digital communication.

3 Related Work

The field of text classification and sentiment analysis has been widely explored in the realm of natural language processing. Various methods, ranging from traditional machine learning techniques to advanced deep learning models, have been employed to classify text into different categories or sentiments. However, most of these methods focus on broad classifications, such as positive, negative, or neutral sentiments, and do not delve into the nuances of emotional tones within the text.

For instance, Pang and Lee (2008) proposed a method for movie review sentiment analysis using machine learning techniques [3]. Similarly, Socher et al. (2013) introduced recursive deep models for semantic compositionality over a sentiment treebank [4]. While these works have significantly contributed to the field of sentiment analysis, they do not specifically address the challenges faced by neurodiverse individuals in understanding emotional nuances and subtext in messages.

On the other hand, some studies have focused on emotion detection in text. For example, Mohammad (2012) presented a method for generating a lexicon for detecting emotions in text [2]. However, these methods often classify text into a predefined set of emotions and do not consider the wide range of tones that can be expressed in a message.

In this project, we aim to bridge this gap by developing a tool that can classify texts based on tone tags. This approach allows for a more nuanced understanding of the emotional context of messages, thereby catering to the specific needs of neurodiverse individuals.

4 Methodology

The methodology of this project was crafted to efficiently classify text based on tone tags. The process encompassed multiple stages, commencing from data acquisition to the application of machine learning models.

Initially, data is gathered from Tumblr. Subsequently, preprocessing steps are applied, including the removal of duplicates and null values, cleaning, filtering for English text only, and an attempt to mitigate the impact of class imbalance and use some Word Sense Disambiguation (WSD) algorithms.

Subsequently, the preprocessed data was inputted into three distinct neural network architectures: RNN, LSTM, and Transformer [5]. These models were selected for their capability to grasp contextual information in text, essential for understanding the emotional nuances present.

The methodology was meticulously devised to ensure the tool's efficacy in text classification based on tone tags, facilitating a deeper comprehension of the emotional context within messages. This methodology holds particular significance for neurodiverse individuals, aiding them in deciphering emotional subtleties within text.

5 Data Preparation

5.1 Data Collection

The first step was to gather data, which was done by parsing posts from the social media platform Tumblr. A list of 40 tone tags and their variations was compiled, which was used to filter and search posts within the Tumblr API based on the platform's internal hashtags.

Table 1. Tone Tags and Explanations

Tone tag	Explanation	Tone tag	Explanation
/a	affectionate	/nm	not mad
/c	copypasta	/npa	not passive aggressive
/cb	clickbait	/nsb	not subtweeting
/f	fake	/nsrs	non-serious
/g, /gen	genuine	/nsx, /nx	non-sexual
/genq	genuine question	/p	platonic
/hf	half joking	/pa	passive aggressive
/hyp	hyperbole	/pos, /pc	positive connotation
/ij	inside joke	/q	quote
/j	joking	/r	romantic
/l, /ly	lyrics	/ref	reference
/lh	light-hearted	/rh, /rt	rhetorical
/li	literal	/s	sarcastic
/lu-a, /lu	little upset	/srs	serious
/m	metaphorical	/x, /sx	sexual intent
/nav	not a vent	/t	teasing
/nbh	nobody here	/th	threat
/neg, /ng	negative connotation	/iron	ironic
/neu	neutral connotation	/nay	not at you
/nf	not forced	/np	nothing personal

The parsed dataset consists of 379191 unique rows with the following contents:

Table 2. Raw Dataset

Field	Description
id	Unique numeric identifier of the row
timestamp	Date the post was written
URL	URL of the post
blogName	The name of the blog where the post is written
title	Post title
tags	Post hashtags
text	Post text

5.2 Data Preprocessing

The raw dataset underwent several transformations to prepare it for the machine learning models. This included removing rows with missing values and duplicates, retaining only columns containing tags and text, selecting target tags for the project, standardizing all variations into a unified format, filtering out non-English texts using the langdetect library, applying the autocorrect library to correct spelling errors in the text, and balancing the dataset.

5.3 Imbalancing

Because of class imbalance and misinterpretation of certain classes, the decision was made to remove classes with fewer than 1000 instances, as well as the "quote" and "literal" classes.

5.4 Word Sense Disambiguation (WSD)

Two WSD algorithms were applied to the datasets. The first algorithm, nltk.wsd.lesk, determined the definition of each word based on the context in the sentence and replaced the word with its definition. The second algorithm, based on the glove dataset, determined the context vector for each sentence [1]. For the second algorithm, we utilized glove.6B.50d embeddings, followed by glove.twitter.27B.50d for another instance of the dataset and compare them.

5.5 Results of Data Preparation

The output is a dataset comprising 19 tags and 83786 unique rows. Each row consists of tags and text data tokenized into sheets, with some datasets including an additional column of context vectors.

The tag "lyrics" has the maximum number of rows. The data is unbalanced. Rows have not undergone manual processing, which does not exclude inaccuracies in text writing and incorrect interpretation of tone tags.

6 Experiments with models and Data

The experiments conducted in this project involved the application of three different neural network architectures: RNN, LSTM, and Transformer. These models were chosen for their ability to learn context-aware representations, which is crucial for understanding the emotional nuances in text.

Each model was trained and evaluated on the six datasets prepared in the data preparation stage. This allowed us to compare the performance of the models across different preprocessing techniques and word sense disambiguation (WSD) algorithms.

The RNN model, despite its simplicity, is capable of processing sequential data, making it suitable for text classification tasks. However, it often struggles with long sequences due to the vanishing gradient problem.

The LSTM model, an extension of the RNN, addresses this issue with its memory cell, which can maintain information in memory for long periods of time. This makes it particularly effective for tasks that involve understanding the context in text.

The Transformer model, on the other hand, uses self-attention mechanisms to weigh the relevance of each word in the text for the task at hand. This allows it to capture both the local and global context in the text, making it highly effective for understanding emotional nuances.

The experiments were designed to evaluate the effectiveness of these models in classifying texts based on tone tags. The results of these experiments are discussed in the following sections.

7 Evaluation

Following experimentation with models and data, the optimal parameters for neural networks were determined. It should be noted that these parameters are significantly constrained by hardware limitations.

Table 3. Compare NN Parameters

Parameter	LSTM	RNN	Transformer
optimizer	Adam	Adam	Adam
loss function	CrossEntropy	CrossEntropy	CrossEntropy
layers number	4	2	3
learning rate	5e-4	1e-5	5e-4
dropout	0.2	0.2	0.2
sequence length	4096	2048	1024

The performance of the models was evaluated using several metrics, including accuracy, precision, recall, and F1-score. These metrics provided a comprehensive understanding of how well the models were able to classify texts based on tone tags.

- **Accuracy** measures the proportion of total predictions that are correct. It is a useful metric when the classes in the dataset are balanced.
- **Precision** measures the proportion of positive predictions that are actually correct. It is a key metric when the cost of a false positive is high.
- **Recall** measures the proportion of actual positives that are correctly identified. It is crucial when the cost of a false negative is high.
- The **F1-score** is the harmonic mean of precision and recall, providing a balance between the two metrics.

Using the majority of the datasets led to an undesirable outcome, where the model consistently predicted only one class, typically the one with the highest number of rows in the dataset (16.8%). An acceptable result was defined as when the model provided predictions that varied across classes and performed better than random chance ($\frac{1}{19}$). The most successful outcomes were achieved on datasets processed with the second WSD algorithm, which generated a context vector for each sentence. Metrics are taken with argument top_k=3. Accuracy weighted.

Best performance metrics for each architecture:

Table 4. Results of Neural Networks

Model	Sweetspot epoch	Acc.	Precision	Recall	F1-Score
LSTM	11	0.526	0.152	0.526	0.224
Transformer	5	0.497	0.141	0.497	0.208
RNN	7	0.425	0.084	0.413	0.109

The RNN, LSTM and Transformer models demonstrated promising results, with their predictions varying across classes and performing better than random chance. This indicates that these models have learned to discern the different tone tags to some extent.

The evaluation results provided valuable insights into the effectiveness of the models and the impact of different preprocessing techniques and WSD algorithms on their performance.

8 Analysis and Observations

Upon analyzing the results, several key observations were made. Firstly, the LSTM and Transformer models demonstrated superior performance compared to the RNN model. This could be attributed to their ability to capture both the local and global context in the text, which is crucial for understanding emotional nuances.

Secondly, the choice of preprocessing techniques and word sense disambiguation (WSD) algorithms had a significant impact on the performance of the models. The datasets processed with the second WSD algorithm, which generated a context vector for each sentence, yielded the most successful outcomes.

However, it was also observed that using the majority of the datasets led to an undesirable outcome, where the model consistently predicted only one class. This highlights the importance of having a balanced dataset for training the models.

Despite these challenges, the models were able to provide predictions that varied across classes and performed better than random chance. This indicates that the models have learned to discern the different tone tags to some extent, thereby demonstrating the feasibility of the project.

As a result of resource limitations and successful parameter selection, the LSTM-based machine learning model architecture demonstrated the best performance.

9 Deployment

The decision was made to deploy the project using the LSTM model, the results of which were mentioned earlier. Streamlit's local version was chosen for deployment due to the large size of the LSTM model that needed to be uploaded. The deployed application consists of a textarea where users can input the text they want to evaluate. The model then processes the input and provides the three most probable evaluations of the text. Users can click on the button corresponding to the desired result, and the text with the corresponding tone tag is copied to the clipboard. The outcome of the model's operation can be observed in the figure below.

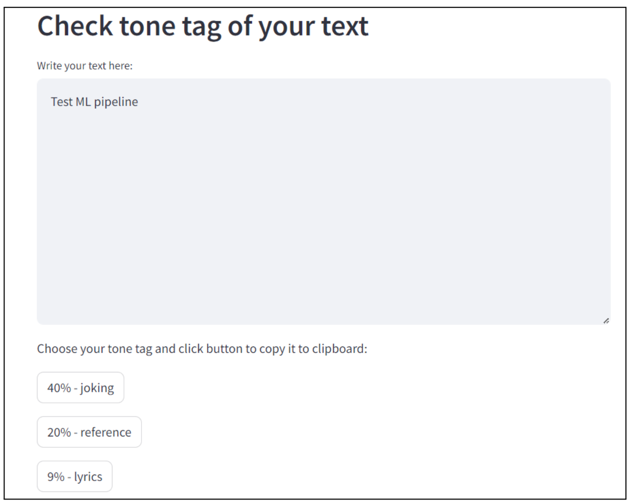

Figure 1. Using an application example

10 Conclusion

In conclusion, this project demonstrates the feasibility of developing a tool that can classify texts based on tone tags. The LSTM and Transformer models, in particular, showed promising results, providing predictions that varied across classes and performed better than random chance. This indicates that these models have learned to discern the different tone tags to some extent.

However, the project also highlighted several challenges, such as the need for a balanced dataset and the impact of different preprocessing techniques and word sense disambiguation (WSD) algorithms on the performance of the models. These insights provide valuable directions for future work in this area.

The project has significant implications for enhancing the communication experience for neurodiverse individuals, who may find it challenging to discern emotional nuances in text. By developing a tool that can accurately discern these nuances, we aim to address a significant need in the field of digital communication.

A demonstration of our project can be found at the following link: [https://github.com/abobafett-dev/advanced-machine-learning-project-iu-2024].

To enhance the project, the following options should be considered:

- Further training the existing LLM model.
- Creating a balanced dataset with manually processed text.

References

[1] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeffrey Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*. 2013.

[2] Saif M Mohammad. #emotional tweets. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics-Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation*. Association for Computational Linguistics, 2012.

[3] Bo Pang and Lillian Lee. Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, 2, 2008.

[4] Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D Manning, Andrew Ng, and Christopher Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, 2013.

[5] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*. 2017.