Sasi Kishore Varma

Sentiment Analysis

1 February, 2025

+ Code + Markdown

Sentiment Analysis Study

In this study, I have explained **Sentiment Analysis** in detail.

Key Steps in the Study:

1. **Dataset Selection:**
    - A sample dataset was chosen to illustrate the concepts of sentiment analysis on a real-world example.
2. **Data Analysis & Visualization:**
    - Conducted a detailed analysis of the dataset and visualized key findings.
3. **Preprocessing:**
    - Prepared the data by applying necessary preprocessing techniques.
4. **Modeling:**
    - Completed the sentiment analysis task using **state-of-the-art models**.
5. **Result Analysis:**
    - Analyzed and interpreted the

Introduction

SENTIMENT ANALYSIS

Table of Contents

Types of Sentiment Analysis

- **Emotion Detection**
- **Multilingual Sentiment Analysis**
- **Graded Sentiment Analysis**
- **Aspect-based Sentiment Analysis**
- **Intent Analysis**

Why Is Sentiment Analysis Important?

Types of Sentiment Analysis

- **Emotion Detection**
- **Multilingual Sentiment Analysis**
- **Graded Sentiment Analysis**
- **Aspect-based Sentiment Analysis**
- **Intent Analysis**

Why Is Sentiment Analysis Important?

The overall benefits of sentiment analysis include:

- **Sorting Data at Scale**
- **Real-Time Analysis**
- **Discovering New Marketing Strategies**

How Does Sentiment Analysis Work?

Sentiment Analysis Approaches

- **Rule-based Approaches**
- **Automatic Approaches**

EDA

- Information of the DATA
- Information of the Problem
- Imports
- Helper Functions
- Read Data
- Visualizations
- Word Cloud
- Target Count
- Token Counts with simple tokenizer
- Token Counts with BERT tokenizer
- Characters Count in the Data
- Reviews Lengths
- Word Counts
- Most Common Words
- Most Common ngrams

MODELS

- A brief information about BERT
- A brief information about XLNET
- A brief information about RoBERTa
- Comparison of Transformer Models

Preprocess for BERT Train

Preprocess for BERT Train

- Train and Validation Split
- BertTokenizer and Encoding the Data
- Creating the Model
- Data Loaders
- Optimizer & Scheduler
- Performance Metrics
- Training Loop
- Test on validation set

ERROR ANALYSIS

INFERENCE

REFERENCES

1. SENTIMENT ANALYSIS

...

Sentiment Analysis

- Sentiment analysis (or opinion mining) is a natural language processing (NLP) technique used to determine whether data is positive, negative, or neutral. Sentiment analysis is often performed on textual data to help businesses monitor brand and product sentiment in customer feedback and understand customer needs.
- Sentiment analysis helps data analysts within large enterprises gauge public opinion, conduct nuanced market research, monitor brand and product reputation, and understand customer experiences. In addition, companies often develop sentiment analysis systems for customer experience management, social media monitoring, or workforce analytics platforms to better understand their own customers.

Types of Sentiment Analysis

Sentiment analysis is aimed at determining the general emotional state of a text. One of these cases focuses on the polarity of a text (positive, negative, neutral) but it also goes beyond polarity to detect specific feelings and emotions (angry, happy, sad, etc.), urgency (urgent, not urgent) and even intentions (interested vs. not interested).

- Let's explain them in more detail

Types of Sentiment Analysis

Sentiment analysis is aimed at determining the general emotional state of a text. One of these cases focuses on the polarity of a text (positive, negative, neutral) but it also goes beyond polarity to detect specific feelings and emotions (angry, happy, sad, etc.), urgency (urgent, not urgent) and even intentions (interested vs. not interested).

- Let's explain them in more detail

Empty markdown cell, double-click or press enter to edit.

Emotion Analysis

- The type of emotion analysis in which emotion types (happiness, frustration, anger, and sadness) are classified is called emotion detection.
- There are some difficulties with this classification. Users can express their feelings with many different words. They can use a word with a bad meaning for happiness. The most difficult examples of classification models here are; For example, the sentence "I connect to customer service too late, it's killing me" is a negative sentence, while the sentence "you are killing me" is positive.

Multilingual Sentiment Analysis

It is the version of Sentiment Analysis systems that provides multi-language support. What is mentioned here is to do sentiment analysis in more than one language.

I usually have two suggestions for this:

My first suggestion is to detect the language of the text with the language classifier and run a sentimen analysis model suitable for this language. The second method is to develop a Multilingual language model and finetune this model and make the model work in many languages.

Graded Sentiment Analysis

If the precision of the mood is important, the categories can be further elaborated. A broader classification can be made, not just positive and negative:

- **Very Positive**
- **Positive**
- **Neutral**
- **Negative**
- **Very Negative**

This classification is often used in reviews and ratings where 5 stars are awarded.

- **Very Positive** = 5 stars
- **Very Negative** = 1 star

Aspect-based Sentiment Analysis

- Generally, when analyzing the emotions of the texts, the focus is on determining whether the comment/opinion is positive or negative. But we do not focus on what is positive or negative in this text.
- To put it more clearly, in the expression "I did not like the product at all, the size is too small", the user is not satisfied with the product and complains about its dimensions. In a normal sentiment analysis, this sentence is classified as negative, but in aspect-based sentiment analysis, the "the size is too small" part is also focused on.

Intent Analysis

Intent analysis focuses on what the user wants to do. Understanding what the user wants to do will allow us to better guide them.

For example, being able to understand that a customer browsing an e-commerce site has a shopping intention also allows us to offer them the right products.

One of the most used areas is in smart assistant systems within applications. It allows us to direct users to the right places within the application in line with their requests, thereby offering a better user experience.

Why Is Sentiment Analysis Important?

Sentiment analysis plays a crucial role in helping businesses understand customer opinions and sentiments. By analyzing customer feedback, reviews, and social media

Why Is Sentiment Analysis Important?

Sentiment analysis plays a crucial role in helping businesses understand customer opinions and sentiments. By analyzing customer feedback, reviews, and social media posts, companies can gain valuable insights into customer experiences, preferences, and pain points.

This allows businesses to improve their products, services, and marketing strategies, ultimately leading to enhanced customer satisfaction, loyalty, and overall brand reputation.

People now share their comments/emotions on social media, e-commerce sites, and many other platforms. A lot of data is created on these platforms.

Often, brands want to know what customers are talking about. Brands/companies make great efforts to quickly identify their customers' expectations and provide them with the right service. This allows them to understand what makes customers happy or disappointed, so they can tailor products and services to meet customer needs.

In addition, brands want to observe the impact of their advertisements on users.

For these reasons, sentiment analysis is becoming more important every day.

The overall benefits of sentiment analysis include:

Sorting Data at Scale

Users make a lot of comments about brands, and it is almost impossible to process them manually. Sentiment analysis enables businesses to automatically classify large amounts of raw data.

The overall benefits of sentiment analysis include:

Sorting Data at Scale

Users make a lot of comments about brands, and it is almost impossible to process them manually. Sentiment analysis enables businesses to automatically classify large amounts of raw data.

Real-Time Analysis

Companies can learn the wishes of their customers by analyzing social media comments in real time. They can identify angry customers and ensure their satisfaction.

Discovering New Marketing Strategies

With more data and information gathered through sentiment analysis, organizations can develop effective marketing strategies. The outcome from the strategies can be measured by customers' positive or negative key messages.

By observing customers' conversations on social media and detecting specific key messages related to your brand, specific marketing campaigns can be designed for target consumers.

How Does Sentiment Analysis Work?

Sentiment analysis works by analyzing text data and determining the sentiment expressed in the text. This can be done using various techniques, including:

1. **Natural Language Processing (NLP):** NLP techniques are used to process and analyze the text. It helps in tokenizing the text, removing stop words, and analyzing sentence structures.

How Does Sentiment Analysis Work?

Sentiment analysis works by analyzing text data and determining the sentiment expressed in the text. This can be done using various techniques, including:

1. **Natural Language Processing (NLP):** NLP techniques are used to process and analyze the text. It helps in tokenizing the text, removing stop words, and analyzing sentence structures.
2. **Machine Learning Models:** Algorithms like Naive Bayes, Support Vector Machines, and deep learning models are trained to recognize sentiment in text based on labeled data (positive, negative, neutral).
3. **Lexicons and Dictionaries:** These are pre-built sets of words and phrases associated with specific sentiment values (positive or negative) that can be used for sentiment classification.

By applying these techniques, sentiment analysis can determine whether a text expresses a positive, negative, or neutral sentiment.

Sentiment analysis works to automatically determine emotional tone thanks to natural language processing (NLP), rule-based methods, and machine learning algorithms.

There are different ways we can do sentiment analysis, depending on how much data you need to analyze, how accurate your model needs to be, and how many resources you have.

We will talk about some of them below.

Sentiment Analysis Algorithms

Sentiment analysis algorithms fall into one of three buckets:

Sentiment Analysis Algorithms

Sentiment analysis algorithms fall into one of three buckets:

- **Rule-based**: These systems automatically perform sentiment analysis based on a set of manually crafted rules.
- **Automatic**: Systems rely on machine learning techniques to learn from data.

Rule-based Approaches

Usually, a rule-based system tries to help determine the subjectivity of the sentence, the polarity, or the subject matter of an idea. The most used tool here is "regex".

These rules usually include the following two NLP techniques:

- Stemming, tokenization, part-of-speech tagging, and parsing.
- Lexicons (i.e. lists of words and expressions).

The working mechanism of these systems is briefly as follows:

- Build a list of polarized words (e.g. bad-good, worst-best, ugly-beautiful, etc.). You can find them as open source.
- The ratio of positive and negative words in a sentence.

Rule-based approaches are now obsolete and not used as much as they used to be. Rule-based approaches fail to detect ironies or exactly how users are feeling. For this reason, automated approaches are gaining more importance now.

Automatic Approaches

These systems don't rely on manually crafted rules, but on machine learning techniques, such as classification. Classification, which is used for sentiment analysis, is an automatic system that needs to be fed sample text before returning a category (e.g. positive, negative, or neutral).

Automatic Approaches

These systems don't rely on manually crafted rules, but on machine learning techniques, such as classification. Classification, which is used for sentiment analysis, is an automatic system that needs to be fed sample text before returning a category (e.g. positive, negative, or neutral).

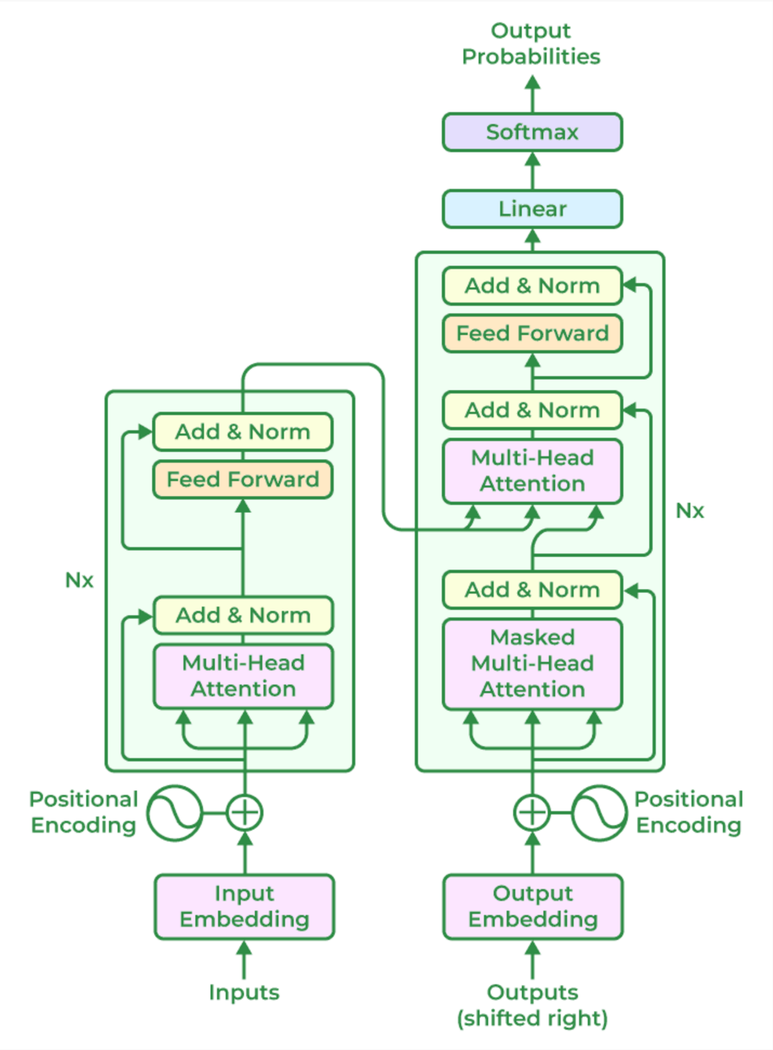
Here's how a machine learning classifier can be implemented:

- **Classification Algorithms**: The classification step usually involves a statistical model like Naïve Bayes, Logistic Regression, Support Vector Machines, or Neural Networks.
- **Naïve Bayes**: A family of simple "probabilistic classifiers" based on applying Bayes' theorem with strong (naïve) independence assumptions between the features (se Bayes classifier).
- **Linear Regression**: A linear approach for modeling the relationship between a scalar response and one or more explanatory variables (also known as dependent an independent variables).
- **Support Vector Machines (SVM)**: A supervised machine learning algorithm that can be used for classification or regression problems. It is mostly used in classification problems. The Support Vector Machine is a boundary that best separates two classes (hyperplane/line).
- **Deep Learning**: (also known as deep structured learning) is part of a broader family of machine learning methods based on artificial neural networks with representation learning. Learning can be supervised, semi-supervised, or unsupervised.

Transformer Architecture:

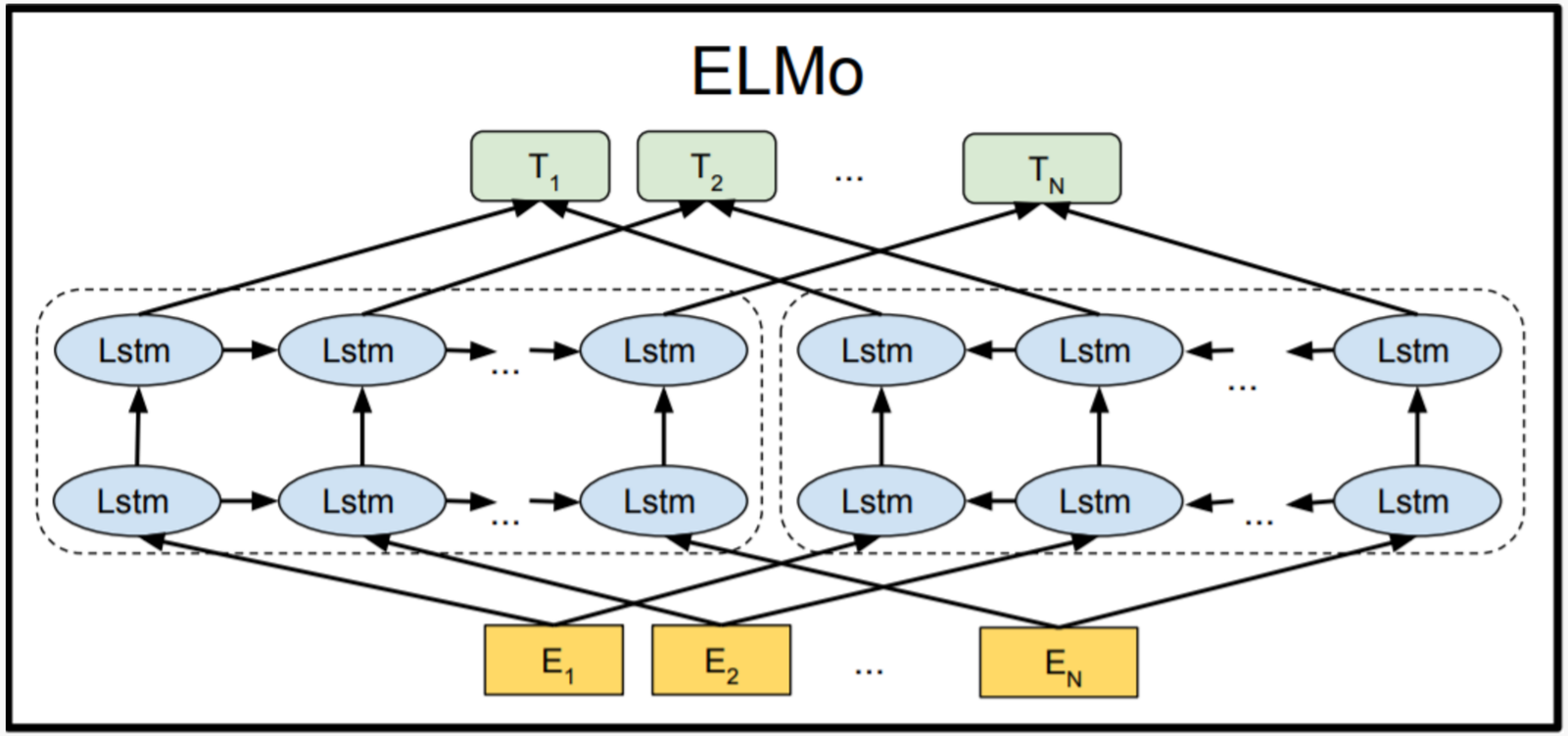
High-Level Overview

- The transformer model is built on an encoder-decoder architecture, where both the encoder and decoder are composed of a series of layers that utilize self-attention mechanisms and feed-forward neural networks. This architecture enables the model to process input data in parallel, making it highly efficient and effective for tasks involving sequential data.
- In a transformer model, the encoder processes the input sequence and generates a set of continuous representations. These representations are then fed into the decoder, which produces the output sequence. The encoder and decoder work together to transform the input into the desired output, such as translating a sentence from one language to another or generating a response to a query.



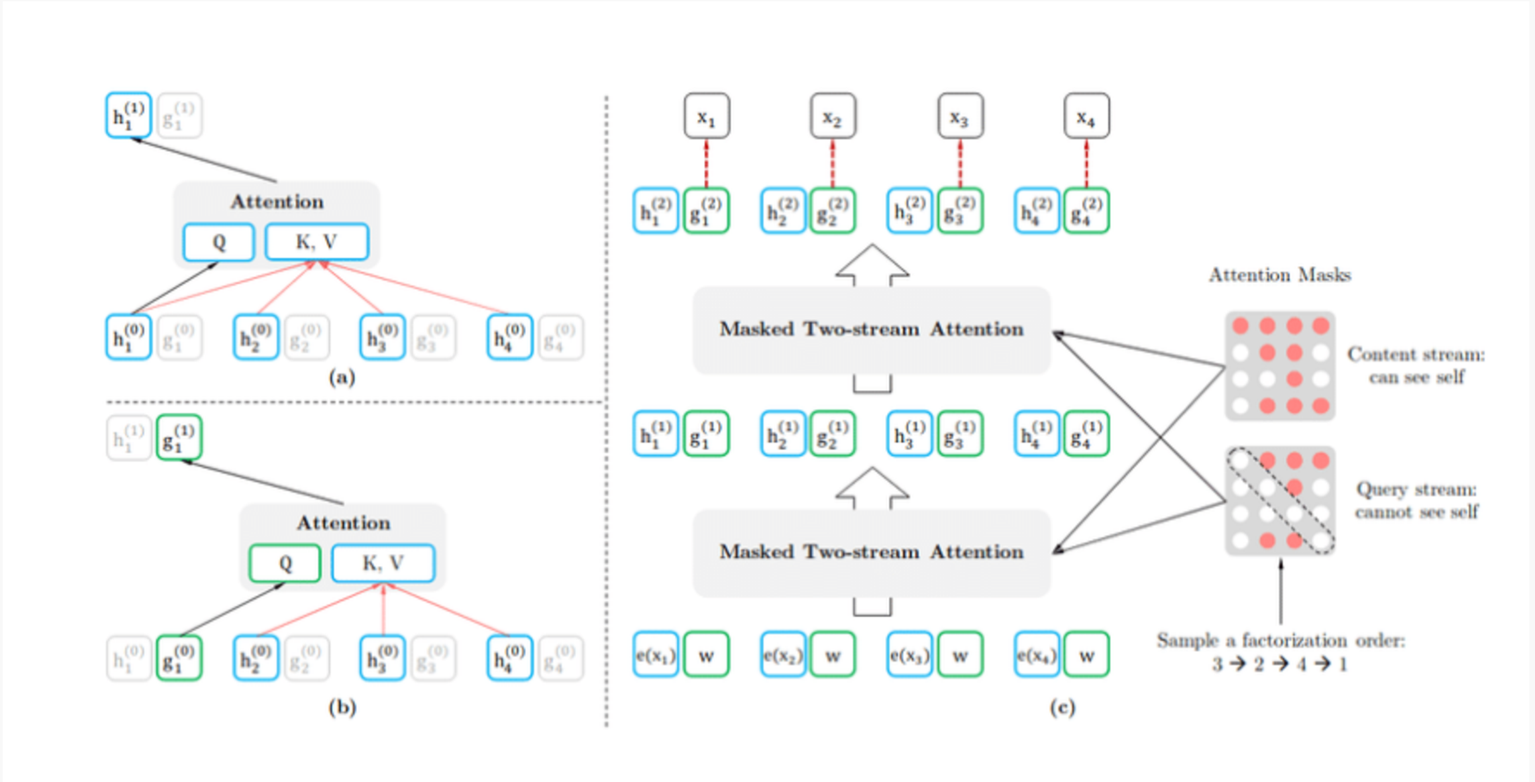
What is BERT?

- BERT stands for Bidirectional Encoder Representations from Transformers. It is designed to pre-train deep bidirectional representations from unlabeled text by jointly conditioning on both left and right context. As a result, the pre-trained BERT model can be fine-tuned with just one additional output layer to create state-of-the-art models for a wide range of NLP tasks.



Architecture of XLNet

- XLNet comprises input embeddings, multiple Transformer blocks with self-attention, position-wise feedforward networks, layer normalization, and residual connections. Its multi-head self-attention differs by allowing each token to attend to itself, enhancing contextual understanding compared to other models.



Thank you