

Chapter: 07

SENTIMENT ANALYSIS AND NLP MODELS AS THE LANGUAGE OF DATA

Mohd Hyder Gouri*

*Faculty, Glocal School of Science and Technology,
Glocal University, Saharanpur, U.P.*

*Correspondence to: hyder@theglobaluniversity.in

Mr. Mohit Kumar

*Faculty, Glocal School of Science and Technology,
Glocal University, Saharanpur, U.P.*

DOI: <https://doi.org/10.52458/9789388996747.nsp2023.eb.ch-07>

Ch.Id:-GU/NSP/EB/EFMLDSP/2023/Ch-07

ABSTRACT

NLP models and sentiment analysis are two core facets of natural language processing (NLP) that are examined in this chapter. It starts by exploring the several approaches used to create NLP models, such as semi-supervised learning, supervised learning, unsupervised learning, and the crucial function of deep learning. It then delves into the field of sentiment analysis, illuminating its importance, variety of approaches, and practical uses. These subjects taken together offer a thorough overview of the rapidly changing field of NLP and its wide variety of applications.

Keywords: *Sentiment Analysis, Deep Learning, Supervised Learning, Unsupervised Learning, Semi-Supervised Learning, Natural Language Processing (NLP)*

7.1 INTRODUCTION

The capabilities of Natural Language Processing (NLP) models have greatly influenced how we interact with and comprehend human language in the age of artificial intelligence and big data. The fundamentals of building supervised, unsupervised, and semi-supervised learning as well as the revolutionary potential of deep learning are covered in Chapter-6.

The emphasis in this chapter moves to sentiment analysis, a crucial area of NLP. Opinion mining, or sentiment analysis, is a technique that makes it possible to extract and understand feelings and viewpoints from textual data. This chapter also delves into the complexities of sentiment analysis, including its methods and wide range of applications.

7.2 LITERATURE REVIEW

The foundation for comprehending and interpreting human language is provided by NLP models. There is a lot of research on this subject that is always being done. NLP supervised learning techniques include named entity recognition and text classification. To create models that can perform these tasks well, methods like Support Vector Machines (SVM), Naive Bayes, Convolutional Neural Networks (CNN), and Recurrent Neural Networks (RNN) are used.

Word embeddings, which make it easier to comprehend the semantic associations between words, and topic modeling, which finds latent patterns in textual data, are two examples of unsupervised learning in natural language processing. This field has benefited greatly from methods like Word2Vec, GloVe, and Latent Dirichlet Allocation (LDA).

Both supervised and unsupervised learning components are included in semi-supervised learning. Techniques like co-training and self-training use labeled and unlabeled data to optimize model performance while minimizing the requirement for large labeled datasets.

The literature on sentiment analysis investigates techniques for interpreting text's emotional undertones. Rule-based methods make use of pre-established lexicons and patterns, while machine learning methods include Naive Bayes, Support Vector Machines, and deep learning models such as Transformers and Recurrent Neural Networks (RNNs). Aspect-based sentiment analysis breaks down sentiment into particular facets or characteristics, providing more in-depth understanding of consumer opinions and product acceptance.

7.3 OVERVIEW OF OPENCV AND COMPUTER VISION

In NLP, supervised learning is giving models labeled data to train them to comprehend and produce human language. The input text and matching output labels – which are frequently produced by text classification, sentiment analysis, named entity recognition, or machine translation, to mention a few – make up this labeled data.

a. Classification of Texts

A basic NLP task is text classification, which is classifying a supplied text into predetermined categories or labels. Applications like sentiment analysis, subject classification, and spam email detection frequently employ this.

Machine learning methods like Support Vector Machines (SVM), Naive Bayes, or deep learning models like Convolutional Neural Networks (CNN) or Recurrent Neural Networks (RNN) are commonly used for text classification.

Identification of Named Entities (NER): Identified Party The process of locating and categorizing certain items (such as names of individuals, groups, or places) inside a text is known as recognition. Information retrieval, answering questions, and many other applications requiring the extraction of structured data from unstructured text depend on this.

Conditional Random Fields (CRF) and Deep Learning Methods like transformers and bidirectional long short-term memory (LSTMs) are frequently used to build NER models.

b. Unsupervised NLP Learning

In NLP, unsupervised learning pertains to the comprehension and manipulation of text without the use of labeled data. When labeled data is hard to come by or unavailable, this method is especially helpful. Unsupervised natural language processing (NLP) employs clustering, word embeddings, and topic modeling.

- **Inserting Words**

Vector representations of words in a continuous vector space are called word embeddings. By capturing the semantic connections between words, these representations enable NLP models to comprehend the context of individual words within a text. Some well-liked word embedding methods include Word2Vec, GloVe, and fastText.

- **Modeling Topics**

One technique to find latent topics in a set of documents is topic modeling. It is extensively utilized for content recommendation, summarization, and document clustering. For topic modeling, Latent Dirichlet Allocation (LDA) is a popular algorithm.

- c. **Partially Supervised NLP Education**

In natural language processing, semi-supervised learning integrates aspects of supervised and unsupervised learning. It makes use of both a significant amount of unlabeled data and a modest amount of annotated data. Multi-view learning, co-training, and self-training are methods used in semi-supervised natural language processing.

- **Educating Oneself**

Self-training entails utilizing a model that has been iteratively trained on a small labeled dataset to label unlabeled data. The training dataset is subsequently expanded with these newly labeled instances, providing more labeled data for subsequent model training.

- **Joint Education**

Multiple data views or features are used in co-training to enhance model performance. It swaps labeled instances to improve each other's performance and trains models independently on various feature sets.

- d. **NLP's Deep Learning**

NLP has been transformed by deep learning, which has made it possible to create strong models for a variety of NLP tasks. Transformers and other deep learning models have produced state-of-the-art outcomes in tasks including sentiment analysis, text production, and language translation.

- **Convertibles**

One class of deep learning models that has become well-known in NLP is called Transformers. Their ability to extract contextual information from input sequences through self-attention mechanisms makes them a great choice for machine translation jobs (e.g., BERT, GPT, and T5).

- **Knowledge Transfer**

In NLP, transfer learning entails pre-training models on sizable corpora and optimizing them for particular tasks. Significant gains in NLP performance have resulted from this method's large labeled dataset reduction requirement.

7.4 COMPREHENDING EMOTION

Understanding the nature of sentiment and the reasons behind its importance in NLP is crucial before diving into sentiment analysis techniques and applications.

a. What Is Sentiment?

Sentiment pertains to the sentiments, thoughts, and emotions conveyed inside a written piece. These feelings might be anything from happiness, love, and enthusiasm to despair, rage, and contempt. It is also possible for sentiment to be neutral, which denotes the lack of strong positive or negative feelings.

b. Why Examine Emotion?

Sentiment analysis in text is important for a number of reasons:

- **Business Insights:** By examining social media sentiment and customer reviews, businesses can obtain insightful information. They can better grasp consumer perceptions of their goods and services thanks to this data.
- **Brand Reputation:** Maintaining and defending a brand's reputation requires constant sentiment monitoring. Unfavorable opinions can damage a business's reputation and reduce sales.
- **Sentiment Analysis:** It is a useful tool for assessing customer happiness and pinpointing potential areas of deficiency in customer service. Sentiment analysis is often used in the political sphere to determine public perception of political figures and programs.
- **Social listening:** Sentiment analysis is a tool used by people and organizations to monitor a range of subjects and conversations on news and social media platforms.

7.5 METHODS OF SENTIMENT ANALYSIS

Sentiment analysis includes a broad range of approaches, from sophisticated machine learning algorithms to rule-based methodologies.

Here are a few essential methods:

- a. **Sentiment Analysis Based on Rules:** Rule-based sentiment analysis determines sentiment using pre-established rules and lexicons. It entails compiling lists of terms and expressions linked to various emotions. The sentiment of a text containing these words or phrases is inferred from their frequency and context of occurrence.
- b. **Sentiment Analysis Using Machine Learning:** Algorithms for machine learning techniques in sentiment analysis are trained on labeled data. Popular techniques include Support Vector Machines (SVMs), Recurrent Neural Networks (RNNs), and Transformers from deep learning models, as well as Naive Bayes. In order to predict sentiment, these models examine text and take into account word order, context, and other linguistic elements.
- c. **Sentiment Analysis Based on Aspects:** Aspect-based sentiment analysis dissects sentiment into certain characteristics of a good, service, or text rather than just categorizing sentiment as a whole. Businesses can utilize this information to determine whether aspects are viewed more favorably or unfavorably.

7.6 USES OF SENTIMENT ANALYSIS

Applications for sentiment analysis may be found in many different fields, and it provides useful insights and automation features.

- a. **Analysis of Customer Feedback:** To better understand consumer happiness, pinpoint areas for development, and improve the customer experience, businesses examine customer feedback.
- b. **Monitoring Social-Media:** Sentiment analysis is widely utilized in tracking public opinion on a variety of issues, including brands, goods, and current events, on social media platforms.
- c. **Management of Brand Reputation:** Businesses monitor online mentions of their brand and take proactive measures to reduce bad sentiment by using sentiment analysis.
- d. **Analysis of Financial Markets:** By examining news and social media data, sentiment analysis is a tool used by financial analysts and investors to estimate market sentiment and forecast market moves.
- e. **Reviews of Goods and Services:** Sentiment analysis is used by e-commerce platforms to sort and present product reviews, assisting users in making wise selections.

The basics, methods, and many applications of sentiment analysis have all been covered in this chapter. Sentiment analysis is a flexible tool with broad implications across many fields since it is essential to comprehending and responding to the feelings and viewpoints conveyed in text. Sentiment analysis will only get better and more significant as NLP techniques advance.

7.7 CONCLUSION

Sentiment analysis and NLP models are complex fields that are essential to information extraction, comprehension, and decision-making in many different fields. Sentiment analysis is a useful tool for customer feedback analysis, social media monitoring, and company reputation management, while natural language processing and understanding are made possible by NLP models.

Both sentiment analysis and NLP models will likely see additional advancements and improvements as technology develops. These developments will increase their precision and expand their uses, making them essential instruments in our data-driven society.

REFERENCES

1. Schütze, H., Raghavan, P., and Manning, C. D. (2008). *Information Retrieval Overview*. Cambridge University Press, Cambridge, UK.
2. Martin, J. H., and D. Jurafsky (2019). *Third Edition of Speech and Language Processing*. Pearson.
3. Lee, L., and B. Pang (2008). *analysis of sentiment and opinion mining*. *Information retrieval: Foundations and Trends®*, 2(1-2), 1-135.
4. B. Liu (2012). *analysis of sentiment and opinion mining*. *Human Language Technologies Synthesis Lectures*, 5(1), 1-167.
5. Khan, M. T., Durrani, M., Ali, A., Inayat, I., Khalid, S., & Khan, K. H. (2016). *Sentiment analysis and the complex natural language*. *Complex Adaptive Systems Modeling*, 4(1), 1-19.
6. Wijayanto, A., Bernardo, J. F. A., & Pamungkas, S. (2021). *Analisis Klasifikasi Kepuasan Penumpang Maskapai Penerbangan Menggunakan Algoritma Naïve Bayes: Analysis of Airline Passenger Satisfaction Classification Using the Naïve Bayes Algorithm*. *Jurnal Sains Komputer dan Teknologi Informasi*, 3(2), 97-103.
7. Dongre, M. S. C. S. (2019). *E-Mail Spam Classification Using Long Short-Term Memory Method*.