VIEW POINT

THE SENTIMENT ANALYSIS PERSPECTIVE ON TEXT ANNOTATION

Abstract

Machine learning (ML) models gather information from data sets. One of these data sets is text. ML models read, comprehend, and analyse information from all the text data sources. You can annotate content from e-commerce websites, customer surveys, online reviews, emails, social media accounts, blog posts, chatbots, and many other sources. This POV explores how text data annotation leads to a deep understanding of public sentiment analysis, as well as ways to collect text data, the five types of text annotations, and general measures to consider for sentiment analysis.

Infosys®
Navigate your next

What is text data annotation?

Text data annotation refers to tagging and labelling text data, including keywords, sentences and phrases, proper nouns, intentions, and sentiments. The purpose of text data annotation is to develop the natural language processing (NPL) capabilities of ML models for easy and natural interaction with AI.[1] NLP, at its core, is an attempt to understand human speech and respond with the help of ML and AI. And, with more and more businesses relying on AI for different business functions – such as contact centre automation and customer service – you need to annotate more text data to power these ML projects.[1][2]

Definition of sentiment analysis

Sentiment analysis (or opinion mining), in NLP parlance, assesses the emotions expressed in a text. It helps analyse customer feedback, product reviews, survey responses, or comments across different sources. It offers insights into customers' feelings – whether those feelings are positive, negative, or neutral – which allows businesses to modify, tweak, and improve their services.[*]

- **Sentiment scoring**

Polarity classification is an integral part of sentiment analysis. It refers to the overall sentiment of the text. You can assign a numerical rating – or a 'sentiment score' – to this polarity based on either individual phrases or the entire text. You have the option to fine-grain the sentiment scoring, going beyond positive, negative, and neutral, as per specific use case. For example, you can choose 5 categories for scoring customer reviews, ranking a 1-star review as 'very negative,' a 2-star review as 'negative,' a 3-star review as 'neutral,' a 4-star review as 'positive,' and a 5-star review as 'very positive.'[3]

- **Aspect-based sentiment analysis (ABSA)**

Aspect-based sentiment analysis (ABSA) goes beyond sentiment scoring, tying sentiment analysis to specific attributes described in the text. For example, while analysing the reviews for a laptop, ABSE can assess sentiment surrounding specific features, such as processor speed or graphics.[3]

ABSA and ML

ML models focus on identifying prevalent patterns in the data; you can train them to analyse text data with high accuracy. This makes ABSA possible even if slightly varying words are present in the reviews. As a result, businesses can analyse and visualise the overall sentiments of customers, changes in sentiment over time, and aspect-based sentiments to improve their product offerings.[3]

For example, a mobile app with a rating of 4.9 out of 5 stars on the app store and external platforms constantly witnessed a drop in its subscriber base. The management was bewildered because there was no correlation between ratings and subscriber trajectory. The company discovered that the general sentiment was negative, using text-based sentiment analysis on app reviews and blogs. Only those who enjoyed the app posted a review, while all other users uninstalled it, presenting a skewed progress signal. The text annotation-based sentiment analysis helped the company further develop the app to boost its popularity and enhance customer satisfaction.

- **ABSA for real-time monitoring**

You can easily monitor customer reviews on social media and other channels in real time with the help of ABSA. This helps businesses identify the issues as they are being reported and respond immediately for a better customer experience.[3]

The significance of sentiment analysis

The power of ML and AI is that it makes human interventions redundant, as well as avoiding the limitations of human and manual operations. In the next sections, we discuss some of the main benefits of sentiment analysis.

- **Consistent analysis to remove human bias**

Sentiments are highly subjective, which use context clues, language variations, and tones to convey meaning. Our experiences and unconscious biases creep in while attempting to understand this meaning. Sentiment analysis eliminates potential human bias and errors using consistent criteria and generates accurate insights. For example, in a review, 'It is a bit expensive, but works well,' one can focus on the positive sentiment linked to the functionality, while others may focus on the negative sentiment linked to price. On the other hand, the sentiment analysis ML model would recognise two different sentiments attached to two attributes.[3]

- **Process data at scale**

With huge quantities of unstructured data available, businesses need to analyse it, and relying on human analysts is not practical. Sentiment analysis helps a business track and monitor hundreds of reviews posted every day, assess the overall polarity, and guide policies for a better customer experience.[3]

- **Save time through automation**

In addition to processing data at scale, sentiment analysis algorithms can analyse megabytes of text data in a matter of minutes instead of taking hours or days, which is case in manual analysis. Businesses can use this time for more value-added activities, such as validating insights or developing an action plan to build a better strategic position.[3]

- **Real-time analysis and insights to act faster**

Automatic sentiment analysis, with the help of ML algorithms, monitor and analyse text in real time. Such real-time insights help a business identify and resolve the issue quickly before it affects other customers.[3]

Ways to approach sentiment analysis

The automated process follows one approach to sentiment analysis while there is another traditional approach too. The two main approaches that we need to understand are rule-based sentiment analysis and automated or ML sentiment analysis.

- **Rule-based sentiment analysis**

The traditional approach to sentiment analysis, rule-based sentiment analysis, relies on a manually created set of rules. This analysis involves NLP techniques — such as lexicons, stemming, tokenisation, and parsing — along with a list of positive and negative words for analysing tokens of text to count the number of positive and negative words and calculate the overall sentiment score. However, as the rule-based analysis does not consider whole sentences, it may miss some complexities of the evolving human language. So, it needs regular reviews and updates for optimal performance.[3]

- **Automated or ML sentiment analysis**

Automated sentiment analysis relies on ML techniques and algorithms to classify sentiment. The quality of the training dataset plays a crucial role in its accuracy and success.

- **Step 1 – Feature extraction:** The first step of automated sentiment analysis is tokenising the text and transforming it into numbers via vectorisation. Recent developments in deep learning also offer the advantage of assigning a similar numeric representation to words with similar meanings for better accuracy.[3]
- **Step 2 – Training and prediction:** The next step is to input sentiment-labelled data (or training set) to the ML algorithm. This step helps the ML model associate tokenised (and vectorised) words with corresponding sentiment.[3]
- **Step 3 – Predictions:** In the final step, the model analyses a new text and predicts a sentiment label with the help of learnings from the training data. This eliminates the need for pre-defined lexicons in rule-based analysis, highlighting the key advantage of automated sentiment analysis.[3]

Collecting text data for sentiment analysis

A large-enough dataset is not sufficient for sentiment analysis applications. The quality of the data and annotation also plays a crucial role. Therefore, it is important to have clear guidelines to collect and annotate data for sentiment analysis. You can collect data for sentiment analysis by any of these methods:[6]

- Built-in APIs of social media platforms, for example, to extract tweets by hashtags via Twitter API
- Web browser plugins to extract information from public websites, such as Web Scraper, a free Google Chrome extension
- With the help of open-source repositories that clean and compile data for direct use, such as, IMDB, Yelp, or Rotten Tomatoes
- Web scrapers that collect specified information from web data, such as Beautiful Soup, a Python package that extracts information from online resources.

You need to follow some general text annotation measures for accurate sentiment analysis, such as:

- Have consistent tagging criteria that represent the problem
- Avoid inadvertent human disagreement in tagging
- Maintain inter-annotation standards[6]

Five types of text annotations

Large, annotated text datasets are necessary to train the ML models for accurate sentiment analysis. The five most common types of text annotations to build a text dataset are equity annotation, entity linking, text classification, sentiment annotation, and linguistic annotation.

- **Entity annotation**

  Entity annotation refers to locating, extracting, and tagging various entities within the body of the text and includes:

  - **Named entity recognition (NER):** Annotating entities with a proper name
  - **Key phrase tagging:** Annotating key phrases (or keywords) in the text
  - **Part-of-speech (POS) tagging:** Annotating functional elements (nouns, adjectives, verbs, and adverbs) in the text

This helps NLP models in learning to identify key phrases, named entities, and parts of speech within the text.[6]

- **Entity linking**

  Entity linking refers to connecting identified entities to a larger information pool. The two types of entity linking are:

  - **Entity disambiguation:** Linking named entities to a knowledge pool about them
  - **End-to-end entity linking:** Undertaking NER followed by entity disambiguation

This helps in improving the search functions and overall user experience.[6]

- **Text classification**

  Compared with entity annotation, text classification labels an entire block of text with a single label based on the subject matter, intent, and sentiment, which includes:

  - **Document classification:** Classifying documents to simplify sorting and recalling the text content
  - **Product categorisation:** Sorting products into intuitive categories (or classes) to improve search functions and user experience
  - **Sentiment annotation:** Classifying text based on emotion or tonality of the text[6]

- **Sentiment annotation**

  Sentiment annotation – also known as opinion mining or sentiment analysis – labels the overall emotion, tonality, sentiment, or opinion within a text body. Sentiment annotation plays a crucial role in detecting sentiments in social media posts, customer reviews, and other content to gauge public opinion and evolve business strategies accordingly.[5]

- **Linguistic annotation**

  Linguistic annotation, also known as corpus annotation, is tagging language data – grammar, phonetic elements, or semantics – in text or audio data. The four types of linguistic annotation are:Document classification: Classifying documents to simplify sorting and recalling the text content

  - **Discourse annotation:** Linking anaphors and cataphors to their antecedent or postcedent subjects
  - **POS tagging:** Annotating functional words in data
  - **Phonetic annotation:** Labelling natural pauses, intonation, and stress in speech
  - **Semantic annotation:** Annotating definitions of words

This helps in building training datasets for various NLP applications, such as chatbots, search engines, and virtual assistants.[5]

Conclusion

In a digital world that wants to make decisions based on real-time data, businesses must leverage massive amounts of data that users input on their platforms. This helps businesses in distinguishing themselves from their competitors. And with the bottom line in mind, implementing sentiment analysis solutions is important to know what customers are talking about you and evolve accordingly.

Text annotation helps you in building quality training datasets so your business can leverage accurate sentiment analysis to respond to customers in real time and gain a competitive edge.

*For organisations on the digital transformation journey, agility is key in responding to a rapidly changing technology and business landscape. Now more than ever, it is crucial to deliver and exceed organisational expectations with a robust digital mindset backed by innovation. Enabling businesses to sense, learn, respond, and evolve like a living organism, will be imperative for business excellence going forward. A comprehensive yet modular suite of services is doing exactly that. Equipping organisations with intuitive decision-making automatically at scale, actionable insights based on real-time solutions, anytime/anywhere experience, and in-depth data visibility across functions leading to hyper-productivity, Live Enterprise is building connected organisations that are innovating collaboratively for the future.

References

1. Feeding the machine: A guide to text data annotation
2. The State of AI in 2021
3. Sentiment analysis | Comprehensive beginners guide
4. Data collection and annotation measures for sentiment analysis
5. What is text annotation? 5 Different types of annotations

For more information, contact infosysbpm@infosys.com

Infosys
Navigate your next