International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 06 | June - 2022 Impact Factor: 7.185 ISSN: 2582-3930

A Review on Overcoming the Challenges of Sentiment Analysis- Tone and Framing

Author

D. Sai Satya vamsi Krishna Raju, Bharath u Jadhav, Praneeth Kumar Reddy

Abstract:

As we all know that how meteoric the social network is growing across the world. By taking this as an advantage many product-based companies have made their products supremacy. The underlaying paradigm of this growth is about the applications of natural language processing. As there were uncountable number of applications in NLP, today in this we will chew over about sentiment analysis- "precisely it's an emotion detector". We all know that the individuals express their feelings by emotions and it's a complex task to detect their emotion. In those situations, using this sentiment analysis is beneficial to find the emotion behind their expression or even behind their sentences. Not only from the emotion, but also extracting subjective information from the sentences you speak. So, what else is needed for the companies to invest after knowing about their customers emotion. Once the sentiment analysis model is defined using various machine learning algorithms, but the model later faces so many challenges which will affect the final prediction. These are the few challenges A we should more concern about- tone related issues, appropriate polarity, exact domain analysis-how to work on different domains, sarcasm behind your text and emojis- these days people more expressing their emotion using emotions than words so, the model might get confused by seeing this emojis. By these you might get know that the challenges are also the most important which we should care about. This paper will apprise you about how to overcome completely from these challenges. Mainly focuses on the tone related issues that sentiment analysis model is facing.

Keywords: *social network, natural language processing, emotion, sentiment analysis, opinion, sentiment, extracting, subjective information, prediction, tone, polarity, domain, sarcasm, apprise, analysis, automatic model and challenges.*

Introduction:

Data is an ever-growing thing in the world of information technology; it can be a collection of facts, observations, or anything else that has been digitised. Data is now measured in trillions of gigabytes. Today, data is inundating every aspect of the technological world, nearly 80% of this data is unstructured. Because the data comes from various new sources such as device logs, server logs, twitter feeds, chat data, blogs, web pages, emails, and social media content. This results in a massive collection of text data created by humans to express themselves to others, making it an important source of data that may contain valuable information. From this different kind of unstructured data there are some areas where public used to express their feelings, ratings and many things related to the products, people and so on. Among these areas there are twitter feeds and social media content which will be useful for our sentiment analysis model to train up on. This paper will concentrate on the one of the techniques of NLP which is sentiment analysis. There are wide range of domains where this sentiment analysis is used but everywhere the models which they are creating are facing

some or the other problems like which I mentioned above in the abstract. By considering this as a main point I started working on the challenges faced by the sentiment analysis model. Among all the challenges which I mentioned in the abstract I choose the “Tone problem” which is quite more problematic. As the tone plays a major role in the time of finding or extracting subjective information for the sentences you spoke. For example, the sentence "I love chocolate" is extremely positive about chocolate as a food. "I despise this new phone," for example, reveals the customer's feelings about the product. The words "love" and "hate" carry a clear sentiment polarity in these two specific cases. A more complex example is the sentence "I do not like the new phone," in which the negation turns the positive polarity of "like" into a negative polarity. The same is true for "I do not dislike chocolate," where negating a negative word like "dislike" results in a positive sentence. Here with an example, you might have understood the exact problem that sentiment analysis model faces. So, tackling this problem is the main aim and the target too. Not only the tackling of the problem but the sequence which we will be using to tackle the problem is much more important as they are not created equal. Because the tone problem it’s facing will also depends upon the domains where the sentiment analysis model is made. It depends on the domain as it works based on different applications.

Background:

As you all know that there is a booming word in today’s world which you often heard about is ‘Natural language Processing’. Day by day NLP is getting and gaining more fame across the world because of its uncountable number of applications and make use of these applications in almost all the fields. For suppose it’s being used in industries, medicine, AI based systems, transportation, software companies and technological companies and many more. All this sentiment analysis, opinion mining and information extraction works under the underlaying paradigm of natural language processing. A more recent trend in text analysis goes beyond topic detection and attempts to identify the emotion behind a text. This is known as sentiment analysis, as well as opinion mining and emotion AI. This is the definition by the author Federico Alberto Pozzi in his book named Sentiment Analysis in Social Network is that – The main aim of sentiment analysis is to create or define a automatic models that extract subjective information form the sentences in natural language, here the subjective information means the opinions and sentiments behind the sentences or the words you speak, in order to create structured and actionable knowledge to be used by decision support model which predicts the final output in the form of polarity- positive, negative and neutral. Unsurprisingly, there has been some confusion among researchers about the distinction between sentiment and opinion, leading to debate over whether the field should be referred to as sentiment analysis or opinion mining. Merriam-Collegiate Webster's Dictionary defines sentiment as "an attitude, thought, or feeling. A feeling-driven judgement, whereas an opinion is a formed view, judgement, or appraisal in one's mind about a specific subject The distinction is subtle, and each of them contains some components of the other According to the definitions, an opinion is more of a person's concrete point of view. A sentiment is more of a feeling than an opinion. For example, "I am concerned about the current political situation" expresses a feeling, whereas "I believe politics is failing" expresses an opinion. In a conversation, we can respond to the first sentence with "I share your sentiment," but for the second sentence, we would normally say "I agree/disagree with you." The underlying meanings of the two sentences, however, are inextricably linked because the sentiment depicted in the first sentence is most likely a feeling caused by the opinion in the second sentence.

This was the portfolio that the google has released based on the keyword sentiment analysis and how it is getting varied from the keyword customer feedback.

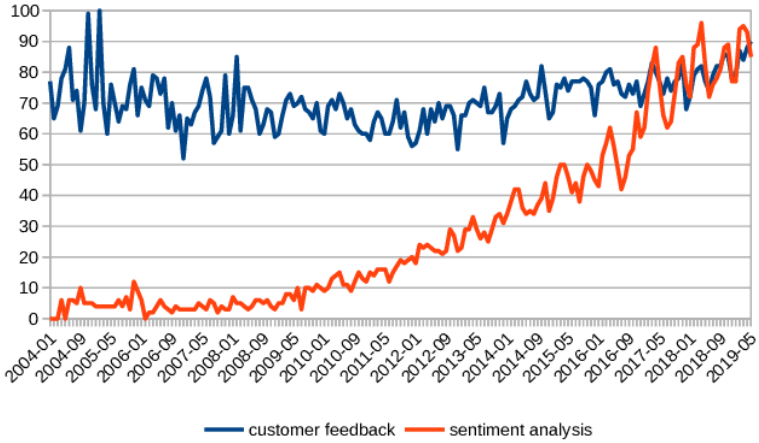

Fig1: google trends of the word strings from 2004 to 2019

From the above fig 1 you can observe that how the sentiment analysis taking over the customer feedback-which is the old technique to find the customer satisfaction on the product. You can see the evolution of sentiment analysis from 2004 to 2019 and how its overtaking the customer-feedback. At starting of the 2004 it faced so many issues because of the poor knowledge regarding the machine learning algorithms and natural language processing techniques. And later when the data science started growing and reached to peak stage then this applications like sentiment analysis had come into an act and kicked up to start. Because of its importance in business and society, sentiment analysis has spread from computer science to management science and the social sciences. In recent years, sentiment analysis-related industrial activities have flourished as well: numerous start-ups have emerged, and many large corporations have built their own in-house capabilities (for example, Microsoft, Google, Hewlett-Packard, IBM, SAP, and SAS Global Communications). Now let's get deep into the evolution of sentiment analysis or opinion mining.

Evolution of sentiment analysis:

The main requirement for any Entrepreneurs is to know about their customer's feedback which will let them either to be succeed or failure. So, it's very important for them to know about customer satisfaction details. In certain period there is an epic rise of social network which led the chance to sentiment analysis to come into the race. The general trend in sentiment analysis research in social networks is to apply techniques inherited from traditional sentiment analysis studied since the early 2000s. And there is one interesting thing which will let us know that the sentiment analysis is growing is that-In recent years, there has been a massive increase in the number of papers focusing on sentiment analysis and opinion mining. According to our data, nearly 7000 papers on this topic have been published, with 99 percent appearing after 2004, making sentiment analysis one of the fastest growing research areas. Here it is mentioned a particular year "2004", why 2004? As it was the period when the data science and natural language processing are already in research and some of the techniques are in use. Using these techniques, researchers started writing up the papers on sentiment analysis which led us to know more knowledge about this sentiment analysis or opinion mining. We discovered that the first academic studies measuring public opinion were conducted during and after WWII, and their motivation was highly political. The modern sentiment analysis explosion occurred only in the mid-2000s, and it focused on product reviews available on the Web, for example. Since then, sentiment analysis has been used in a variety of other fields, including financial market forecasting, responding to terrorist

attacks. Furthermore, research overlapping sentiment analysis and natural language processing has addressed many problems that contribute to sentiment analysis's applicability, such as irony detection and multi-lingual support. Sentiment-analysis systems are broadly classified as either knowledge-based or statistics-based. While knowledge bases were initially more popular for identifying emotions and polarity in text, sentiment analysis researchers have recently been increasingly using statistics-based approaches, with a special emphasis on supervised statistical methods. Recent studies, for example, use microblogging text or Twitter-specific features like emoticons, hashtags, URLs, @symbols, capitalizations, and elongations to improve sentiment analysis of tweets. Tang et al created a convolutional neural network-based method for obtaining word embeddings for the most used words in tweets. These word vectors were then fed into a convolutional neural network to analyse sentiment analysis.

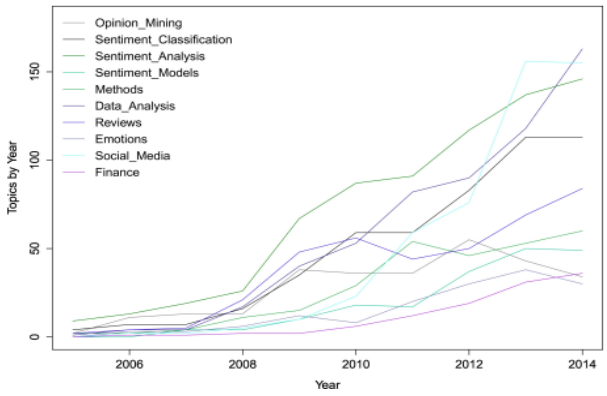

Fig.2: Evolution of the 10 identified topics

The above Figure depicts the progression of the identified topics. Most topics, like the keywords in Figure 2, show a significant increase over time. Opinion Mining and Emotions are the exceptions, but their scopes are rather limited. According to LDA, the most researched topics are Sentiment Classification, Sentiment Analysis, Data Analysis, and social media, all of which are core topics found in the keyword co-occurrence map.

A Closer Look at the Problem:

As we already discussed in the introduction about the problem which we will investigate it. But let's take a close look into the problem. There were many challenges faced by the sentiment analysis. Tone is one of the challenges that will discuss about. It is sometimes important to understand the tone of the text besides understanding what people are talking about. While predicting the sentiment or the opinion behind the sentence the model will get affected by this problem. If the model directly predicting the speech, there will be no issues as it can identify the tone behind the customers live speech. The data that will be given as input can be anything either live speech, sentences, or the paragraphs. But if the input data is sentences or the paragraphs then it will face tone problem. For suppose in some cases, if the sentences which we are about to speak are positive, but we will tell in an angry tone which leads to be an issue while the predicting the sentiment behind the sentences. Not only while extracting sentiment or opinion it will face issue while labelling the text into polarity.

Literature:

As we looked closer into the problem and how the issue is affecting sentiment analysis model and of course the result too. The only solution for this problem based on tone is to keep a "tone detector"- it can detect a wide range of emotions in the writing by using machine learning algorithms. And this tone detector should keep before the sentiment analysis model. The output from this tone detector should given as input to the sentiment analysis model. Machine learning algorithms can help to set up the tone detector. Not only the process how this tone detector works but also other considerations will be covered in this paper which includes how to choose a corpus of text documents and other applications for text classification software. The procedure:

1) The Evolution of Text Classification:

One of the first applications of text classifiers was spam detection. Blocking email addresses is ineffective when spammers can easily create new ones and blocking emails with specific keywords is ineffective when spammers do not use those keywords. Engineers began using machine learning to automate the spam detection process, developing open-source software that could also be used to detect the tone of an article.

Human readers classified messages as spam or legitimate, resulting in the creation of a corpus of labelled email messages. Using this corpus as an input in a machine learning algorithm, researchers developed a model that could automatically determine whether a message was spam or not.

Human readers can label documents in a variety of ways, including positive or negative, formal or informal, objective or subjective. The machine learning algorithm can build models that detect these features using these labelled documents.

2) Choosing Text Documents:

It is critical to choose the right documents to use for text classification. Before the machine learning algorithm can classify the text documents, they must be labelled. This step of the process may have a greater impact on classification results than using a more powerful algorithm.

It is also a step where many technology companies attempt to save money. Some companies will pay high salaries to the machine learning engineers who develop the algorithm but outsource the document tagging to a cloud platform where freelance contractors earn $4 per hour.

Alexandria Technology, an alternative data provider, was recently nominated for an award by financial publisher Benzinga for its text classification software. This company hired financial professionals to label the documents analysed by its machine learning algorithms, resulting in better performance than other fintech's' software.

Because of the environmental, social, and governance (ESG) requirements that fund managers must follow, text classification is becoming increasingly important in investing. Companies are including social responsibility reports with their financial statements and reading through all of the environmental data can take a long time.

Consumers are also discussing corporate social responsibility on social media networks, and they frequently consider a company's actions when making purchases. Using text classification and sentiment analysis software can help to accelerate ESG analysis and may even provide useful trading signals.

3) Conducting the Analysis:

There are several steps involved in configuring the text classification algorithm. The documents are first collected and added to a corpus. Tweets, newspaper articles, journal articles, text messages, and forum posts are all examples of data sources. Tweets are a popular source of real-time sentiment data.

The researchers then clean the data. This section frequently consists of stemming words, changing every word to lower case, and removing punctuation. To speed up the analysis, researchers may also remove stop words, which are commonly used words like 'the' and 'is' that do not provide unique information. Stop words, on the other hand, can be useful in certain types of text analysis, so they aren't always removed.

Because the machine learning algorithm cannot perform calculations on text, it must be converted into a vector before it can be analysed. A function called CountVectorizer in the Sklearn Python module can convert a corpus of text documents into a vector. The vector is an extremely large array of numbers.

Sklearn also has machine learning features like the Naive Bayes classifier. If you pass the vector and the list of document tags to this function, it will generate a model that can be used to classify other types of text documents that lack tags.

4) Process conclusion:

Text classification software has several applications, including adjusting an article's emotional pitch and measuring consumer and investor perceptions of a company's environmental initiatives. Using the Sklearn module, you can write your own text classification script in Python, and tutorials are available to show you how. It's a good idea to look for tutorials on Naive Bayes implementation or sentiment analysis.

However, you'll also need to collect a corpus of labelled documents, and the corpus included in a machine learning tutorial may not be suitable for a specialised task like using sentiment analysis on environmental or financial documents. If movie reviews are used as the corpus, the machine learning model will be best suited to analysing other movie reviews and may not be as effective when analysing financial documents.

So, if you're using an app or another data provider that provides features like tone detection or sentiment analysis, think about the corpus they're using as well as the quality of their machine learning algorithm.

Conclusion:

The challenges of sentiment analysis are more important, and we should have a concern about how to tackle them or to overcome them. These days so many MNC’s using sentiment analysis for their product reviews and to increase their companies value. The work presented here is only applicable to corpus (writings on specific issues) of text documents. It can be made available for different issues by creating a context dictionary for that specific issue. The work presented here is not applicable to real-time data, which may be worked out in the future. Even though sentiment analysis is a relatively new subject, a few observations can be made. The most obvious is that the field has grown at an exponential rate. This is true for most keywords,

but the keyword analysis revealed that some, such as Social Networking, have a faster growth rate than others, and some new keywords, such as Twitter, have recently gained popularity. Cooccurrence maps are useful for identifying patterns and similarities between keywords and authors. The map of frequently used keywords confirms that much emphasis has been placed on reviews and social media, but its true value lies in its utility in identifying new research opportunities and possibilities. As a result, it can assist researchers in determining where they can make the greatest contribution to the scientific community.

References:

Pozzi, Federico, et al. *Sentiment analysis in social networks*. Morgan Kaufmann, 2016.

Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. "Thumbs up? Sentiment classification using machine learning techniques." *arXiv preprint cs/0205070* (2002).

Reforgiato Recupero, Diego, et al. "Sentilo: frame-based sentiment analysis." *Cognitive Computation* 7.2 (2015): 211-225.

Liu, Bing. "Sentiment analysis and opinion mining." *Synthesis lectures on human language technologies* 5.1 (2012): 1-167.

Ahlgren, Oskar. "Research on sentiment analysis: the first decade." *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*. IEEE, 2016.

Cambria, Erik, et al. "SenticNet 6: Ensemble application of symbolic and subsymbolic AI for sentiment analysis." *Proceedings of the 29th ACM international conference on information & knowledge management*. 2020.

Poria, Soujanya, et al. "Fusing audio, visual and textual clues for sentiment analysis from multimodal content." *Neurocomputing* 174 (2016): 50-59.

Chu, Chao-Ping, Li-Te Shen, and Shaw-Hwa Hwang. "A new algorithm for tone detection." *AASRI Procedia* 8 (2014): 118-122.

Chandra, Jonathan Kevin, Erik Cambria, and Andrea Nanetti. "One belt, one road, one sentiment? A hybrid approach to gauging public opinions on the New Silk Road initiative." *2020 International Conference on Data Mining Workshops (ICDMW)*. IEEE, 2020.

Pozzi, Federico Alberto, et al. "Challenges of sentiment analysis in social networks: an overview." *Sentiment analysis in social networks* (2017): 1-11.

Ebrahimi, Monireh, Amir Hossein Yazdavar, and Amit Sheth. "Challenges of sentiment analysis for dynamic events." *IEEE Intelligent Systems* 32.5 (2017): 70-75.