Social Listening through Sentiment Analysis

Ms. Mini Ulanat
Systems Manager
Cochin University of Science and Technology
Chair WIE AG, IEEE Kerala Section
miniu@ieee.org

Abstract

Sentiment analysis, otherwise known as opinion mining is a much quoted but often misunderstood term. Sentiment is a feeling or emotion, an attitude or opinion conveyed . In this article we are looking are methods and procedures to understand the sentiment with special reference to social media. The challenges and the commonly used tools are also discussed.

Sentiment Analysis – An introduction

The exponential growth of World Wide Web and social media is generating massive volume of unstructured data through social interactions. This has increased the need for analysing the content shared over the media. This channel opened up new areas to explore by addressing the core questions in social sciences with the ability to collect and process data. Social media channels like Twitter and FaceBook engage users and influencers and this interactions have become a part of our daily life routine.

Sentiment analysis, otherwise known as opinion mining is a much quoted but often misunderstood term. This is the process of understanding the emotional tone beneath an expression, for understanding of the the attitudes, opinions and emotions expressed. Sentiment analysis is extremely useful in social media monitoring as it allows us to gain an understanding of the wider public opinion behind the topics. The ability to understand insights from social data is found helpful and is being widely adopted by organisations world wide.

Social Sentiment analysis use natural language processing (NLP) techniques to analyse social conversations online and determine deeper context as they apply to a topic, brand or theme. This computational task of automatically determining the feelings behind the expression can be a simple distinction (positive vs. negative), or can also be more fine-grained by identifying the emotion like fear, joy or anger expressed. Our net sentiment score and brand passion index show how users feel about your brand and compares across your competitors.

Like all the other fields, advances in machine learning and Deep learning in particular, has brought significant advances in the area of Sentiment Analysis also. The natural language processing methods including statistical methods and text analysis are used to extract the general sentiments into positive, negative or neutral sentiments.

Users share posts, photos, videos, comments, live events etc. through social media. Companies and organisations use this channel to promote products and engage with customers. Social sentiment analysis accept these contents like posts, tweets, status updates etc. , analyse the sentences and tries to find the sentiment to a given set of documents. The comments and reviews can be categorised as positive, negative or neutral.

Sentiment Analysis Methods

The sentiment analysis system can classified under one of these three mentioned below:

1) Rule-based systems
2) Machine learning techniques
3) Hybrid systems that combine both rule based and machine learning approaches.

Rule Based Systems

The rule based system uses a dictionary of words labelled by sentiment to determine the sentiment of a sentence. To make it better sentiment scores will be combined with additional rules to mitigate sentences containing negations, sarcasm, or dependent clauses. The rules-based sentiment analysis are very simple, hence makes it a good option for basic document-level sentiment scoring of predictable text documents.

The rules are made using inputs like:

- NLP techniques like *stemming*, *tokenization*, *part of speech tagging* and *parsing*.
- Dictionary resources like lexicons (i.e. lists of words and expressions).

A simple rule-based implementation is as follows:

- The list of defenite polarised words are defined (e.g. negative words like *bad*, *worst*, *ugly*, etc and positive words like *good*, *best*, *beautiful*, etc).
- In a document, the number of positive words and negative words that appear in the text are counted.
- The sentiment is calculated based on the count. If the count of positive word appearances is greater than the number of negative word appearances, it is treated as positive sentiment, or else it is treated as a negative sentiment. Otherwise, return neutral.

This system is very naïve as the combination of words are not considered in a sequence. But the challenge is that a rules-based system must contain a rule for every word combination in library which needs manual intervention. The advanced processing will make this system get very complex quickly. It is difficult to maintain due to addition of new rules to support new expressions and vocabulary.

Machine Learning Models

It is practically not possible to create a sentiment analysis rule set to account for every potential meaning. Hence to overcome the above said difficulties, machine learning techniques are used to automate the low-level text analytics functions including Part of Speech tagging. Machine learning models are trained to identify nouns by feeding it a large volume of text documents containing pre-tagged examples. The model learns using supervised and unsupervised machine learning techniques. This is a data driven approach using labelled corpus of their text and their sentiments to predict. Machine learning models create model by training with large amount of data

The classification step usually involves a statistical model like Naïve Bayes, Regression, Support Vector Machines, or Neural Networks like Deep Learning methods:

Hybrid Models

A combination of machine learning with traditional rules to make up for the deficiencies of each approach is Hybrid model.

Rules-based sentiment analysis is an effective way to build a basis for PoS tagging and sentiment analysis. But the disadvantage is that these rule sets quickly increase in size to become unmanageable. This is where machine learning can step in to shoulder the load of complex natural language processing tasks, such as understanding double-meanings.

Most hybrid sentiment analysis systems combine machine learning with software rules across the entire text analytics function stack, from low-level tokenization and syntax analysis all the way up to the highest-levels of sentiment analysis.

Sentiment Analysis- procedure

Information Extraction, the basic step in the processing of textual information for extracting main components and relationship. Each document is tokenised and annotated. These processed words are basic lexical units and denoted as tags. A configurable stop word list and syntactic filter is applied to to refine the selection to most relevant lexical unit. A graph of tag co-occurrences are created.

The steps involved can be describes as shown below:

Data Collection and Pre-processing

Users express their opinions at various social network sites, blogs, review sites etc. in different format.. The large temporal data is generated through social conversations. These data are from different sources and hence is heterogeneous in nature. This need to be cleaned and should be converted to structured format from unstructured format. Data cleansing includes reformatting, de-duplication, merging and filtering.

The text undergoes pre-processing task where the text is prepared for processing with the help of linguistic tools like tokenisation, sentence spitting, morphological analysis etc. The text document is broken down into its component parts like sentences, phrases, tokens and parts of speech. Part of Speech tagging is the process of identifying the structural elements of a text document, such as verbs, nouns, adjectives, and adverbs.

Sentiment Detection and Classification

Sentiment analysis, an NLP task, can be modelled as a classification problem by subjectivity classification and polarity classification. A sentence may be classified as subjective or objective. Subjective sentences express information as

opinions, judgement or speculation. Objective sentences are factual, hence may be ignored in this context. Polarity classification is the process of classifying a sentence expression as positive, negative or neutral opinion. The scope of Sentiment analysis can be at different levels – document level, sentence level or aspect level.

Challenges

The human language is complex. People express their opinions in multiple ways, which are complex at times. The challenge of Sentiment Analysis is the language technologies. It is extremely difficult to teach a machine to analyse the grammatical nuances, cultural variations, slang and misspellings that occur in online mentions. A machine understanding the context, while analysing is also needed many times.   The main challenges are :

- *Sarcasm, irony and implications* - Expressions where negative comments using positive words could be wrongly interpreted as a very positive expression
- *Anaphora resolution* – referencing back, resolving what a pronoun or noun phase is referring to eg - It is his product.
- *Negation Handling* – eg- Fuel consumption is slightly high (double negation)
- *Abbrevations and Emojis*   new social media expressions os short ungrammatical utterances like Lol etc.
- *Word ambiguity* - the word sentiment changes with context. Eg; that was not a comedy

Sentiment Analysis Tools

Sentiment analysis used the combination of natural processing tools, text analytics, computational social sciences to understand the feelings as already mentioned. There are a variety of open source text analytic tools used in NLP for information extraction and classification which can be applied to Sentiment Analysis. Some of the popular open source tools are :

- *Stanford's CoreNLP* - in a cohesive library which provides a part-of-speech (POS) tagger, a named entity recognizer (NER), a parser, a coreference resolution system, sentiment analysis, and bootstrapped pattern learning tools.
- *Python's NLTK* – Natural Language Toolkit - is a platform available for Python which is capable of doing textual tokenisation, parsing, classification, stemming, tagging, semantic reasoning and other computational linguistics.
- *R packages* - The RSentiment package helps in analyzing the sentiment of a sentence and assign a score to it. The function calculate_sentiment predicts the sentiment of sentences which may be classified into six categories: positive, negative, very positive, very negative, sarcasm and neutral.
- *Weka* – An open source software, which is a collection of machine learning algorithms for data mining tasks

Applications

The business applications of sentiment analysis are very broad and extremely powerful. The ability to extract insights from social data and giving feedback is currently a part of customer service.

Sentiment Analysis or opinion mining is becoming an essential part of the market research and customer service. This gives an insight into opinion of the products or services and a comparison with competitors too. Sentiment analysis reveals the overall customer experience and get into more granular level responsed.

This helps in understanding your brand, product, or company as viewed by your customers and stakeholders. Commercial organisations can also use sentiment analysis to measure the impact of a new product, ad campaign, or consumer's response to recent company news on social media.

Simple examples of usage can be:

- Movie review: Is this review positive or negative?
- Product review: What do people think about the new product?
- Politics: How do people view about an issue? What is the opinion about this candidate?
- Prediction: Predict stock markets or election outcomes or market trends

Future and Way forward

Sentiment Analysis is hot trending topic in the field of Natural Language Processing (NLP) and Machine Learning. Sentiment analysis has come a long way from the simple keyword matching to interpret its tone. Initial implementation

was simple logic-based algorithms to identify words as either positive or negative. The current automated systems understand the context and attribute true meaning , learning and getting smarter over time.

It is understood that in the   "next-generation" of sentiment analysis, along with emotions in the written content, voice and facial expressions are also measurable. This can lead to personalisation based on facial recognition with the image of customer.

Despite all the challenges and potential problems, the application of artificial intelligence in sentiment analysis is helping in improving communication which provides businesses with more intelligence to act proactively. Some of the current problems faced in Sentiment analysis can be overcomed by improved accuracy and consistency in text mining techniques.

In conclusion, sentiment analysis matters a lot to business as it makes your communication smarter and more efficient, helps to make better decisions and measure the impact customer communication.   Sentiment analysis has evolved from a simple naive technology to a mainstay in customer service, with more business adding sentiment data into their routine processes.

Reference:

1) B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," Foundations and Trends in Information Retrieval, vol. 2, nos. 1–2, 2008, pp. 1–135.
2) B. Liu, "Sentiment Analysis and Opinion Mining", Morgan and Claypool, 2012.
3) M.Q.Hu and B.Liu, "Mining and Summarizing Customer Reviews," Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD-2004), Seattle, 22-25 August 2004, pp. 168-177.
4) B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs Up? Sentiment Classification Us-ing Machine Learning Techniques," Proc. Ann. Conf. Empirical Methods in Natural Language Processing, Assoc. for Compu-tational Linguistics, 2002, pp. 79–86.
5) E Cambria, B Schuller, Y Xia, C Havas "New avenues in opinion mining and sentiment analysis"
   - IEEE Intelligent Systems, 2013

About the author

Ms. Mini Ulanat is a technology evangelist with a career spanning about three decades in software & infrastructural development, implementation and management of enterprise IT projects. She is a recipient of several awards and accolades including three prestigious international fellowships -- Chevening Fellowship (UK) , Fulbright Scholarship (USA) and CICC Scholarship (Japan). Active in Professional organizations, she has held leadership positions in the local and national committees. She is currently the chair of Women in Engineering (WIE), IEEE Kerala Section. Prior to that she was the chair- IEEE Kochi Sub section, IEEE Computer Society, Kerala Section and National Student Coordinator of Computer Society of India.

Her off the job activities involve driving diversity of leadership in technology by promoting and encouraging women in technology to aspire and reach greater heights through many voluntary activities. She is also a part of other communities and initiatives like Women Techmakers, GDG. She is philanthropically engaged by supporting organization working towards making positive difference in the life of less privileged girls as a mentor and teacher. She is a speaker, moderator, sessions chair for numerous forums and has been serving in various program committees for national and international conferences, workshops and seminars. Ms. Mini Ulanat has been actively involved in organizing and co-coordinating conferences. She is a resource person at all levels from community development programmes at grass root to the scientist of national institutions, involved in reshaping and continuing education programmes to train and retrain outside traditional channels using innovative techniques.

Microsoft and Washington University scientists have developed the first fully automated DNA data storage system that codes digital information into manufactured DNA. The $10,000 prototype encoded "hello" into DNA in liquid form and translated it back into digital information in 21 hours. Microsoft said that natural DNA found in old bones preserved information for several thousands of years.

A 'robot watchman' is being tested in a residential community in Beijing since December 2018 to replace human night patrol. The 'first-of-its-kind' robot "Meibao" integrates facial recognition, man-machine communication and infrared thermal imagery to monitor illegal activities and also provides useful information like weather updates. Chinese media reported the robot was welcomed by watchmen.