

Overview of Tools and APIS for Sentiment Analysis

Prof. Pranali Kosamkar¹ Nikita Bhandari² Rupali Avhad³ Ashwini Bhaisare⁴ Akshata Bandagale⁵

¹Professor ^{2,3,4,5}Student

^{1,2,3,4,5}Department of Computer Engineering

^{1,2,3,4,5}MIT, Pune

Abstract— To the emergent accessibility of opinionated texts like tweet, blog posts and product review websites, a Sentiment Analysis has sprung up for opinion & text analysis in the past decade. Social networking websites let users build social connections with family, friends, and coworkers. Users can also build profiles for storing and sharing various types of content with others, including photos, videos, and messages. To encourage this interaction and provide richer content, social-networking sites expose their networks to Web services in the form of online application programming interfaces. These APIs allow third-party developers to interface with the social-networking site, access information, and build social applications that aggregate, process, and create content based on users' interests. The paper gives an overview of the different API, tools used for different social media. Starting from this overview the paper provides a classification of (i) approaches with respect to different dataset and (ii) different techniques used for sentiment analysis.

Key words: Sentiment analysis, API, classification, neural network, opinion, machine learning, subjectivity, objectivity

I. INTRODUCTION

Twitter/Facebook are a popular microblogging service where users create status messages (called "tweets/comments"). These tweets/comments sometimes express opinions about different topics. Consumers can use sentiment analysis to research products or services before making a purchase. Marketers can use this to research public opinion of their company and products, or to analyze customer satisfaction. Organizations can also use this to gather critical feedback about problems in newly released products. There has been a large amount of research in the area of sentiment classification. Sentiment analysis is a kind of text classification that classifies texts based on the Sentimental orientation (SO) of opinions they contain. Sentiment analysis of product reviews has recently become very popular in text mining and computational linguistic research. Sentiment analysis (also known as opinion mining) refers to the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials. Sentiment analysis is widely applied to reviews and social media for a variety of applications, ranging from marketing to customer service. Sentiment analysis aims to determine the attitude of a speaker or a writer with respect to some topic or the overall contextual polarity of a document. The attitude may be his or her judgment or evaluation, affective state, or the intended emotional communication. First eliminate objective sentences, and then use remaining sentences to classify document polarity (reduce noise).

A. Types of Sentiment Analysis

1) Document-Level of Sentiment Analysis

In document-level of sentiment analysis, each document focuses on a single entity or event and contains opinion from a single opinion holder. The opinion here are can be classified in to two simple classes: Positive or negative (probably neutral). For example: A product review: "I bought a new phone few days ago. It is a nice phone, though it is a little big. The touch screen is good. The voice clarity is better. I simply love the phone". Considering the words or phrases used in the review (nice, good, better, love), the subjective opinion is said to be positive. The objective opinions are measured using the star or poll system, where 4 or 5 stars are positive and 1 or 2 stars are negative.

2) Sentence-Level of Sentiment Analysis

To have more refined view of different opinions expressed in the document about the entities, we have the sentence level. This level of sentiment analysis – filters out those sentences which contain no opinion and – determines whether the opinion on the entity is positive or negative.

3) Aspect Based Sentiment Analysis

Document level and sentence level sentiment analysis works well when they refer to a single entity. However, in many cases people talk about entities that have many aspects or attributes. They will also have different opinions about different aspects. It often happens in product review and discussion forums. For example: "I am a Nokia phone lover. I like the look of the phone. The screen is big and clear. The camera is fantastic. But, there are few downsides too; the battery life is not up-to the mark and access to Whatsapp is difficult". Categorizing the positive and negatives of this review hides the valuable information about the product. Therefore, the Aspect –based sentiment analysis focuses on the recognition of all sentiment expressions within a given document and the aspects to which the opinions refer.

B. Different Aspects

As sentiment analysis is being used for different aspects deciding on a platform and figuring out how sentiment fits into workflow is essential as:

- Demystifying accuracy: As machines will never be able to measure sentiment as well as humans, and even humans don't agree 100% of the time. The number of sentiment types is also part of the equation. Some platforms offer three sentiments, some offer four, and some offer more than five. The more you increase the number of sentiment types, the less accurate (but more information rich) your results become.
- Isolating content types: A lot of social media mentions are neutral in nature and some social media sources tend to skew higher on the neutral scale. For example, a higher percentage of updates are neutral on Twitter than any other medium. Depending on the source you are looking at, your sentiment results will differ this should be

expected. Make sure your sentiment platform allows you to isolate results by content type.

- Sentiment override: Because automated sentiment is not going to be 100% accurate, you, the user, need to have some kind of override control. When picking a tool, ensure that it allows you to override sentiment, and toss irrelevant results.
- Entity level vs. article level sentiment: Until recently, the industry default has been able to measure sentiment at the level of the article. Over time, some platforms have developed ways to measure sentiment on the level of the entity (entity level analysis can measure the sentiment of an entity or multiple entities within an article even if the overall sentiment of the article is different).

C. Sentiment Measurement

- Influence: Because social media mentions are plentiful, prioritization tools must continue evolving. Of the 10,000 tweets and blog posts about your brand, how do you pick the top 50 to focus on? If you need to neutralize the mentions that hurt your brand the most, you should drill down into negative mentions, identify the content coming from the most influential people in your industry, understand how far each tweet traveled, and how many people were impacted by this Content. If someone blogs and tweets the same negative mention about you, how do you account for that? How do you quantify the multiplied effect of cross-platform communication originating from the same person? Including influencer analytics alongside sentiment measurement is becoming a standard of the social media monitoring industry.
- Reputation: Taking the influencer concept a step further, each notable user should have a social media reputation profile. If someone's negative sentiment indexes higher than average (i.e. that person hates everything equally), then that person's negative sentiment should be somewhat discounted in statistics, we toss outliers like these out of the consideration set. Moreover, your reputation and influence on one channel should carry over into other channels.
- Intensity: As far as sentiment algorithms are concerned, part of a successful prioritization process is going to be identifying the intensity of each mention. "I really hate product X and will never buy it" is quite different from "Product X is running a little slow today." Ability to cross-reference intensity, influence, trajectory, velocity and sentiment of each social media mention will drive us towards a reliable priority system.

D. Tools for Sentiment Analysis

- Melt-water: Assess the tone of the commentary as a proxy for brand reputation and uncover new insights that help you understand your target audience.
- Google Alerts: A simple and very useful way to monitor your search queries. I use it to track "content marketing" and get regular email updates on the latest relevant Google results. This is a good starting point for tracking influencers, trends and competitors.
- People Browser: Find all the mentions of your brand, industry and competitors and analyze sentiment. This tool allows you to compare the volume of mentions before, during and after your marketing campaigns.

- Google Analytics: A powerful tool for discovering which channels influenced your subscribers and buyers. Create custom reports, annotations to keep uninterrupted records of your marketing and web design actions, as well as advanced segments to breakdown visitor data and gain valuable insights on their online experiences.
- Hootsuite: A great freemium tool that allows you to manage and measure your social networks. The premium subscription provides enhanced analytics at a very reasonable 5.99USD per month.
- Tweetstats: This is a fun, free tool that allows you to graph your Twitter stats. Simply enter your Twitter handle and "let the magic happen."
- Facebook Insights: If you have more than 30 Likes on your Facebook Page you can start measuring its performance with Insights. See total page Likes, number of fans, daily active users, new Likes/Unlikes, Like sources, demographics, page views and unique page views, tab views, external referrers, media consumption and more!
- Pagelever: This is another tool for measuring Facebook activity. Pagelever gives you the ability to precisely measure each stage of how content is consumed and shared on the Facebook platform.
- Social Mention: The social media equivalent to Google Alerts, this is a useful tool that allows you to track mentions for identified keywords in video, blogs, microblogs, events, bookmarks, comments, news, Q&A, hash tags and even audio media. It also indicates if mentions are positive, negative, or neutral.
- Marketing Grader: Hubspot's Marketing Grader is a tool for grading your entire marketing funnel. It uses over 35 metrics to calculate your grade by looking at if you are regularly blog posting, Tweeting, updating on Facebook, converting visitors into leads, and more. It's a full funnel way to help you measure your inbound marketing initiatives.

E. Apis for Sentiment Analysis

- Skyttle: API is designed to turn any text into constituent terms (meaningful expressions), entities (names of people, place and things), and sentiment terms. Languages supported are English, Spanish, French, German, Chinese, Swedish, Greek, Czech, Italian and Russian
- Sentiment Analysis Spanish: Sentiment analysis for Spanish language of any given tweet.
- Nlp Tools: Text processing framework to analyze Natural Language. It is especially focused on text classification and sentiment analysis of online news media (general purpose, multiple topics).
- Tweet Sentiments: Returns the sentiment of Tweets. Two online APIs call the Twitter API to analyze Tweets from a given Twitter user or Tweets returned by a Twitter search query. The offline API analyzes texts of Tweets you've already got, one Tweet at a time.
- Fluxifi NLP: Cloud based Natural Language Processing API. Includes Sentiment and Language Detection.
- Alchemy API: provides advanced cloud based and on premise text analysis infrastructure that eliminates the expense and difficulty of integrating natural language processing systems into your application, service, or data processing pipeline.

- Starget sentiment analysis: This is a short text (a tweet or a single sentence) sentiment classification API. It has two types of analysis: one for finding more (but less accurate) sentiment snippets and another one for finding more accurate sentiment (but missing some difficult cases).
- Textalytics Media Analysis: API analyzes mentions, topics, opinions and facts in all types of media. This API provides services for: Sentiment analysis extracts positive and negative opinions according to the context.
- Free Natural Language Processing Service: 100% free service including sentiment analysis, content extraction, and language detection.
- Bitext Sentiment Analysis: The purpose of this service is to extract opinions from text. An opinion represents the subject an author is writing about and a sentiment score that classifies how positively or negatively the author feels towards that subject. Deep Linguistic Analysis is used to identify the subject the author is discussing.

II. LITERATURE SURVEY

Range of feature selectors and feature weights with both Naïve Bayes and Support Vector Machine (SVM) classifiers was evaluated. It includes the two new feature selection methods SWNSS & SWNPD & compares them at a number of selection thresholds with PD. It also introduces three feature weighting methods SWN-SG, SWN-PG & SWN-PS and compares their performance with the standard & popular FF, FP, TF-IDF methods. Their results show that it is possible to maintain a state-of-the-art classification accuracy of 87.15% while using less than 36% of the features. Overall, FP was the most successful feature weighting method for both SVM and Naïve Bayes[1].

Rule-based classification, supervised learning & machine learning was a new combined method. This method tested on movie reviews, product reviews & my-space comments. The result show that a hybrid classification can improve the classification effectiveness in terms of micro and macro averaged F1.F1 measure both precision & recall of classifier's effectiveness into account. Multiple classifiers in a hybrid manner can result in a better effectiveness in terms of micro & macro-averaged F1 than any individual classifier [2].

Novel approach was used to identify feature specific expressions of opinion in product reviews with different features & mixed emotions. A method that represents the feature and corresponding opinions in the form of graph where they use dependency parsing to capture the relations between the features & their associated opinions was proposed. This idea was to capture the association between any specific feature & the expressions of opinion that come together to describe that feature. This was done by capturing the short range & long range dependencies between the words using dependency parsing. A system that extracts potential features from a review & clusters opinion expressions describing each of the features was developed. It finally retrieves the opinion expression describing the user specified feature [3]. They focuses on the comparative study(1997-2012) of different sentiment classification techniques performed on different data set domain such as web discourse, reviews .The main aim was to show which sentiment classification technique used on what type of data

set. Basically, machine learning techniques were used for feature based sentiment and other for bag of words [4].

It introduced a method for capturing patient experience from free text comments posted online. Machine learning techniques were used to understand patient's unstructured comments about their care. WEKA tool was used. Machine learning approach had two components as 1) Pre-Processing: in which data from patient comments are split into manageable units to build representation of data. 2) Classification: Algorithm decide which category each comment falls [5]. Twitter messages as dataset with emoticons are used as noisy labels using distant supervision learning method. In this paper machine learning algorithm (Naïve bayes, maximum entropy, SVM) are compared which gives accuracy greater than 80%. Training dataset used is tweets from twitter (April 6,2009 to June 25, 2009)[6]. Neural network approach is used. Two types of dataset are used, which are movie review dataset (2000 reviews), Sentence polarity dataset (5331 positive reviews and 5331 negative reviews). Neural network is implemented in c++ and Lua using Torch7 machine learning library [7].

Events of cricket video are extracted and automatically classified them into two classes; namely wicket and hit and attached the concept tag (keyword) as the class name of the event. Seven hours of cricket videos dataset are used which contains 18 wickets and 130 hits at the semantic concept level. For football video, 8 hours of video are used, containing 10 goals and 33 misses. News is classified using rule based classifier. Videos of news channels "Aaj Tak", "BBC" are even classified[8]. Text categorization is used where first label the data as subjective or objective and machine learning classifier are applied. Document level polarity classification is considered to gather subjective sentences, where 5000 movie review are collected from www.rottentomatoes.com and objective data collected from www.imdb.com [9]. Opinion mining is performed on newspaper articles from quotations (reported speech) . Annotated quotations are extracted from news provided by EMM news gathering engine. 99 quotes are selected and results were fed into EMM News Brief & EMM-News Explorer. Bag of Words Approach is used [10].

III. CONCLUSION

Sentiment analysis, also known as opinion mining, is the analysis of the feelings (i.e. attitudes, emotions and opinions) behind the words using natural language processing tools. The overview of various tools and APIs for sentiment analysis makes developers easy to access information of social-networking sites. As sentiment analysis is complex process, it takes time to understand our needs and also to decide which tool, API and technique to be used for this. This paper gives various options for tools and API used for sentiment analysis as well as helps to solve related issues.

REFERENCES

- [1] Tim O'Keefe, Irena Koprinska, Feature Selection and Weighting Methods in Sentiment Analysis, Proceedings of the 14th Australasian Document Computing Symposium, Sydney, Australia, 4 December 2009.

- [2] Rudy Prabowo, Mike Thelwall, Sentiment Analysis: A Combined Approach, University of Wolverhampton Wulfruna Street WV1 1SB Wolverhampton, UK.
- [3] ubhabrata Mukherjee, Pushpak Bhattacharyya, Feature Specific Sentiment Analysis for Product Reviews, Dept. of Computer Science and Engineering, IIT Bombay.
- [4] Shailendra Kumar Singh, Sanchita Paul and Dhananjay Kumar, Sentiment Analysis Approaches on Different Data Set Domain: Survey, International Journal of Database Theory and Application Vol.7, No.5 (2014).
- [5] SA for capturing patient experience from free-text comments posted online.
- [6] Alec Go, Richa Bhayani, Lei Huang, Twitter Sentiment Classification using Distant Supervision, Stanford University, Stanford, CA 94305.
- [7] Soumith Chintala, Sentiment Analysis using neural architectures, New York University, New York, NY 10012.
- [8] M. H. Kolekar, Student Member, IEEE and S. Sengupta, Member, IEEE, Semantic Concept-driven Sports and News Video Indexing and Retrieval, Indian Institute of Technology Kharagpur, (India) July 22, 2009.
- [9] A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts, Bo Pang and Lillian Lee Department of Computer Science Cornell University Ithaca, NY 14853-7501 pabo,llee@cs.cornell.edu.
- [10] Alexandra Balahur, Ralf Steinberger, Erik van der Goot, Bruno Pouliquen, Mijail Kabadjov, Opinion Mining on Newspaper Quotations, 2009 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technologies.

