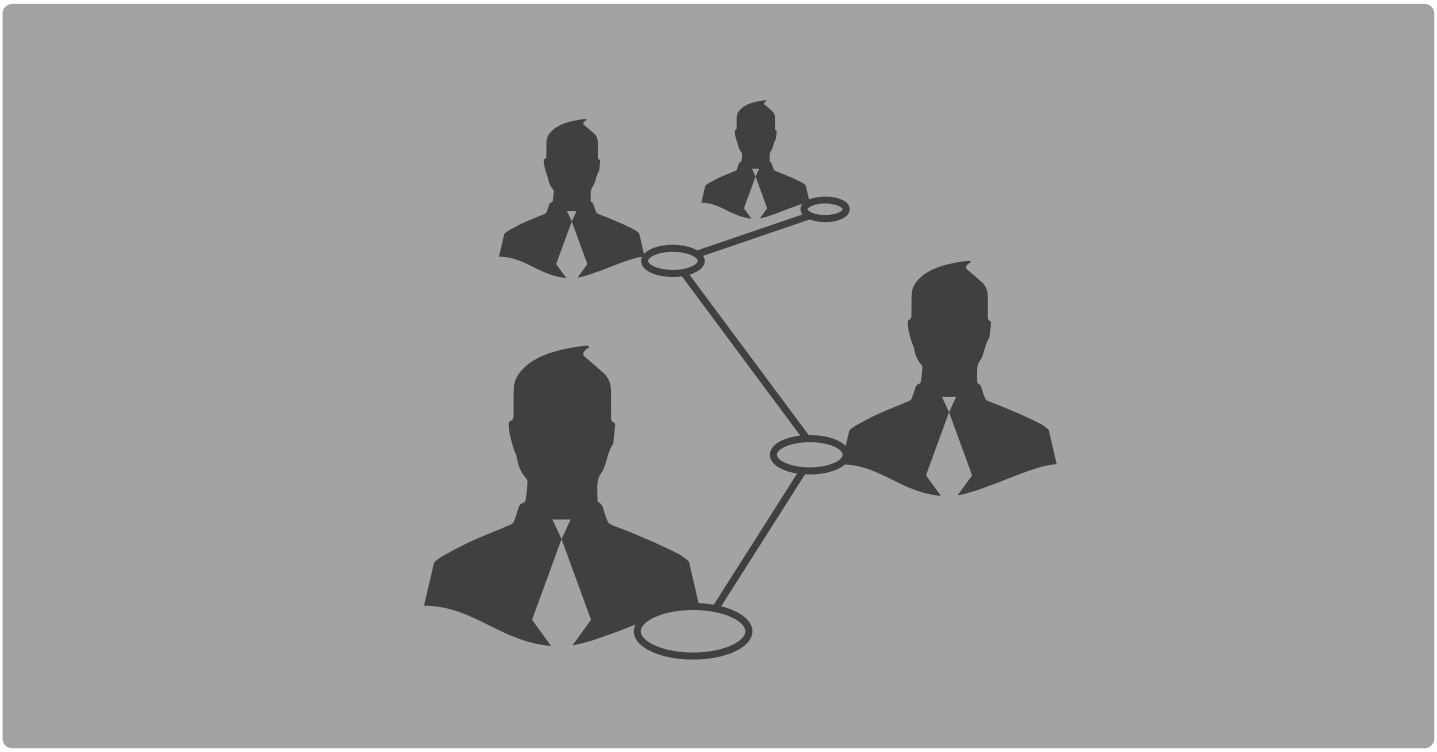
WHITEPAPER

Sentiment Analysis and its Challenges



There is a huge explosion of social media and the rapid growth in users using smart phone devices to log in to social media platforms. This gives ample opportunities for a customer to express their feelings and attitudes about anything and everything at any given point of time. The opinion of a social media user will be in the form of chats, reviews, shares and tweets which will prove to be very valuable to businesses to gain insights on what people's perception is about a brand, product or service. This will help the businesses to improve upon their products and service based on informed decisions, which in turn will result in better promotion of the brands.

The key to success for a business with respect to sentiment data lies in the ability to mine into vast sources of data which is unstructured and across various social media platforms. This can be done using sophisticated technology like Natural Language Processing (NLP) which produces a sentiment analysis for a text resulting in knowing whether the text is positive, negative or neutral.

© eMudhra Limited. All Right Reserved.

HOW BUSINESSES CAN FIND VALUE OUT OF SENTIMENT ANALYSIS

Enhance Customer Experience

Sentiment Analysis falls into three categories – positive, negative or neutral. Through sentiment, organizations can detect the tone and temperaments of each and every word found in a posting and categorize those sentiments. Equipped with this information, business can know what they are doing right with respect to products, services and customer support – positive sentiment; similarly businesses can work on certain aspects to improve based on the negative sentiments.

Gain Competitive Edge

Sentiment Analysis not only helps companies to improve their businesses but also to gauge their competitors and how they stack up against the competition in the market. For example, a company that has 20 percent of negative sentiment may view that as acceptable but when compared to its competitor having 10 percent of negative sentiment, then the 20 percent does not look good.

Gain Business Intelligence

Sentiment Analysis data provides powerful and rich information about existing and future customers. The insights provided can lead to targeting unknown potential customers. To do this sentiment analysis alone is not enough, this has to be backed by human insights too.

Building Brand Perception

Branding is all about the kind of perception it has developed over the years. Sentiment Analysis allows the business to quantify the brands perception that existing and future customers may have about their products and services, their customer experience, online content, marketing and social campaigns. Access to negative sentiments helps the brand to revitalize and infuse some energy into the brand.

SENTIMENT ANALYSIS AND WHY IT CANNOT ACHIEVE 100 PERCENT ACCURACY

Factors which make sentiment analysis inaccurate:

- **Context:** a positive or negative sentiment word can have the opposite connotation depending on context. "I've done a great job" may be interpreted as a positive statement. However, in "my internet provider does a great job when it comes to stealing money from me", doing a great job is no longer a positive thing, based on the context ("stealing money from me").

- **Sentiment Ambiguity:** a sentence with a positive or negative word doesn't necessarily express any sentiment. For example, "can you recommend a good tool I could use?" doesn't express any sentiment, although it uses the positive sentiment word "good". Likewise, sentences without sentiment words can express sentiment too. So, "this browser uses a lot of memory" doesn't contain any sentiment words, although it clearly expresses a negative sentiment.

- **Sarcasm:** a positive or negative sentiment word can switch sentiment if there is sarcasm in the sentence. "Sure, I'm happy for my browser to crash right in the middle of my coursework" is obviously a sarcastic (and negative) statement, even though it has the positive word "happy". We can detect the sarcasm mainly from how the sentence starts with "sure", and the context (we know for a fact that a browser crashing is negative).

- **Comparatives:** social listening tools often misunderstand comparative statements. For example, what's the sentiment of "Pepsi is much better than Coke"? If you're reporting for Pepsi, then this is definitely a positive statement. However, if you work for Coca Cola and you're reporting back to the company, then this statement would be negative. Most social listening tools aren't intelligent enough to "pick sides" when they find comparative statements like the above, leaving them to pick the sentiment based on keywords. So, the previous example would be tagged as "positive" as it contains a positive keyword, "much better", regardless of who you're reporting for.

- **Regional Variations:** a word can change sentiment and meaning depending on the language used. This is often seen in slang, dialects, and language variations. An example is the word "sick", which can change meaning based on context, tone and language, although clear to the target audience ("That is a sick song!" vs. "I'm not feeling well at all, I might be sick"). An example of a regional variation can be found between British and American English for words like 'quite', 'rather', 'pretty': in British English those words take the meaning of "fairly", while in American English they take the meaning of "very". This can sometimes be misunderstood in day-to-day conversations too, so it's no wonder that tools may find this problematic.

- **The biggest threat to accuracy in sentiment analysis today is human concordance:** this is the degree of agreement among humans (or between humans and machines). Numerous studies have shown that the rate of human concordance is between 70% and 80%.

SOME PRACTICAL EXAMPLES

Standard engines use a dictionary based approach, looking for the presence of 'keywords' to mark a sentence positive or negative. Unfortunately, this is not enough to provide insights to a decision maker who wants to use the information to make marketing decisions. After all, how would you treat a somewhat complex sentence, e.g. "I love the food, but the service is bad'. Is the sentence positive, negative or neutral? Now imagine trying to aggregate even 50 sentences like this to derive insights, let alone 500 or 5,000.

For companies trying to use social media to accurately measure sentiments about their brand, that of competitors, and use the information of product enhancement, accuracy is critical.

Here are some examples of how inaccurate engines can lead to completely misleading results:

Example sentence	What typical engines do	What PRISM does
I love the screen of the phone.	Mark it as a positive sentence	Mark it as a positive sentence, assign it a score, and link the positiveness to 'screen'
I love the screen, but hate the price!	Mark it as a neutral sentence	Mark it as positive for 'screen' but negative for 'price', while assigning marks to each attribute
This model is like the earlier model!	Mark it as positive (seeing the word 'like' in the sentence)	Mark it as neutral
I cannot say this is nice / This is not good.	Mark it as positive	Mark it as negative (recognizes negation)
I prefer Brand X over Brand Y	Mark it as positive	Mark it as Positive only for Brand X and negative for Brand Y

Behind the accuracy level of PRISM, lies the patent pending clause level sentiment approach, which allows sentiments to be extracted at a clause level, rather than relying on the presence or absence of keywords in a sentence. This sentiment extraction approach requires deep expertise in Natural Language Processing techniques, and that is what differentiates the engines from typical engines that rely on open source dictionaries and dashboards.

ABOUT EMUDHRA

eMudhra Limited is a Certifying Authority licensed by Controller of Certifying Authorities, under Government of India. eMudhra operates under the guidelines set by Information Technology Act. With more than one million certificates issued, eMudhra caters to all kinds of subscribers who use Digital Certificates for Income Tax, MCA (ROC), Tenders, Foreign Trade, Banking, Railways and many other needs.

Web: *http://www.emudhra.com/products/data-analytics/prism.html*