Sentiment Analysis in Unstructured Data

Anurag MIGLANI

February 17, 2017

1 SYNTHETIC SUMMARY

The objective of the systematic study conducted at Octopeek, Enghien-Les-Bains, France is to determine human sentiments from microblog posts i.e. Tweets posted on social network Twitter. Tweets are used because they contain valuable information about people's behavior on different products and services. This study was carried at Data Science Department of Octopeek.

Sentiment Analysis is a class of text classification in Natural Language Processing. Categorizing the text into positive and negative categories about its context is called Sentiment Analysis.

With the help of sound supervised machine learning algorithms, my study has made it possible to use data on the social networking micro-blogging site like Twitter, interpret the patterns behind it; thus improving the credibility of any subject discussed be it a product, policy or service.

The subject has its relevance because of the opportunities and benefits it offers from different organizations and governments. Nothing beats real time feedback from real stakeholder for any organization be it a business, government or an NGO. Growing interest in Sentiment Analysis is not unique to any particular genre. Instead, it is an open area of interest for many domains, such as policy making, financial services, healthcare, manufacturing, etc.

This study is critical from a business perspective view for Octopeek because it limits the gap between comprehension and operation for its client, and provides them with a framework to access and explore in real-time, what consumers and stakeholders are thinking about their product/services/policies. And thus, what measures can be taken to improve it.

Making this work multilingual would increase the feasibility for Octopeek's global client tail. Currently, this analysis is coherent to typical sentiment analysis so by using different datasets for different domains would influence the credibility of the model.

Further, a business use case regarding public sentiment in most recent US presidential elections is shown. In this use case, Donald Trump and Hillary Clinton's public opinion is examined and how it changes with real life event happening in US politics. Recently a video was released featuring Donald Trump speaking against women; this shifted the reaction of public towards Donald as shown by the use case.

Conducting extensive surveys implicates massive money flow, these days companies use social media and specifically NLP tools such as Sentiment Analysis as they are cost & time efficient. In this study of Sentiment Analysis, I discuss its literature, definition, execution into the real world and more in detail.