Unit - III

Sentiment Analysis Overview

* Sentiment
* Sentiment classification
* sentiment analysis

Sentiment: is a difficult word to define. It is often linked to or confused with other terms like belief, view, opinion & conviction

* It suggests a settled opinion reflective of one's feelings.
* sentiment analysis is different from simply finding topics in text.

Feelings (vs) Facts:

When we look for sentiment, we're looking for opinions or feelings - like whether something is good or bad.

* In topic categorization we just want to know what to know what the text is about like 'sport' without needing to know how people feel about it.

Emotion strength

With sentiment, we often want to measure how +ve or -ve a text is - not just that it is +ve or -ve. This is different from topic analysis, where it's enough just to know the subject.

Sentiment classification

* usually deals with two classes, a range of polarity, or even a range in strength of opinion
* working with just a few categories seems simple, but it's hard because each category includes many different topics, uses, & types of documents

Sentiment Analysis

* It has many names. It's often referred to as opinion mining, Subjectivity analysis & appraisal extraction, with some connections to affective computing
* As a field of research, sentiment analysis is closely related to computational linguistics natural language processing & text mining.
* Sentiment analysis is trying to answer the question. "what do people feel about a certain topic?" by digging into opinions of many using a variety of automated tools

In business, especially in marketing & customer relationship management, sentiment analysis seeks to detect favorable & unfavorable opinions towards specific products & services using large numbers of textual data sources

* current trends are to implement analytic methods to consider both implicit & explicit Sentiments
* Sentiment polarity is a particular feature of text that Sentiment analysis primarily focuses on.
* It is usually dichotomized into two-three but polarity can also be thought of as a range.
* A document containing several opinionated statements would have a mixed polarity overall, which is different from not having a polarity at all.

Sentiment Analysis Applications

1. voice of customer (VOC)
2. voice of the Market (VOM)
3. voice of the employee (VOE)
4. Brand Mangement
5. Financial Markets
6. politics
7. Government Intelligence
8. Other interesting Areas

* Traditional Sentiment analysis relied on Surveys & focus groups, which were expensive & slow; using only small participant samples
* Now text analytics automates large-scale data collection & processing usually up to analyze both facts & opinions from social media & the web efficiently
* Sentiment analysis is perhaps the most popular application of text analytics, tapping into data-sources like tweets, online communities, web logs, chatroom etc

voice of the customer (VOC)

* It is an integral part of an analytic CEM & customer experience management systems.
* As the enabler of VOC, sentiment analysis can access a company's product & service reviews to better understand & better manage the customer complaints & praises
* A software company may detect the new buzz regarding the bugs found in their newly released product early enough to release patches & quick fixes to alleviate the situation.
* focus of VOC is individual customers, their service- & support-related need, wants & issues
* VOC draw data from the full set of customer touch points, including, e-mails, survey, call center notes & social media & individual customer profiles captured in enterprise operational systems

* VOC mostly driven by sentiment analysis is a key element of customer experience management initiatives, where the goal is to create an intimate relationship with the customer.

Voice of the Market (VOM)

It is about understanding aggregate opinions & trends. Its about knowing what stakeholders - customers, potential customers, influencers, whoever are saying about your products & services. A well-done VOM analysis helps companies with competitive intelligence & product development & positioning

Voice of the Employee (VOE)

It has been limited to employee satisfaction surveys. Text analytics in general is a huge enabler of assessing the VOE; using rich, opinionated textual data is an effective & efficient way to listen to what employees are saying; happy employees empower customer experience efforts & improve customer satisfaction.

Brand Management

It focuses on listening to social media, where anyone can post opinions that can damage or boost your reputation. There are a number of relatively new launched start-up companies that offer analytics-driven brand management services for others.

Brand management is product & company focused. It attempts to shape perceptions rather than to manage experiences using sentiment analysis techniques

Financial Markets

Predicting the future values of individual stocks has been an interesting & seemingly unsolvable problem. What makes a stock move up (or) down is anything but an exact science. Many belive that the stock market is most sentiment driven, making it anything but rational..

∴ The use of sentiment analysis in financial markets has gained significant popularity. Automated analysis of market sentiments using social media, news, blogs & discussion groups seems to be a proper way to capture the market movements.

* If done correctly, sentiment analysis can identify short-term stock movements based on the buzz in the market, potentially impacting liquidity & trading.

Politics

* Opinions matter a great deal in politics
* Because political discussions are dominated by quotes, sarcasm, & complex references to persons, organizations & ideas, politics is one of the most difficult, & potentially fruitful areas for sentiment analysis

* By analyzing the sentiment on election forum one may predict who is more likely to win or lose.
* Sentiment analysis can help understand how voters are thinking & can clarify a candidate's position on issues
* Sentiment analysis can help political organizations, campaigns & news analysts to better understand which issues & positions matter the most to voters
* The technology was successfully applied by both parties to the 2008 & 2012 American Presidential election campaigns

Government Intelligence

* It is another application that has been used by intelligence agencies. eg
* ex- It has been suggested that one could monitor sources for increases in hostile communications
* Sentiment analysis can allow the automatic analysis of the opinions that people submit about pending policy (or government-regulation) proposals
* Monitoring communications for spikes in negative sentiment may be of use to agencies like Homeland security

Other Interesting Areas

* Sentiments of customers can be used to better design e-commerce site, better place advertisement & manage opinion-or-~~ine~~ review oriented search engines
* e-commerce sites
* Better place to advertisements
* Manage opinion-of-review-oriented search engine
* Sentiment analysis can help with e-mail filtration by categorizing & prioritizing incoming mails

Sentiment Analysis process

* Because of the complexity of the problem there is no readily available standardized process to conduct sentiment analysis
* It is based on the published work in the field of sensitivity analysis so far a multi-step, simple logical process, as seems to an appropriate methodology for Sentiment Analysis
=> These logical steps are iterative & experimental in nature, & once completed & combined, capable of producing desired insights about the opinions in the text collection.

Step ① : Sentiment Detection
② : N-p polarity classification
③ : Target identification
④ : Collection & Aggregation.

Sentiment Detection

* After the retrieval & preparation of the text documents, the first main task in sentiment analysis is the detection of objectivity
* The goal is to differentiate btw a fact & an opinion, which may be viewed as classification of text as objective or subjective. This may also be characterized as calculation of O-P polarity
* If the objectivity value is close to 1, then there is no opinion to mine, The process goes back & grabs the next text data to analyze
* Usually opinion detection is based on the examination of adjectives in text
* ex: The polarity of "what a wonderful world" can be determined relatively easily by looking at the adjective

N-P polarity classification

* The second main task is that of polarity classification
* Given an opinionated piece of text, the goal is to classify the opinion as falling under one of two opposing sentiment polarities, or locate its position on the continuum b/w these two polarities
* When viewed as a binary features, polarity classification is the binary classification task of labeling an opinionated documents as expressing either an overall positive or an overall negative opinion
* To identification of N-P polarity one should also be interested in identifying the strength of the sentiment.
* Most of this research work done on product or movie reviews - where the distinctions of "positive" & "negative" are quite clear
* Other task, such as classifying news as "good" or "bad" present some difficulty
* An article may contain negative words without explicitly using any subjective words or terms
* These classes usually appear intermixed when a document expresses both positive and negative sentiments
* Then the task can be to identify the main sentiment of the document
* Still, for lengthy texts, the tasks of classification may need to be done at several levels: term, phrase, sentence & perhaps, document level.
* It is common to use the outputs of one level as the inputs for the next higher layer
* Several methods used to identify the polarity & strengths of the polarity are explained in the next section.

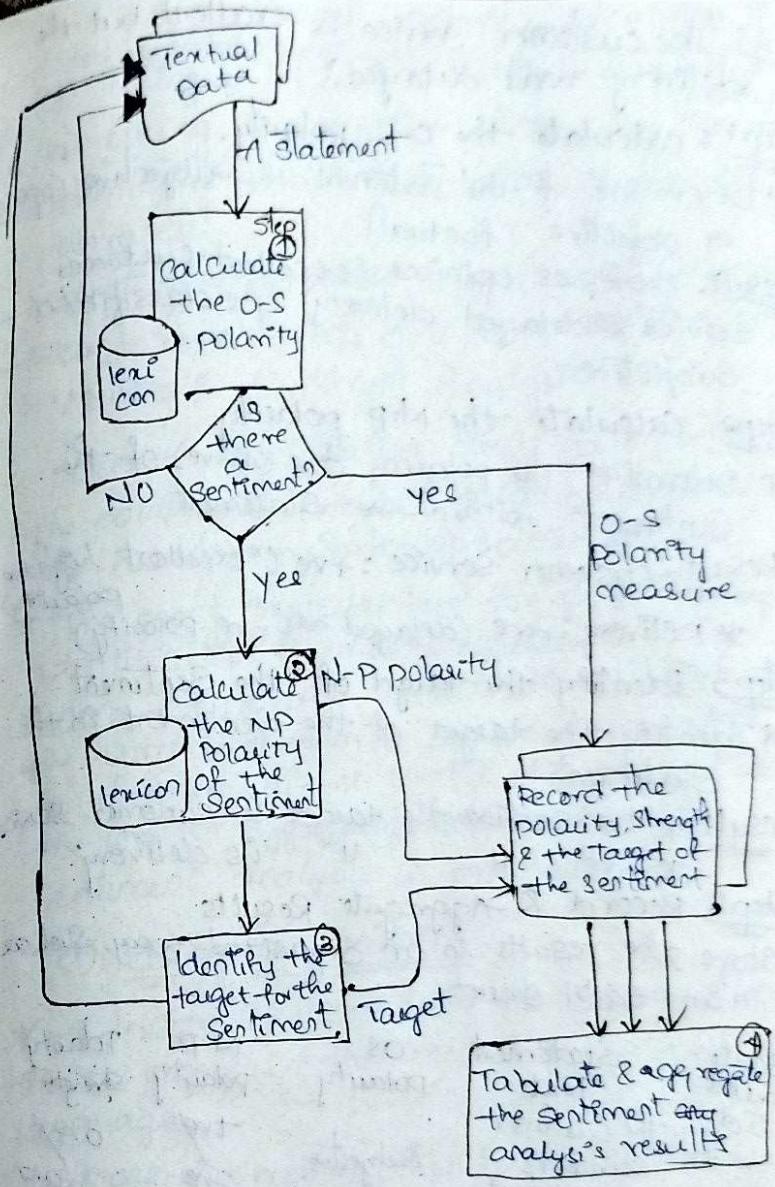
Target Identification

* The goal of this step is to accurately identify the target of the expressed sentiment
* The difficulty of this task depends largely on the domain of the analysis

* Even though it usually easy to accurately identify the target for product or movie reviews, because the review is directly connected to the target, it may be quite in other domains.
* For instance, lengthy, general-purpose text such as web pages, news articles & blogs do not always have a predefined topic that they are assigned to & often mention many objects, any of which may be deduced as the target
* Sometimes there is more than one target in a sentiment sentence, which is the case in comparative texts
* A subjective comparative sentence orders objects in order of preference.
* ex "This laptop computer is better than my desktop PC"
* These sentences can be identified using comparative adjectives & adverbs, superlative adjectives & other words.
* Once the sentences have been retrieved, the objects can be put in an order that is most representative of their merits.

Collection & Aggregation

* Once the sentiments of all text data points in the document are identified & calculated. They are aggregated & converted to a single sentiment measure for the whole document
* This aggregation may be as simple as summing up the polarities & strengths of all texts or as complex as using semantic aggregation techniques from NLP to come up with the ultimate sentiment



A Multi-step process to Sentiment Analysis

Ex: "The customer service is excellent, but the delivery was delayed."

Step1: calculate the O-S polarity

* Determine if the statement is subjective (opinion) or objective (factual)

Result: expresses opinions (excellent customer service & delayed delivery) it is classified as subjective

Step2: Calculate the N-P polarity

* Determine the polarity (+ve (or) -ve) of the sentiment within the statement

Result: customer service: +ve ("excellent" high the polarity)

* Delivery: -ve (delayed has -ve polarity)

Step3: Identify the target of the sentiment

* Identify the target of the sentiment in the statement.

Result: * +ve sentiment's target is customer service
* -ve " " is delivery

Step4: Record & Aggregate Results

Store the results in an organized way, such as in an excel sheet.

Statement Id	Statement Text	OS polarity	N-P polarity	Polarity strength
1.	The customer service is excellent but delivery was delayed	Subjective	+ve / -ve	0.9 / 0.7

Target

Customer service / Delivery

Methods of polarity Identification

Two dominant techniques used for identification of polarity of the word.

(a) using a lexicon

(b) using a collection of Training Documents

→ Identifying semantic orientation of sentences & phrases

→ Identifying semantic orientation of documents

→ large Textual Data sets for predictive Text Mining & Sentiment Analysis

* polarity identification deals with identifying the polarity of a text & it can be made at the word, term, sentence (or) document level
* The most granular level for polarity identification is at the word level
* once the polarity identification is made at the word level, then it can be aggregated to the next higher level & then the next until the level of aggregation desired from the sentiment analysis is reached.

(a) Using a lexicon

* A lexicon is essentially the catlog of words their synonyms, & their meanings for a given language.
* lexicons for many other languages, there are several general-purpose lexicons created for English
* The most popular general-purpose lexicon is wordNet, created at princeton university, which has been extended & used by many researchers & practitioners for sentiment analysis

Process