TEXT ANALYSIS WITH PYTHON

Lesson 4 – 6/17/2021

a.y. 2020-2021

Università Bocconi
MILANO

IT Education Center
www.unibocconi.eu/itec

Lesson content

SENTIMENT ANALYSIS

- Issues about sentiment detection

- Lexicon-based methods

- Rule-based analysis methods

- Machine Learning based approach

Università
Bocconi
MILANO

Attendance registration

To track your presence in class (wherever you are), please:

- either use the app on your smartphone or tablet

- or go to this web page ➔ www.unibocconi.it/attendance

using

- your own yoU@B credentials

- today's six-digit code*

If you have problems with the app:

- try to log out and then log in again

- if the problem persists, notify our tutor via chat who will manually register your presence

abcdef

((*) It will be shown in the classroom and will remain active only for 10 minutes starting from the first of you who will register

Università
Bocconi
MILANO

Sentiment analysis

- Sentiment Analysis, or Opinion Mining, is a sub-field of Natural Language Processing (NLP) that tries to identify and extract opinions from a given text (i.e., detect whether the opinion expressed towards a brand, a product or a service is positive, negative or neutral)

- The aim is to gauge sentiments, evaluations, attitudes and emotions of a speaker/writer based on the computational treatment of the subjectivity in a text

- A limit is the difficulty of recognizing ambiguous attitudes, multiple sentiments in the same text, irony, sarcasm, use of similes/metaphors/hyperboles, and so on

- Another serious challenge is posed by the kind of language used in social media to express sentiments by using short forms, figurative speech, slangs, memes and emoticons

Università Bocconi
MILANO

It helps in …

- Managing critical posts on social media

- Improving the Customer Experience

- Analyzing product launches

- Assessing the impact of sponsorships and CSR (Corporate Social Responsibility) activities

- Discovering new market trends

- Maintaining the quality of the service on a national and/or global scale

- Monitoring the popularity of Management

Sentiment Analysis enables companies to make sense out of data. Thus they are able to elicit vital insights from a vast unstructured dataset without having to manually indulge with it

Università
Bocconi
MILANO

How does Sentiment work? - 1

Lexicon-based methods: *detection and evaluation of keywords*

- Lexicon: set of words identified as keywords in messages; a positive or negative value is assigned to each word
- Each time the tool detects these keywords, it assigns a positive or negative value to the entire message

  Limits: it is not possible to contextualize the message, the level of accuracy of the analysis generally varies from 50% to 80%, sarcasm not detected

- Improvement ➔ Adding rules to create wider association patterns
- It provides valid results only if aimed at identifying the macro trend of highly discussed topics or if applied on a large scale

Università
Bocconi
MILANO

Liu and Hu opinion lexicon

▪ Is one of most used lexicons in sentiment tools

▪ Is a list of positive and negative words, it contains around 6800 opinion words or sentiment words for English language

▪ This list was composed over many years

Hu and Liu, Mining and summarizing customer reviews, KDD-2004

Università Bocconi MILANO

How does Sentiment work? - 2

Rule-based analysis methods: *customized categories*

- Users can establish categories to manually classify a few results, which constitute the training set, and the rules that algorithms will then have to follow

- It offers a higher level of precision of results

    Limits: requires a significant investment in time, and very strict parameters for the classification of results, this approach returns a limited number of results

Università
Bocconi
MILANO

VADER (Valence Aware Dictionary and sEntiment Reasoner)

- Is a lexicon and rule-based sentiment analysis tool
    - The lexical features (e.g., words) are labelled according to their semantic orientation as either positive or negative
- Is specifically attuned to sentiments expressed in social media, performs very well with emojis, slangs, and acronyms in sentences

- VADER not only computes the Positivity and Negativity score, but also tells us about how positive or negative a sentiment is in the range [-4; 4]
- VADER provides four different elements of sentiment once inserted a word or phrase:
    - the negative polarity, the positive polarity, the neutrality
    - and a value called **compound** which calculates the sum of all the lexicon ratings which have been normalized between -1 (most extreme negative) and +1 (most extreme positive)

Hutto, Gilbert, VADER: A Parsimonious Rule-based Model for Sentiment
Analysis of Social Media Text, 2014

Università
Bocconi
MILANO

VADER key points

VADER analyses sentiments primarily based on:

- **Punctuation**: the use of an exclamation mark (!) increases the magnitude of the intensity without modifying the semantic orientation

- **Capitalization**: using upper case letters to emphasize a sentiment relevant word in the presence of other non-capitalized words, increases the magnitude of the sentiment intensity

- **Degree modifiers**: also called intensifiers, they impact the sentiment intensity by either increasing or decreasing the intensity (e.g. "extremely", "marginally")

- **Conjunctions**: use of conjunctions like "but" signals a shift in sentiment polarity, with the sentiment of the text following the conjunction being dominant. "The food here is great, but the service is horrible" has mixed sentiment, with the latter half dictating the overall rating

- **Preceding Tri-gram**: by examining the tri-gram preceding a sentiment-laden lexical feature, we catch nearly 90% of cases where negation flips the polarity of the text. A negated sentence would be "The food here isn't really all that great"

Università
Bocconi
MILANO

How does Sentiment work? - 3

Machine Learning based approach: *Sentiment & AI*

- Algorithms classify the contents on the base of entire sentence analysis (e.g., they are able to contextualize a tweet, a post or an article, and to accurately interpret the opinion of customers)

- They use **deep learning models** capable of simulating the cognitive functions of the human brain: technology is able to distinguish and understand complex linguistic structures, as well as entire sentences and simple forms of sarcasm and irony

    Limits: the accuracy of the results increases as well as the training set increases

Università
Bocconi
MILANO

AI tools

- NLTK SentimentAnalyzer
  machine learning approach with several modules and functions using both Liu and Vader lexicons

- TextBlob
  text processing Python library. The sentiment property returns a named tuple of the form Sentiment (polarity, subjectivity)

- Stanford CoreNLP
  deep learning model computes the sentiment based on how individual words change the meaning of longer phrases

Università
Bocconi
MILANO