Aspect Based Sentiment Analysis

Seminários 2015/2016

Ricardo Fusco

Escola de Ciências e Tecnologia
Universidade de Évora

January 15, 2016

UNIVERSIDADE DE ÉVORA

Summary

UNIVERSIDADE DE ÉVORA

Sentiment Analysis

- The sentiment analysis topic has been quite a "buzzing" one. And has lots of applications.
- It's a Natural Language Processing task used to analyze the polarity/sentiment linked to a given set of words/texts.
- Aspect based sentiment analysis refers to the identification of sentiments/opinions/polarity expressed in several aspects of entities or features in a given text.

Figure 1: Sentiment Analysis

UNIVERSIDADE DE ÉVORA

Problems That May be Encountered

- Dependence on the topic, meaning that sentiment prediction algorithms from a particular domain may be less accurate than others.
- It can be a quite hard and complicated task translating into a challenge for language technologies making it difficult to achieve acceptable results
- Disagreements by annotators on the classification given to an aspect, sentence or text can lead to problems.
- Multilingualism and cross-domain portability can also bring problems and make the task harder.
- Shorter and badly written texts can make the task increasingly harder.

UNIVERSIDADE DE ÉVORA

Differences from other types of Sentiment Analysis

- The possibility of capturing sentiments linked to certain specific objects of interest.
- The aspect based sentiment analysis methodologies are capable of identifying sentiment on a more fine grained way.
- Leads to error avoidance by processing terms which may be out of context.
- This more specific and detailed sentiment identification is less adequate for a more general sentiment analysis of a document/text.

Figure 2: Sentiment Analysis Polarity Classification

UNIVERSIDADE DE ÉVORA

Steps in Sentiment Analysis I

The sentiment analysis process is an intricate one which involves different steps to actually be able to analyze aspect based sentiments in a text.

- In a first phase it is necessary to collect the data from any user generated content anywhere in the web.
- Then the next steps are the sentiment identification and classification processes.
- Non-subjective information is removed and subjective information is maintained.

UNIVERSIDADE DE ÉVORA

Steps in Sentiment Analysis II

- After the useless information removal process comes the aspect extraction process and then the sentiment classification and polarity analysis phase.
- This data may be present in many different ways because of the different languages, different language syntaxes and semantics, contexts, slangs, etc.
- There's a need to clean the text identifying and removing all irrelevant content from said text.
- Extract and classify the data using natural language processing methods.

UNIVERSIDADE DE ÉVORA

Methods I

In order to do sentiment analysis one can use several methods, the main ones are:

- *Document level sentiment analysis*
  - Takes a document and evaluates it as a whole (as an overall theme) stating which is the associated sentiment.
- *Sentence level sentiment analysis*
  - Takes the sentences of a document and evaluates them stating for each sentence which is the associated sentiment.
- *Aspect level sentiment analysis*
  - Takes specific aspects of entities from sentences in documents and classifies them with a sentiment.

UNIVERSIDADE DE ÉVORA

Methods II

In order to do the sentiment classification process one can use several methods, the main ones are:

- *Machine learning approach*
  - A supervised or unsupervised learning process is done retrieving the features from the text.
  - Learns the model finding the polarity of sentiments based on trained and test data sets.
- *Subjective lexicon based approach*
  - Assigns a score to a list of words classifying each word with a specific sentiment.
  - Doesn't require any previous training of the data.
- *Hybrid approach*
  - This is an approach that uses both machine learning approaches and subjective lexicon based approaches.

UNIVERSIDADE DE ÉVORA

Methods III

- These are the mainly used methods within the 3 main ones discussed earlier.

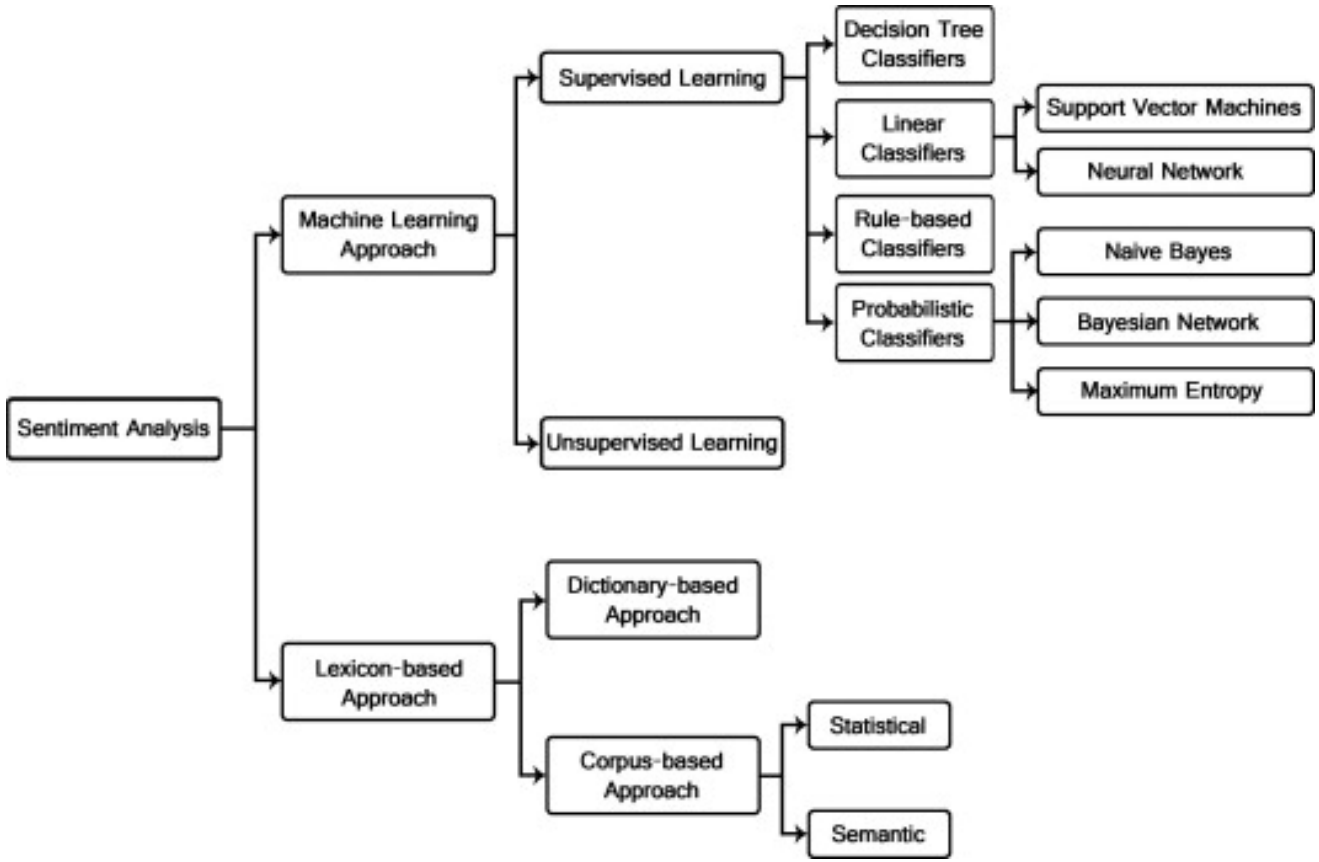

Figure 3: Sentiment Classification Techniques

UNIVERSIDADE DE ÉVORA

Advantages/Disadvantages - Machine Learning Approaches

- Advantages regarding ease of adaptation and training of data models for specific scenarios and contexts
- It is more adequate for older sets of data because the newer data creates the need to have labeled data.
- The labeled data is mostly obtained by collecting human judgments on pieces of unlabeled data and this "labeling process" can be quite expensive.

Figure 4: Machine Learning

UNIVERSIDADE DE ÉVORA

Advantages/Disadvantages - Subjective Lexicon Based Approaches

- Makes use of a sentiment dictionary holding several opinion words which are then associated with the data in order to determine the polarity.
- It provides a wider term coverage than other kinds of approaches.
- The lexicons hold a limited/restricted number of words, which may result in problems extracting sentiment from more dynamic environments.
- Some of this kind of approaches can be quite difficult and time consuming tasks

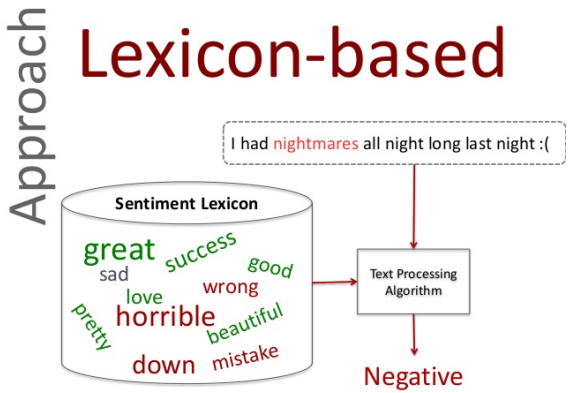

Figure 5: Machine Learning

UNIVERSIDADE DE ÉVORA

Advantages/Disadvantages - Hybrid Approaches

- May bring some potential improvements in terms of the sentiment classification process performance.
- These approaches carry significantly lesser sensibility to changes in the topic field/domain.
- Allows for an identification and assessment of sentiments at the concept/aspect level.
- The assignment of neutral sentiments to texts which may have quite a large amount of "noise"

UNIVERSIDADE DE ÉVORA

Tools Used I

Regarding the tools used for Sentiment Analysis, there are several of them that can be considered. Here are some of them:

- EMOTICONS
- Twitter Part-of-Speech Tagging tool can be used here for the sentiment analysis considering emoticons, smileys, nouns, adjectives, etc, from tweets
- Sentiment140 - tool for classifying tweets as positive, negative and neutral with an API that can be used
- NLTK - Natural Language Toolkit for building Python programs in order to be able to work with human language data and has a big amount of lexical resources.
- LIWC - Linguistic Inquiry and Word Count tool used for obtaining dictionary and sentiment classified categories.
- SentiStrength - Used for estimating the strength of positive and negative sentiment in short texts both for formal and informal language.

UNIVERSIDADE DE ÉVORA

Tools Used II

- SentiWordNet - It is a lexical resource publicly available for supporting sentiment classification and aspect mining applications based in the WordNet dictionary.
- WordNet - Dictionary which collects verbs, adjectives, nouns and puts them into sets called synsets.
- SenticNet - Natural language processing tool that recognizes the polarity and emotion at the semantic level (concept level semantic analysis).
- NRC - Sets of human-provided words with their correspondent sentiment tags.
- ANEW - Affective Norms for English Words word list
- WordNet-Affect - It is an extension of WordNet Domains, including a subset of synsets suitable to represent affective concepts correlated with affective words.
- ISEAR (International Survey On Emotion Antecedents And Reactions)- data set that contains reports on seven different emotions each by close to 3000 respondents in 37 countries on all 5 continents.

UNIVERSIDADE DE ÉVORA

Real World Applications

- Analysis of sentiment using texts is a relatively new and quickly growing area of study and application.
- It has great application value on several areas.
- Business
- Politics
- Real world events monitoring
- Public actions
- Finance
- Security

UNIVERSIDADE DE ÉVORA

Conclusions & Future Perspectives

- The way people express their opinions and sentiments has changed in unthinkable ways.
- Great technological innovation
- More and more means of communication like social networks, blogs, web communities, wikis and many other online collaborative means.
- Explosion of digital user generated content.
- Filtering and "purification" of data and knowledge retrieved from the unstructured and unlabelled information remains a central issue.
- There are lots of tools being used for these purposes making it relatively easy to try these approaches.
- Challenge: Data ambiguity, for instance, the cases where irony or sarcasm is identified within text may prove to be particularly difficult to separate from other contexts.

UNIVERSIDADE DE ÉVORA

Conclusions & Future Perspectives

The End

UNIVERSIDADE DE ÉVORA