Analysis on ABSA (Aspect Based Sentiment Analysis)

Ricardo Fusco

Universidade de Évora, Évora, Portugal
January 15, 2016
ricardo.fusco2@gmail.com

Abstract

Sentiment analysis or opinion mining has been quite a rising subject for some time gathering more and more interest and a widely used Natural Language Processing (NLP) and Data Mining task for analyzing the polarity/sentiment linked to a given set of words/texts, most specifically Aspect Based Sentiment Analysis in which a polarity(feeling, opinion, etc) is extracted from a particular aspect which can be a concept, a topic, entity, etc. This task of classifying and analyzing a text in natural language processing into a specific polarity/sentiment may be a quite strenuous and intricate one due to differences in text interpretation and to how the text was written. This is an essential task in natural language understanding which brings a vast array of real applications. The massive, and growing, amount of user generated opinions brings all kinds of service providers, companies, and areas of expertise a great need to analyze those opinions in order to take better conclusions out of the data and lead to better decisions overall.

Contents

1 Sentiment Analysis at the Aspect Level

The sentiment analysis task has been quite a developing one using Natural Language Processing and Data Mining techniques and seen as an extremely important task for several knowledge areas and businesses. Most of the current approaches to this task usually try to analyze the polarity/sentiment linked to a given set of words/texts, most specifically Aspect Based Sentiment Analysis in which a polarity(feeling, opinion, etc) is extracted from a particular aspect which can be a concept, a topic or entity.

As said before the feature/aspect based sentiment analysis refers to the identification of sentiments/opinions/polarity expressed in several aspects of entities or features, and classification of the polarity of a given text or feature/aspect in some document as positive, negative, or neutral assuming a sort of binary composition of opinions is assumed: for/against, like/dislike, good/bad etc.

However, an opinion may also be categorized into a neutral, positive or negative sentiment in a more basic level, and in a more advanced "beyond polarity" classification the aspects/features can be classified as emotional states like for instance "happy", "sad", "angry", "disgusted", "surprised", amongst others. The entity can be a product, a service, a person, a topic which has some aspects/features that represent the attributes of said entity.

1.1 Problems that may be encountered

A recurring problem with the several types of sentiment analysis is that it can be highly dependent on the topic meaning that sentiment prediction algorithms from a particular domain may be less accurate than others. It can be a quite strenuous and intricate task posing a big challenge for language technologies making it difficult to achieve acceptable results, the task of automatically classifying a text written in a natural language into a polarity/sentiment (positive, negative or neutral) can turn out to be quite complicated due to disagreements by annotators on the classification given to an aspect, sentence or text.

Multilingualism and cross-domain portability can also bring some problems being a very discussed challenge in the field. In the case of short texts for instance in social networks like Twitter, Instagram or Facebook in which texts tend to be shorter and, some times, badly written, the task becomes increasingly harder.

1.2 Differences from other types of Sentiment Analysis

One of the advantages of feature/aspect based sentiment analysis is the possibility of capturing sentiments linked to certain objects of interest. Distinct features can lead to different sentiment responses, for instance a restaurant can have great food but have terrible customer service. The aspect based sentiment analysis methodologies are capable of identifying sentiment on a more detailed and specific level leading to error avoidance by processing terms which may be out of context. This more specific and detailed sentiment identification is, however, not proper for the more general sentiment analysis of a document/text, just for specific objects inside the text.

2 Steps in Sentiment Analysis

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

The first step is usually the data collection necessary from any user generated content be it from websites like forums, blogs or social networks, opinion articles, news articles, emails, etc, and the preparation and treatment of the text/data gathered. 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 here to clean the text identifying and removing all irrelevant content from said text and use natural language processing methods to extract and classify the data.

Then the next steps are the sentiment identification and classification processes, sentences with factual information (non-subjective information) are removed and the non-factual information (subjective information) is maintained in order to proceed to the aspect extraction process and then to the sentiment classification and polarity analysis phase classifying aspects as positive, negative, neutral, etc.

Finally the last step is to show all the meaningful information that was taken from aspect based sentiment analysis of a text, this final information may be shown as charts or graphs, showing sentiment timelines, frequencies, percentages, etc.

3 Sentiment Analysis and Classification Methods

In order to do sentiment analysis one can use several methods, the main ones are: *document level sentiment analysis* which takes a document, for instance like an opinion document, and evaluates it as a whole (as an overall theme), stating if it expresses a positive, negative or neutral sentiment; *sentence level sentiment analysis* which takes the sentences of a document and evaluates them stating for each sentence if they express a positive, negative or neutral sentiment; *aspect level sentiment analysis* which takes specific aspects of entities from sentences in documents and classifies them with a sentiment, there can be different opinions for different aspects regarding the same entity. With the aspect level sentiment analysis one can focus its attention in the opinion on the aspect itself not focusing too much on the language construction of the text and the opinion here is composed of a sentiment attached to an opinion target.

Regarding the sentiment classification process there are several methods available like the holistic lexicon based approach, semantic orientation, machine learning, opinion polling, etc. The several methods can be seen in the next figure.

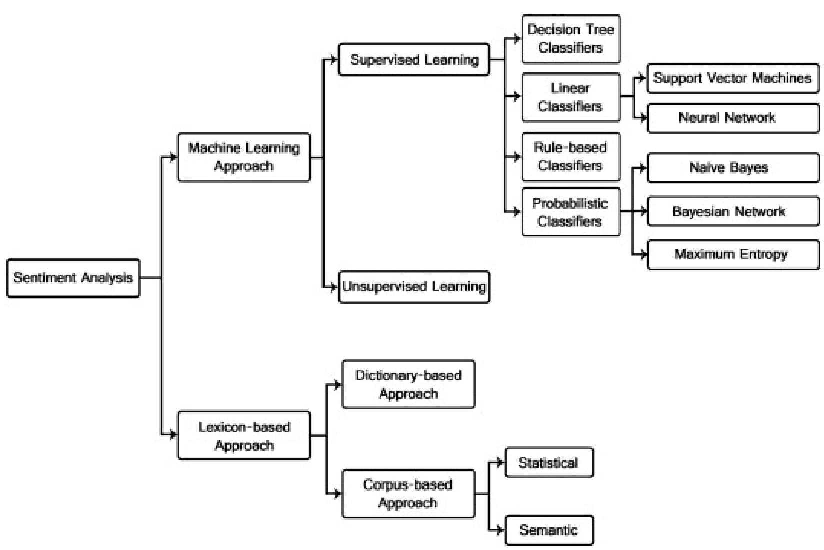

Figure 1: Sentiment Classification Techniques [9]

From the several approaches available for sentiment extraction there are some which are the most common. In the *machine learning approach* there can be performed a supervised (requires labeled data in order to train classifiers) or unsupervised (doesn't require the data to be previously trained) learning process by retrieving the features from the text and learning the model finding the polarity of sentiments based on trained and test data sets. The *subjective lexicon based approach* doesn't require any previous training of the data, assigning a score to a set of words in a list classifying each word with a specific sentiment(positive, negative or neutral). The *Hybrid approach*, n-gram modeling for instance makes use of uni-gram, bi-gram, tri-gram or any of these combined for the classification. This is an approach that uses both machine learning approaches and subjective lexicon approaches.

Within the *machine learning* methods the most used ones are *bayesian networks, naive bayes classification, maximum entropy, neural networks* and *support vector machine.* Amongst the *subjective lexicon* based approaches the most commonly used are the *Dictionary based approach, corpus based approach, Novel Machine Learning Approaches* and *Ensemble Approaches.*

3.1 Advantages/Disadvantages Between the Mainly Used Methods

Using machine learning sentiment classification approaches brings advantages in terms of easiness of adaptation and training of data models for specialized scenarios and contexts, but, on the other hand it just makes more sense for older data, because the newer data creates the need to have labeled data which means that every piece of said data (unlabeled data) must have a meaningful *label/class* or *tag* that gives some sort of desired information about every piece, and this labeled data is often obtained by collecting human judgments on pieces of unlabeled data and this "labeling process" can be quite expensive instead of having just raw unlabeled data.

Regarding the subjective lexicon based sentiment classification approaches, it makes use of a sentiment dictionary which holds several opinion words and they are then linked with the data in order to determine the polarity, it provides a wider term coverage than other kinds of approaches, but, on the other hand the lexicons hold a limited/restricted number of words which, in more dynamic sets, may result in problems extracting sentiment from these kinds of sets/environments, and the use of some of this kind of approaches can be quite difficult and time consuming tasks.

With the use of hybrid approaches, which are a combination of both the subjective lexicon based and machine learning ones, 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, there is a good compatibility in terms of lexicon based and learning based approaches combining these two approaches in a satisfactory way and allows for an identification and assessment of sentiments at the concept/aspect level, but, on the other hand it also carries its disadvantages like, for instance, the assignment of neutral sentiments to texts which may have quite a large amount of "noise", in other words, which may have irrelevant words for the chosen theme thus failing to deduct any sentiment in said irrelevant words.

4 Tools Used

Regarding the tools used for Sentiment Analysis, there are several of them that can be considered. One of the most used and most tempting tool to use for sentiment analysis is EMOTICONS contained in the texts mainly in social networks like Twitter or Facebook. A *Twitter Part-of-Speech Tagging* tool can be used here for the sentiment analysis considering emoticons, smileys, nouns, adjectives, etc, from tweets. There is also another tool for classifying tweets as positive, negative and neutral called *Sentiment140* which has an API for this purpose that can be used.

NLTK is used to deal with natural language processing as it is a solid toolkit/platform for building Python programs in order to be able to work with human language data. It possesses a big amount of lexical resources like WordNet along with several "text processing libraries for parsing", "tokenization", "tagging", "classification", "stemming", "semantic reasoning" and "wrappers for industrial-strength NLP libraries". This toolkit is available for Windows, Mac OS X and Linux, and it is and open source project.

LIWC is a *Linguistic Inquiry and Word Count* tool and a powerful research and learning one used for obtaining dictionary and sentiment classified categories. The LIWC program basically receives a text and counts the percentage of words which transmit different kinds of sentiments or opinions, thinking styles, social concerns, emotions.

There is a tool that is also used regarding the estimation of the strength of positive and negative sentiment in short texts both for formal and informal language, it's called *SentiStrength* and it uses a sentiment lexicon approach for attributing scores to positive and negative phrases in a given text. Besides this tool there is a similar one called *SentiWordNet* which is a lexical resource publicly available for supporting sentiment classification and aspect mining applications, using semi-machine learning approaches, and it is based on a dictionary called *WordNet* which collects verbs, adjectives, nouns and puts them into sets called synsets. In order to recognize the polarity and emotion at the semantic level (concept level semantic analysis) there is a natural language processing tool called *SenticNet* for this purpose.

If sets of human-provided words with their correspondent emotional tags are needed one can fetch *NRCs* which are basically large sets of said words with the sentiment tags attached. And other of the mainly used tools for concept-level sentiment analysis use knowledge bases such as *ANEW*, *WordNet-Affect* and *ISEAR* amongst others.

5 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 areas such as business, politics, real world events monitoring, public actions, finance, security, it can be used to analyze differences of opinions regarding some specific products. There is a huge amount of news items, articles, tweets, blogs about an even bigger amount of themes and different areas of expertise. For instance in the stock market, a sentiment analysis system make use of the gathered data and attach the corresponding sentiment about it as a single score which can, in turn, be used by an automated trading system much like the Stock Sonar (`http://datamarket.azure.com/dataset/digitaltrowel/sentimentservice`) that shows us the daily positive and negative sentiments about each stock by using graphs containing said sentiments and the respective price for the stock.

6 Conclusion

The way people express their opinions and sentiments has changed in unthinkable ways at the course of the past years thanks to the great technological innovation bringing about great changes, bringing more and more means of communication like social networks, blogs, web communities, wikis and many other online collaborative means. It all contributed to an explosion of digital user generated content, as shown in the next figure, bringing whole new possibilities and new areas of expertise in order to improve other ones.

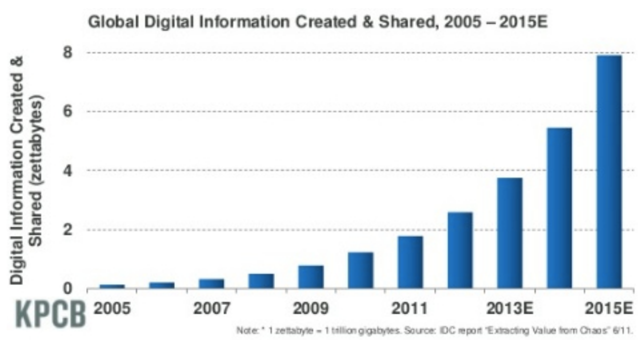

Figure 2: Digital User Generated and Shared Content

The filtering and "purification" of data and knowledge from the huge amounts of unstructured and unlabelled information found on the Web remains a central issue for many businesses, areas of expertise and individuals. And there are lots of tools being used for these purposes making it relatively easy to try on the aspect based sentiment analysis task using any theme/-topic.

One of the main challenges in applying these tools and approaches may be the 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 but surely the approaches and tools used shall continuously evolve in the future.

References

[1] SemEval-2014 Task 4. *Task Description: Aspect Based Sentiment Analysis (ABSA)*. http://alt.qcri.org/semeval2014/task4/, 2014.

[2] Abdulaziz Alghunaim, Mitra Mohtarami, Scott Cyphers, and James Glass. *A Vector Space Approach for Aspect Based Sentiment Analysis*. https://groups.csail.mit.edu/sls/publications/2015/Alghunaim_NAACL-15.pdf, 2015.

[3] Oscar Araque, Ignacio Corcuera, Constantino Roman, Carlos A. Iglesias, and J. Fernando Sanchez-Rada. *Aspect based Sentiment Analysis of Spanish Tweets*. http://ceur-ws.org/Vol-1397/gsi.pdf, 2015.

[4] Erik Cambria. *An Introduction to Concept-Level Sentiment Analysis*. http://sentic.net/concept-level-sentiment-analysis.pdf, 2013.

[5] Alessia D'Andrea, Fernando Ferri, Patrizia Grifoni, and Tiziana Guzzo. *Approaches, Tools and Applications for Sentiment Analysis Implementation*. http://www.ijcaonline.org/research/volume125/number3/dandrea-2015-ijca-905866.pdf, 2015.

[6] Jussi Karlgren, Magnus Sahlgren, Fredrik Olsson, Fredrik Espinoza, and Ola Hamfors. *Usefulness of Sentiment Analysis*. https://www.diva-portal.org/smash/get/diva2:589140/FULLTEXT01.pdf.

[7] Bing Liu, Minqing Hu, and Junsheng Cheng. *Opinion Observer: Analyzing and Comparing Opinions on the Web*. https://www.cs.uic.edu/~liub/publications/www05-p536.pdf, 2005.

[8] MeaningCloud. *An Introduction to Sentiment Analysis (Opinion Mining)*. https://www.meaningcloud.com/blog/an-introduction-to-sentiment-analysis-opinion-mining-in-meaningcloud/, 2015.

[9] Walaa Medhat, Ahmed Hassan, and Hoda Korashy. *Sentiment analysis algorithms and applications: A survey*. http://ac.els-cdn.com/S2090447914000550/1-s2.0-S2090447914000550-main.pdf?_tid=c3b67146-baef-11e5-97ed-00000aacb35f&acdnat=1452797651_7f5cd61d5e195f1eb5e1274a5c874211, 2013.

[10] AKayhan Moharreri, Royden Lewis, and Steven Ware. *Sentiment analysis using emoticons*. http://web.cse.ohio-state.edu/~moharrer/PDFs/Moharreri_Ware_Lewis.pdf.

[11] Andrius Mudinas, Dell Zhang, and Mark Levene. *Combining Lexicon and Learning based Approaches for Concept-Level Sentiment Analysis*. http://wan.poly.edu/KDD2012/forms/workshop/WISDOM2012/camera_ready/a5-mudinas.pdf, 2012.

[12] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. *Thumbs up? Sentiment Classification using Machine Learning Techniques*. http://www.cs.cornell.edu/home/llee/papers/sentiment.pdf, 2002.

[13] Ioannis (John) Pavlopoulos. *Aspect Based Sentiment Analysis*. http://www.aueb.gr/users/ion/docs/pavlopoulos_phd_thesis.pdf, 2014.

[14] Maria Pontiki, Dimitrios Galanis, Haris Papageorgiou, Suresh Manandhar, and Ion Androutsopoulos. *SemEval-2015 Task 12: Aspect Based Sentiment Analysis*. http://www.aclweb.org/anthology/S15-2082, 2015.

[15] NLTK Project. *NLTK*. http://www.nltk.org/.

[16] José Saias. *Senti.ue: Tweet Overall Sentiment Classification Approach for SemEval-2014 Task 9*. http://www.aclweb.org/anthology/S14-2095, 2014.

[17] Ankit Singh and Md. Enayat Ullah. *Aspect Based Sentiment Analysis*. http://home.iitk.ac.in/~enayat/files/absa_report.pdf, 2015.

[18] Mike Thelwall and Kevan Buckley. *Topic-Based Sentiment Analysis for the Social Web: The role of Mood and Issue-Related Words*. http://scitsc.wlv.ac.uk/~cm1993/papers/SentiStrength3Preprintx.pdf, 2012.

[19] Bo Wang and Min Liu. *Deep Learning for Aspect-Based Sentiment Analysis*. https://cs224d.stanford.edu/reports/WangBo.pdf, 2015.

[20] Wikipedia. *Sentiment analysis*. https://en.wikipedia.org/wiki/Sentiment_analysis.