

Original Research Paper

Sentiment Analysis of French Tweets based on Subjective Lexicon Approach: Evaluation of the use of OpenNLP and CoreNLP Tools

Abdelkader Rhouati, Jamal Berrich, Mohammed G. Belkasmı and Toumi Bouchentouf

Team SIQL, Laboratory LSEII, ENSAO Mohammed First University, Oujda, Morocco

Article history

Received: 14-11-2017

Revised: 21-4-2018

Accepted: 19-6-2018

Corresponding Author:

Abdelkader Rhouati
Team SIQL, Laboratory LSEII,
ENSAO Mohammed First
University, Oujda, Morocco
Email:
abdelkader.rhouati@gmail.com

Abstract: Nowadays, sentiment analysis is becoming a very important issue of research. This paper present experimentation on sentiment analysis based on subjective lexicon method. This experimentation is tested over French tweets using "Public Opinion Knowledge (POK)" platform. POK is a platform consists in getting public opinion orientation from text extracted from social network and blogs, which we have developed and presented in previous papers. There are three algorithms as classifiers, which are based on Natural Language Processing Tools. The first is based on OpenNLP, the second on CoreNLP and the third on dependency analysis implemented by CoreNLP. Each classifier consists of three steps, which are Part of Speech Tagging (POS), word polarity classification and sentiment classification algorithm. On the one hand, the results are used to evaluate the use of OpenNLP and CoreNLP, on other, they draw to make a comparison between lexicon and machine-learning approaches. So, experimentation leads us to conclude that tools of sentiment analysis based on lexicon are much performant than those based on machine learning and they can reach a rate of precision of 70% and F-measure of 0.7. Also, we conclude that CoreNLP is more efficient than OpenNLP by 3% of precision, this fact is due to the efficiency of Part of Speech tagging algorithms.

Keywords: Sentiment Analysis (SA), Natural Language Processing (NLP), Opinion Mining, OpenNLP, CoreNLP, Part of Speech (POS) Tagging, Public Opinion Knowledge, Machine Learning

Introduction

Social Networks have evolved and become the first source of information. People post a real-time message about their sentiments on subjects that concern them. All those messages constitute a huge amount of information that can be used in many fields and many ways, for example to get feedback of product and service, to measure the popularity of a brand or even as aid of politician to make adequate decisions. In fact, this kind of information represents the public opinion. Today the most popular way to get public opinion is by doing surveys, which requires much effort and time. In addition, most of the time surveys do not reflect the reality of what people think. So, our main goal is to propose a new way of getting public opinion based on social-network content and a Big Data approach (Rhouati *et al.*, 2016b).

Solving the problem of the public opinion through social network content leads us to a more precise problem

which is the measure of sentiment expressed by a given sentence or text. This is called Sentiment Analysis.

Sentiment Analysis may be approached by several ways. In this paper, we will interest on subjective lexicon method as one of the useful technique. We describe our a model for sentiment classification of French tweets into positive, negative and neutral categories. We present an experimentation of this model using two Natural Language Processing tools, which are openNLP and coreNLP. The experimentation leads to a comparative study of the two solutions.

Our model follows three steps three steps. The first one consists of Part of Speech tagging on tweets using Natural Language Processing Tool. This step transforms a tweet on grammatical structure by tagging every word depending on his role and position in the sentence. The second step is to determine the polarity of each word, this is means that for every word we indicate if it is negative, positive or neutral. This is done using a

polarity word dictionary. The third and last step, is a sentiment classifier. This step applies an algorithm classification to deduce the sentiment of tweet using the two inputs POS Tag and word's polarity.

The rest of the paper is organized as follows. In section 2, we make a review of existing research on sentiment analysis. In section 3, we describe the platform of experimentation, the tools and the approach applied. In section 4 and 5 we present respectively the results and the analysis of these results approach. We conclude and give future directions to further research in section 6.

A Review of the Literature

Sentiment Analysis (SA) (Dhuria, 2015), also named as Opinion-Mining, is a classifier of polarity of a given text or document. In fact, it's a technique to determine if an opinion, expressed by a given text, is positive, negative or even neutral. This field of research have many area of application as the improvement of product's quality, give recommendation, aid of making personal or political decision.

The Existing Work in Sentiment Analysis

Sentiment Analysis can be approached from different angles (Prabowo and Thelwall, 2009). Some researchers focus on assigning sentiment to entire document, others work on finding the sentiments of words (Hatzivassiloglou and McKeown, 1997), expressions (Wilson *et al.*, 2005; Kim and Hovy, 2004), sentences (Pang and Lee, 2004) and even a topics (Yi *et al.*, 2003). Various technical approaches were developed for this purpose (Mohamed Hussein, 2016; Prabowo and Thelwall, 2009; Vinodhini and Chandrasekaran, 2012), the most known are:

- Subjective lexicon: this approach is based on machine translation using specific dictionaries, in which every word is assigned to a score that determine its polarity (positive, negative or neutral) (Liu, 2010; Melville *et al.*, 2009). This technique uses Natural Language Processing (NLP) to understand the human language expressed by text
- Machine learning: This technique can be Supervised, Semi-Supervised or unsupervised learning (Turney, 2002). This approach gets its model from extracting features from the data itself. Several methods exist such as Naive Bayes (NB), Maximum Entropy (ME), Support Vector Machine (SVM) (Joachims, 1998) and Deep learning (LeCun *et al.*, 2015)

Subjective Lexicon Method using Natural Language Processing Techniques and Dictionary

The final goal of Natural Language Processing (NLP) Techniques is to make the computer able to understand

human language (Allen, 1987). It's a very difficult field for different reasons which are related all to ambiguity in languages. In fact, the Ambiguity of a language is due to many levels as phonetic, lexical, syntax, semantic or pragmatic (the use of irony or metaphor). To understand that lets look to the sentence "*il est à cheval sur ses principes*" (he is riding on his principles). If you depend only on the meaning of words while analyzing this sentence the opinion it conveys will be neutral, but if you try to consider the metaphor used, the meaning will then be that "he respect his principles" and it's clearly a positive opinion. So today and to make things easier, many researches are based only on the syntactic representation of text using techniques that are based mainly on analyzing words (Cambria and White, 2014).

Natural Language Processing (NLP) involves several different techniques which useful to sentiment analysis. The most common technique is Part of Speech (POS) Tagging (Toutanova and Manning, 2000; Toutanova *et al.*, 2003; <https://nlp.stanford.edu/software/tagger.shtml>). Part of Speech (POS) tagging is a process of labeling words, for example determine if the word "Joli" is adjective, noun or verb. Efficient tagging must consider also the word's context, such as surrounding words and its position in the sentence.

Sentiment Analysis require the ability to calculate the sentiment of each word. Many existing dictionaries provides this functionality for French Language as "Feel" dictionary (<http://advanse.lirmm.fr/feel.ph>) that we have used on our experimentations and SentiWordNet (Esuli and Sebastiani, 2006) which is not so efficient in French language. Generally, these dictionary associates three scores of polarity of each word: 0 for neutral, 1 for positive and -1 for negative. After that a "Sentiment Classification Process" must be applied to texts which use the polarity and POS Tag of each word to determine the sentiment of the whole text.

The main challenge of Sentiment analysis is finding the orientation of words which can be positive or negative. However, some words change orientation depending on contexts and sentences. For example, the word "*Joli*" (Beautiful) in the sentence "*Cette photo est jolie*" (this picture is beautiful) expresses a positive opinion, but the same word in the sentence "*Cette photo n'est pas jolie*" express the opposite opinion. At the same time, the spelling mistakes and abbreviation of words must be handled and all that could retrieve the correct orientation. This problem will not be treated in this article. We focus our work on the use of Subjective Lexicon method using static dictionaries of words polarity.

Platform of Experimentation

Our experimentations have been done on a supervised and manually constructed dataset and using BigData platform called "Public Opinion Knowledge (POK)". In this chapter, we give more details.

DataSet

We chose to work on data from Twitter for several reasons. First, the tweets are limited to 140 characters, which contains an average of 14 words. Sentiment analysis in tweets is simpler view that tweets are shorter. The second reason is the availability of data. With the usage of Twitter API, we can collect a million tweets for training purpose. The third and last reason is that tweets contain acronyms, abbreviations and elongated words. Other features such as URL, image, hashtags, punctuations and emoticons are included as well. Most of these features affect the accuracy of analysis process as they are not proper text that can be found in dictionary. This is a perfect context for our study, since it contains quite a variety of texts.

For our work, we use a supervised dataset. We used Twitter API to retrieve more than 3000 tweets and we processed to a manually classification to finally have two databases of French tweets: The first one contains 1998 positive tweets and the second covers 898 negative tweets.

Public Opinion Knowledge (POK) Platform

The "Public Opinion Knowledge" platform is based on approach based on Big Data (Rhouati *et al.*, 2016b). This approach is explained in the Fig. 1.

So, the implemented approach is conducted through four steps:

- Data source: consists of the extraction of data from several web sources
- Data management: consists in modeling data and proceed to store it on a NoSQL storage platform
- Modeling: consists in using a Web mining process to analyze data

- Result: consists in visualizing the results and distinguish the positive and negative opinion of people

The functional design of the POK platform (Rhouati *et al.*, 2016a) is based on distributed computing to addressing the problem of massiveness of data to process. After the step of extracting data (articles and comments) from blogs on the web and directly from database of CMS, then saving this data in a Big Data database, we will apply an algorithm of Web Mining to deduce the public opinion from all stored data. To optimize the Web Mining treatment, a distributed system of several machines will be used. The Fig. 2 explains the implementation of 4 steps of the approach.

Natural Language Processing Tools

In our tests, we use the following two Natural Language Processing tools for a comparative study.

The Stanford CoreNLP

The Stanford CoreNLP `citernanning2014 stanford` (<https://stanfordnlp.github.io/CoreNLP/>) is a natural language parser developed by The Stanford Natural Language Processing Group. This tool uses probabilistic methods to work out parsing for sentences. It makes possible to represent sentences in a grammatical structure. The Part of Speech tags used for coreNLP is on Table 1 and the techniques used are detailed on (Klein and Manning, 2003; Klein and Manning, 2003).

To illustrate the operation of the parser, the Fig. 3 is an example of analysis of the sentence "*Cette photo est jolie*" (this picture is beautiful).

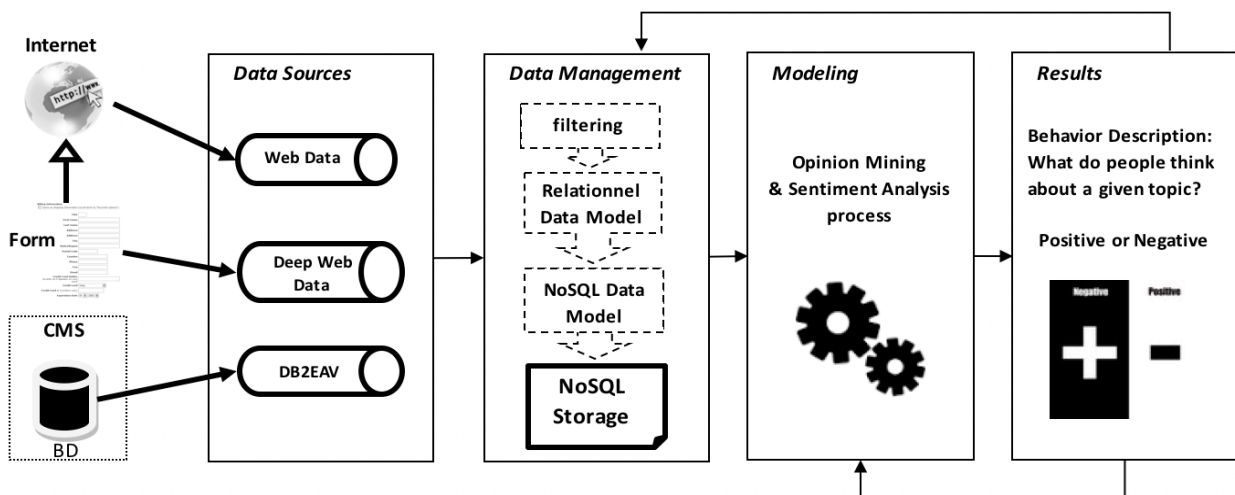


Fig. 1: A new approach to get Public Opinion from web content based on Big Data

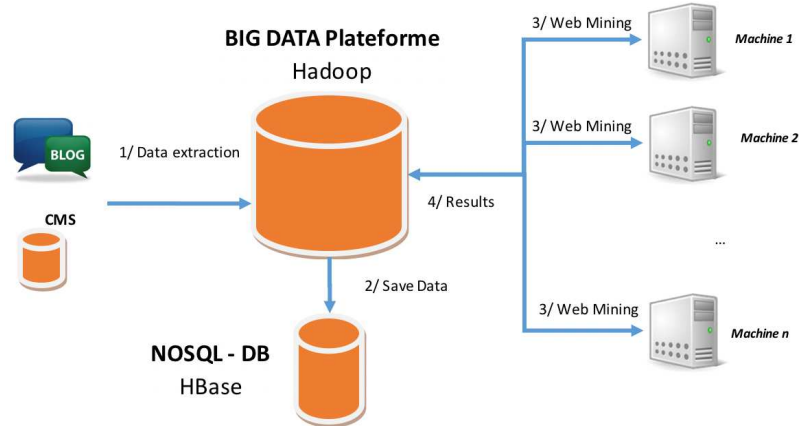


Fig. 2: The functional design of POK platform using a distributed system

Table 1: Part of Speech Labels

| labels/tagds | CoreNLP | OpenNLP |
|---|---------|---------|
| adjective | ADJ | ADJ |
| Interrogative adjective | ADJWH | ADJWH |
| Adverb | ADV | ADV\ |
| Interrogative adverb | ADVWH | - |
| conjunction and subordinating conjunction | CC | CC |
| weak clitic pronoun | CL | - |
| subject clitic pronoun | CLO | CLO |
| reflexive clitic pronoun | CLR | CLR |
| object clitic pronoun | CLS | CLS |
| subordinating conjunction | CS | CS |
| determiner | DET | DET |
| interrogative determiner | DETWH | DETWH |
| foreign word | ET | ET |
| interjection | I | I |
| noun | NC | NC |
| proper noun | NPP | NPP |
| preposition | P | P |
| prefix | PREF | PREF |
| strong pronoun | PRO | PRO |
| relative pronoun | PROREL | PROREL |
| interrogative pronoun | PROWH | PROWH |
| verb | V | V |
| infinitive verb | VINF | VINF |
| imperative verb | VIMP | VIMP |
| past participle verb | VPP | VPP |
| present participle verb | VPR | VPR |
| subjective verb | VS | VS |
| punctuation | PUNC | PONCT |

The vigor of the coreNLP tool is the ability to perform a dependency analysis on a given text. The dependency analysis uses the sentence tree to determine the dependency of the different words of other words in the same sentence. This analysis is used to improve the sentiment algorithm classification. So, the sentiment of a sentence can be determined by the sum of the polarities of all group of dependent words, instead of the sum of the polarities of separated words.

Sentence Cette photo est jolie

POS Tagging

Cette / DET
 photo / NC
 est / V
 jolie / ADJ
 . / PUNC

Parse

(ROOT
 (SENT
 (NP (DET Cette) (NC photo))
 (VN (V est))
 (AP (ADJ jolie))
 (PUNC .)))

Fig. 3: An example parse of sentence by CoreNLP

The Fig. 4 shows an example of this technique in practice on the phrase "*Le poulet grillé préparé par elle est dégueulasse*" (The grilled chicken prepared by her is disgusting).

Apache OpenNLP

The Apache OpenNLP library (<http://opennlp.apache.org/>) is a toolkit for the processing natural language text. It supports the most common Natural Language Processing tasks, such as tokenization and part-of-speech tagging and uses probabilistic methods for parsing text. In brief OpenNLP Offers the required features to build more advanced text processing services.

OpenNLP uses a universal Part-of-Speech Tagset (Petrov and McDonald, 2011) detailed in Table 1.

To illustrate the operation of the parser, the Fig. 5 is an example of analysis of the sentence ""*Cette photo est jolie*" (this picture is beautiful).

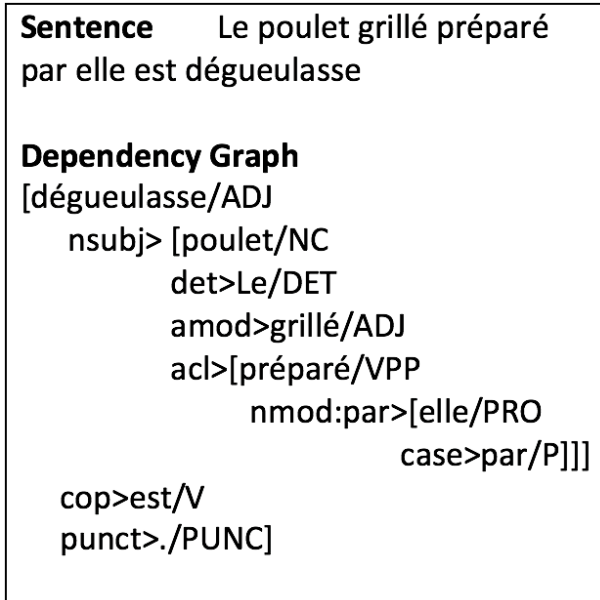


Fig. 4: An example output of dependency analysis on sentence using CoreNLP

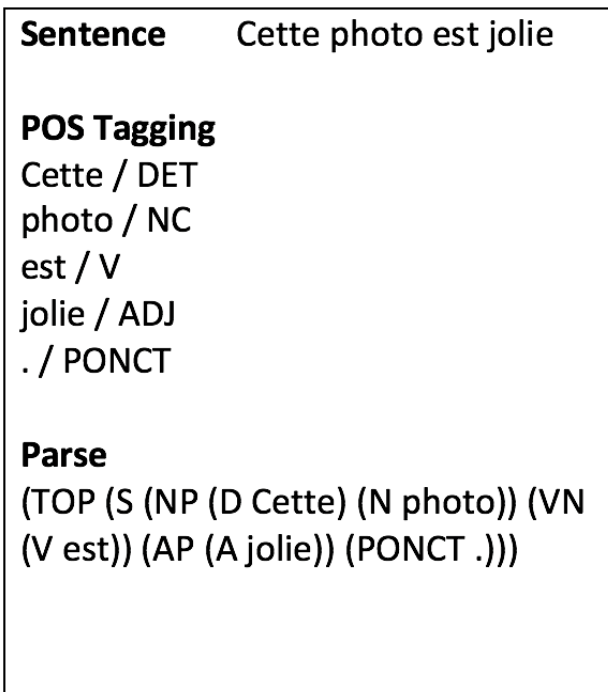


Fig. 5: An example output of parsing sentence using openNLP

Brief Comparative Discussion: CoreNLP vs. OpenNLP

Overall, OpenNLP and CoreNLP offer the same basic functionalities of a natural language processing tool, as shown in the Fig. 2. In terms of Part of Speech tagging, the results produced by these two tools are very similar. The difference between their results may be due to their own tokenizer. Stanford CoreNLP tokenizer is better in handling punctuation. As a result, Stanford CoreNLP can have superior accuracy.

At the level of Training API the OpenNLP is easier in use when it comes to existing models. But, if you want to build a new model from a given dataset set using a training process, both should be at the same level of complexity.

Experiments

Test Scopus

Our experimentation consists in the analysis of feelings of French tweets. Indeed, we apply a sentiment algorithm classification, based on lexicon approach, on two supervised databases. The first database containing only positive tweets and the second containing negative tweets. The tests are done using the BigData platform POK (Rhouati *et al.*, 2016a). In addition, we have implemented two Natural Language Processing tools, CoreNLP (<https://stanfordnlp.github.io/CoreNLP/>) and OpenNLP (<http://opennlp.apache.org/>), on POK platform and we have done the tests twice with both tools and using the same classification algorithm in order to get out with a comparative study. Another test is done with a new classification algorithm based on a dependency feature offered by CoreNLP tools. This feature is unfortunately not existing on OpenNLP. The Table 2 illustrates the functionalities offered by each tool.

Finally, we have applied a machine-learning algorithm using WEKA (2017) for the same databases to verify the efficient of our proposed system compared to other techniques.

Table 2: Features comparison between openNLP and CoreNLP applied on French language

| Features | CoreNLP | OpenNLP |
|--------------------|---------|---------|
| Tokenize/Segment | X | X |
| Sentence Split | X | X |
| Part of Speech | X | X |
| Lemma | | X |
| Named Entities | | X |
| Dependency Parsing | X | |
| Sentiment Analysis | | |

Results

Analysis and Future Works

Table 3 and 4 show how the use of Natural Language Processing Tools can impact the results of sentiment analysis. In fact, with the same algorithm classifier we had better results using CoreNLP comparing with openNLP. We see an improvement of 3%. The precision is respectively 68% and 49% for positive and negative dataset using CoreNLP and 67% and 41% using OpenNLP. It's related to the efficiency of POS Tagging of each Natural Language Processing tool.

Applying either a new algorithm classifier based on dependency analysis features offered by CoreNLP, gives more than 6% of improvement as shown on table 5. So, the precision of this new algorithms is 72% for positive dataset and 56% for negative one. We also did analysis of the same datasets using Weka and a machine learning classifier. The results are presented in Tables 6 and 7.

Finally, all results are resumed in Table 8 and Figure 6. In addition and for more precision, we calculate the recall and F-measure of each classifier. The classifier that uses dependency analysis of text has the same F-measure as classifier using C4.5 decision tree. Both classifiers have a F-Measure nearby 0,7. However, the classifier using SVM is more efficient, with F-Measure equals to 0,777. We Conclude that dependency based classifier is efficient as C4.5 machine learning classifier. However, to improve this result and come closer to SVM Classifier, two axes are to be explored. First Axe is related to Part of Speech (POS) Tagging tool. This is essential and can impact positively or negatively the results. The second one is about the dictionary. Lexicon classifiers are mainly based on polarity of words, that is retrieved from specific dictionaries. A word which is not found in the dictionary is considered as a neutral one, which misleads the analysis of the sentence or the text. This last axe will be the subject of our next work. We will focus in our future work on how to enrich the initial dictionary with a new word that we meet during the analysis process.

Table 3: Results of Sentiment Classification from POK Platform using OpenNLP

| Dataset | Nbr Positive | Nbr Negative | Nbr Neutral | Precision |
|----------|--------------|--------------|-------------|-----------|
| positive | 1343 | 277 | 376 | 67% |
| negative | 306 | 377 | 215 | 41% |

Table 4: Results of Sentiment Classification from POK Platform using CoreNLP

| Dataset | Nbr Positive | Nbr Negative | Nbr Neutral | Precision |
|----------|--------------|--------------|-------------|-----------|
| positive | 1365 | 351 | 280 | 68% |
| negative | 325 | 442 | 131 | 49% |

Table 5: Results of Sentiment Classification from POK Platform using CoreNLP with dependency analysis

| Dataset | Nbr Positive | Nbr Negative | Nbr Neutral | Precision |
|----------|--------------|--------------|-------------|-----------|
| positive | 1446 | 440 | 110 | 72% |
| negative | 367 | 508 | 23 | 56% |

Table 6: Results of Sentiment Classification from Weka using SVM classification

| Dataset | Nbr Positive | Nbr Negative | Nbr Neutral | Precision |
|----------|--------------|--------------|-------------|-----------|
| positive | 1740 | 257 | 0 | 87,1% |
| negative | 376 | 522 | 0 | 58,1% |

Table 7: Results of Sentiment Classification from Weka using C4.5 decision tree classification

| Dataset | Nbr Positive | Nbr Negative | Nbr Neutral | Precision |
|----------|--------------|--------------|-------------|-----------|
| positive | 1698 | 299 | 0 | 76.7% |
| negative | 517 | 381 | 0 | 56% |

Table 8: Summarized Result for the different Sentiment Analysis experimentation

| Systems | Precision ¹ | Recall ² | F-measure ³ |
|-------------------------------|------------------------|---------------------|------------------------|
| POK using OpenNLP | 59% | 74.9% | 0.660 |
| POK using CoreNLP | 62% | 72.8% | 0.669 |
| POK using CoreNLP dependency | 67% | 71,6% | 0.692 |
| Weka using C4.5 decision tree | 70.3% | 71.8% | 0.706 |
| Weka using SVM | 77.5% | 78.1% | 0.777 |

¹This is a measure of how often a sentiment analysis is correct

²This is a measure of how many documents with sentiment were rated as sentimental.

³THIS is a combination of precision and recall. For more information (Makhoul *et al.*, 1999)

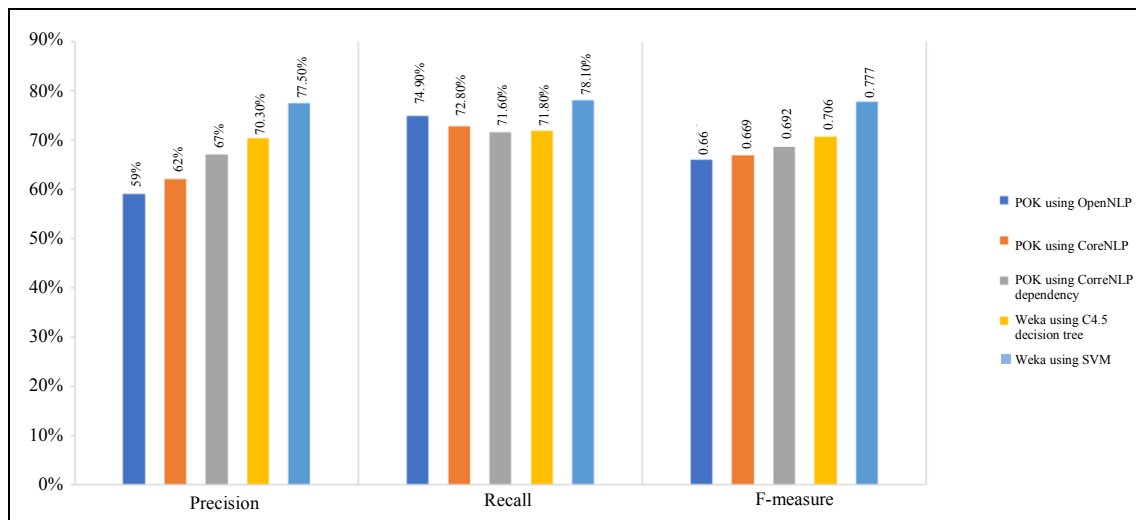


Fig. 6: Summarized Result for the different Sentiment Analysis experimentation

Conclusion

In this paper, we presented the general context and problematic of getting public opinion from text on the web while focusing on sentiment analysis based on lexicon approach. Basically, this approach consist of implementing a classification algorithms that use Natural Language Processing tool and Sentimental dictionary to determine if a given text express a positive, negative or neutral opinion. The main goals behind this work is to evaluate this approach. We also presented the experimentations made; using Public Opinion knowledge Platform and Natural Language Processing Tools applied on French Tweets. So, in the last part we have discussed the results of this experimentations and carried out a comparative study between lexicon approach and machine learning approach. We draw a conclusion that the lexicon approach is as efficient as machine learning techniques as C4.5 decision tree. However, we quoted two axes to improve our technique to have more efficient results: revise of Part of Speech tagging techniques used by Natural Language Processing and look for ways to have dictionaries with more intelligence and able to include new words not taken into account.

Acknowledgement

I would like to thank especially Pr. Toumi Bouchentouf for his precious guidance during the realization of this work, as well as all members of the SIQL research team.

Author's Contributions

Abdelkader Rhouati: Done in all experiments, coordinated the data-analysis and contributed to the writing of the manuscript.

Jamal Berrich: Participated in all experiments, coordinated the data-analysis and contributed to the writing of the manuscript.

Mohammed G. Belkasmi: Participated in all experiments, coordinated the data-analysis.

Toumi Bouchentouf: Supervised all the work and defined the plan and the axes of research.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

References

- Allen, R.B., 1987. Several studies on natural language and back-propagation. Proceedings of the IEEE 1st International Conference on Neural Networks, IEEE Piscataway, NJ, pp: 341.
- Cambria, E. and B. White, 2014. Jumping nlp curves: A review of natural language processing research. IEEE Computational Intelligence Magazine, 9: 48-57.
- Dhuria, S., 2015. Sentiment analysis: An approach in natural language processing for data extraction. Int. J. New Innovations Engineering Technology.
- Esuli, A. and F. Sebastiani, 2006. Sentiwordnet: A high-coverage lexical resource for opinion mining. Institute of Information Science and Technologies (ISTI) of the Italian National Research Council (CNR).
- Hatzivassiloglou, V. and K.R McKeown, 1997. Predicting the semantic orientation of adjectives. Proceedings of the 8th Conference on European Chapter of the Association for Computational Linguistics (ACL' 09), Association for Computational Linguistics, pp: 174-181.

- <http://advanse.lirmm.fr/feel.php>
<http://opennlp.apache.org/>
<https://nlp.stanford.edu/software/tagger.shtml>
<https://stanfordnlp.github.io/CoreNLP/>
- Joachims, T., 1998. Text categorization with support vector machines: Learning with many relevant features. *Machine Learning: ECML-98*, pp: 137-142.
- Kim, S.M. and E. Hovy, 2004. Determining the sentiment of opinions. *Proceedings of the 20th International Conference on Computational Linguistics*, Aug. 23-27, ACM, Geneva, Switzerland. DOI: 10.3115/1220355.1220555
- Klein, D. and C.D. Manning, 2003. Accurate unlexicalized parsing. *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1*, Association for Computational Linguistics, pp: 423-430.
- Klein, D. and C.D. Manning, 2003. Fast exact inference with a factored model for natural language parsing. *Advances in Neural Information Processing Systems*, pp: 3-10.
- LeCun, Y., Bengio, Y. and G. Hinton, 2015. *Deep Learning Nature*, 521: 436-444.
- Liu, B., 2010. Sentiment analysis and subjectivity. *Handbook Natural Language Processing*, 2: 627-666.
- Makhoul, J., F. Kubala, R. Schwartz and R. Weischedel, 1999. Performance measures for information extraction. *Proceedings of the DARPA Broadcast News Workshop*, pp: 249-252.
- Melville, P., W. Gryc and R.D. Lawrence, 2009. Sentiment analysis of blogs by combining lexical knowledge with text classification. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, (KDDM' 09)*, ACM, pp: 1275-1284.
- Mohamed Hussein, D.M.E.D., 2016. A survey on sentiment analysis challenges. *J. King Saud University-Engineering Sci.*
DOI: 10.1016/j.jksues.2016.04.002
- Pang, B. and L. Lee, 2004. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. *Proceedings of the 42nd annual meeting on Association for Computational Linguistics*, Association for Computational Linguistics, pp: 271.
- Petrov, S., D. Das and R. McDonald, 2011. A universal part-of-speech tagset. *arXiv preprint arXiv:1104.2086*.
- Prabowo, R. and M. Thelwall, 2009. Sentiment analysis: A combined approach. *J. Informetrics*, 3: 143-157.
- Rhouati, A., E.H. Ettifouri, M.G. Belkasmı and T. Bouchentouf, 2016a. Public opinion knowledge (pok) platform based on apache hadoop: To get public opinion from french content published on the web/csm. *Proceedings of the 2016 2nd International Conference on Cloud Computing Technologies and Applications (CloudTech' 16)*, IEEE, pp: 132-139.
- Rhouati, A., E.H. Ettifouri, M.G. Belkasmı and T. Bouchentouf, 2016b. Get the public opinion from content published on the web/csm: new approach based on big data. *Proceedings of the Mediterranean Conference on Information and Communication Technologies (ICT' 16)*, Springer, pp: 413-421.
- Toutanova, K. and C.D. Manning, 2000. Enriching the knowledge sources used in a maximum entropy part-of-speech tagger. *Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora: Held in Conjunction with the 38th Annual Meeting of the Association for Computational Linguistics-Volume 13*, Association for Computational Linguistics, pp: 63-70.
- Toutanova, K., D. Klein, C.D Manning and Y. Singer, 2003. Feature-rich part-of-speech tagging with a cyclic dependency network. *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*, Association for Computational Linguistics, pp: 173-180.
- Turney, P.D., 2002. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, Association for Computational Linguistics, pp: 417-424.
- Vinodhini, G. and R.M. Chandrasekaran, 2012. Sentiment analysis and opinion mining: A survey. *International J.*, 2: 82-292.
- WEKA, 2017. *Weka 3: Data Mining Software in Java*. WEKA the University of Waikato, New Zealand.
- Wilson, T., P. Hoffmann, S. Somasundaran, J. Kessler and J. Wiebe *et al.*, 2005. Opinionfinder: A system for subjectivity analysis. *Proceedings of hlt/emnlp on Interactive Demonstrations (HID' 05)*, Association for Computational Linguistics, pp: 34-35.
- Yi, J., T. Nasukawa, R. Bunescu and W. Niblack, 2003. Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. *Proceedings of the Third IEEE International Conference on Data Mining, (ICDM '03)*, IEEE, pp: 427-434.