Ontology-Based Sentiment Analysis and Community Detection on Social Media: Application to Brexit

Moudhich Ihab
LIST Department of
Computer Science faculty of
sciences and techniques
Tangier, Morroco

Loukili Soumaya
LIST Department of
Computer Science faculty of
sciences and techniques
Tangier, Morroco

Bahra Mohamed
LIST Department of
Computer Science faculty of
sciences and techniques
Tangier, Morroco
bahra002@gmail.com

Hmami Haytam
LIST Department of
Computer Science faculty of
sciences and techniques
Tangier, Morroco
hmami.h@gmail.com

Fennan Abdelhadi
LIST Department of
Computer Science faculty of
sciences and techniques
Tangier, Morroco
afennan@gmail.com

ABSTRACT

Sentiment Analysis and Community Detection are two of the main methods used to analyze and comprehend human interactions on social media. These domains expanded immensely with the rise of social media, as it provided a free and ever-increasing quantity of data. Domain ontologies are of great assistance in collecting specific data, as it describes the domain's features and their existing relationships. Therefore, we utilize them in collecting subject-specific data on social media. This paper describes the framework we've designed in order to understand, in depth, the impact of a subject on social media users, and also to evaluate the difference between the Lexicon Approach and the Machine Learning Approach, by assessing the strengths and weaknesses of each. This framework also aims to deeply understand the connections that exist between users, depending on their point of view on a particular subject. The resulting framework not only analyzes textual data (by taking into account the negation and sentence POS tags), but also visual one, such as images. In order to test the framework, we chose to analyze the Brexit phenomenon by collecting ontology-based data from Twitter and Reddit, and it had some promising results.

Keywords

Sentiment Analysis; Opinion Mining; Community Detection; Classification; Lexicon; Machine Learning; Social Media Analysis; Ontology

SCA2019, October 2–4, 2019, CASABLANCA, Morocco

ACM ISBN 978-1-4503-6289-4/19/10…$15.00
https://doi.org/10.1145/3368756.3369090

1. INTRODUCTION

Currently, in the space of a full day, 2.5 quintillion bytes of data is generated. Knowing that this pace is increasing and that the majority of this mass is shared online, we can immediately see the interest of analyzing this information for our own purposes. Social networks, in particular, have grown monumentally over the last decade, and now, in just one minute, 456,000 tweets are sent, 527,760 photos are shared on Snapchat, and 4,146,000 videos are viewed on YouTube, just to name a few.

Numbers do not lie: we live in an information sharing era. Internet users no longer hesitate to share their views online, or to offer their recommendations about a service, or just generally to express their opinions with others. In parallel, we all acquired the reflex to look up what is said about a subject in particular before forming our own opinions. That is exactly why analyzing sentiments based on social media is efficient; social network users express themselves freely and without taboos, which allows to collect honest and accurate feelings.

It is also quite interesting that the community detection matter totally shifted nowadays, because the geographical limits between individuals simply vanished through social media. This allows to study a new type of relationships and communities formed through the user interactions.

But before analyzing the sentiment or detecting communities, it is key to collect the correct data of the subject we're interested in. That is where ontologies come in handy; they allow us to represent the subject and its entities, and therefore gather only information that is relevant.

2. RELATED WORK

Sentiment analysis of social media data has been proposed before, by Abhishek Kaushik et al. [1]. They worked on mining general web content and analyzing the sentiment that it holds.

The different approaches of this field have been studied thoroughly, by Walaa Medhat et al. [2]. Their work evolves mainly around comparing all the different sub-methodologies used in the field; whether it's under the machine learning approach or the Lexicon-based one.

For the Lexicon approach, it has been discussed and implemented by Fatehjeet Kaur Chopra et al. [3] in their paper. They basically worked on creating an algorithm that analyzes the sentiment contained in sentences, using the Lexicon approach.

Sentiment Analysis using Machine Learning has also been proposed several times. The work of Amol S. Gaikwad et al. [4] is an illustration of it. They focused on developing a Naïve Bayes classifier, that divides data into positive or negative. They also heavily worked with ontologies. They relied on the domain ontology to identify the feature they were meaning to analyze if it was present.

There's also the paper submitted by Chen-Kai Wang et al. [5]. They contributed to the subject by developing a Recurrent Neural Network for classification of diseases based on symptoms contained in tweets.

For community detection over Social Media, Kwan Hui Lim et al. [6] discussed it in their paper that describes their work aiming to find communities with common interests using following links of celebrities.

Cataldo Musto et al. [7] proposed a solution to extract sentiments from tweets through a Framework. They developed a multi-level Framework that collects data, contextualizes them and analyzes them. They published two study cases: one of violence in Italy, and the other about the L'aquila Earthquake.

Another framework with the same idea is by Adil Bouktaib et al. [8]. Their intend was to predict the results of French elections in 2017 strictly based on the citizens' sentiments about the candidates. Their results were accurately correct.

The concept of smart cities is very popular at the moment, and has been for the last few years. That is because it allows a better management of the city and entails that every aspect of it is improved thanks to the system. Using sentiment analysis for that goal has been explored by Kaoutar Ben Ahmed et al. [9], as they discussed the state-of-the-art of this particular topic.

3. MAIN CONCEPTS

In this chapter, we'll discuss a number of concepts that we explored and incorporated into our Framework.

3.1. Sentiment Analysis

Xiang et al. [10] define sentiment analysis as "a special type of text mining with the focus on identification of subjective statements and contained opinions and sentiments, particularly in consumer-generated content on the Internet".

Sentiment analysis is a subfield of natural language processing that provides a methodology for detecting the sentiment in a non-structured information on computer. The analyzed text can be an entire document, a sentence or only a part of a sentence [11]. Typically, we perform a preliminary treatment on the text with tools such as: stemming, tokenization, speech tagging, entity extraction and relationship extraction. Then there are other steps like the subjectivity / opinion detection, the detection of polarity, the detection of intensity, the detection of a specific emotion, or sentiment sensing about a specific aspect.

There are three main approaches to implement sentiment analysis: either a lexicon based one [12], or by applying machine learning [13], or by combining two in what we call the hybrid approach [14]. They all have the same goal; detecting the sentiment contained in a text. But they are way different; not only in their methodologies, but also in their level of difficulty.

3.1.1. Lexicon Based Sentiment Analysis

In sentiment analysis, a lexicon is a dictionary of words and their correspondent polarity scores. It can either be domain-specific and have a theme, or be general and contain as much terms as possible. The main idea behind the lexicon-based approach is the following : build a lexicon (or use an existing one), create a bag-of-words from the text we want to analyze, preprocess it by these procedures : normalization, stop-words elimination, stemming, part-of-speech tagging, etc …, and finally calculate the sentiment score, which equals the average of each word's score from the lexicon.

As for creating the lexicon itself, it can be done either manually, or with statistical and/or semantic methods.

3.1.2. Machine Learning Based Sentiment Analysis

Machine Learning is an application of artificial intelligence that allows systems to learn and improve automatically from the experience without being explicitly programmed. For the learning process, the method can either be supervised (i.e., creating an inferred function from labeled data), or unsupervised (i.e., learning from its mistakes). For Sentiment Analysis, we create a model by training the classifier with labeled examples. That is, if we find a fitting dataset for training.

3.1.3. Hybrid Sentiment Analysis Approach

The last method consists of combining the two previous ones in order to obtain the best accuracy in the results. The idea is to have the best of both approaches. That is because, in case we don't have a dataset for the machine learning approach, or it's inaccurate, it allows to train the model using data scored with the Lexicon method. As a result, we have a model that trained with a supervised algorithm, so it is very precise, combined with the stability of the Lexicon results.

3.2. Community Detection

The existence of communities in a network corresponds to the presence of groups of vertices (nodes) more strongly connected to each other than with the others vertices of the graph; these are related classes, which have a higher density than the graph in its totality.

The identification of this type of structure represents advantages on several levels. On one hand, it exists in many real networks and, most of the time, it has a concrete meaning in terms of organization. In social networks, communities can represent groups of individuals with common interests, common activities, etc. In the Web, groups of pages or highly connected sites often deal with themes and the detection of these communities can improve search engines.

3.3. Ontologies

The definitions of the concept of ontology are numerous in computer science. The most quoted one is: "An ontology is an explicit specification of a conceptualization" [15] - Gruber.

Conceptualization is an abstract simplified view of the world. Which means that it's based on concepts, objects, and other entities that are alleged to exist in the area of interest, as well as between them.

For the specification part of the definition, it means a formal and declarative representation. In the data structure representing the ontology, the type of concepts used and the constraints of their use are explicitly declared.

Needless to say, that as a representation readable by the machine, the ontology is not a runnable program.

4. PROPOSED FRAMEWORK

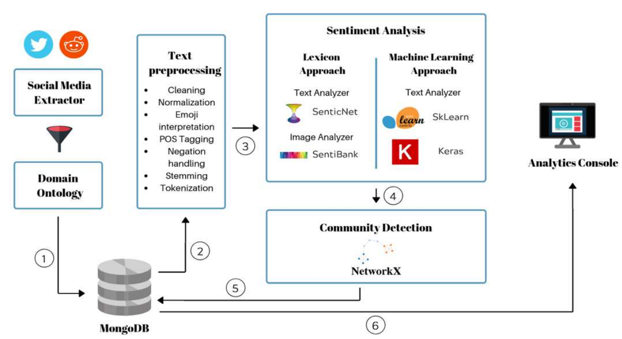


Figure 1. Global architecture of proposed Framework

The proposed Framework is modular and multilevel. This means that it consists of five modules plus the database, each with a specific mission to accomplish, using a set of technologies. In a typical pipeline, the designed Framework is capable of massive data extraction using the domain's ontology for accuracy, of filtering and preprocessing it (tokenization, stemming, negation handling, etc.), identifying the sentiment that it holds with the Lexicon based approach and Machine Learning one, detecting the formed communities, and finally displaying the results on an analytics console, in multiple formats.

The fig. 1 is a simplified representation of the global architecture of the proposed Framework, along with the technologies and/or methods used in each module.

The following section is a description of each module of the proposed Framework.

4.1. Social Media Extractor

This module is the entry point of the Framework. It is in this component that the extraction of data from social networks takes place. It establishes the connection with Twitter and Reddit, captures any content that meets the defined criteria, and stores it into the database.

Before starting the data extraction, it is key to define the domain of interest's ontology, in order for the Framework to gather the right data, that is related to the subject.

The process of extracting data from Social Media happens through the official Twitter and Reddit APIs.

4.2. Text preprocessing

The purpose of this module is to transform the collected content in the previous layer, to noiseless data with a unified format. The reason we apply these changes is that in the Lexicon based approach of Sentiment Analysis, the dictionaries are predefined, and so we have to provide clean terms to the algorithm, so it can find their scores.

We start off by removing any existing URLs, extra whitespaces and tab spaces, and any screen names, or numbers. The next step is to substitute emojis by their meaning, along with any abbreviations (e.g. OMG becomes oh my god), and contractions (e.g. haven't becomes have not). Afterwards, we normalize the entire text into lower case.

After the first phase, we proceed to determine which words are affected by negation in a sentence [16]. We did that so we can invert their sentiment scores in the Lexicon method.

To determine the negation scope, we define a number of principles to follow:

- We detect the start of the negation scope through negation cues, such as: no, not, never.

- We rely on part-of-speech tagging and coordinating conjunctions (e.g. and, or) to determine if the term is affected by negation.

- We only invert the score of verbs, adjectives and nouns, because they're the meaning holders, and also because the rest likely will not exist in the dictionary.

e.g. for the sentence 'i do not like or love you.', the words 'like' and 'love' are the ones that are affected by the negation. We got that because first, there is the negation cue 'not'. Right after it, we have a verb, followed by a coordinating conjunction 'or', and a verbal phrase. The first verb, 'like', is going to be inverted due to the cue 'not'. The second one, 'love', because it is the first verb of the verbal phrase, that is preceded by a negated verb (i.e. same nature) and a conjunction.

Lastly, we remove any remaining stop words (e.g. the, on, in), and punctuation, and we lemmatize the text.

Lemmatization is a lexical treatment that consists of returning the canonical neutral form of a term, the one that exists in the dictionary.

4.3. Sentiment Analysis

In this module, we analyze the sentiment held by every text, following the two approaches (i.e. Lexicon and Machine Learning).

4.3.1. Lexicon based Approach

We used the Lexicon based Approach for two types of data: textual and visual.

Text Sentiment Analysis: To calculate the sentiment score of a text, we proceed as following: the algorithm first tokenizes the text. It retrieves the sentiment score of each term from the SenticNet dictionary [17]. For the terms that are affected by the negation, it inverts their polarity. The average of all scores obtained is calculated. It assigns the total score to the text.

```
Algorithm: calculateTextSentimentScore
  Input:  preprocessed text (preprocessed_text)
          list of inverted words (inverted_terms)
  Variables: terms, term, score, counter, scores_total
  Output: text's sentiment's score (text_sentiment)
terms ← tokenize(preprocessed_text);
scores_total ← 0;
counter ← 0;
foreach term ∈ terms do
    if existsInDictionnary(term) then
        score ← getScoreFromDictionnary(term);
        counter ← counter +1;
        if term ∈ inverted_terms then
            score ← score * (-1);
        scores_total ← scores_total + score;
text_sentiment ← scores_total / counter;
return text_sentiment;
```

Images Sentiment Analysis: For each image, we extract all the bigrams 'noun, adjective' that it contains, using the SentiBank [18] visual sentiment ontology framework. These obtained terms are then processed like the previous textual data, in order to get their average score using SenticNet.

```
Algorithm: calculateImageSentimentScore
  Input: image
  Variables: features, feature, score, counter, scores_total
  Output: image's sentiment's score (image_sentiment)
features ← extractFeaturesSentibank(image);
scores_total ← 0;
counter ← 0;
foreach feature ∈ features do
    score ← calculateTextSentimentScore(feature);
    scores_total ← scores_total + score;
    counter ← counter +1;
image_sentiment ← scores_total / counter;
return image_sentiment;
```

4.3.2. Machine Learning Approach

For the machine learning approach, we created a recurrent neural network that was trained with over one million labeled tweets. The label is either 'positive' or 'negative'.

The RNN architecture combines the following layers:

- Embedding: transforms data into same-size dense vectors.

- LSTM: Long Short-Term Memory networks that learn long-term dependencies, which make them perfect for usage in natural language processes [19].

- Dense: since the output of a LSTM layer is not a SoftMax [20], it is important to add the dense layer.

Each text's sentiment is then predicted by the generated model.

The quality of a model is assessed based on its accuracy, which formula is:

$$Accuracy = \frac{correct\ amount\ of\ guesses}{total\ amount\ of\ guesses}$$

In our case, the accuracy is equal to: 90%.

4.4. Community Detection

At this module, we analyze data to identify the relations that exist between users, depending on their interactions on social media.

We first create a network which nodes are users and edges the retweets existing between them. We run the Edge Betweenness [21] algorithm by Girvan and Newman, which returns a list with the formed communities.

The way this algorithm works is by calculating the Betweenness Centrality score between all edges, and then removing the ones with the highest score. We reiterate this operation until no edge can be removed.

$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

Where σ_{st} is the total number of shortest paths from node s to node t and $\sigma_{st}(v)$ is the number of those paths that pass through v.

4.5. Analytics Console

The final layer is a mean for the user to take advantage of and visualize the results of the previous modules. It displays the gathered data and the results of both sentiment analysis and community detection.

The user can consult the generated reports, charts and graph from his own account. Reports and maps are used to interpret the results. This means that there is the percentage of negative, positive and neutral data, a comparison between different methods, an evolution over time, and the feeling according to the regions of the world.

For the network, it is a representation of the formed communities, as well as the details of each one if you click on it. The details of each community consist of the number of the community, the number of nodes and edges it has, the influencer nodes, and the more recurrent tweets.

5. BREXIT ANALYSIS CASE STUDY

The Brexit is a term that was born in 2016, and refers to the United Kingdom's exit from the EU (e.i. European Union). It was the former Prime Minister David Cameron who took this initiative, by organizing a referendum back in 2016, in which 51.9% of citizens chose to leave the EU. The whole process is still ongoing, and there are a lot of people who spoke up against the Brexit itself. The fact that there are individuals who want the Brexit to take place, and others who would rather remain in the EU, makes this topic highly fitting to test the Framework. That is because not only it provoked a lot of mixed feelings within the population, but it also separated the country into communities who share not only the same point of view, but also the approach they see fitting to solve the problem.

The first step was to define the keywords which we were going to use to collect Brexit related data on Twitter and Reddit. The selected keywords are highly important and are game changers for the accuracy of the Framework, because it is crucial to gather all the relevant data, and to not omit any important ones. We started off by studying the different political parties in the UK and their position on the matter to build our domain ontology. From there, we gathered politicians' usernames, the parties' official usernames, and their used hashtags. Also, it is important to collect data using hashtags that are proper to each side of the movement, such as "Remainer", "LeaveEU", "RevokeArt50", etc.

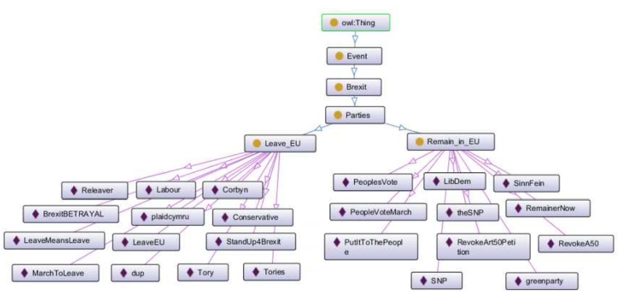


Figure 2. Brexit Domain's Ontology

We started the data collection on March 1st 2019 until current day, both through Twitter API and Reddit API. We managed to collect over 5 million tweets, retweets, quotes, replies, posts and comments, which were all stored in a mongo data base.

Going forward, we submitted the data to the proposed framework. The results have different forms, so we can understand in depth the subject.

Sentiment analysis – Overall pie charts:

The first charts represent the negative and positive quantity of data that is related to the Brexit, both on Twitter and Reddit, and with both approaches – Lexicon based and Machine Learning. The color green is mapped to positive, blue to negative, and yellow to neutral sentiment.

Figure 3. Sentiment Analysis Approaches Comparison – Overall Brexit Sentiment

As it can be seen on fig.3, the majority of data is negative in this topic, which clearly indicates that people strongly dislike the Brexit idea and prefer to stay within the EU. In order for us to see which approach is more accurate and closer to real results, we looked up recent polls that were conducted by NatCen, which is the largest independent social institute in the UK, which full title is "National Centre for Social Research". The following table synthesizes the comparison between our Framework's results with both approaches, and NetCen's polls answer to the question: "If there was another referendum on Britain's membership of the EU, how would you vote?". All results are from the same time frame.

Table 1. Proposed Framework's Results Comparison with NatCen's results

	Leave the EU	Remain in the EU
NatCen's results	44.45%	55.55%
Proposed Framework – Lexicon Based Approach results	35.01%	64.99%
Proposed Framework – Machine Learning results	45.12%	54.88%

The results show that the Machine Learning approach is more accurate, as its percentages are very close to the ones we retrieved from actual polls.

Sentiment Analysis - Results on Map

To further our understanding of the Brexit subject, and to understand which region are most likely to vote "Remain" in case of a second referendum, we generated the following map on the proposed Framework.

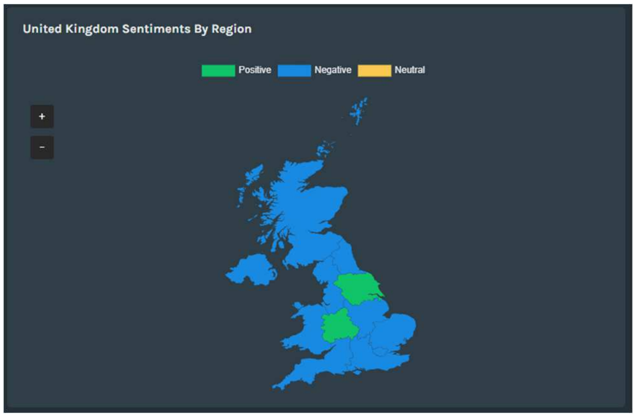
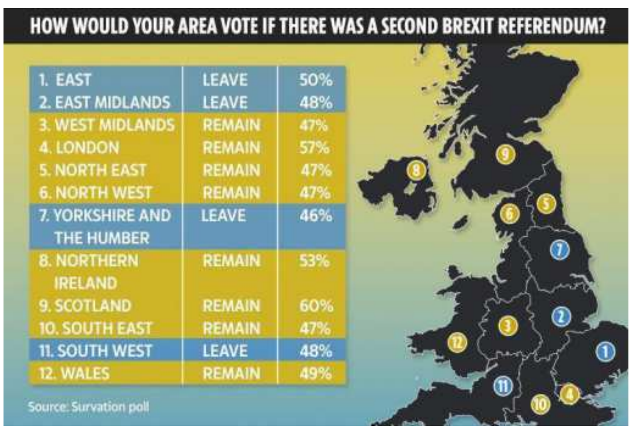


Figure 4. Proposed Framework's Sentiment Analysis Results by Region on Map

As seen on fig.4, most regions' sentiments about the matter is negative, except for two regions: "East Midlands" and "Yorkshire and The Humber". The results generated by the proposed Framework are similar to the ones established by Survation, which is a polling and market research agency with British origins. Except for one region, "East", that has a 50% leave vote.



HOW WOULD YOUR AREA VOTE IF THERE WAS A SECOND BREXIT REFERENDUM?

1. EAST	LEAVE	50%
2. EAST MIDLANDS	LEAVE	48%
3. WEST MIDLANDS	REMAIN	47%
4. LONDON	REMAIN	57%
5. NORTH EAST	REMAIN	47%
6. NORTH WEST	REMAIN	47%
7. YORKSHIRE AND THE HUMBER	LEAVE	46%
8. NORTHERN IRELAND	REMAIN	53%
9. SCOTLAND	REMAIN	60%
10. SOUTH EAST	REMAIN	47%
11. SOUTH WEST	LEAVE	48%
12. WALES	REMAIN	49%

Source: Survation poll

Figure 5. Suvation's Poll Results by Region on Map

Sentiment Analysis – Overtime stacked bars

The following figure (i.e. fig.6) represents the overtime variation of sentiments for the Brexit. The first bar is the proposed Framework results, and the second is the same results published by Survation. We used the results of Machine Learning approach, as it is more accurate.

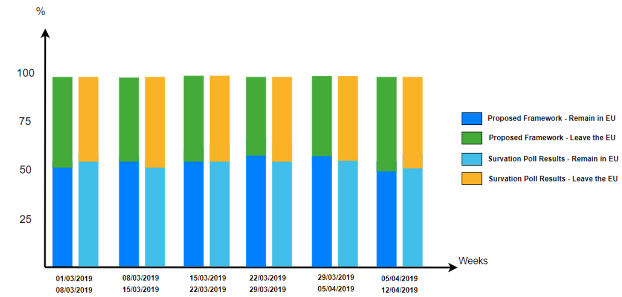


Figure 6. Sentiment Analysis Over Time – Framework's results comparison with Survation's poll

Even though there is a variation in weeks' results, they all clearly indicate that people strongly dislike the Brexit, and thus want to remain in the EU.

Community Detection results

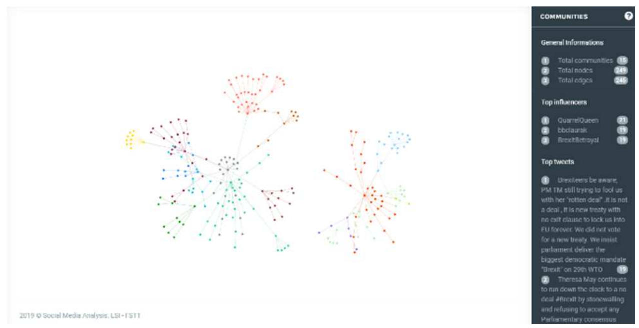


Figure 7. Community Detection results

The figure above (fig.7) represents the formed communities within social media users who are invested in the Brexit subject. It is interesting that we have two separate subgraphs; each composed of several communities formed. This is explained by the fact that all the communities of the first subgraph (on the right), for example, are against Brexit. The difference between them is that each has a different argument, or the intensity of their remarks is of a different degree. For the second subgraph, all users want the Brexit to take place, and are disappointed to see that it is not executed. There are communities that feel betrayed and present this argument, others that are more practical and who organize petitions to sign, etc.

N.B: If the "Leaver" community is much bigger, it is because we have added conditions to the Framework, such as limiting the number of tweets and retweets to be analyzed, thus a relatively small amount of random data is used, for the community detection.

6. CONCLUSION

The proposed Framework integrates many approaches and methods to analyze sentiments and detect communities that

emerge following the rise of a subject. The study case that we implemented proved that the results are accurate and on point.
As a perspective for this framework, there are many ideas we can execute to enhance the performance of our framework and have even better results than the one we currently have.
Regarding the Lexicon-based approach in the sentiment analysis section of the framework, it could be improved by combining multiple Lexicons, so we can ensure the presence of nearly every term we want to analyze. Therefore, the accuracy will automatically get better, because the analysis would take in consideration the majority of words. Also, in case a word isn't found in any Lexicon, we could lookup the score of its synonym.
Another interesting perspective is to combine between the two approaches in Sentiment Analysis. Using a hybrid one would definitely increase the accuracy, as it takes the best of each. Training data that has been labeled using a Lexicon to train a model would be ideal.
For community detection, it would be quite interesting to explore other algorithms, to see which ones of them is the quickest and most accurate. Having used the Edge Betweenness algorithm, our next step will consist of implementing infoMap algorithm [22], as they both treat weighted and directed edges. Using machine learning for this part of the Framework is also something we are looking forward to achieve.

REFERENCES

[1] Kaushik, A., Kaushik, A., & Naithani, S. (2015). A Study on Sentiment Analysis: Methods and Tools. International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064

[2] Walaa Medhat, Ahmed Hassan, Hoda Korashy (2014). Sentiment analysis algorithms and applications:A survey. Ain Shams Engineering Journal (2014) 5, 1093–1113

[3] Fatehjeet Kaur Chopra, Rekha Bhatia (2016). Sentiment Analyzing by Dictionary based Approach. International Journal of Computer Applications (0975 – 8887) Volume 152 – No.5, October 2016

[4] Amol S. Gaikwad, Anil S. Mokhade, (2017). Twitter Sentiment Analysis Using Machine Learning and Ontology. International Journal of Innovative Research in Science, Engineering and Technology, Volume 6, Special Issue 1, January 2017, ISSN (Online) : 2319 – 8753 ISSN (Print) : 2347 – 6710

[5] Chen-Kai Wang, Onkar Singh, Zhao-Li Tang and Hong-Jie Dai (2017). Using a Recurrent Neural Network Model for Classification of Tweets Conveyed Influenza-related Information, International Workshop on Digital Disease Detection using Social Media 2017 (DDDSM-2017), pages 33–38

[6] Lim, Kwan Hui & Datta, Amitava. (2012). Finding Twitter Communities with Common Interests using Following Links of Celebrities. MSM'12 - Proceedings of 3rd International Workshop on Modeling Social Media. 10.1145/2310057.2310064.

[7] Cataldo Musto, Giovanni Semeraro, Marco de Gemmis, Pasquale Lops (2014). Developing Smart Cities Services through Semantic Analysis of Social Streams. Copyright is held by the International World Wide Web Conference Committee (IW3C2).

[8] M.Bahra, A.Bouktib, H.Hanafi, M.El Hamdouni, A.Fennan (2017). Sentiment analysis in social media with a semantic web-based approach: Application to the French presidential elections 2017. Innovations in Smart Cities and Applications. Proceedings of the 2nd Mediterranean Symposium on Smart City Applications

[9] Kaoutar Ben Ahmed, Atanas Radenski, Mohammed Bouhorma, Mohamed Ben Ahmed (2016). Sentiment Analysis for Smart Cities: State of the Art and Opportunities. ISBN: 1-60132-439-1, CSREA Press

[10] Ko, C.-H. (2018) Exploring Big Data Applied in the Hotel Guest Experience. Open Access Library Journal, 5: e4877.

[11] Seema Kolkur, Gayatri Dantal and Reena Mahe (2015). Study of Different Levels for Sentiment Analysis. International Journal of Current Engineering and Technology E-ISSN 2277 – 4106, P-ISSN 2347 – 5161

[12] Fazal Masud Kundi, Aurangzeb Khan, Shakeel Ahmad, Muhammad Zubair Asghar (2014). Lexicon-Based Sentiment Analysis in the Social Web. ISSN 2090-4304 Journal of Basic and Applied Scientific Research

[13] Ali Hasan, Sana Moin, Ahmad Karim and Shahaboddin Shamshirband (2018). Machine Learning-Based Sentiment Analysis for Twitter Accounts. Math. Comput. Appl. 2018, 23, 11; doi:10.3390/mca23010011

[14] Kolchyna, Olga & Souza, Thársis & Treleaven, Philip & Aste, Tomaso. (2015). Twitter Sentiment Analysis: Lexicon Method, Machine Learning Method and Their Combination.

[15] Thomas R. Gruber (1993). Toward Principles for the Design of Ontologies Used for Knowledge Sharing. In International Journal Human-Computer Studies 43, p.907-928.

[16] Maral Dadvar, Claudia Hauff and Franciska de Jong. Scope of Negation Detection in Sentiment Analysis.

[17] Erik Cambria, Soujanya Poria, Rajiv Bajpai and Bjorn Schuller. SenticNet 4: A Semantic Resource for Sentiment Analysis Based on Conceptual Primitives

[18] Yuhai Yu, Hongfei Lin, Jiana Meng and Zhehuan Zhao (2016). Visual and Textual Sentiment Analysis of a Microblog Using Deep Convolutional Neural Networks. Algorithms 2016, 9, 41; doi:10.3390/a9020041

[19] Wenpeng Yin, Katharina Kann, Mo Yu and Hinrich Schutze (2017). Comparative Study of CNN and RNN for Natural Language Processing. arXiv:1702.01923v1 [cs.CL] 7 Feb 2017

[20] Lei Zhang, Shuai Wang and Bing Liu. Deep Learning for Sentiment Analysis: A Survey

[21] Xavier Polanco, Eric San Juan. Text Data Network Analysis Using Graph Approach. I International Conference on Multidisciplinary Information Sciences and Technology, Oct 2006, Mérida, Spain. pp.586-592. ffhal-00165964f

[22] William Deitrick, Wei Hu Machine Learning-Based Sentiment Analysis for Twitter Accounts, Journal of Data Analysis and Information Processing, 2013, 1, 19-29