DEPARTAMENT DE FILOLOGIA ANGLESA I DE GERMANÍSTICA

The use of linguistic knowledge in sentiment analysis tools

Treball de Fi de Grau

Natalia Ruiz Martínez

Supervisor: Dr. Ana Fernández Montraveta

Grau en Estudis Anglesos

January 2021

ACKNOWLEDGEMENTS

Firstly, I would like to thank my TFG supervisor Dr. Ana Fernández Montraveta, for letting me take this topic. Thanks to her, I have been able to research something related to my future career intentions, business. Not only this but also for the support and patience I have received from her all this way.

Moreover, I am deeply thankful to all my friends who have accompanied me during this process. I am especially grateful to my dearest friend Jordi, who has been with me during all the years studying this degree, being both my rock and my fortress. We have accomplished this together.

Finally, and most importantly, I would like to thank my family and my partner. They have been the ones who have always believed in me and have always stood by my side. It is an honour for me to finish the degree with this TFG and make them proud. Also, I dedicate it to my little nieces Noa and Leire, and to my stepdaughter Nayara; I have always told them that, with effort and constancy, they would be able to accomplish whatever they wish in their lives.

TABLE OF CONTENTS

INDEX OF FIGURES

ABSTRACT

In recent decades, business companies have drawn their attention to sentiment analysis, a growing field due to the many applications and utilities it has to offer. Sentiment analysis is an automatic process in which natural language processing (NLP) techniques and biometrics are used to identify and extract, in a systematic way, affective states and opinions from texts published on the Internet. Many useful tools help these companies figure out the satisfaction of customers with a product or a brand. These tools interpret and sort out emotions within text data using text analysis techniques, which allow these companies to improve and implement the necessary changes. This work aims to provide a proper background for sentiment analysis, also known as opinion mining, and analyse these tools considering the linguistic knowledge they require, see how they work, and propose any relevant improvements, if possible.

Keywords: Sentiment Analysis, linguistics, customers' feedback, tools, business.

1. INTRODUCTION

Sentiment analysis dates back to the 1950s. At that time, it was mainly used on written paper documents and concerning public opinion. It was not until the 1990s that these texts started to be analysed by the computational linguistics community. Nowadays, this "science" is one of the fastest-growing research areas and widely used to detect and extract subjective information from reviews, survey responses, opinions, and attitudes, inter alia, by using several techniques, methods, or tools.

Information mining is beneficial for companies willing to identify and better target their message towards their potential customers. Interpreting people's emotions is essential for corporations since the consumer can express his or her thoughts and feelings more openly than ever before. By analysing customers' feedback, sentiment analysis can identify its polarity, which means describing the opinion expressed as positive, neutral, or negative, among other sentiment classifications.

Moreover, sentiment analysis, also known as opinion mining, has a powerful position in social media tracking. Nowadays, the vast majority of social media tools present the user with the option of analysing the text with some sentiment analysis running in the background. Another important factor affecting its popularity is that accuracy rates of sentiment analysis for social media monitoring keep improving, up to 75% accuracy (source KDnuggets). Sentiment analysis also allows companies to discover which influencers are satisfied with their brand or which influencers they have to try to change their perception of the brand if they are unsatisfied with it.

In what follows, this paper will address the basics of sentiment analysis (Section 2) and how it works (Section 3). Afterwards, Section 4 presents different sentiment analysis methods and some of the tools used in this area (Section 5). My main objective in this work is to linguistically analyse these tools and propose any significant improvements in the area, if feasible.

2. THE BASICS OF SENTIMENT ANALYSIS

In this section, we are presenting the basics of sentiment analysis. Generally speaking, sentiment analysis usually has three different phases (Bibi 2017) (although sometimes there might be an optional fourth phase):

1. In the first phase, comments on the social network being analysed are collected, and each text document is broken down into parts, depending on the system. These units of analysis can be sentences, phrases, tokens, or parts of speech. For example, let's consider the following two sentences as units of analysis:

   (1) The awful location and the dire quality of the restaurant made the hotel have horrible reviews.

   (2) The bad location and the poor quality of the restaurant made the hotel have low reviews.

2. In the second phase, sentences are split, and the system identifies each sentiment-bearing phrase and component. In (3) and (4), we observe how sentence (1) and sentence (2) have been divided into phrases, respectively:

   (3) Awful location | dire quality | horrible reviews.

   (4) Bad location | poor quality | low reviews.

Although both sentences are negatively talking about a hotel, there is a clear difference between them since sentence (1) is more negative than (2). The human reading detects this difference by looking and interpreting sentiment-bearing phrases, that is, sentences that carry an opinion or a tone. Sentiment-bearing phrases usually have the form of adjective-noun combinations, like the examples that we have seen before.

3. The third phase consists of assigning a sentiment mark to each phrase and component. Values range from (-1) to (+1). In this way, we start to get numerical results. It also gives the user to create graphics to make results more visual (see Figure 1 and Figure 2 below, which are the visual representation of sentences (1) and (2)).

   Each sentence is assigned a positive, a neutral, a negative, and a compound score in this phase. The compound score is the name given to the measure that estimates all the ratings assigned to the lexical items or phrases. Below we provide the numerical representation of sentences (1) and (2), respectively:

   1) "The awful location and the dire quality of the restaurant made the hotel have horrible reviews. "

   { ‘neg’ : 0 . 897 , ‘neu’ : 0 . 103 , ‘pos’ : 0 . 0 , ‘compound’ : -0 . 9973 }

   2) "The bad location and the poor quality of the restaurant made the hotel have low reviews."

   { ‘neg’ : 0 . 725 , ‘neu’ : 0 . 275 , ‘pos’ : 0 . 0 , ‘compound’ : -0 . 8675}

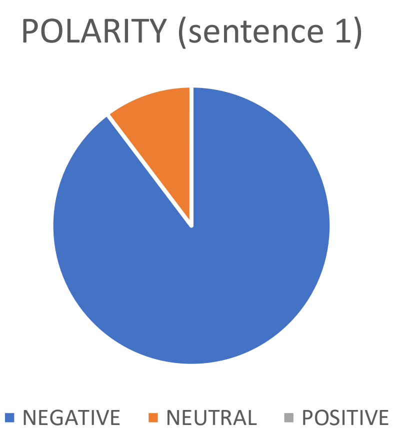

Figure 1: Analysis in sentence (1)

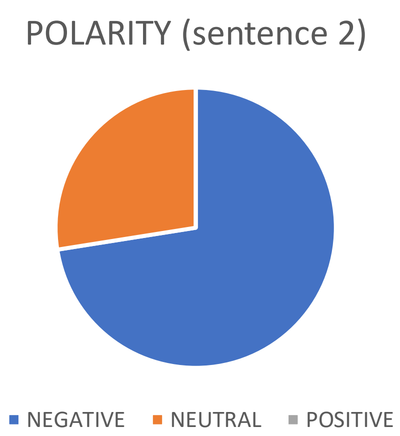

Figure 2: Analysis in sentence (2)

4. Finally, the last phase, which is not present in every system, combines the scores to provide multi-layered sentiment analysis. This way, an integrated perspective of the analysis of a group of sentences can be provided (see Figure 3).

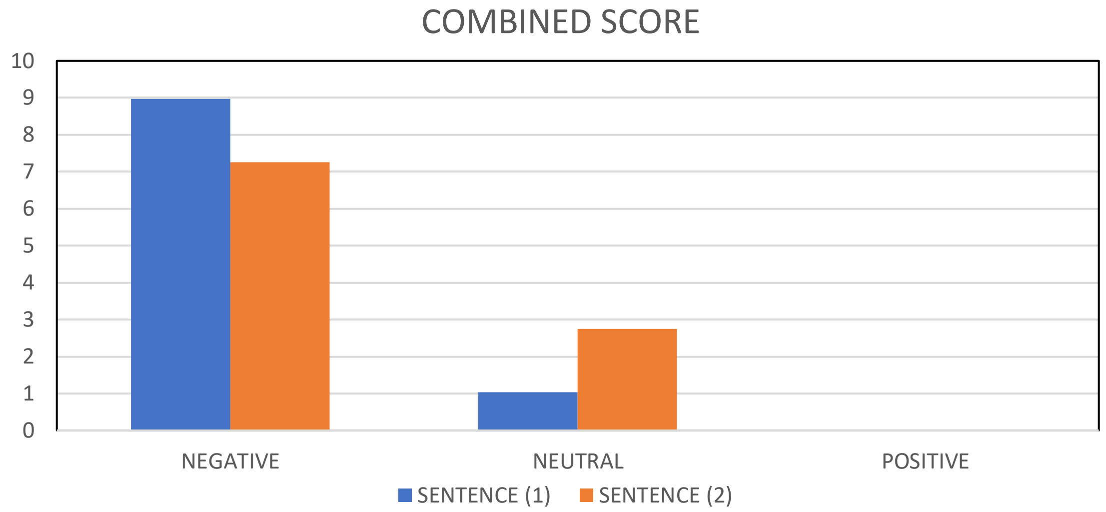

Figure 3: Combined score

2.1. Sentiment classification

Sentiments can be classified in different ways depending on how businesses want to interpret customer feedback. Companies can define and tailor the categories that meet their sentiment analysis needs better. Hence, this paper details just some of the most popular types of sentiment analysis:

- Sentiment classification according to polarity:
    - Positive
    - Neutral
    - Negative

If the company is willing to have polarity precision, polarity categories could be increased using intensifiers such as 'Very Positive' or 'Very Negative.' This system is known as "Fine-grained Sentiment Analysis." The use of these labels can be seen in satisfaction surveys, like the ones that universities or schools send students at the end of every academic year.

- Sentiment classification according to feelings or emotions:
    - Happy
    - Sad
    - Angry

  We are more familiar with the use of this terminology, given that social networks such as Facebook or LinkedIn established this classification a long time ago. This fact enables the customer to interact and show their opinion on a particular topic or product.

- Sentiment analysis according to interests:
    - Interested
    - Not interested

  This type of classification can be seen in dating apps such as Tinder or Grinder.

- Sentiment Analysis according to based-aspects:

When analysing sentiments of texts such as product reviews, a specific company may be interested in particular aspects of a product, willing to know if these comments are positive, neutral, or negative. For example, this type of analysis can determine that a sentence such as "The durability of this lipstick is really short" carries a negative opinion on the feature 'lipstick durability.'

2.2. Negation handling in sentiment analysis

Jiménez-Zafra et al. (2019) reported that negation detection is one of the most crucial tasks within the sentiment analysis context. In fact, all the seminal works in the field agree with this, stating that it is necessary to identify the sequence of words affected by negation (i.e., the scope).

The negation scope goes from the word next to the negation to other words following it, which might affect their polarity (Farooq et al. 2016). In the sentence "these heels are not pretty, but they are easy-wearing," the scope of negation only affects the word immediately after "not." However, in the sentence "these heels are not what I was looking for," for example, all the elements in the wh-clause fall under the scope of negation. These examples show that the scope of negation is not immovable and can vary according to the linguistic features, such as the punctuation marks, part of the speech (POS), or even the conjunctions (Farooq et al. 2016).

When negation and its scope have been properly identified, the next issue to determine is how negation affects the dictionary values for sentiment words. One procedure is to provide the opposite value of the lexical items' polarity in the scope of the negative element (Taboada 2016). For example, in a structure where dictionary words have information about both strength and polarity, if the term "great" has the value of +3, "not great" will have the value of -3. This procedure is known as *switch negation* (Saurí 2008). However, it is not always that easy, since in a highly positive sentence, the negation downtones the sentence, rather than having a reversal effect (Taboada 2016). In the case of "fantastic" being assigned a value of +5, "not fantastic" does not carry the value of -5, since it is not that severely bad. Since there are no specific rules to undergo this process, there is another option called the *shift negation*. The shift negation is the process in which the negated item is altered in the scale by a specific amount, but it is not assigned the exact opposite mark as the original term. For example, in the case of "fantastic" being a +5, "not fantastic" would count +1.

In addition, it is also worth commenting on the practice known as markedness (Taboada 2016). A negative word will always have a higher mark than a positive one, given that they have a more significant impact. It is difficult to accurately identify negative sentiment in a sentence because we don't use these terms as frequently as the positive ones and because negative evaluation is couched in negative terms (Pang & Lee 2008). One way to deal with this issue is to always give a higher mark to negative terms than to positive ones (Taboada et al. 2011).

Finally, another issue that faces difficulties within the recognition of negation is the notion of irony. Taking into account that irony is sometimes hardly recognised by humans, it is even more difficult for a machine to detect it. There are certain patterns that can be used to detect it, but what has resulted very useful is to take tweets with the tag #sarcasm on it and use them to learn ironic features. These tags are generally used when the author and the receiver are not familiar with each other, making it easier to understand whenever there is a lack of context (Bamman & Smith 2015). However, this procedure is far from perfect, which means that when analysing a sentence that includes irony or sarcasm, the results may not be accurate enough due to the lack of conclusive research on the field.

3. TYPES OF POLARITY DETECTION

As stated before, polarity is whether a text is positive or negative, including a big range of variations within these two options. The most common way to analyse sentiment is by using polarity. According to Jiménez-Zafra et al. (2019), there are two ways to detect polarity: lexicon-based and machine learning. Nevertheless, thanks to the research I have carried out, I have been able to find a third one: the hybrid approach.

3.1. Lexicon-based methods

In order to carry out the lexicon-based method, dictionaries of negative and positive words are gathered. These dictionaries detect polarity and strength (e.g., “bad” is negative while “horrible” is strongly negative). When a new text is being analysed, the system extracts all the words that are included in the dictionaries and provides them with a polarity and strength tag by using different rules. Moreover, these words or expressions receive a score by the system that goes from -1 to 1. Thus, I will list some of the best well known lexicon-based methods.

- SO-CAL, the Semantic Orientation CALculator

As reported by S.M. Jiménez-Zafra et al. (2019), this system was explicitly designed for customer reviews, but it also works well on other texts. This structure contains dictionaries in which words are classified according to their belonging to one of these four categories: adverbs, adjectives, nouns, and verbs). The SO-CAL dictionaries are composed of more than 5000 words in English and about 4200 words in Spanish. The process of sentiment analysis in SO-CAL starts with two assumptions (Taboada et al., 2011). The first assumption is that individual words have something known as "prior polarity," which means that each word has a semantic orientation independent of its context; the second determines that a numerical value can be used to express this semantic orientation. Moreover, this system considers intensification by words such as *really* or *quite* by giving them an associated percentage that increases or decreases the mark of the word next to it.

- SentiWordNet

This lexical resource is publicly available for research purposes and widely used. According to Musto et al. (2014), this system provides three numerical sentiment marks that correspond to positivity, negativity, and neutrality. It is the outcome of the automatic glossary of WordNet's synsets (Baccianella et al. 2010). A Synset is a set of one or more synonymous concepts defined in WordNet, a machine-readable lexical database organized by meanings. WordNet has been developed at Princeton University. It also understands that there might be some terms that can hold different opinion-related properties.

- WordNet-Affect

This system is suitable to recognize the expression of affective knowledge. It was also developed starting from the ontology WordNet, just as the SentiWordNet. Basically, it works by labelling those words that express affection with A-LABELS, i.e., those affective concepts found in synsets that are used as the system labels (Musto et al., 2014). These A-LABELS are independent of the domain hierarchy. This method covers all the English lexicon and is very concise when it comes to providing large amounts of conceptual distinctions (Strapparava et al. 2004).

- MPQA Subjectivity Lexicon

The acronym MPQA stands for Multi-Perspective Question Answering. Subjectivities clues are compiled through different sources (Jagdale et al., 2018). These clues are also labelled with polarity (positive, negative, and neutral), along with their part-of-speech tagging, also known as POS-tagging. These tags also carry intensity, by using the labels strong and weak.

- SenticNet

This method analyses the polarity of structures. It does not entirely trust word co-occurrence frequencies but accounts for emotion recognition by checking the connotative and denotative information in the text. SenticNet provides a wordmark for 14,000 common-sense concepts, which are classified within twenty-four basic emotions. These emotions are defined in a model called 'Hourglass of Emotions' (see Figure 4), which assumes an hourglass shape (Cambria et al., 2011).

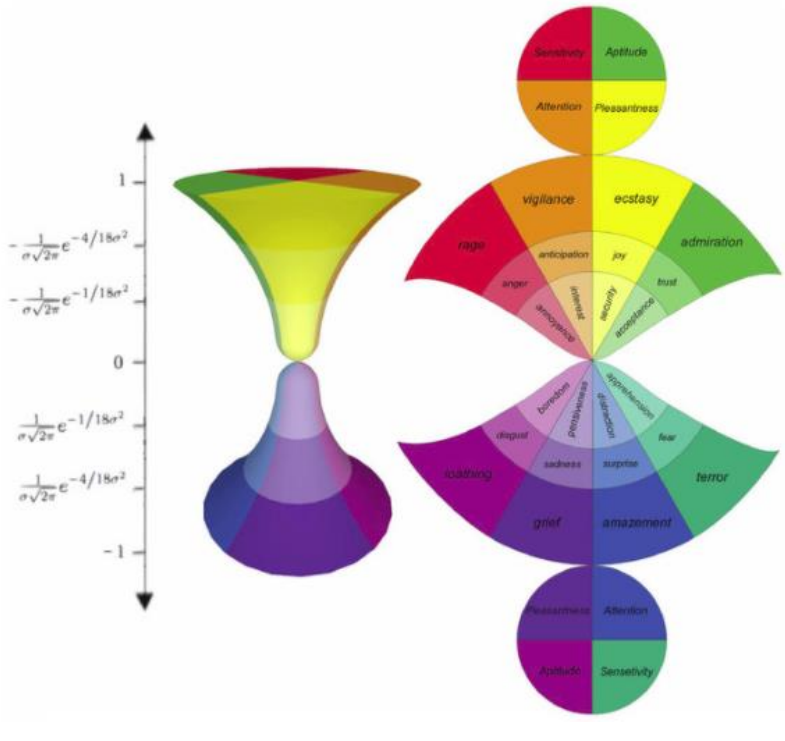

Figure 4. The Hourglass of Emotions.

3.2. Machine-learning methods

Besides the models that use linguistic (mostly lexical) knowledge to guide the analysis (Section 3.1), machine-learning methods do not require any linguistic knowledge at all. These programs learn, from texts already tagged as positive or negative, how to distinguish feeling. As Kennedy and Inkpen (2006) state, standard content usually includes annotation about parts of speech or punctuation and n-grams, which are adjoining sequences of 'n' items from a specific text or speech. These items can be words, letters, syllables, phonemes, or base pairs, depending on the application.

In the following section, we present a list of some of the best-known machine-learning methods with a very brief explanation since this is not the topic of this work.

- Bayesian Networks

It is a probabilistic graphical method that understands the dependence between variables explicitly with direct edges in an acyclic graph model. All absent relations outline the conditional independencies in the procedure. These models can be learned from experts' data, and they can estimate the probabilities for casual or subsequent events (Chickering 1995).

- Naïve Bayes Classification

This classification approach is widely used. It assumes that the value of a particular feature of a class is unrelated to the value of any other feature (Ren et al., 2009). It is based on class probability and conditional density estimation. Nevertheless, conditional density estimation is gathered from uncertain data points, which require different methods (Ren et al., 2009). It is commonly agreed that this system is only appropriate when the level of the inputs is high.

- Maximum Entropy

This method can be used to estimate any probability distribution. It is better than some other machine-learning methods because this system makes no inherent conditional independence assumptions between random variables (Wang 2010). The main objective of Maximum Entropy is to find out the best probability disposal among preceding test data (D'Andrea et al., 2015).

- Neural Networks

As Kim et al. (2016) stated, this system has shown better results than other machine-learning methods for object detection and recognition. It is bottomed on a group of natural and artificial neurons used for mathematical and computational analysis methods (D'Andrea et al., 2015). Nevertheless, this process requires ample computation and a big dataset in order to be trained (Kim et al., 2016).

- Support Vector Machine

Tong & Koller (2001) affirm that with this approach, the user has access to a wide range of unlabelled instances and can request the labels for some of them. Moreover, it is possible to analyse evidence and patterns that are to be beneficial for classification and regression analysis (D'Andrea et al., 2015). The maximum margin hyperplane is to be found and represented by the own system.

3.3. Hybrid approach

This method combines both lexicon-based and machine-learning techniques to improve the performance of sentiment classifiers (D'Andrea et al., 2015). Basically, it makes use of natural language processing (NLP), i.e., a sentiment lexicon reinforced with the help of SentiWordNet, and distorted sets to determine the semantic orientation, intensity, and polarity for sentences, and machine-learning techniques used for learning the expression of sentiments (Appel et al. 2016). Nevertheless, a list of systems using hybrid approach methods cannot be provided. At present, no hybrid systems have been fully implemented, and, therefore, such a proposal would require the user to use and combine results from both lexicon-based and machine-learning methods.

4. TOOLS

It is undeniable that if a company is willing to know their customers' opinions on a specific product, they will get better and faster results using sentiment analysis software, which allows them to automatize the process. Sentiment analysis software permits users to reach and analyse millions of opinions. In the current market, there are many tools that perform this type of analysis. Obviously, in order to use most of them, you have to pay a fee. Next, a present a list in which I summarize five of the best-known tools in this area.

- Lexalytics

Lexalytics works with social comments, reviews, surveys, and other text documents using Natural Language Processing. It analyses different types of texts and tries to explain why a customer reacts to a product in a specific way. It also provides expanded utilities such as theme extraction, categorization, and intention detection. These extra features offer a better understanding of what is said about their business. The main difference with other similar tools is that Lexalitics also informs why customers feel in a certain way, instead of only how they think. It is a payment product and offers several possible profiles.

- Brandwatch

One of this tool's main characteristics is that it allows the user to search on the Internet using the company's logo. It automatically finds the websites where this logo has been uploaded, finding, this way, all the possible reviews of the brand or customers' opinions. This app is more related to pictures than to texts. It also has additional metrics such as aggregate followers, the latest activity, or mention volume. It uses a combination of customer data, web analytics data, and social data to provide a detailed summary of a company's situation. The price is 650€ per month.

- Social Searcher

This social media monitoring platform is a good option for companies with low budget since it allows you to search up to 100 keywords, hashtags, or usernames per day. It will inform you whether the comments about it are positive or negative. This app monitors social networks and local websites such as Flickr, VK, Vimeo, and others. The app's price varies since it goes from free to 16€ per month if you are interested in the Professional Plan.

- Rosette

This application is a good option for international companies since it can analyse texts in over 30 different languages. Besides, it also analyses shorthand or slang, which is a handy functionality. It makes use of Natural Language Processing (NLP) technologies, which in turn use the most convenient methods for each task, mixing machine learning methods with lexical and linguistic knowledge. However, and precisely because of the additional functionalities it presents, it is also relatively expensive. Prices go from 80€ to 800€ per month.

- Clarabridge

This app includes customer experience management. Its software for sentiment analysis is quite complete and complex and can analyse grammar, context, industry, and source. Moreover, it presents an option of recording and analysing calls. This tool uses Artificial Intelligence (AI) powered machine learning algorithms trained on thousands of customer interactions. The price varies according to the pack the company is interested in, but it is slightly more expensive than other analysing tools.

5. CONCLUSIONS

As stated at the beginning of this work, sentiment analysis is a fast-growing field with many future opportunities since companies know that customers' feedback is essential for their progress. Although it is a relatively new area, as we have seen, there are already many studies about the topic. Nevertheless, there are still some issues that have to be improved.

This is the case of irony and sarcasm recognition. In negation detection, as we have seen, it is a limitation that must be overcome. From my perspective, thousands of texts with samples of irony and sarcasm, appropriately tagged, have to be gathered so an approach based on machine learning could be taken. It is a very complex subject given that there are no stipulated rules that can be taught to a machine, even more considering that every person can express irony in a different way. The concepts of irony and sarcasm are precisely based on the inadequacy of an answer. This interpretation would require the software understanding semantically the text, which up to now is not possible.

This leads me to the next point, which is sentiment analysis according to location. Up to my knowledge, this approach has not been studied yet, not even considered. I propose that considering geographic information about the writer of a comment would provide competent results in sentiment analysis. It would not be an easy nor fast task, though, since it would require gathering samples of local businesses with a low ratio of foreign people. Moreover, socio-economic and socio-cultural features could also be considered, creating profiles and big databases. I believe that by analysing texts taking into consideration the user's profile would definitely increase the accuracy of opinion mining.

Moreover, the hybrid approach has not been studied in depth. Studies state that this process would have the potential to improve sentiment classification performance, rather than lexicon-based or machine learning methods acting separately. By increasing research on this method, opinion mining tools would finally introduce the hybrid approach in their systems, making them more efficient.

Furthermore, I will conclude by saying that sentiment analysis is a relatively new field. For this reason, and its complexity, I believe that nowadays only big companies have access to it. While working on this dissertation, I have realized that opinion mining companies should enlarge their range of action. If they adapted prices and made sentiment analysis comprehensible for small companies, they would increase their benefits, and micro-businesses would improve their outcomes, too.

Finally, I would like to comment on a final issue I have faced while working on this dissertation. Sentiment analysis falls in the intersection of several areas such as Natural Language Processing, Machine Learning, and Linguistics. Since my field of knowledge is Linguistics, I have tried to simplify the explanations so that the reader did not encounter many difficulties when interpreting them. This means that some formulas have not been included. Nevertheless, I hope I have successfully described the main foundations and trends of sentiment analysis.

REFERENCES

Appel, Orestes; Chiclana, Francisco, Carter, Jenny; Fujita, Hamido. (2016). A Hybrid Approach to Sentiment Analysis with Benchmarking Results. In: Fujita, Hamido; Ali, Moonis; Selamat, Ali; Sasaki, Jun; Kurematsu, Masaki. *Trends in Applied Knowledge-Based Systems and Data Science.* (2016) Springer International Publishing.

Appel, Orestes; Chiclana, Francisco; Carter, Jenny; Fujita, Hamido. (2018). Successes and challenges in developing a hybrid approach to sentiment analysis. *Applied Intelligence*. *Vol.* 48, 1176-1188. 10.1007/s10489-017-0966-4.

Baccianella, Stefano; Esuli, Andrea; Sebastiani, Fabrizio. (2010). SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. *Proceedings of Seventh International Conference on Language Resources and Evaluation-LREC 2010.* European Language Resources Association (ELRA). 2200-2204.

Bamman, David & Smith, Noah. (2015). Open Extraction of Fine-Grained Political Statements. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing.* 76-85. 10.18653/v1/D15-1008.

Bibi, Maryam. (2017). Sentiment Analysis at Document Level. *Communications in Computer and Information Science*. *Vol* 628.

Cambria, Erik, Livingstone, Andrew & Hussain, Amir. (2011). The Hourglass of Emotions. 1. In: Esposito Anna, Esposito Antonietta M., Vinciarelli Alessandro, Hoffmann Rüdiger, Müller Vicent C. (eds) *Cognitive Behavioural Systems. Lecture Notes in Computer Science*, *Vol* 7403. Springer, Berlin, Heidelberg. *Vol* 7403, 144-157. 10.1007/978-3-642-34584-5_11.

Chickering, David; Geiger, Dan; Heckerman, David. (1995). Learning Bayesian networks: Search methods and experimental results. *Preliminary Papers of the Fifth International Workshop on Artificial Intelligence and Statistics.*

D'Andrea, Alessia; Ferri, Fernando; Grifoni, Patrizia; Guzzo, Tiziana. (2015). Approaches, Tools and Applications for Sentiment Analysis Implementation. *International Journal of Computer Applications*. *Vol* 125, 26-33. 10.5120/ijca2015905866.

Farooq, Umar; Mansoor, HasanM Nongaillard, Antoine; Ouzrout, Yacine; Qadir, Muhammad Abdul. (2016). Negation Handling in Sentiment Analysis at Sentence Level. *Journal of Computers*. *Vol* 12, 470-478. 10.17706/jcp.12.5.470-478.

Jagdale, Rajkumar; Shirsat, Vishal; Deshmukh, Sachin. (2018). Review on Sentiment Lexicons. *3rd International Conference on Communication and Electronics Systems (ICCES)*. Coimbatore, India. *Vol* 12, 1105-1110.10.1109/CESYS.2018.8723913.

Jiménez-Zafra, Salud María; Cruz-Díaz, Noa; Taboada, Maite; Martín-Valdivia, Maria. (2020). Negation detection for sentiment analysis: A case study in Spanish. *Natural Language Engineering*, *Vol* 1, 1-24. 10.1017/S1351324920000376.

Kennedy, Alastair; Inkpen, Diana. (2006). Sentiment classification of movie reviews using contextual valence shifters. *Computational Intelligence*, *Vol* 22(2), 110-125. doi:10.1111/j.1467-8640.2006.00277.x

Kim, Jihun; Kim, Jonghong; Jang, Gil-Jin; Lee, Minho. (2016). Fast learning method for convolutional neural networks using extreme learning machine and its application to lane detection. *Neural Networks. Vol* 87, 109-121. 10.1016/j.neunet.2016.12.002.

Musto, Cataldo; Semeraro, Giovanni; Polignano, Marco. (2014). A comparison of lexicon-based approaches for sentiment analysis of microblog. *CEUR Workshop Proceedings*. *Vol* 1314, 59-68.

Pang, Bo; Lee, Lillian. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval. Vol* 2, 1-135. 10.1561/1500000011.

Ren, Jiangtao; Lee, Sau; Chen, Xianlu; Kao, Ben; Cheng, Reynold; Cheung, David. (2009). Naive Bayes Classification of Uncertain Data. 944-949. 1083-1086.10.1109/ICDM.2009.90.

Saurí, Roser. (2008) *A Factuality Profiler for Eventualities in Text*. Ph.D. Dissertation thesis. Brandeis University, Waltham

Strapparava, Carlo; Valitutti, Alessandro. (2004). WordNet-Affect: an Affective Extension of WordNet. *Proceedings of the Fourth International Conference on Language Resources and Evaluation – LREC, Vol* 4, Lisbon: Portugal.

Taboada, Maite; Brooke, Julian; Tofiloski, Milan; Voll, Kimberly; Stede, Manfred. (2011). Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics. Vol* 37, 267-307. 10.1162/COLI_a_00049.

Taboada, Maite. (2016). Sentiment Analysis: An Overview from Linguistics. *Annual Review of Linguistics*. *Vol* 2, 325-347. 10.1146/annurev-linguistics-011415-040518.

Tong, Simon; Koller, Daphne. (2001). Support Vector Machine Active Learning With Applications To Text Classification. *The Journal of Machine Learning Research. Vol* 2, 45-66. 10.1162/153244302760185243.

Wang, Hui; Wang, Lin; Yi, Lixia. (2010). Maximum Entropy framework used in text. classification. *IEEE International Conference on Intelligent Computing and Intelligent Systems, Xiamen,* 2010. 828-833. doi: 10.1109/ICICISYS.2010.5658639.

Webgraphy

Pandey, Pandey. (2018). *Simplifying Sentiment Analysis using VADER in Python (on Social Media Text)*. Medium. Retrieved 17 January 2021, from https://medium.com/analytics-vidhya/simplifying-social-media-sentiment-analysis-using-vader-in-python-f9e6ec6fc52f

Sims, Scott. (2015). *Sentiment Analysis 101*. KDnuggets. Retrieved 10 August 2020, from https://www.kdnuggets.com/2015/12/sentiment-analysis-101.html

Sentimnent analysis explained. Lexalyties. Retrieved 7 September 2020, from https://www.lexalytics.com/technology/sentiment-analysis

Brownlee, Jason. (2019). Machine Learning Mastery. Retrieved 14 January 2020, from https://machinelearningmastery.com/introduction-to-bayesian-belief-networks/

MonkeyLearn Inc. (2020). *Sentiment analysis: A definitive guide.* .Monkey Learn. Retrieved 10 August 2020, from https://monkeylearn.com/sentiment-analysis/

Mitchell, Sierra. (2020). *Sentiment Anlaysis – Chabots Life.* News Akmi, Retrieved 10 August 2020, from https://newsakmi.com/news/tech-news/ai/sentiment-analysis-chatbots-life/

SenticNet. Retrieved 14 January 2020, from https://sentic.net/

Barysevich, Aleh. (2020*). 10 Sentiment Analysis Tools to Measure Brand Health.* SocialMediaToday. Retrieved 29 December 2020, from https://www.socialmediatoday.com/news/10-sentiment-analysis-tools-to-measure-brand-health/575334/

Gupta, Shashank. (2018). *Sentiment Analysis: Concept, Analysis and Applications*. Towards Data Science. Retrieved 7 September 2020, from https://towardsdatascience.com/sentiment-analysis-concept-analysis-and-applications-6c94d6f58c17

Rao, Prashanth. (2019). *Fine-grained Sentiment Analysis in Python (Part 1).* Towards Data Science. Retrieved 12 January 2021, from https://towardsdatascience.com/fine-grained-sentiment-analysis-in-python-part-1-2697bb111ed4