RUPKATHA JOURNAL
ON INTERDISCIPLINARY STUDIES IN HUMANITIES
ISSN 0975-2935

About the Journal

Journal DOI	https://dx.doi.org/10.21659/rupkatha
Journal Home	www.rupkatha.com
Indexed by	Scopus Web of Science: Emerging Sources Citation Index (ESCI) DOAJ
Journal Metrics	CiteScore 2020: 0.2 | SJR 2020: 0.162 | SNIP 2020: 0.193 | JCI 2020: 0.50

About the Issue

Issue	**Volume 14, Number 2, 2022**
Editor	Tirtha Prasad Mukhopadhyay
Issue DOI	https://doi.org/10.21659/rupkatha.v14n2
TOC	https://rupkatha.com/v14n2.php

About the Article

Title	**Sentiment Analysis for a Humanist Framework: How Emotions are Recognized and Interpreted in the Age of Social Media**
Author/s	**Rafael Guzman Cabrera**
Affiliation	Department of Electrical Engineering, University of Guanajuato, Mexico.
Author ID	None
Funding	**No funding received. Published free of any charge.**
Article DOI	**https://doi.org/10.21659/rupkatha.v14n2.01** **Pages:** 1-9
Abstract	https://rupkatha.com/v14n201
Full-text PDF	https://rupkatha.com/V14/n2/v14n201.pdf
Article History	**First Published: 18 June 2022**
Article Impact	**Check Dynamic Impact**
Copyright	Aesthetics Media Services
Licensing	Creative Commons Attribution Non-Commercial 4.0

Sentiment Analysis for a Humanist Framework: How Emotions are Recognized and Interpreted in the Age of Social Media

Rafael Guzman Cabrera
Department of Electrical Engineering, University of Guanajuato, Mexico.
Email: guzmanc81@gmail.com

Abstract
Language is in constant evolution – this theory has been demonstrated most aptly and comprehensively by Marshall McLuhan. Specialisation in the different areas of knowledge, especially technology, has contributed to this process. Technological advances and the development of so-called intelligent devices allow interaction through voice interfaces, text, or gesture and in its most advanced forms by means of the incorporation of artificial intelligence-generated linguistic communications in human-machine interfaces. In recent years, the ways of communication or watching news have changed, now we do it by means of the internet and through different options of the social networks. We interact with people and react to their communications by means of divergent ways of language formation. It is increasingly common to express opinions through social networks and the internet. So much so that now we know that it is possible to analyse a person's sentiment from his or her communications of opinion issued in social networks? The question is, can we determine, for example, whether the opinion has a positive or negative emotive charge only by analysing the written or inscribed texts of such formats of communication? This paper presents a brief description of how technological evolution has created an x-factor of language, that is expressed, appropriated and re-used in machine learning modules, artificial intelligence, and automatic sentiment analysis.

Keywords: Artificial intelligence, Language evolution, Sentiment analysis.

1. Introduction

The evolution of the human language is one of the most important and interesting post-humanist questions about the human ability to think and interact with the world and the environment (Nowak, Komarova, & Niyogi, 2002). The earliest records of language come from the Denisova cave inhabitants of southern Siberia, some 175,000 years ago (Barnard, 2016). We don't know if they spoke or developed a language or protolanguage. A protolanguage is a language reconstructed on coincidences and common features of a *family* of original languages. There are several theories about the first stages of protolanguage (Tecumseh and Donald 2010). Generally, four models are recognized for a protolanguage template: lexical, musical, mimetic and gestural. Again, every language, whether spoken or written, evolves to have grammar as a defining feature. Grammar is essentially syntax: the part of language that lies between the sound system that makes up speech (phonology) and the part that carries meaning and is called semantics. Heine and Kuteva (2007) propose a six-stage scheme for the evolution of grammar: nouns, verbs, adjectives and adverbs, pronouns, and demonstratives and finally negations. Their model further suggests that language evolved gradually, and that the lexicon evolved before syntax. Since the early

1990s, there has been an increasingly productive study of language, with advancements in many different sectors, and an encouraging increase in exchange and interdisciplinary collaboration (Fitch 2005). Currently we normally see the way we interact with other people, in our communications includes not only the older methods of communication that were available to humanity for thousands of years but also forms or signals derived from communications technologies, some of which include communications across formats like social networks or e-mail correspondences. But what kind of technologies did we leave behind to get here? When did we stop using other ways to communicate? Where did postal mail go, and telegrams and faxes that even relatively recently were used on a daily basis?

We have witnessed a technological revolution that has put our reach of technological resources far in advance so that we have changed our way of interacting with others in so many ways. As a consequence of this change there has also been a change in the nature or structure of language (especially inscription); now we can use emoticons or abbreviations that literally say nothing, but we still use them to express ourselves. The paper mode of communications has changed from paper material surfaces or inscribable surfaces to digitally simulated platforms or screens. Surprisingly, we went from talking on mobile phones to writing text messages through social networks in platforms that we now call social media. Yet it also suggests a new medium of communication such as some of those that McLuhan had barely begun to identify (McLuhan 2003). When we actually talk and interact with a person, either in person or by means of using some technological resource, we also perceive their mood or their "sentiment" either by looking at their gestures, expressions, modulation, and tone of voice, or a whole range of other characteristics that we use to express ourselves. The big question is: is it possible to perceive such feelings from a written text-format alone, like a text that incorporates not just words, but extra morphological semantics like those engendered through emoticons, GIFs, memes, visual codes, digits etc, or new sets of phraseology? The rest of this paper tackles the question of the languages in the latest media, specifically social networks, which constrain us to meet and interact with people by looking at the textual equivalent of their emotions and not at their physical bodies. What are the written expressions and resources that help us to identify the feeling of a person through a written expression? This question also leads us to directly understand *how* a systematic classification and understanding of emotional cues might be undertaken so that machine learning modules can predict these emotions?

2. ***Artificial Intelligence***

The evolution of technology had a decisive impact on the way we live today, particularly the development of computer hardware. Just 70 years ago, researchers wondered if a machine could ever think for itself. Over time the question was changed to whether it could come to think by being manipulated by physical symbols sensitive to the structure that they had. In those times they managed to understand the great power of systems that were governed by established rules, but what if the systems were automated? Automation could turn a reading process from being an abstract computational system into a real physical system (Fernández-López, 2011). To determine if a machine uses artificial intelligence or to put it in simple words if a machine is intelligent, Alan M. Turing's proof was taken as a reference (Millican, 2021), which indicates that any recursively computable function can be calculated in a finite time by means of a machine that manipulates simple symbols. This was Turing's universal machine. A Turing machine is a device that negotiates with symbols on a strip of tape according to a table of intervening rules. Despite its simplicity, a Turing machine can be adapted to simulate the logic of any computer algorithm and is particularly useful in procuring the functions of a central processing unit within a computer. This implies that a symbol manipulating machine should be able to have intelligent consciousness, where positive results could be obtained since these machines could perform a series of cognitively intelligible activities, as for example the solution of algebraic problems, or of arithmetic, or engagement in meaning interpretative human dialogue or games like checkers and chess. Thanks to the emergence of larger hardware memories, we could evolve more efficient and faster machines that could go ahead and engage with human language systems.

For its part Hubert L. Dreyfus, one of the main characters who argued against the fact that a machine could have its own consciousness, published a book in the 1970s where he criticised the modules of machine cognition (or interpretation) and mentioned that the consciousness was reserved to the capabilities and common sense that people possess, Dreyfus didn't deny that a machine could be made to think, but said that this could be based only on the manipulation of symbols, that is, by means of programs (Su & Luvaanjalba, 2021). In the 80s Jon Searle proposed a thought experiment called 'the Chinese room' which posits that a machine is incapable of thinking, since the human mind doesn't function like a computer program, nor can a computer program behave like a human mind (Tabares Cardona, 2021). The Chinese room consists of a room, isolated from the outside, in which there is a person who doesn't know the Chinese language but who, through a hole, can receive sheets of paper with texts written in this language, and if inside the room the person has manuals and dictionaries with which he is able to relate the characters to write a response, without having to study the language but applying rules then, for each set of input characters, the person would be capable of issuing an answer without understanding the language. In the same way, a machine will work with inputs and obtain outputs, even if it doesn't 'understand' them. Therefore, a machine that applies rules is incapable of having consciousness, but we humans can also be, retroactively arguing, a Chinese room full of rules.

The main objective of the Chinese room is to deny that the mind is similar to a computer program, demonstrating that a machine can perform an action without understanding what it does and why it does so, since its logic only operates with symbols without understanding the content involved. Such a machine could easily pass the Turing test by pretending that the machine understands the language. Artificial intelligence consists of a simulation of some activities of the nervous system by means of machines: this refers to the fact that some of the processes that are performed in the brain can be analysed as computational processes. An example would be that rule-guided machines wouldn't have the distractions of goals to be achieved- as it happens to human beings who are always faced with emotional distractions and destinies of their interactions. These destinies may be simple happinesses from a stream of pain or simple tirednesses. The interface between the brain and the computer allows measuring brain activities, processing and creating communication channels with the environment. We can define a system capable of translating aspects of the nervous system into a model of interactions with the virtual world.

3. ***Machine Learning***

Learning refers to a broad spectrum of situations in which the learner increases his knowledge or skills to accomplish a task. Learning applies inferences to certain information and constructs an appropriate representation of some relevant aspect of reality or some process (Moreno, 1994). A common metaphor around machine learning – within Artificial Intelligence – is to consider problem solving as a type of learning that consists – once a type of problem has been solved – in being able to recognize the problematic situation and react using the learned strategy (Klahr & Kotovsky, 2013). A classic example is the problem of the farmer, who, accompanied by a fox, a goose and a sack of grain must cross a river on a barge in which there is only the and one more, but if he leaves the goose with the fox, the fox will eat it and if he leaves the grain with the goose, the goose will eat it. Here the problem must be recognized, and decisions made that allow everyone to reach the other side of the river. In this sense, we have different classifications or types of learning, we will briefly describe the most used in the state of the art: supervised, unsupervised, and deep learning.

Supervised learning (Nguyen Cong, Rivero Pérez, & Morell, 2015) has the purpose of obtaining a distance metric function, usually represented mathematically as the Mahalanobis distance between two instances and their corresponding classes for a specific application, and based on using information from the training set. Most algorithms that learn a distance function try to solve an optimization problem with constraints. On the other hand, unsupervised learning (Tello & Informáticos, 2007) obtains a model that fits the observations, because there is no *a priori* knowledge. A usual problem of this type of learning falls on decision-making itself, and whether they are correct or not, for this, grouping techniques with logic are used. Data collected is like other data, and thus can be treated collectively as a group. Clustering is a form of unsupervised classification where, in contrast to the supervised group, the class labels are not known (there are no previously defined classes) and the number of groups may not be known either. Fuzzy clustering is a method frequently used in pattern recognition (Fan, Zhen, & Xie, 2003). In recent years, deep learning has been widely used. It consists of a set of algorithms that attempt to model high-level abstractions using computational architectures. Such structures may support nonlinear and iterative transformations of data expressed in matrix or tensor form. In simple terms, deep learning implies the mastery, transformation, and use of this knowledge to solve real problems (Valenzuela Granados, 2021). Independently of the type of learning, the objective is the same: *to have a system that is capable of learning from experience and one that can include the conditions of the environment to successfully perform its task.* When we talk about the identification of sentiments in written texts it is important, in this sense, to have instances manually labelled by an expert, that allow machine learning techniques to identify trends, associations, patterns, and collocations in the text that allow associating these features with the type of sentiment labelled in the instance under study.

4. ***Social Networks***

Currently, microblogging websites have become digital spaces of varied information, where users post information in real-time and opinions are expressed by means of texts that implicitly carry an emotional charge. Statements thus become a positive or negative opinion about people, products, or services. Several companies, organisations and institutions have made use of this type of media to obtain feedback, promote themselves, or to turn the opinion of users into an improvement network (Rani, Gill, & Gulia, 2021). Being able to know the opinions of the users of a product or service will guide the decision-making to achieve an improved sales profile of a company, by identifying areas of opportunity and improvement within it. Twitter in recent years has recorded a growth in the so-called "social panoramas", used in a transmission system, as well as conversation tools. Twitter is the social network that is currently used for the development of numerous investigations of sentiment analysis (also known as opinion mining), where sentiment analysis is defined as the process of determining opinions based on attitudes, valuations, and emotions about specific topics. In this context, an opinion is a positive or negative evaluation of a product, service, organisation, person, or any other type of entity about which some feeling can be expressed (Cambria, Xing, Thelwall, & Welsch). Due to the importance of sentiment analysis for business and society, it has been extended from computer science to management and social sciences (Coba, Barrera, & Sánchez, 2022). Since, if opinions on the network are successfully collected and analysed, they allow not only to understand and explain many complex social phenomena, but also to predict them. The emotions that users express in Tweets are related to the person's sentiment, and the polarity (positive, negative, and neutral) is the measure of the emotions expressed in a phrase. Generally, the polarity goes from negative (-1) to positive (1) through neutral (0), where this last value means that no sentiment or opinion has been expressed.

5. ***Sentiment Analysis***

Khamphakdee & Seresangtakul (2021) describe sentiment analysis as a task that is responsible for identifying and classifying different points of views and opinions about something without being specific: it can be an object, a person, an activity, etc. Analysis is based on Natural Language Processing (NLP). The main objective is the analysis of opinions and their classification based on the identified sentiment: positive or negative. There is also the possibility that they don't exist and would be classified as neutral. The possible applications can be as useful as they would be different. In recent years such analysis has been a very attractive and interesting field of research, creating a classification set that can be performed in the polarity of sentiment as mentioned above added to this can be added a classification of primary sentiments such as joy, sadness, anger, fear, and others. Antonakaki and colleagues (2021) present some techniques used for the review of sentiment analysis, such as those which will help us to automatically determine the emotional polarity in a text with Artificial Intelligence, i.e., develop programs or learning algorithms and knowledge generation capable of learning to solve problems.

The authors in Jiménez-Zafra, Cruz-Díaz, Taboada, & Martín-Valdivia (2021) tell us about the ways of adapting a semantic orientation system to be able to perform the analysis of sentiment in a new language, building support vector machine (SVM) classifiers. We must bear in mind that a classification system, used to find 'feelings' in written expressions, based on machine learning, can be trained in any language. Another technique used for sentiment analysis review is Semantic Orientation, which oversees extracting opinions (Appel, Chiclana, Carter, & Fujita, 2021). Appel and colleagues explain that the semantic orientation of a word can become positive when it is shown with praise words, or negative when a criticism-word is identified. Semantics uses a learning technique that doesn't necessarily need to be supervised since it doesn't require initial training. This type of unsupervised learning uses different lexical rules in sentiment classification.

There are also 3 levels of classification for sentiment analysis:

- Document-based
- Sentence-based
- Word/phrase-based.

The first level is document-based, where the document is understood in a unique way and the whole document is thus classified according to a feeling for the whole document. The sentence-based level is responsible for classifying each sentence in a document or text: machine learning is generally used to detect subjective sentences. Finally, the word/phrase level is essential since the word is the smallest unit containing meaning in the entire text and is therefore indicative of the most detailed of the levels. In the Sentiment Analysis method, a machine learning approach based on a training and testing, using one set of collections to differentiate between text features (training) and another for classifier accuracy (testing) may be used. Our research has repeatedly used such techniques. Some of the classifiers we have used were Support Vector Machine (SVM), Nayve Bayes (NB) and Maximum Entropy (ME). Nayve Bayes is a classifier commonly used to classify text documents based on a probability model, for estimating the probability of a given group with a text document as input. The Support Vector Machine (SVM) classifier is also proposed for solving problems in pattern recognition. It is a learning model with algorithms that is responsible for data analysis. The two classifiers were top-rated in the machine learning approach to data mining and sentiment classification.

Sentiment analysis starts with the collection of data on a website or social network, mostly by taking advantage of the data that already exists publicly. The data can be classified according to the input of information from such sources as forums, blogs, articles, news, or social networks. For forums, the research is based on publications, and for this data collection is based on the access information of the users since they must be registered to be able to participate in them. A main advantage here is that most of the forums are dedicated to a single topic. Reviews focus a lot on opinions that describe good and bad attributes whether in products or services, such as movies. In social sentiment analysis classification depends a lot on the use of keywords in the texts. To finish with this part of the methodologies implemented to carry out sentiment analysis in texts, I want to refer to two projects in which I had the opportunity to participate. In Sánchez, Cabrera, Carrillo, & Castro (in preprint 2022) we conducted analysis of sentiments, with a methodology that allowed us to identify the polarity of a text in Spanish according to the emotion of its authors: this polarity could be identified with 3 labels: positive, negative, and neutral, and the emotions that could be identified being of 5 kinds: anger, fear, joy, sadness and love.

As the first point of the methodology, use is made of the corpus labeled SemEval 2018 "Task1: Affect in tweets", first a cleaning process of the tweets is performed, eliminating: emoticons, punctuation signs and special signs to subsequently separate the tweets into words, and using POS (part of speech), we place a label and word lemma (base form of the word). With this information a text classification model is created. This model receives (matches with input signal) an instance and categorizes it as: anger, fear, joy, sadness, or love, corresponding to the emotion that was identified for each instance. This is possible because the training corpus is labelled according to the emotion and can be used to train the system; once the system is trained it can receive new instances and identify the emotion. Once the emotion has been identified, polarity identification is performed, whose objective is to obtain a positive, negative, or neutral classification. This stage is performed through the extraction of the POS tags, here a search is performed for each lemma within the ML-Senticon lexicon, to obtain its respective positive or negative classification. Another research (Guzmán Cabrera & Hernández Farias, 2020) presents an exploration of diverse lexical resources that support the task of sentiment analysis. For the development of the methodology as a first point a series of experiments based only on the content of the tweets was presented in our projects. For this we used five configurations, in each one the pre-processing to be performed was increased, the first of them was without performing any type of pre-processing, the second consisted of tokenizing the text, eliminating empty words, conversion to lowercase and to terms that exceed a frequency threshold. Two approaches to lexical resources were used, the first one was a basic approach based on the creation of lists of terms associated with two polarities: positive and negative. And the other approach labelled a word with a score that reflects its value with respect to a particular aspect. The authors in our group selected a set of fourteen lexical resources divided into two main groups, those that include information strongly related to sentiment and emotions and those in which psycholinguistic information was also considered. It is undoubtedly a very exciting area of explorations and there is much more to write about. The important thing is to show that both the identification of sentiment and polarity can be performed in written texts and that these resources become necessary given the popularity of social networks and the daily posting of opinions on them. Surely language will continue to evolve, and, in a few years, everyone would be discussing some other strategies for performing sentiment identification.

Conclusion

Computational sentiment analysis betokens a process that helps us determine the emotion with which a series of words is defined, and it consists of evaluating attitudes and opinions from word-tokens to obtain information that helps in identifying the reaction of users for a product or service, or by extension any piece of communication. In general, the idea of sentiment analysis was partly elaborated for the development of better products and services, based on the opinions that were found in the different areas of communications. Yet a lot remains to be discovered. But the final take for any interpretative process is to understand how any thinking entity, b it a machine or human arrives at the meaning of texts, what kind of flow chart is really relevant and expedient and how such insights change our notion of interpretation in the academic theoretical literature. What do machines teach us about reading?

Declaration of Conflict of Interests

The author(s) declared no potential conflicts of interest.

Funding

No funding has been received for the publication of this article. It is published free of any charge.

References

Antonakaki, D., Fragopoulou, P., & Ioannidis, S. (2021). A survey of Twitter research: Data model, graph structure, sentiment analysis and attacks. *Expert Systems with Applications, 164*, 114006.

Appel, O., Chiclana, F., Carter, J., & Fujita, H. (2021). A Fuzzy Approach to Sentiment Analysis at the Sentence Level *Fuzzy Logic* (pp. 11-34): Springer.

Bellet, A., Habrard, A., & Sebban, M. (2013). A survey on metric learning for feature vectors and structured data. *arXiv preprint arXiv:1306.6709*.

Cambria, E., Xing, F., Thelwall, M., & Welsch, R. Sentiment Analysis as a Multidisciplinary Research Area.

Carpenter, E., & McLuhan, M. (1956). The new languages. *Chicago Review*, *10*(1), 46-52.

Coba, J. A. A., Barrera, L. F. A., & Sánchez, K. P. M. (2022). Perspectivas del Big data. *AlfaPublicaciones, 4*(1.1), 514-531.

Fan, J.-L., Zhen, W.-Z., & Xie, W.-X. (2003). Suppressed fuzzy c-means clustering algorithm. *Pattern Recognition Letters, 24*(9-10), 1607-1612.

Fasce, E. (2007). Aprendizaje profundo y superficial. *Rev Educ Cienc Salud, 4*(1), 2.

Fernández-López, M. (2011). ¿Pueden pensar las máquinas?/Mariano Fernández López.

Fitch, W. (2005). The evolution of language: a comparative review. *Biology and philosophy, 20*(2), 193-203.

Heine, B., & Kuteva, T. (2007). *The genesis of grammar: A reconstruction* (Vol. 9): Oxford University Press.

Jiménez-Zafra, S. M., Cruz-Díaz, N. P., Taboada, M., & Martín-Valdivia, M. T. (2021). Negation detection for sentiment analysis: A case study in spanish. *Natural Language Engineering, 27*(2), 225-248.

Khamphakdee, N., & Seresangtakul, P. (2021). Sentiment Analysis for Thai Language in Hotel Domain Using Machine Learning Algorithms. *Acta Informatica Pragensia, 10*(2), 155-171.

Klahr, D., & Kotovsky, K. (2013). *Complex information processing: The impact of Herbert A. Simon*: Psychology Press.

Kumar, P., & Sarin, G. (2022). WELMSD–word embedding and language model based sarcasm detection. *Online Information Review*.

Logan, R., & Heyer, P. (2001). The sixth language: Learning a living in the internet age. *Canadian Journal of Communication*, *26*(4), 566.

Marshall McLuhan. *Understanding Media: The Extensions of Man*; 1st Ed. McGraw Hill, NY; reissued by Gingko Press, 2003 ISBN 1-58423-073-8.

Millican, P. (2021). Alan Turing and Human-Like Intelligence. *Human-Like Machine Intelligence*, 24.

Moreno, A. (1994). Aprendizaje automático: Edicions UPC.

Nguyen Cong, B., Rivero Pérez, J. L., & Morell, C. (2015). Aprendizaje supervisado de funciones de distancia: estado del arte. *Revista Cubana de Ciencias Informáticas, 9*(2), 14-28.

Nowak, M. A., Komarova, N. L., & Niyogi, P. (2002). Computational and evolutionary aspects of language. *Nature, 417*(6889), 611-617.

Rani, S., Gill, N. S., & Gulia, P. (2021). Survey of Tools and Techniques for Sentiment Analysis of Social Networking Data. *International journal of Advanced computer Science and applications, 12*, 222.

Sánchez, B. P., Cabrera, R. G., Carrillo, M. V., & Castro, W. M. Identifying the polarity of a text given the emotion of its author. Journal of Intelligent & Fuzzy Systems, (Preprint), 1-9.

Su, B.-c., & Luvaanjalba, B. (2021). *The Effect of Hubert Dreyfus's Epistemological Assumption on the Philosophy of Artificial Intelligence.* Paper presented at the International Conference on Human-Computer Interaction.

Tabares Cardona, M. A. (2021). *Aproximaciones al argumento de la habitación china de John Searle.* Manizales.

Tecumseh Fitch, W, Arbib, M.A., & Donald M. (2010). A molecular genetic framework for testing hypotheses about language evolution *The evolution of language* (pp. 137-144): World Scientific.

Tello, J. C., & Informáticos, S. (2007). Reconocimiento de patrones y el aprendizaje no supervisado. *Universidad de Alcalá, Madrid*.

Valenzuela Granados, W. L. (2021). *La inteligencia emocional y su influencia en el proceso de aprendizaje.* Universidad de Guayaquil: Facultad De Filosofía, Letras Y Ciencias.