ISSN 1870-4069

Emotion Mining Mechanism over Texts in Social Media

Luis Casillas, Alejandro Ramirez

University of Guadalajara, Computer Science Department, Mexico
luis.casillas@cucei.udg.mx
alejandro_ramirez_munoz@protonmail.com

Abstract. Nowadays societies are clearly bound to social networking through the Internet. It is common to find out people posting remarks, quotes or moods in social media. Leaders, politicians, celebrities, and ordinary people have posted on social media as a regular in their expression range. Even organizations generate postings on social networks. Humans' manifestation is always linked to the emotions they have. Authors believe that it is possible to discover the sentiment expression from texting. This proposal consists of a model to gather, classify, and emotionally-assess the texting in social networks. The model collects text posts from social media, processes the postings, and generates an emotional assessment. Such evaluation could be bound to the mood of postings' authors. The emotional-assessment consist of inference mechanisms based on knowledge coming from affective dictionaries and automated reasoning.

Keywords: emotion mining, emotions classification, affective dictionary, affective awareness, fuzzy classification.

1 Introduction

Social networks have become a popular mean for informal communication. Along the time, diverse social media have appeared and disappeared. Twitter[1] is one of the most popular social networks, with huge traffic on a daily basis. In social networks, members share large volumes of diverse opinions related to events occurring in their environment during real time.

Society is clearly bound to social media through the Internet. Leaders, politicians, celebrities, and common people have posted on social networks as a regular in their expression range. Even organizations generate postings on social networks. Humans' manifestation is always linked to the emotions they have.

According to Power [21] in their Social Media Benchmark Study:
“67% of consumers have used a company's social media site for servicing, compared with 33% for social marketing. Younger consumers (18-29 years old) are more likely to use brands' social media sites for servicing interactions (43%) than for marketing (23%).”

A second edition of the Social Media Benchmark Study from Power [22] express:
"Nearly one-third (29%) of social media users get recommendations about a product or service from friends and family exclusively through social media. The most frequently

[1] "Twitter" is a registered trademark of Twitter, Inc.

used social media marketing channel is Facebook (29%), followed by YouTube (19%) and Twitter (11%) The most frequently used social media servicing channel is Facebook (84%), followed by Twitter (34%) and YouTube (25%)[2]."

The contributions and messages available in social media tend to follow informal patterns regarding their organization, sequence, and structure. Humans usually express freely, aiming at a human public. People convey ideas without hoping a machine should understand them. Nevertheless, in order to achieve an automated assessment of the human's expression, a machine should understand the messages as a human would do. Such a goal implies binding an emotional awareness of humans' discourse.

Humans' expression is deeply bound to emotional machinery. As Ekman [9] asserts when trying to organize his 20 years-old research, by presenting the emotions' understanding through families and features. Emotions are not isolated states, but families organizing common responses. Any emotion can be measured by nine features as universality, presence in other primates, psychological distinction, events succession, response coherence, onset speed, duration, appraisal, and unbidden occurrence.

Emotions have clearly defined evolutionary goals. As Plutchik [20] establishes in his long-term research, which involves the nature of emotions and is based on defining evolutionary profits to every response in order to achieve survival effects. This approach requires understanding emotions as a "complex chain of loosely connected events", instead of a simple feeling state.

In this context, a feeling could be assumed as a psychological state derived from an opinion or emotion, for example, fear would be a feeling while panic would be an emotion, love is a feeling and passion is an emotion.

According to Yadollahi, Shahraki, & Zaiane [25], "text sentiment analysis has been an attractive topic of study since the mid-1990s". Based on Medhat, Hassan, & Korashy [14], Sentiment Analysis (SA) or Opinion Mining (OM) refers to a computational study of people's opinions, as well as their attitudes and emotions toward an entity. That entity could represent individuals, events or topics.

The work [25] provides a taxonomical structure to Sentiment Analysis (SA). First, the SA is split into Opinion-mining (OM) and Emotion-mining (EM). The OM implies Subjectivity Detection, Opinion Polarity Classification, Opinion Spam Detection, Opinion Summarization, and Argument Expression Detection. EM involves Emotion Detection, Emotion Polarity Classification, Emotion Classification, and Emotion Cause Detection.

The SA is a broad area of study, in which there is no universal analysis technique for all the fields it covers. The diversity of languages can involve greater or lesser complexity from language to language. Depending on the type of problem to address, they must be reviewed the advantages and disadvantages of techniques to decide which strategy suits the requirements of the problem.

Medhat et al. [14] have classified the existing SA techniques under types of approach: Machine Learning (ML), Lexicon Based (LB), and Hybrid. The ML approach involves the already defined machine learning algorithms by supervising the linguistic features in a text. The LB approach is based on a sentiment lexicon, which is

[2] "Twitter" is a registered trademark of Twitter, Inc., "YouTube" is a registered trademark of Google Inc., and "Facebook" is a registered trademark of Facebook, Inc.

a collection of known and precompiled sentiment terms, these techniques might involve a dictionary or corpus. Finally, the hybrid approach involves some creative mixture of ML and LB.

Our proposal consists of an emergent and hybrid strategy to identify and determine the emotional perception that users have about specific events, situations, or people around them. The proposed model performs semantic analysis to determine the presence of emotions in the texts extracted from a social network, chats, emails, etc. In most cases, these texts will not follow grammar and spelling rules, the suggested model could preprocess the information to discriminate data that does not contribute value to the classification process.

The actual study has used data coming from social media, which are highly accessed due to the available information and their services. However, the proposed model could easily be operated over text coming from any social media, text-based chats, web-based blogs, emails, etcetera. Hence, this model must not be bound to a specific social network.

This proposal consists of a scheme for gathering, classify and emotionally-assess text postings in social networks. Although the system can be used on emails, blogs, chats, etc. The system collects text and generates an emotional assessment to the posts. Such an assessment might be bound to the mood of the post's author. The emotional-assessing is provided by an inference mechanism, which is based on knowledge coming from affective dictionaries and automated reasoning.

The paper is organized as follows: "Introduction" presents the study object as some bases from the theoretical frame, "Related work" focuses on the most remarkable projects in sentiment and emotion analyses, "Methodological approach" section presents the analysis strategy for this study, "Applying the model" relates the experience of seizing the strategy over some famous quotes. Finally, the "Conclusion" and "References" sections close to the current proposal.

2 Related Work

The sentiment and emotion analyses have been done for over 15 years. One of the first successful efforts was made by Pang, Lee, & Vaithyanathan [18] through a proposal to classify documents by sentiment instead of the topic. This challenge was interestingly managed through machine learning algorithms, mainly Naive Bayes, Maximum Entropy Classification, and Support Vector Machines. In the same perspective, involving thumbs up or down to phrases, and at the same time Turney [23] presents a simple unsupervised learning algorithm for classifying reviews. That reviews' classification was made by predicting the average semantic orientation of phrases in reviews, involving adjectives or adverbs. Phrases had positive semantic-orientation when they included "good" associations, and negative semantic-orientation when comprised "bad" associations. These preliminary efforts are somehow the base of the impressive opinion-mining available nowadays.

Bifet & Frank [2] developed an approach involving novel algorithms and the use of data-mining tools. These authors used Weka[3] and MOA[4] to support online learning-software based on examples. The training datasets were collected from a popular social network. The goal is to extract features using text filters. Those features are used as analysis-classes to produce sentiment awareness. This model would face challenging scenarios when unbalanced data streams are presented, due to the presence of a scarce list of classes. Such an issue is handled by authors through the use of sliding windows. Unfortunately, this approach has a deep dependency on specific tools, including the social network selected for the study.

Liang & Dai [12] created a system with an architecture that is able to automatically analyze the sentiments of collected messages. These authors' strategy consisted of collecting a set of messages and cataloged them. Those cataloged messages provided the system with the ability to filter, extract and figure out the sentiment direction, as positive or negative. Unfortunately, their approach is tied up to preconceptions for stored messages, implying a lack of flexibility.

The efforts from Feidakis & Daradoumis [10] implied a literature review regarding emotional learning and emotion assessment. This study allowed a deeper understanding of emotions and their impact on the learning processes. Throughout this review, the authors involve an approach to classifying sentiments along with rings as neuro-bio-cognitive, emotional, and socio-cultural. That classifying machinery deals with emotions from biological, cognitive, social, and cultural bases. The present study includes some classifying effort to sentiments and emotions gathered from social-networks texting.

Musto, Semeraro, & Polignano [16] performed a lexicon-based classification. Their experiment was based on two datasets, SemEval-2013[5] and Stanford Twitter Sentiment (STS)[6]. The SemEval-2013 dataset consists of 11,435 tweets already divided into training (8,180 Tweets) and test data (3,255). The tweets have been manually processed. They were tagged and classified as positive, neutral, and negative. The STS dataset has more than 1,600,000 Tweets, already divided into training and testing, but the set of tests is smaller than the training (only 359 Tweets). The approach from these authors needed a preliminary training step to define analysis thresholds. This implies an undesirable consumption of machine resources.

Mukherjee & Bala [15] proposed a practical approach to detect sarcasm in costumers' posts. This is a highly focused field to research. There are few open studies, although the corporative marketing-areas would have plenty of related studies. These authors offer a solution divided into several stages, the first consists of obtaining test data. They gathered 15 thousand tweets having the hashtag #sarcasm. All the retweets were eliminated. Using filters, authors manually classified the Tweets. There were found 2,600 sarcastic texts and 2,400 non-sarcastic posts. Both datasets were used as training packages using cross-validation. The second stage is described as the extraction

[3] Waikato Environment for Knowledge Analysis (Weka) is a suite of machine learning software written in Java, developed at the University of Waikato, New Zealand.

[4] Massive Online Analysis (MOA) is a free open-source software project specific for data stream mining with concept drift. It is written in Java and developed at the University of Waikato, New Zealand.

[5] SemEval-2013 is a challenging task aimed at Sentiment Analysis on Twitter.

[6] The Stanford Twitter sentiment corpus consists of two different sets, training, and test. The training set has 1.6 million tweets automatically labeled as positive or negative based on emotion.

of characteristics. Features extraction becomes essential when dealing with classification problems. Finally, these authors applied two classification algorithms: Naive Bayesian and Maximum Entropy. Unfortunately, their approach has deep connections to specific hashtags and needs non-automated (manual) classification.

Ohmura, Kakusho, & Okadome [17] developed an intensive strategy to assess massive volumes of tweets to produce awareness about public moods. The analysis consisted of Latent Dirichlet Allocation (LDA), which assumes a categorical and multinomial distribution to words. LDA deals with the universe of words, involving diverse sources. This perspective allows the discovery of global trends. These trends are assessed through a six-dimensional mechanism involving daily time series of public mood ("calm," "alert," "sure," "vital," "kind," and "happy"). Our proposal handles every text-asset independently, enabling specific sentiment-analysis.

Another interesting approach to give an affective-analysis has been developed by Cycil, Perry, & Laurier [5], which focuses on conflict resolution in special conditions. Specifically, in the confined scope of passengers in a car. These authors examine the efforts of parents while managing multiple stresses during the driving experience, along with the challenges of distractions from media use and disputes in the car. This kind of emotional scenario can be analyzed through an ethnographic study, involving conversations analysis. Automobiles have become a communication platform. Besides, mobile devices could improve the social interactions among passengers, by presenting relaxing expressions.

The reviewed projects included an important selection of data-mining techniques. Most of the results are accurate and fulfill the purposes. However, some negative aspects can be highlighted. These projects are restricted to certain scopes or they are not open-source. In most of the cases, systems were implemented under specific guidelines and constraints. Thus, they are strongly coupled systems. Besides, it is common that the architecture of the system is not fully described or available.

Our proposal is based on a methodology that provides flexibility to the system. The model is weakly coupled and is enabled to exchange its components in future versions. This proposal follows some principles and techniques that gave good results in reviewed projects. Applications are not limited to specific scopes or situations. Any communication scenario involving words can be measured by this affective analyzer.

3 Methodological Approach

Most of the emotion-classification systems provide results in three categories: negative, neutral and positive. The present project involves the use of fuzzy logic to classify among 27 possible emotions. This approach assumes that emotions cannot be classified with specifically defined limits from qualitative and quantitative variables such as age or maturity.

To capture some elements about the semantic dimension, it is performed a grammatical tagging over the messages that were gathered from social media. But as mentioned before, the proposed model can be applied to diverse text constructions as emails, chats, blogs, etcetera.

As regards social networks, the information produced by people could be hard to analyze since users will likely disregard correct grammatical structures. The grammatical categorization of words is a helpful technique to discriminate valuable information from the noise that does not contribute anything and keeps vainly occupying processing time.

Hence, a Part of Speech (PoS) tagging is performed over each word and punctuation mark in the text. Along with this process, text tokens will be bind to morphosyntactic labels. Different labeling systems use different sets of labels, but usually, a label describes a word class and some specific characteristics of that word class, for instance, number and gender. The number of labels varies between dozens and hundreds.

Carlberger & Kann [6] highlight two problems that should be addressed when developing a grammar tagger:

1. Finding out all the possible labels for each word. The complexity of this task will increase when the words are unknown to the tagger since it must infer a possible label, otherwise, the tagger must fail on tagging and report a not-found result. When all the words to tag are known, the complexity decreases.
2. Choosing a single label for the word in the specific position of the sentence. This is known as syntactic disambiguation, and it must be solved on every ambiguous word in the sentence. Ambiguous words are frequently present in most languages.

Due to the large scale to develop a grammatical tagger, the present project conferred the tagging tasks to the Stanford CoreNLP [13]. A powerful, flexible, and extensible natural language processing tool.

Another methodological aspect involved is the use of affective dictionaries: packages of words evoking feelings, affective response, emotions, emotive actions, etc. According to Warriner et al. [24], "information about the affective meanings of words is used by researchers working on emotions and moods, word recognition and memory, and text-based sentiment analysis. Three components of emotions are traditionally distinguished: valence (the pleasantness of a stimulus), arousal (the intensity of emotion provoked by a stimulus), and dominance (the degree of control exerted by a stimulus)". Some psychologists conceptualize the emotion along two dimensions: "valence" and "arousal"; excluding the "dominance". Lewis et al. [11] established that "arousal invokes a single axis of intensity increasing from neutral to maximally arousing. Valence can be described variously as a bipolar continuum, as independent positive and negative dimensions, or as hedonic value (distance from neutral)."

The present project has involved the three affective components: valence, arousal, and dominance to perform the proposed sentiment analysis over social media.

It is very common that this type of research makes use of the affective dictionary ANEW from Bradley & Lang [4], which consists of 1034 words. Based on the diversity of messages, the number of words in ANEW may be sufficient to obtain acceptable results. In the present project, it is assumed that there is an important diversity of messages since different topics can be exposed, as well as the presence of expressions regarding temporary nature. Besides, the scope of social networks implies diverse social and cultural situations. Hence, the affective dictionary from [24] was selected. This extended dictionary consists of 13915 words, which has been assessed under the same principles of ANEW. The ANEW is inside in the extended version from [24].

The third main stage in the proposed strategy is fuzzy logic. This classification technique is one of the most widely used inference methods in many areas. Fuzzy logic allows simulating diverse procedures of human reasoning in knowledge-based systems.

It deploys models that allow dealing with the uncertainty happening in cognitive human-processes performed by computers. Although fuzzy logic was already used in the projects [1, 7], the actual approach has considered a completely new classification understanding over the input variables, as well as for the implied emotion categories in the output. This new, experimental, and from-scratch organization has implied a simpler, dynamic, and accurate processing of data.

The input variables are valence, arousal, and dominance, which are the analysis dimension produced by the affective dictionary from [24]. Every input variable is fed with data coming from the text analysis. This procedure consists in selecting from the analyzed text those words found in the affective dictionary. All the emotion ratings per word are summarized and averaged. Averaged values are settled in the corresponding input variables. Now, every input variable is fuzzified to produce a qualitative representation of the stored average. As shown in figures 1 and 2.

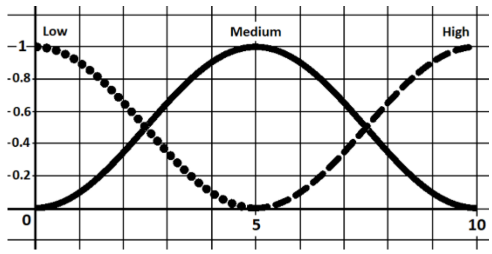

Fig. 1. Membership functions to produce fuzzy representation for input variables arousal and dominance: Low, Medium, and High. The shown curves and their actual dimensions are presented in this figure for demonstration purposes. The classifying system uses specific and floating limits, based on the indicators for centrality and dispersion of the data from the affective dictionaries.

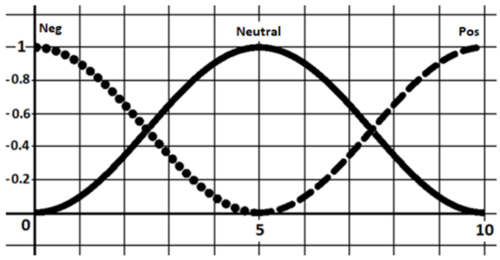

Fig. 2. Membership functions to produce fuzzy representation for the input valence: Negative, Neutral, and Positive. The shown curves and their actual dimensions are presented in this figure for demonstration purposes. The classifying system uses specific and floating limits, based on the indicators for centrality and dispersion of the data from the affective dictionaries.

Membership functions in figures 1 and 2 are produced by the formulas (1), (2), and (3). Formula (1) is known as Z function, formula (2) is known as soft-lambda (sΛ) function, and formula (3) is known as S function. The Z function has been bound to the linguistic-value "Low" in figure 1, and the linguistic-value “Negative” in figure 2. The sΛ function has been bound to the linguistic-value "Medium" in figure 1, and the linguistic-value “Neutral” in figure 2. The S function has been bound to the linguistic-value "High" in figure 1, and the linguistic-value “Positive” in figure 2. These formulas are based on the proposal from [8].

$$Z=1,\ \text{if } x < \text{left};\quad Z=(1+\cos(((x-\text{left})/(\text{right}-\text{left}))*\pi))/2,\ \text{if left} \le x \le \text{right};\quad Z=0,\ \text{if } x > \text{right} \tag{1}$$

$$s\Lambda=0,\ \text{if } x < \text{left or } x > \text{right};\quad s\Lambda=(1+\cos(((x-\text{center})/(\text{center}-\text{left}))*\pi))/2,\ \text{if left} \le x < \text{center};\quad s\Lambda=(1+\cos(((x-\text{center})/(\text{right}-\text{left}))*\pi))/2,\ \text{if center} \le x \le \text{right} \tag{2}$$

$$S=0,\ \text{if } x < \text{left};\quad S=(1+\cos(((x-\text{right})/(\text{right}-\text{left}))*\pi))/2,\ \text{if left} \le x \le \text{right};\quad S=1,\ \text{if } x > \text{right} \tag{3}$$

The borders “left”, “center” and “right” in formulas (1), (2), and (3) refer to specific limits that can be settled dynamically in the model, according to actual centrality and dispersion data from the general average and the standard deviation shown by the dataset to classify.

Once the input variables are fuzzified, the resulting linguistic-values are used to produce the corresponding fuzzy output. Figure 3 shows the 3D crossing structure for input variables. The 27 inner regions produced by this crossing process are bound to the qualitative results for output variable: emotion.

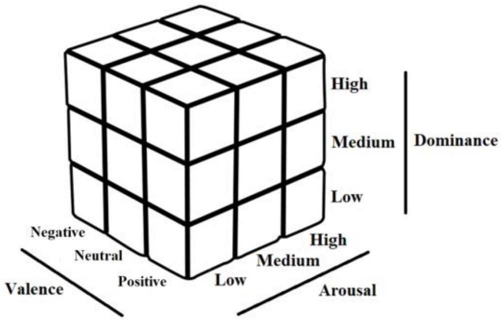

Fig. 3. This is the 3D classifying structure. It involves the three input variables as dimensions. Every variable has three segments. Segments refer to linguistic values.

Regarding the classification process, the input variables are fed through the sum and averaging of accumulated values. These values are collected from the identified words in the affective dictionary. These calculations produce sharp and precise numbers. Due to computational complexity to deal with these numerical representations, data are fuzzified. The fuzzy version allows a simplified decision scheme when predicting and classifying the emotional state. Instead of an unlimited continuum of emotions, the actual emotional state is selected from a limited set with 27 alternatives. These alternatives are the emotions set up by Boulic et al. [3]: Scared, Afraid, Empathic, Embarrassed, Surprised, Delighted, Furious, Vigilant/Alert, Excited, Depressed/Sad, Doubtful/Pessimistic, Compassionate, Unhappy, Bored/Sad, Tired, Peaceful, Anxious, Happy, Neutral, Sleepy, Glad, Angry, Alert/Willing, Sulky, Annoyed, Malicious, and Vicious.

It is true that humans have emotional manifestations as an infinite continuum. However, it is also true that computers have real problems when detecting and handling humans' emotions. The present proposal gives a simplified scheme to predict emotional profiles for social-networking users, based on their text postings. Nevertheless, any specific diagnosis for the emotional situation would need professional assessing performed by humans. The automated predictions made by the proposed system aims at producing some guidelines to measure and, in some cases, feedback users.

The project [3] is aimed at producing facial expressions for avatars and they directly use the actual values from valence, arousal, and dominance to produce adjustments on the facial components of avatars. The proposal in our study is aimed to produce a qualitative response to fuzzified values from the input, by processing the text according to an affective dictionary. Tables 1, 2, and 3 explain the classification rules from linguistic representations of input variables. These tables represent the 3D classification layers as 2D ensembles. The sliced variable is dominance, its organization is also inspired by the distribution set up at [3], although some reorganization was made over the original proposal.

Table 1. Classification rules among Valence, Arousal, and Dominance. Low Dominance.

Dominance: Low		**Arousal**		
		Low	**Medium**	**High**
Valence	**Negative**	Bored/Sad	Depressed/Sad	Afraid
	Neutral	Tired	Doubtful/Pessimistic	Scared
	Positive	Peaceful	Compassionate	Empathic

Table 2. Classification rules among Valence, Arousal, and Dominance. Medium Dominance.

Dominance: Medium		**Arousal**		
		Low	**Medium**	**High**
Valence	**Negative**	Anxious	Unhappy	Embarrassed
	Neutral	Sleepy	Neutral	Surprised
	Positive	Glad	Happy	Delighted

4 Applying the Model

The analysis performed by the proposed model is fed from social media elements. Authors have decided to apply the emotional-classification to classical quotes. These quotes have been written or spoken by famous authors or personages. Social media is currently under scrutiny, due to the privacy-risks that can be implied to the users of these services. Hence, authors have decided to apply the emotion mining analysis over famous quotes. The very same emotional classification can be applied to any text from social-networks postings, mini-blogs, emails, etc. when the texts' authors have agreed on the emotion-mining.

Table 3. Classification rules among Valence, Arousal, and Dominance. High Dominance.

Dominance: High		**Arousal**		
		Low	**Medium**	**High**
Valence	**Negative**	Sulky	Angry	Furious
	Neutral	Annoyed	Alert/Willing	Vigilant/Alert
	Positive	Vicious	Malicious	Excited

In table 4 is presented a detailed analysis of one quote. Quotes are entered into the CoreNLP, which proceeds to tokenize and tag every phrase. The CoreNLP tagging is based on the treebank approach from [19]. All words are searched in the affective-dictionary, but only some of them use to be there. Affective-dictionaries have limited corpus, due to the massive effort to produce affective-data for words. Those found words give data to the input variables. From the amounts collected in summarized input variables, the fuzzy classifier produces qualitative representations.

Finally, these qualitative inputs imply a specific emotion bounded to the quote. The automated process for emotion-implying is decorated by a certainty level, which is based on membership satisfaction during the classifying stage.

As shown in table 4, a sentence is automatically processed to produce a prediction of the emotion beating in the quote's author when the phrase was created. As well as the certainty bound to such prediction. Thus, the quote "you know you're in love when you can't fall asleep because reality is finally better than your dreams" from Dr. Seuss, was written under a HAPPY emotional state. In addition, this prediction has a 70.8% certainty. This incomplete-certainty can be explained by the fuzzy classification process when values had no full membership to the involved sets.

Unfortunately, the analysis does not include all the words in the sentence, but as explained before: affective-dictionaries have limited sets of words. A similar analysis was performed over some other famous quotes. Table 5 holds the results from these analyses.

Input variables are now fuzzified into three values. This approach simplifies dealing with data and decision making. Nevertheless, there is a significant loss of information. A highly precise value is fuzzified into one of three alternatives. A future approach for this solution would be based on a more precise splitting for variables arousal and dominance. If the semantic differential scale has five levels as "very low", "low", "medium", "high", and "very high", 75 regions would result by crossing five values for arousal, five values for dominance, and three values for valence.

The 27 emotions defined at [3] would be the same, but now they would be spread along the 75 regions. There would be a repetition of emotions in the regions, but the system could provide improved performance with precise responses. Future efforts in this research line would produce an assertive distribution of emotions in these 75 regions.

Table 4. An example of system's operation over a quote.

Quote	
"You know you're in love when you can't fall asleep because reality is finally better than your dreams." Dr. Seuss	
Words found in the affective dictionary:	**Results**
Word: "know", PartOfSpeech: VERB_SINGULAR_PRESENT_NONTHIRD_PERSON :: Arousal: 3.24,Valence: 6.82, Dominance: 5.78 Word: "love", PartOfSpeech: NOUN :: Arousal: 5.36, Valence: 8, Dominance: 5.92 Word: "fall", PartOfSpeech: VERB :: Arousal: 4.24, Valence: 3.89, Dominance: 3.83 Word: "asleep", PartOfSpeech: ADVERB :: Arousal: 2, Valence: 6.5, Dominance: 4.33 Word: "reality", PartOfSpeech: NOUN :: Arousal: 4.1, Valence: 5.73, Dominance: 7.26	**Numerical results:** Averaged Arousal: 3.78 Averaged Valence: 6.19 Averaged Dominance: 5.42 **Fuzzy classification:** Fuzzy Arousal: **Medium** Fuzzy Valence: **Positive** Fuzzy Dominance: **Medium** ********************** *Implied Emotion:* ***HAPPY*** *Certainty level:* ***70.8%***

Table 5. More examples of emotion-mining on quotes.

Quote	Results
"You cannot escape the responsibility of tomorrow by evading it today." Abraham Lincoln	*Implied Emotion:* ***WILLING*** *Certainty level:* ***55.86%***
"A revolution is a struggle to the death between the future and the past." Fidel Castro	*Implied Emotion:* ***AFRAID*** *Certainty level:* ***81.09%***
"The keenest sorrow is to recognize ourselves as the sole cause of all our adversities." Sophocles	*Implied Emotion:* ***SLEEPY*** *Certainty level:* ***67.3%***
"Jealousy is that pain which a man feels from the apprehension that he is not equally beloved by the person whom he entirely loves." Joseph Addison	*Implied Emotion:* ***EMBARRASED*** *Certainty level:* ***54.22%***
"The true soldier fights not because he hates what is in front of him, but because he loves what is behind him." G.K. Chesterton	*Implied Emotion:* ***ALERT*** *Certainty level:* ***74.12%***

Regarding the current use of famous quotes as the target of the system's operation. The reader is invited to consider that these quotes were expressed by quote author in social media, even though some of those authors would not have agreed with the experience of using social networks as most people do. Nevertheless, the system is currently enabled to extract, process, and emotionally classify assertions, conveys, arguments, postings, blogging, etcetera. All of them regular nowadays expression mechanisms under the ceiling of modern Information and Communications Technologies (ICT).

5 Conclusion

Humans' expression is deeply bound to emotional machinery. Emotions have clearly-defined evolutionary goals. Nowadays social-media has become a standard expression mechanism. The academic/education environments, the collaborative projects, the business arena, and almost any scope with human interaction are clearly bound to the emotional states of people.

Authors of this proposal believe that any awareness about the current emotional state of people might be useful. There are diverse sources of emotional awareness, such as human' faces, handwriting, sweating, etcetera. This proposal has focused on the word choice during persons' text-based assertions. This analysis is based on the grammatical categorization and the affective assignment for words in text messages.

Affective indicators define sentiment loads to words and expressions. The system developed as part of this proposal is enabled to recover, process, tag, and affectively assess text-based arguments. Nevertheless, the authors of this study have considered that sentiment analysis over social-media implies an important responsibility. Text elements project more than people may want to reveal. Hence, any analysis over expressions in social networks must be carefully handled.

There must be always a formal permission-request to analyzed people, as well as a formal consent for that analysis. That is the reason this proposal has processed famous-quotes in the public domain, instead of specific extracts from social-media. Any analysis of specific text from social-media would run the same way as shown for famous-quotes. The selected quotes for this study do not represent any specific claim from the authors of this study. They are only test cases.

The current analysis gives an awareness about emotions that could be laying-beneath the arguments in quotes. Although the system is currently enabled to extract, process, and emotionally classify assertions, conveys, arguments, postings, blogging, etcetera.

Besides, some other components might be discovered through a syntax analysis. Authors believe that adverbs and adjectives could involve some active modification over the already defined numbers for the input variables. This active response, as well as considering the 75 regions previously mentioned will lead the future efforts of current research.

References

1. Arguedas, M., et al.: A model for providing emotion awareness and feedback using fuzzy logic in online learning. Soft Computing 22(3), 963–977 (2018)
2. Bifet, A., Eibe, F.: Sentiment knowledge discovery in twitter streaming data. International conference on discovery science. Springer, Berlin, Heidelberg (2010)
3. Boulic, R., et al.: Towards the Instantaneous Expression of Emotions with Avatars. Cyberemotions. Springer, Cham, 255–278 (2017)

4. Bradley, M.M., Lang, P.J.: Affective norms for English words (ANEW): Instruction manual and affective ratings 30(1), Technical report C-1, the center for research in psychophysiology, University of Florida (1999)
5. Chandrika, C., Perry, M., Laurier, E.: Designing for frustration and Disputes in the family Car. International Journal of Mobile Human Computer Interaction (IJMHCI) 6(2), 46–60 (2014)
6. Carlberger, J., Kann, V.: Implementing an efficient part-of-speech tagger. Software: Practice and Experience 29(9), 815–832 (1999)
7. Casillas, L., Peña, A., Gutierrez, A.: Towards an Automated Model to Evaluate Collaboration through Non-Verbal Interaction in Collaborative Virtual Environments. Intelligent Systems: Concepts, Methodologies, Tools, and Applications. IGI Global, pp. 1570–1586 (2018)
8. del Brío, B.M., Sanz Molina, A.: Redes neuronales y sistemas difusos. Alfaomega Ra-Ma (2002)
9. Ekman, P.: An argument for basic emotions. Cognition & emotion 6(3-4), 169–200 (1992)
10. Feidakis, M., Thanasis, D.: A framework for designing computer supported learning systems with sensibility. International Journal of e-Collaboration (IJeC) 9(1), 57–70 (2013)
11. Lewis, P.A., et al.: Neural correlates of processing valence and arousal in affective words. Cerebral cortex 17(3), 742–748 (2006)
12. Liang, P.W., Bi-Ru, D.: Opinion mining on social media data. In: 2013 IEEE 14th International Conference on Mobile Data Management Vol. 2 (2013)
13. Manning, C. et al.: The Stanford CoreNLP natural language processing toolkit. In: Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations. (2014)
14. Medhat, W., Hassan, A., Korashy, H.: Sentiment analysis algorithms and applications: A survey. Ain Shams engineering journal 5(4), 1093–1113 (2014)
15. Mukherjee, S., Pradip, K.B.: Detecting sarcasm in customer tweets: an NLP based approach. Industrial Management & Data Systems 117(6), 1109–1126 (2017)
16. Cataldo, M., Semeraro, G., Polignano. M.: A comparison of lexicon-based approaches for sentiment analysis of microblog posts. Information Filtering and Retrieval 59 (2014)
17. Ohmura, Masahiro, Koh Kakusho, Okadome, T.: Tweet sentiment analysis with latent dirichlet allocation. International Journal of Information Retrieval Research (IJIRR) 4(3) 66–79 (2014)
18. Pang, Bo, Lillian Lee, Vaithyanathan, S.: Thumbs up? Sentiment classification using machine learning techniques. Proceedings of the ACL-02 conference on Empirical methods in Natural Language Processing - Volume 10. Association for Computational Linguistics (2002)
19. Prasad, R. et al.: The Penn Discourse TreeBank 2.0. In: LREC (2008)
20. Plutchik, R.: The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. American scientist 89(4), 344–350 (2001)
21. Power, J.D.: Poor social media practices can negatively impact a businesses' bottom line and brand image (2013)
22. Power, J.D.: Positive Automotive Social Media Experience Impacts Purchase Decisions across All Generations. April 10 - September. http://www.jdpower.com/press-releases/2014-social-media-benchmark-study-auto (2014)
23. Turney, P.D.: Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th annual meeting on association for computational linguistics. Association for Computational Linguistics (2002)
24. Warriner, A.B., Kuperman, V., Brysbaert, M.: Norms of valence, arousal, and dominance for 13,915 English lemmas. Behavior research methods 45(4), 1191–1207 (2013)

25. Yadollahi, A., Shahraki, A.G., Zaiane, O.R.: Current state of text sentiment analysis from opinion to emotion mining. ACM Computing Surveys (CSUR) 50(2), p. 25 (2017)