ARTICLE IN PRESS

Future Generation Computer Systems xxx (xxxx) xxx

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Improved emotion recognition in Spanish social media through incorporation of lexical knowledge

Flor Miriam Plaza-del-Arco [a,*], M. Teresa Martín-Valdivia [a], L. Alfonso Ureña-López [a], Ruslan Mitkov [b]

[a] Department of Computer Science, Advanced Studies Center in ICT (CEATIC), Universidad de Jaén, Campus Las Lagunillas, 23071, Jaén, Spain
[b] Research Group in Computational Linguistics, University of Wolverhampton, UK

ARTICLE INFO

Article history:
Received 29 April 2019
Received in revised form 5 September 2019
Accepted 20 September 2019
Available online xxxx

Keywords:
Emotion recognition
Machine learning
Affective features
Emotion lexicon
Natural language processing
Social media

ABSTRACT

Emotions play an important role in human intelligence and behaviour and are a major vehicle for communication. Therefore, the integration of emotions in computational models can improve the human–computer interaction systems. In this paper, we present a study of different machine learning approaches to automatically recognise emotions in messages written in Spanish on social media. Although the computational treatment of emotion is more difficult than other sentiment analysis tasks, the baseline of some machine learning algorithms achieve an acceptable accuracy showing that it is possible to tackle the problem using some basic natural language processing techniques. In this study we have experimented with the integration of knowledge from different affective lexical resources. We conclude that the incorporation of lexical affective features leads to improvement over most baseline figures with significant improvement. Indeed, we observe that the use of resources generated particularly for emotion recognition in other languages than English is a promising approach to enhance basic machine learning systems. Particularly, we used a Spanish lexical resource and we notice that it always improves the results. In the best case, it improves 6.15% of the results obtained using the Naive Bayes classifier.

1. Introduction

Emotions play an adaptive, social or motivational role in the life of human beings, as they account for different characteristics indicative of human behaviour, such as emotional state, level of interest or alertness. Studying patterns of human emotions and how people feel is essential in various applications such as public health and safety, business intelligence, emergency response and e-learning environment. At other times, they serve as an effective means of communication that crosses the language barrier.

In recent years, there has been a growing interest in automatically detecting and generating emotions in texts, with a number of promising studies reported [1]. Therefore, affective computing is a key element of the advancement of Artificial Intelligence. The aim of affective computing or artificial emotional intelligence is to enable computers to recognise the emotional states and behaviour of a human developing systems and applications that can analyse, recognise and adapt to the user's emotional states [2].

Human emotions can be expressed through verbal communication (speech, textual data) or non-verbal communication (facial expressions, gestures). Currently, more and more users are using message platforms, social media, blogs, or forums to communicate with others so text is a particularly important source of data with emotional content on the Web. In fact, social media contain a large corpus of public real-time data that is rich with emotional content. Therefore, the large growth of emotion-rich textual data requires automate identification and analysis of people's emotion expressed in text [3].

However, the development of systems able to automatically analysing natural language with the objective of understand its emotional content is a very hard process. In fact, several studies agree that detecting and analysing emotions in text is a complex problem and the interpretation depending on the context [4–6].

Accordingly, the present study is framed within the problem of text-based emotion mining. The main contributions are as follows:

1. Detecting and analysing the emotion expressed in social media messages. In the affective computing area, we find several works focus on polarity classification. However, we find few studies related to emotion classification due to its complexity.

* Corresponding author.
E-mail addresses: fmplaza@ujaen.es (F.M. Plaza-del-Arco), maite@ujaen.es (M.T. Martín-Valdivia), laurena@ujaen.es (L.A. Ureña-López), r.mitkov@wlv.ac.uk (R. Mitkov).

https://doi.org/10.1016/j.future.2019.09.034
0167-739X/

2. We focus on Spanish texts mainly due to the following reasons: It is a major language used in social media; Most research on this subject so far handles English documents; It is important to study its syntactic and semantics in order to develop accurate systems in this language and not to use systems adapted for other languages.
3. Developing a baseline system based on different supervised classifiers and comparing the results obtained. A number of Machine Learning (ML) classification methods are employed for emotion text categorisation in order to study the one that achieves the best results.
4. Studying whether the integration of external knowledge improves a baseline Spanish emotion recognition system based on Machine Learning (ML) approach. We employ lexical features provided by different emotion lexicons and we integrated them into baseline systems seeking to establish whether it improves emotion classification.
5. Comparing multilingual affective lexical resources in our systems. We want to study if the use of features derived from other lexicons adapted in Spanish with a simple translation is enough to improve the emotion classification systems. We also want to show the importance of generating affective lexicons for a specific language.

The rest of the paper is organised as follows. Sections 2 and 3 discusses the background and related work on emotion classification in text. The resources used to develop our systems are described in Section 4. The proposed methodology to detect and analyse emotion in text is presented in Section 5. Section 6 reports the experimental results of our systems. Finally, we conclude our paper with a discussion of the results in Section 7.

2. Background

One of the first studies related on affective computing is the work of Picard [7]. She put forward the idea of training computers to identify human emotions and argues why computers would need the ability to recognise them. Building affective computers requires a multi-modal processing since a human being can express emotions from a wide range of behavioural cues. Researchers have been working on different human sources of information such as gestures [8], speech [9,10], movements [11], facial expression [12,13] or physiological signals (brain electrical activity, respiratory rate, hear rate, salivation). With emotion recognition in text considered to be the most recent branch of affective computing. In fact, social networks provide a huge source of textual emotional expressivity. This is one of the reasons why many researchers from areas such as Natural Language Processing (NLP), Artificial Intelligence (AI) or psychology are interested in this new field of application.

Online social networks and microblogging platforms (e.g., Facebook, Youtube, Twitter) have become an unprecedented global phenomenon that allows people to establish connections and communicate with each other or with organisations and public figures that maintain a presence in services. In these platforms, people can publish a series of messages that include their current activities, thoughts, feelings, photos, videos and so on. This information may contain indicators of emotions of individuals such as anxiety, happiness or depression. Thus, most of the textual posts or messages contain subjective and emotional information that could be extracted and analysed in order to build an automated system capable of detecting human emotion from text.

Opinion mining and Emotion mining from text are part of the Sentiment Analysis (SA) area, but they have different objectives [14]. Opinion mining is concerned with the study of opinions expressed in texts, whereas Emotion mining is related to the study of emotions based on predefined emotion models according to psychological emotion theories.

Emotion detection from text provides different applications in almost every aspect of our day-to-day life, such as monitoring mental health of people like depression [15], improving business strategies according to the preferences of consumers [16], tracking public emotions during elections and prediction based on these emotions [17,18], detecting potential criminals or terrorist from analysing the emotions of people after a terrorist attack or crime [19], identifying if the headline of a news is safe or unsafe for incorporating advertisements [20] or making efficient e-learning systems to improve student motivation [21].

3. Related work

In order to understand emotions, it is necessary to review the literature from psychology, philosophy, sociology, neuroscience, biology and other fields over time. Several models emerged to cover all possible human emotions. Perhaps the most popular and widely used approaches related to emotion recognition is the Ekman's emotion theory [14] which specifies six basic human emotions: *anger*, *fear*, *sadness*, *happiness*, *surprise* and *disgust*. These emotions are characterised as universal, as they are expressed in the same way across different cultures an eras.

Several approaches can be found for textual emotion detection. In fact, they have been classified into several categories by different researchers. The methods established by these researchers can be grouped into four categories: *Keyword-based Method*, *Lexicon-based Method*, *Machine learning Method*, *Hybrid Method* and *Deep learning Method* [19,22]. The idea in the *Keyword-based Method* is to find out patterns similar to emotion keywords and match them. *Lexicon-based approach* emotion detection approach is an unsupervised technique that classifies text using a lexicon (a knowledge-base with text labelled according to emotions) appropriate for the input dataset. *Machine learning methods* are used for textual emotion detection in which a model is designed to train a classifier with a part of the dataset and then test the classifier with the rest of the data. *Hybrid approach* for emotion detection in text combines any two or all three methods defined to achieve the benefit of multiple methods and reach the maximum level of accuracy. *Deep learning models* have been used to develop end-to-end systems in many tasks including speech recognition, text classification, and image classification. It has been shown that such systems automatically extract high-level features from raw data [23,24]. Variations of Recurrent Neural Networks, such as Long Short Term Memory networks (LSTM) [25] and BiLSTM [26] have been effective in modelling sequential information.

Recently, approaches which employ Deep learning for emotion classification in text have been proposed. Zahiri and Choi [27] propose a novel Hierarchical LSTMs for Contextual Emotion Detection (HRLCE) model. Moreover, they combined a novel transfer learning model called BERT and their method in an ensemble achieving better results in the classification. Authors in [28] develop a novel deep learning-based system that addresses the emotion classification problem in text. They propose a novel method to transform it to a binary classification problem and exploit a deep learning approach to solve the transformed problem. The key component of the system was the embedding module, which used three embedding models and an attention function. Chatterjee et al. [22] tackle the problem of emotion detection in English textual dialogues. The input user utterance in the model is fed into two LSTM layers using two different word embedding matrices. One layer uses a semantic word embedding, whereas the other layer uses a sentiment word embedding. These two layers learn semantic and sentiment feature representation and encode sequential patterns in the user utterance.

A lot of research on textual emotion classification is based on building and using emotion lexicons. For example, Strapparava and Mihalcea [29] construct a dataset of news titles annotated for emotions, and propose a methodology for fine-grained and coarse-grained evaluations. Neviarouskaya et al. [30] focus on textual affect sensing and visualisation in virtual communication environments applying a rule-based approach at a sentence-level. Rao et al. [31] propose an efficient algorithm and three pruning strategies to automatically build a word-level emotional dictionary for social emotion detection. Bandhakavi et al. [32] study the problem of emotion feature extraction using the knowledge of domain-specific lexicons and general purpose emotion lexicons. Related to shared tasks, most of the works presented in the well-known competition SemEval [33] use affect lexicons and conclude that they are very valuable source of information because they provide prior information about the type of emotion associated with each term of the text. Moreover, in WASSA-2017 Shared Task on Emotion Intensity it was demonstrated that using features from affect lexicons is useful for emotion mining tasks [34]. However, we can find very few affect lexicons available and most of them have been generated for English [14]. For example, WordNet-Affect (WNA) [35], ANEW [36], Linguistic Enquiry and Word Count (LIWC) [37] or NRC Word-Emotion Association Lexicon (Emolex) [38] are some of the most commonly used by researchers.

At the same time, resources and studies focused on other language different from English are very sparse. In the case of Spanish, there are studies related to opinion mining. For example, Molina-González et al. [39] study the integration of domain information for a Spanish polarity classification system and Martínez-Cámara et al. [40] perform a study of different features and machine learning algorithms for classifying the polarity of Spanish Twitter posts. However, for emotion recognition practically there are no resources and works. To the best of our knowledge, there is only one lexicon of emotions created specifically for Spanish called SEL (Spanish Emotional Lexicon) [41] that has been used in a few studies. There are also some attempts to adapt English resources to Spanish. For instance, Redondo et al. [42] work on the adaptation of ANEW and evaluated it in the dimensions of valence, arousal and dominance, Plaza-del-Arco et al. [43] adapt the NRC Affect Intensity Lexicon [44], and in [45] the authors propose a method to use WordNet-Affect in Spanish demonstrating the difficulty of the emotion detection task and showing some interesting features in the lexicon approaches.

4. Resources

In this section, we describe the resources used to carry out our experiments. Specifically, we used a dataset called Affect in Tweets and three different emotion lexicons including Spanish Emotion Lexicon, Improved Spanish Opinion Lexicon and NRC Word-Emotion Association Lexicon.

- **Affect in Tweets (AIT) Dataset**. We use the dataset provided by the organizers in SemEval 2018 Task 1: Affect in Tweets [46]. The dataset associated with EI-oc subtask is composed by a set of tweets that belong to an emotion E (*anger*, *fear*, *joy*, and *sadness*). It is available in English, Arabic and Spanish. The AIT dataset was partitioned into training, development and test sets for experiments as described in Table 1.
- **Spanish Emotion Lexicon (SEL).**[1] [41] SEL contains 2,036 Spanish words. Each word is associated with the measure of Probability Factor of Affective use (PFA) with respect to at least one basic emotion of Ekman: *anger*, *fear*, *sadness*, *joy*, *surprise*, and *disgust*. An example can be seen in Table 2. Words were marked manually by 19 annotators implementing certain agreement thresholds. It is important to note that SEL was developed specifically for Spanish emotions and it has been used in some works [47,39].
- **Improved Spanish Opinion Lexicon (iSOL)**[2] [48]. This lexicon is composed by a list of polarity Spanish terms. It was generated by translating into Spanish the Bing Liu English Lexicon [49] and correcting the translation manually. iSOL contains 2,509 positive words and 5,626 negative words. Some examples of different terms can be seen in Table 3. This lexicon has been successfully used in lots of works but most of them are focused on polarity classification for Spanish sentiment analysis [50,40,51].
- **NRC Word-Emotion Association Lexicon (EmoLex)**[3] [38]. The NRC Emotion Lexicon, commonly known as EmoLex, contains a list of English words associated to one or more of the following emotions: *anger*, *fear*, *sadness*, *joy*, *anticipation*, *trust*, *surprise*. It is possible than a single word is associated to more than one emotion. This lexicon was built specifically for English but it is also available for more than one hundred languages (including Spanish) since it was automatically translated using Google Translate.[4] Examples can be seen in Table 4. The Spanish version of EmoLex has been applied in some works but the results are not very encouraging, probably because the resource has not been specifically created for Spanish and it loses accuracy in the translation process.

Table 1
Number of tweets per emotion in the SemEval-2018 AIT Dataset.

Dataset	Train	Dev	Test	Total
Anger	1166	193	627	1986
Fear	1166	202	618	1986
Joy	1058	202	730	1990
Sadness	1154	196	641	1991

Table 2
Example of SEL words with the associated emotion and PFA.

Word	Anger	Fear	Sadness	Joy	Surprise
Animar	0	0	0	0.797	0
Atormentar	0.664	0	0.53	0	0

Table 3
Example of iSOL words with the associated polarity.

Word	Polarity
Triunfante	Positive
Virtud	Positive
Fastidio	Negative
Infierno	Negative

Table 4
Example of EmoLex words with the associated emotion.

Word	Anger	Fear	Sadness	Joy	Anticipation	Disgust	Trust	Surprise
Amor	0	0	0	1	0	0	0	0
Abandonar	0	1	1	0	0	0	0	0

1 https://mailman.uib.no/public/corpora/2012-December/016707.html

2 http://sinai.ujaen.es/wp-content/uploads/2013/05/isol.tar.gz

3 https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm

4 https://translate.google.es/?hl=es

Please cite this article as: F.M. Plaza-del-Arco, M.T. Martín-Valdivia, L.A. Ureña-López et al., Improved emotion recognition in Spanish social media through incorporation of lexical knowledge, Future Generation Computer Systems (2019), https://doi.org/10.1016/j.future.2019.09.034.

5. Experiments

5.1. Methodology

Our systems perform classification of four basic emotions (*anger*, *fear*, *sadness* and *joy*) of an input tweet written in Spanish. For this reason, we use the Spanish AIT Dataset described in Section 4. We trained the models on the train and dev sets. Then, we tested our models using a different set, the test set.

Scikit-learn [52], a free software machine learning library for Python was used in our experiments.

Given the inherently unstructured nature of text data, they needed careful preparation before they can be analysed. With a view to preparing data for text classification, we preprocessed the tweets of AIT Dataset following the next steps: the tweets were tokenised using NLTK TweetTokenizer[5] and all letters were converted to lowercase.

The accuracy of a learning system depends on its representation of the problem. With regard to the text categorisation task it is important to transform the document, which is mostly a string of characters, into an appropriate representation for the learning classifier. In our experiments, in order to train a classifier from labelled data, we represent each tweet as a vector of numerical features using the Frequency Term weighting (TF) which converts the text document collection into a matrix of integers generating a sparse matrix of the counts. Thus, a set of features that illustrate the emotion expressed by each tweet is needed. Moreover, we use single words or unigrams as the baseline features for comparison. Other features were explored in the next section.

We have developed different machine learning approaches for the task of emotion classification. We use the first one as a baseline model and to this end we have developed different basic ML classifiers as can be seen in Fig. 1. The second one is an approach integrating external knowledge with specific affective features derived from lexicons, the workflow is shown in Fig. 2. The two types of systems implemented will be presented in Sections 5.3 and 5.4.

5.2. Experimental study

In this section, we focus on study two different approaches: a baseline model base on machine learning algorithms and another model that uses affective features from lexicons.

5.3. Baseline ML systems

A number of ML classification methods have been applied for text categorisation. To classify emotion are discussed four different models. We select *Support Vector Machine* and *Logistic Regression* as decision boundary classifiers, *Multilayer Perceptron* as a feedforward artificial neural network and *Naive Bayes* as a probabilistic classifier. Fig. 1 illustrates a general pipeline of the baseline system.

- *Support Vector Machine (SVM)*[6] [53] is a statistical classification method. It is one of the most well known classifier since it has been showed to be highly effective and accurate for text categorisation. An advantage of this classifier is that with small amount of training data, it works well [54]. In this paper, the linear *SVM* is used. The output of a linear *SVM* can be represented as:

$$u = \vec{w}^{\,t} \cdot \vec{x}^{\,t} - b \tag{1}$$

where $\vec{w}^{\,t}$ is the normal vector to the hyperplane, and $\vec{x}^{\,t}$ is the input vector.
- *Logistic Regression (LR)*.[7] [55] The Logistic Regression model finds a linear combination of the features. This classifier is useful for the emotion classification because only some features like adjectives or adverbs usually describe an emotion [56]. It computes the probability of an event occurrence utilising a logit function which equation is:

$$f(x) = \frac{L}{1 + e^{-k(x-x_0)}} \tag{2}$$

where L is the curve's maximum value, e refers to the natural logarithm base, x_0 is the x-value of the sigmoid's midpoint and k is the logistic growth rate of the curve.
- *Multilayer Perceptron (MLP)*[8] [57] is a feed forward neural network that exploit potential non-linear relationships in the sentiment features [58]. Some advantages of this classifier are that it acts as a universal function approximator and it can learn each and every relationship among input and output variables.
- *Naive Bayes (NB)* is a probabilistic classifier method based on Bayes' theorem [59]. Naive Bayes has been successfully applied to document classification in many studies [60]. In this project, multinomial Naive Bayes (multinomialNB) classification model is used. This model is suitable for classification with discrete features like word frequency information in document, where a document is a sequence of words obtained from vocabulary 'V'. The probability of a document given its class, can be obtained using the multinomial distribution show in Eq. (3):

$$P(d_i|c_j;\theta) = P(|d_i|)|d_i|! \prod_{t=1}^{|V|} \frac{P(w_t|c_j;\theta)^{N_{it}}}{N_{it}!} \tag{3}$$

where $P(d_i|c_j;\theta)$ is the probability of document 'd' for each class 'c'. $P(|d_i|)$ is the probability of document 'd' and $P(w_t|c_j;\theta)$ is the probability of occurrence of a word 'w' in a class 'c'.

5.4. Approach integrating affective knowledge

Feature selection plays an important role in emotion classification. We argue that the use of emotional external knowledge can enhance the classification of emotions and study the effectiveness of different affective lexical features in the algorithms mentioned in Section 5.3. These features are provided by the lexicons explains in Section 4. Fig. 2 illustrates the general pipeline of the approach. Given a sentence, first we preprocess it using Standford core NLP. Then, we compute the TF scheme in order to obtain how frequently an expression (term, word) occurs in a document. To incorporate the affective lexical features we check the presence of lexicon terms in the sentence and we obtain a vector that represent each emotional category (anger, fear, sadness and joy). Finally, to carry out the classification, the concatenation of the TF sentence representation and the word-based features are used as input to the different algorithms (*SVM*, *LR*, *MLP*, *MultinomialNB*).

Then, we discuss how the features can be derived from affective lexical features.

1. **SEL**. We verify the presence of lexicon terms in the sentence and then we compute the sum of the intensity value of the sentence words grouping them by the emotional

5 https://www.nltk.org/api/nltk.tokenize.html

6 https://scikit-learn.org/stable/modules/generated/sklearn.svm.LinearSVC.html#sklearn.svm.LinearSVC

7 https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html

8 https://scikit-learn.org/stable/modules/neural_networks_supervised.html

Please cite this article as: F.M. Plaza-del-Arco, M.T. Martín-Valdivia, L.A. Ureña-López et al., Improved emotion recognition in Spanish social media through incorporation of lexical knowledge, Future Generation Computer Systems (2019), https://doi.org/10.1016/j.future.2019.09.034.

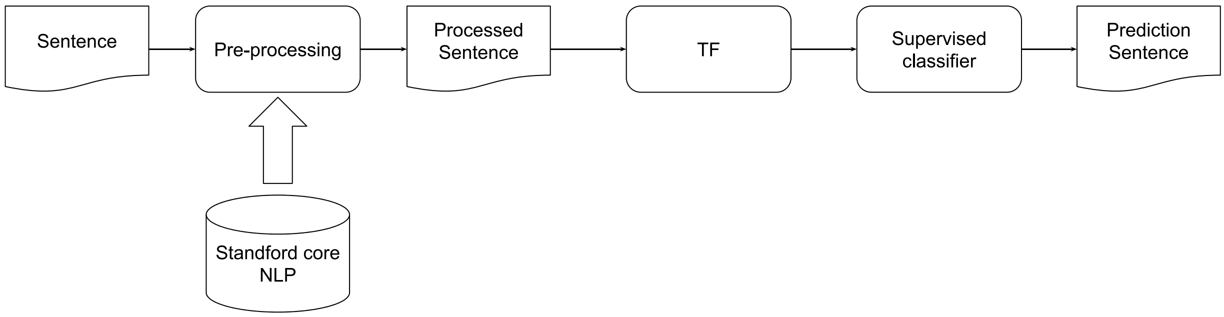

Fig. 1. General scheme of the baseline system.

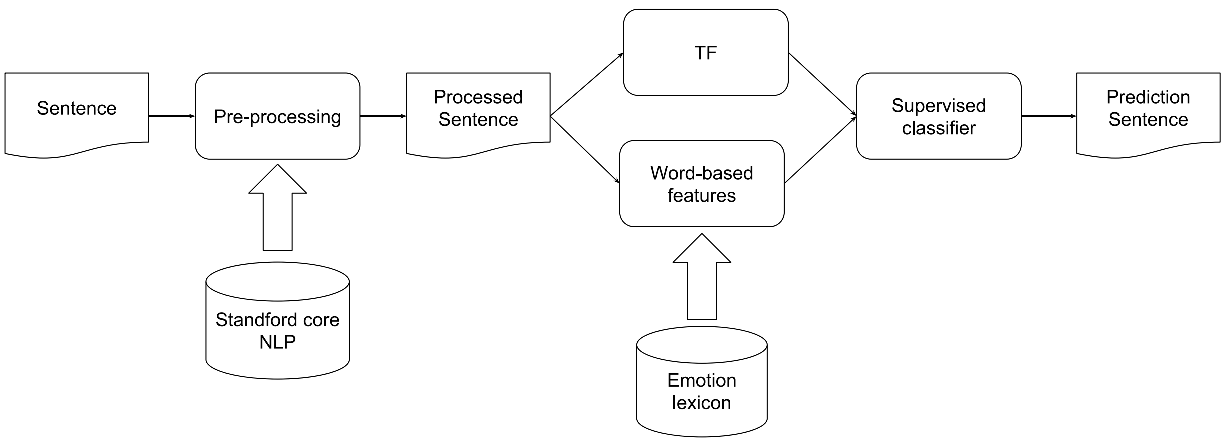

Fig. 2. Baseline system including affective lexical features.

category (*anger*, *sadness*, *fear* and *joy*). On the basis of this, we generating a vector of four values (four emotions) for each sentence.

2. **iSOL**. We check the presence of lexicon terms in the sentence and we assign 1 as Confidence Value (CV). Then, we sum the CVs of the words whose sentiment is the same obtaining a vector of polarity for each sentence. As a result, we have a vector of two values (positive and negative).
3. **EmoLex**. As we handle emotions in Spanish, we have used the Spanish version of this resource. To perform feature extraction, we identified the presence of lexicon terms in the sentence and assign 1 as CV if the word is present, 0 otherwise. Summing the CVs of the words whose emotion is the same and generate a vector of emotions for each sentence. As a result, we obtain a vector of eight values (eight emotions).

6. Results

In this section, we report and discuss the performance of our systems for the Spanish emotion recognition task in the AIT Dataset. In order to evaluate and compare our systems, we employ the usual metrics in text classification, called Precision (P), Recall (R), F-score (F_1) and Accuracy (Acc).

6.1. Baseline ML results

A summary of results for the baseline ML systems on the dataset described in Section 4 is presented in Table 5. *SVM* performs best on F_1 score for each emotion class as well as on Macro F_1 (75%), (see Fig. 3). On the other hand, *MLP* recorded the lowest results with Macro F_1 of 63%.

It should be noted that in most classifiers we obtained the best score on *joy* emotion. This may be because it is the only positive emotion belonging to the set of emotions and it is therefore easier to distinguish from the other negative emotions (*anger*, *fear*, *sadness*). This fact is not only a challenge for automatic systems but also for humans because in many cases a negative polarity sentence can express complementary negative emotions. For instance, when somebody says: "I spent the whole year studying and in the end I didn't pass the most important exam. I feel like I've wasted my time, I don't know what else to do", he is conveying two types of emotions: *anger*, because he has not achieved the results he expected and *sadness* since he does not know what else to do. In case of our classification, and as illustrated in Fig. 3, the most complementary and difficult to analyse emotions are *fear* and *sadness* which is reflected in their lowest F1 score.

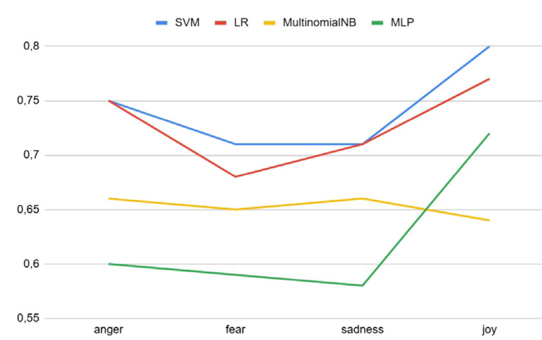

Fig. 3. F1-score of each classifier by emotion.

6.2. Results with systems integrating affective knowledge

In this subsection, we describe the results obtained by the different classifiers including affective features. We used SEL, iSOL

Table 5
Baseline results with different classifiers.

	Anger			Fear			Sadness			Joy			Macro-Avg			
	P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1	Acc
SVM	0.75	0.75	0.75	0.73	0.69	0.71	0.75	0.67	0.71	0.74	0.86	0.80	0.74	0.74	0.74	0.74
LR	0.75	0.74	0.75	0.74	0.63	0.68	0.77	0.66	0.71	0.69	0.87	0.77	0.74	0.72	0.72	0.73
MultinomialNB	0.56	0.78	0.66	0.62	0.68	0.65	0.65	0.68	0.66	0.91	0.50	0.64	0.69	0.66	0.65	0.65
MLP	0.60	0.61	0.60	0.57	0.61	0.59	0.56	0.59	0.58	0.77	0.68	0.72	0.62	0.62	0.63	0.63

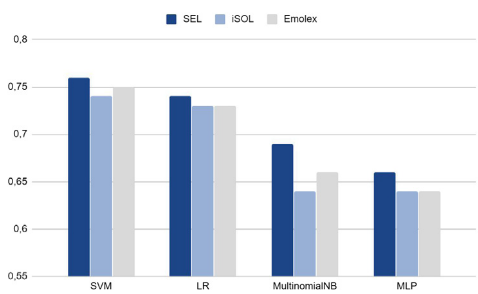

Fig. 4. Accuracy of each classifier including affective features.

and EmoLex lexicons to extract affective features from tweets and then we integrated them into the baseline ML systems discussed in Section 5.3. The first resource is an emotion lexicon created specifically for Spanish, the second one is a polarity lexicon built specifically for Spanish as well, and the last one is an emotion lexicon translated into Spanish but generated particularly for English.

Tables 6, 7, 8, 9 show the performance of the different systems described in Section 5.4. As it can be seen, most systems improve baseline results when they incorporate affective features from lexicons. It is important to note that the baseline systems that obtain lowest scores in the classification (*Multinomial* NB and *MLP*) are significantly improved when the lexical affective features are added.

Table 10 summarises the improvement in terms of percentage of the different classifiers when using a lexicon. It can be observed that SEL is the lexicon which offers the best results in classification because when it is included in each classifier, the improvement of accuracy score is more noticeable compared to the other lexicons (see Fig. 4). In the case of iSOL lexicon, we only observe a slight improvement for MLP classifier. We conjecture that the main reason for this is the fact that iSOL is a polarity lexicon and thus, no emotions are recognised but only positivity or negativity. Regarding the EmoLex lexicon, improvements are observed in the two base classifiers that obtained worse results (*MultinomialNB*, *MLP*). In summary, features extracted from SEL are significantly outperformed by those extracted using the other lexicons. The accuracy performance improvements of all the features extracted using SEL over those using iSOL and Emolex is nearly 2.72%, 1.37%, 6.15% and 4.76% on SVM, LR, MultinomialNB and MLP classifier. We believe SEL is the best lexicon because it is the only has been particularly generated for Spanish.

It is worth noting that when lexical affective features are integrated into classifiers, the incorporation of affective lexical features leads to improvement over most baseline figures with significant improvement noted over the lowest scores. For this reason, we conclude that the compilation of resources for a specific language is a prerequisite to improve the emotion recognition systems.

As mentioned in Section 1, there are very few works that deal with emotion classification in Spanish. The most similar work we have found in this language in order to compare our results is the one organised by SemEval-2018 Task 1: Affect in Tweets [33]. Specifically, subtask 5 (E-c) was about classifying emotions in English or Spanish. Given a tweet, it consisted of classify it as *neutral* or *no emotion* or as one, or more, of eleven given emotions that best represent the mental state of the tweeter. The best result in Spanish in this task was obtained by the MILAB_SNU team with a macro-average F_1 of 40.7 and a micro-average F_1 of 55.8. Most of the top-performing teams relied on deep neural network representations of tweets (sentence embeddings) as well as features derived from existing sentiment and emotion lexicons. As it can be observe, the results obtained are very low for the task. However, in case of English the best team obtained a macro-average F_1 of 52.8 a micro-average F_1 of 70.1. If we compare the results obtained in English and Spanish, we can assume that emotion detection in English has been more feasible than in Spanish.

Another recent task related to emotion classification was the one organised in SemEval-2019 Task 3: EmoContext Contextual Emotion Detection in Text [61]. Given a textual dialogue in English along with two turns of context, its consists of classify the emotion of user utterance as one of the emotion classes: *Happy*, *Sad*, *Angry* or *Others*. The highest ranked submission achieved 79.6 F_1 score and the F_1 average based on the results of all participants was 65.9. If we compare our results with the results obtained in this task in English, we can see that they follow the line of the state-of-the-art related to the emotion classification task which has involved an effort because Spanish is a language with more semantic and morphological richness and complexity than English. For this reason, it is important to study its syntactic and semantics in order to develop accurate systems in this language and not to use systems adapted from other languages.

7. Conclusion

This paper discusses the problem of emotion recognition in Spanish text. We develop and evaluate two different ML approaches and study the impact of integrating knowledge from different emotion lexicons seeking to establish whether it improves emotion classification.

Our experiments show that the use of resources for a specific language improve the emotion recognition systems. In particular, the best performance results are obtained when we use the SEL lexicon. The results also show that the use of other lexicons adapted in Spanish with a simple translation does not always improve the classification. The mapping emotion concept to lexical word certainly depends on cultural differences which may be inferred from the pragmatic use of words various contexts and situations and not from out of context entries in dictionaries. On the basis of these observations, we can safely say that there is a pressing need to compile emotional lexicons languages other than English. The availability of a suitable emotion lexicon and incorporating its lexical knowledge into an emotion recognition system would be likely to improve performance. Therefore, research should be focused not only on English resources but also on resources for other languages.

As future work, we will continue working on emotion classification in Spanish. Emotion mining from text is still in its infancy and yet has a long way to proceed. Indeed, most of the research

Please cite this article as: F.M. Plaza-del-Arco, M.T. Martín-Valdivia, L.A. Ureña-López et al., Improved emotion recognition in Spanish social media through incorporation of lexical knowledge, Future Generation Computer Systems (2019), https://doi.org/10.1016/j.future.2019.09.034.

Table 6
Results with *SVM* classifier and affective lexical features.

	Anger			Fear			Sadness			Joy			Macro-Avg			
	P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1	Acc
SEL	0.78	0.76	0.77	0.76	0.68	0.72	0.76	0.68	0.72	0.73	0.88	0.80	0.76	0.75	0.75	0.76
iSOL	0.76	0.74	0.75	0.73	0.68	0.70	0.76	0.66	0.71	0.74	0.87	0.80	0.75	0.74	0.74	0.74
Emolex	0.76	0.74	0.75	0.75	0.68	0.72	0.76	0.67	0.72	0.72	0.87	0.79	0.75	0.74	0.74	0.75

Table 7
Results with *LR* classifier and affective lexical features.

	Anger			Fear			Sadness			Joy			Macro-Avg			
	P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1	Acc
SEL	0.79	0.74	0.77	0.78	0.64	0.70	0.78	0.66	0.71	0.68	0.91	0.78	0.76	0.74	0.74	0.74
iSOL	0.75	0.73	0.74	0.75	0.64	0.69	0.76	0.65	0.70	0.69	0.88	0.77	0.74	0.73	0.73	0.73
Emolex	0.74	0.74	0.74	0.75	0.63	0.69	0.78	0.64	0.70	0.68	0.88	0.77	0.74	0.72	0.72	0.73

Table 8
Results with *MultinomialNB* classifier and affective lexical features.

	Anger			Fear			Sadness			Joy			Macro-Avg			
	P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1	Acc
SEL	0.62	0.77	0.69	0.67	0.68	0.67	0.68	0.68	0.68	0.84	0.63	0.72	0.70	0.69	0.69	0.69
iSOL	0.59	0.76	0.66	0.59	0.64	0.62	0.67	0.60	0.63	0.75	0.58	0.65	0.65	0.64	0.64	0.64
Emolex	0.62	0.66	0.64	0.59	0.69	0.64	0.67	0.62	0.64	0.76	0.65	0.70	0.66	0.66	0.66	0.66

Table 9
Results with *MLP* classifier and affective lexical features.

	Anger			Fear			Sadness			Joy			Macro-Avg			
	P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1	Acc
SEL	0.64	0.66	0.65	0.62	0.62	0.62	0.59	0.64	0.62	0.78	0.72	0.75	0.66	0.66	0.66	0.66
iSOL	0.61	0.63	0.62	0.58	0.62	0.60	0.58	0.59	0.59	0.77	0.70	0.73	0.64	0.63	0.63	0.64
Emolex	0.60	0.64	0.62	0.58	0.68	0.62	0.62	0.55	0.58	0.78	0.69	0.74	0.64	0.64	0.64	0.64

Table 10
% accuracy score improvement per classifier using affective features.

	SEL	iSOL	Emolex
SVM	**2.72**	−1.33	1.35
LR	**1.37**	0	0
MultinomialNB	**6.15**	−1.53	1.53
MLP	**4.76**	1.58	1.58

has been conducted in English but Spanish resources and works are scarce despite being one of the main spoken languages in the world. Furthermore, we will study how to add more relevant linguistic information, as the influence of emojis or negation in the classification process.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work has been partially supported by Fondo Europeo de Desarrollo Regional (FEDER), LIVING-LANG project (RTI2018-094653-B-C21) and REDES project (TIN2015-65136-C2-1-R) from the Spanish Government.

References

[1] C. Strapparava, Emotions and NLP: Future directions, in: Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, 2016, p. 180.

[2] R.A. Calvo, S. D'Mello, Affect detection: An interdisciplinary review of models, methods, and their applications, IEEE Trans. Affect. Comput. 1 (1) (2010) 18–37.

[3] M. Hasan, E. Rundensteiner, E. Agu, Automatic emotion detection in text streams by analyzing Twitter data, Int. J. Data Sci. Anal. 7 (1) (2019) 35–51.

[4] S.N. Shivhare, S. Khethawat, Emotion detection from text, arXiv preprint arXiv:1205.4944, 2012.

[5] M.T. Khan, M. Durrani, A. Ali, I. Inayat, S. Khalid, K.H. Khan, Sentiment analysis and the complex natural language, Complex Adapt. Syst. Model. 4 (1) (2016) 2.

[6] T. Gadekallu, A. Soni, D. Sarkar, L. Kuruva, Application of sentiment analysis in movie reviews, in: Sentiment Analysis and Knowledge Discovery in Contemporary Business, IGI Global, 2019, pp. 77–90.

[7] R.W. Picard, Affective computing. 1997, Massachusetts Institute of Technology, 1997.

[8] S. Mitra, T. Acharya, Gesture recognition: A survey, IEEE Trans. Syst. Man Cybern. C 37 (3) (2007) 311–324.

[9] A.B. Ingale, D. Chaudhari, Speech emotion recognition, Int. J. Soft Comput. Eng. 2 (1) (2012) 235–238.

[10] G. Trigeorgis, F. Ringeval, R. Brueckner, E. Marchi, M.A. Nicolaou, B. Schuller, S. Zafeiriou, Adieu features? end-to-end speech emotion recognition using a deep convolutional recurrent network, in: Acoustics, Speech and Signal Processing, ICASSP, 2016 IEEE International Conference on, IEEE, 2016, pp. 5200–5204.

[11] J. Reilly, J. Ghent, J. McDonald, Modelling, classification and synthesis of facial expressions, in: Affective Computing, InTech, 2008.

[12] H. Soyel, H. Demirel, Facial expression recognition using 3D facial feature distances, in: International Conference Image Analysis and Recognition, Springer, 2007, pp. 831–838.

[13] D. Dagar, A. Hudait, H. Tripathy, M. Das, Automatic emotion detection model from facial expression, in: Advanced Communication Control and Computing Technologies, ICACCCT, 2016 International Conference on, IEEE, 2016, pp. 77–85.
[14] A. Yadollahi, A.G. Shahraki, O.R. Zaiane, Current state of text sentiment analysis from opinion to emotion mining, ACM Comput. Surv. 50 (2) (2017) 25.
[15] M. De Choudhury, M. Gamon, S. Counts, E. Horvitz, Predicting depression via social media, ICWSM 13 (2013) 1–10.
[16] E. Cambria, Affective computing and sentiment analysis, IEEE Intell. Syst. 31 (2) (2016) 102–107.
[17] E. Diaz-Aviles, C. Orellana-Rodriguez, W. Nejdl, Taking the pulse of political emotions in latin america based on social web streams, in: Web Congress, LA-WEB, 2012 Eighth Latin American, IEEE, 2012, pp. 40–47.
[18] M. Anjaria, R.M.R. Guddeti, Influence factor based opinion mining of Twitter data using supervised learning, in: Communication Systems and Networks, COMSNETS, 2014 Sixth International Conference on, IEEE, 2014, pp. 1–8.
[19] K. Sailunaz, M. Dhaliwal, J. Rokne, R. Alhajj, Emotion detection from text and speech: a survey, Soc. Netw. Anal. Min. 8 (1) (2018) 28.
[20] F.M. Plaza-del-Arco, E. Martínez-Cámara, M.T.M. Valdivia, L.A.U. López, SINAI en TASS 2018: Inserción de Conocimiento Emocional Externo a un Clasificador Lineal de Emociones (SINAI at TASS 2018: Lineal Classification System with Emotional External Knowledge), in: Proceedings of TASS 2018: Workshop on Semantic Analysis At SEPLN, TASS@SEPLN 2018, Co-Located with 34nd SEPLN Conference, SEPLN 2018, Sevilla, Spain, September 18th, 2018, 2018, pp. 125–130, http://ceur-ws.org/Vol-2172/p15_sinai_tass2018.pdf.
[21] C. Suero Montero, J. Suhonen, Emotion analysis meets learning analytics: online learner profiling beyond numerical data, in: Proceedings of the 14th Koli Calling International Conference on Computing Education Research, ACM, 2014, pp. 165–169.
[22] A. Chatterjee, U. Gupta, M.K. Chinnakotla, R. Srikanth, M. Galley, P. Agrawal, Understanding emotions in text using deep learning and big data, Comput. Hum. Behav. 93 (2019) 309–317.
[23] Y. Kim, Convolutional neural networks for sentence classification, arXiv preprint arXiv:1408.5882, 2014.
[24] A. Prakash, C. Brockett, P. Agrawal, Emulating human conversations using convolutional neural network-based IR, arXiv preprint arXiv:1606.07056, 2016.
[25] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.
[26] M. Schuster, K.K. Paliwal, Bidirectional recurrent neural networks, IEEE Trans. Signal Process. 45 (11) (1997) 2673–2681.
[27] S.M. Zahiri, J.D. Choi, Emotion detection on TV show transcripts with sequence-based convolutional neural networks, in: Workshops At the Thirty-Second AAAI Conference on Artificial Intelligence, 2018.
[28] M. Jabreel, A. Moreno, A deep learning-based approach for multi-label emotion classification in tweets, Appl. Sci. 9 (6) (2019) 1123.
[29] C. Strapparava, R. Mihalcea, Learning to identify emotions in text, in: Proceedings of the 2008 ACM Symposium on Applied Computing, ACM, 2008, pp. 1556–1560.
[30] A. Neviarouskaya, H. Prendinger, M. Ishizuka, Textual affect sensing for sociable and expressive online communication, in: International Conference on Affective Computing and Intelligent Interaction, Springer, 2007, pp. 218–229.
[31] Y. Rao, J. Lei, L. Wenyin, Q. Li, M. Chen, Building emotional dictionary for sentiment analysis of online news, World Wide Web 17 (4) (2014) 723–742.
[32] A. Bandhakavi, N. Wiratunga, D. Padmanabhan, S. Massie, Lexicon based feature extraction for emotion text classification, Pattern Recognit. Lett. 93 (2017) 133–142.
[33] S.M. Mohammad, F. Bravo-Marquez, M. Salameh, S. Kiritchenko, SemEval-2018 task 1: Affect in tweets, in: Proceedings of International Workshop on Semantic Evaluation, SemEval-2018, New Orleans, LA, USA, 2018.
[34] S.M. Mohammad, F. Bravo-Marquez, WASSA-2017 shared task on emotion intensity, arXiv preprint arXiv:1708.03700, 2017.
[35] C. Strapparava, A. Valitutti, et al., Wordnet affect: an affective extension of wordnet, in: Lrec, vol. 4, Citeseer, 2004, pp. 1083–1086.
[36] M.M. Bradley, P.J. Lang, Affective Norms for English Words, ANEW: Instruction Manual and Affective Ratings, Tech. Rep., Citeseer, 1999.
[37] J.W. Pennebaker, R.J. Booth, M.E. Francis, LIWC2007: Linguistic Inquiry and Word Count, liwc. net, Austin, Texas, 2007.
[38] S.M. Mohammad, P.D. Turney, Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon, in: Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, Association for Computational Linguistics, 2010, pp. 26–34.
[39] M.D. Molina-González, E. Martínez-Cámara, M.T. Martín-Valdivia, L.A. Ureña-López, A spanish semantic orientation approach to domain adaptation for polarity classification, Inf. Process. Manage. 51 (4) (2015) 520–531.
[40] E. Martínez-Cámara, M.T. Martín-Valdivia, L.A. Ureña-López, R. Mitkov, Polarity classification for spanish tweets using the COST corpus, J. Inf. Sci. 41 (3) (2015) 263–272.
[41] G. Sidorov, S. Miranda-Jiménez, F. Viveros-Jiménez, A. Gelbukh, N. Castro-Sánchez, F. Velásquez, I. Díaz-Rangel, S. Suárez-Guerra, A. Treviño, J. Gordon, Empirical study of machine learning based approach for opinion mining in tweets, in: Mexican International Conference on Artificial Intelligence, Springer, 2012, pp. 1–14.
[42] J. Redondo, I. Fraga, I. Padrón, M. Comesaña, The spanish adaptation of ANEW (affective norms for English words), Behav. Res. Methods 39 (3) (2007) 600–605.
[43] F.M. Plaza-del-Arco, M.D. Molina-González, S.M. Jiménez-Zafra, M.T. Martín-Valdivia, Lexicon adaptation for spanish emotion mining, Proces. Leng. Nat. 61 (2018) 117–124.
[44] S.M. Mohammad, Word affect intensities, in: Proceedings of the 11th Edition of the Language Resources and Evaluation Conference, LREC-2018, Miyazaki, Japan, 2018.
[45] F.M. Plaza-del-Arco, S.M. Jiménez-Zafra, M. Martin, L.A.U. Lopez, SINAI at SemEval-2018 Task 1: Emotion recognition in tweets, in: Proceedings of the 12th International Workshop on Semantic Evaluation, 2018, pp. 128–132.
[46] S. Mohammad, F. Bravo-Marquez, M. Salameh, S. Kiritchenko, Semeval-2018 task 1: Affect in tweets, in: Proceedings of the 12th International Workshop on Semantic Evaluation, 2018, pp. 1–17.
[47] S. Frenda, G. Bilal, et al., Exploration of misogyny in spanish and english tweets, in: Third Workshop on Evaluation of Human Language Technologies for Iberian Languages, IberEval 2018, vol. 2150, Ceur Workshop Proceedings, 2018, pp. 260–267.
[48] M.D. Molina-González, E. Martínez-Cámara, M.-T. Martín-Valdivia, J.M. Perea-Ortega, Semantic orientation for polarity classification in spanish reviews, Expert Syst. Appl. 40 (18) (2013) 7250–7257.
[49] M. Hu, B. Liu, Mining and summarizing customer reviews, in: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2004, pp. 168–177.
[50] M.-T. Martín-Valdivia, E. Martínez-Cámara, J.-M. Perea-Ortega, L.A. Ureña-López, Sentiment polarity detection in spanish reviews combining supervised and unsupervised approaches, Expert Syst. Appl. 40 (2013) 3934–3942.
[51] F.M. Plaza-del Arco, M.T. Martín-Valdivia, S.M. Jiménez, M.D. Molina-González, E. Martínez-Cámara, COPOS: corpus of patient opinions in spanish. application of sentiment analysis techniques, Proces. Leng. Nat. (57) (2016) 83–90.
[52] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al., Scikit-learn: Machine learning in Python, J. Mach. Learn. Res. 12 (Oct) (2011) 2825–2830.
[53] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (3) (1995) 273–297.
[54] A.M. Ruiz, A. Cornet, K. Shimanoe, J.R. Morante, N. Yamazoe, Effects of various metal additives on the gas sensing performances of TiO2 nanocrystals obtained from hydrothermal treatments, Sensors Actuators B 108 (1–2) (2005) 34–40.
[55] S. Shalev-Shwartz, S. Ben-David, Understanding Machine Learning: From Theory to Algorithms, Cambridge university press, 2014.
[56] A. Gupta, H. Sahu, N. Nanecha, P. Kumar, P.P. Roy, V. Chang, Enhancing text using emotion detected from EEG signals, J. Grid Comput. (2018) 1–16.
[57] S. Haykin, Neural Networks, A comprehensive Foundation, second ed., Prentice-Hall Inc, 1999.
[58] S.F. Crone, C. Koeppel, Predicting exchange rates with sentiment indicators: An empirical evaluation using text mining and multilayer perceptrons, in: Computational Intelligence for Financial Engineering & Economics, CIFEr, 2104 IEEE Conference on, IEEE, 2014, pp. 114–121.
[59] C. Catal, U. Sevim, B. Diri, Practical development of an eclipse-based software fault prediction tool using naive Bayes algorithm, Expert Syst. Appl. 38 (3) (2011) 2347–2353.
[60] A. McCallum, K. Nigam, et al., A comparison of event models for naive bayes text classification, in: AAAI-98 Workshop on Learning for Text Categorization, Vol. 752, (1) Citeseer, 1998, pp. 41–48.
[61] A. Chatterjee, K.N. Narahari, M. Joshi, P. Agrawal, SemEval-2019 task 3: EmoContext contextual emotion detection in text, in: Proceedings of the 13th International Workshop on Semantic Evaluation, 2019, pp. 39–48.

Please cite this article as: F.M. Plaza-del-Arco, M.T. Martín-Valdivia, L.A. Ureña-López et al., Improved emotion recognition in Spanish social media through incorporation of lexical knowledge, Future Generation Computer Systems (2019), https://doi.org/10.1016/j.future.2019.09.034.

Flor Miriam Plaza-del-Arco is currently a Ph.D. student in Computer Science at the University of Jaén (Spain). She received an M.S. degree in Computer Science at the University of Jaén (Spain) in 2018. Her current research interest is related to Sentiment Analysis, especially in Emotion Mining.

M. Teresa Martín-Valdivia is full professor in the Department of Computer Science at University of Jaén (Spain). Her current research interests include Natural Language Processing, Machine Learning algorithms, Sentiment Analysis and Text Mining techniques. In addition, she is the Treasurer of SEPLN Committee since 2007 and editor of the Procesamiento de Lenguaje Natural Journal. She is author or co-author of more than 100 scientific publications. She is a technical reviewer for several journals and a member of the Program Committee of several major conferences.

Prof. L. Alfonso Ureña is full professor in the Department of Computer Science at University of Jaén (Spain). He is author of over 200 publications on various topics of Natural Language Processing (NLP). Editor-In-Chief of the Journal of Procesamiento del Lenguaje Natural. Also, he is President of Spanish Society for Natural Language Processing (SEPLN). Programme Chair and keynote speaker of several major international conferences. Dr Ureña is Director of the Research Institute in Information and Communication Technologies (University of Jaén).

Prof Ruslan Mitkov's extensively cited research covers 250 publications on various topics of Natural Language Processing (NLP). He is author of the monograph *Anaphora resolution* (Longman), sole Editor of *The Oxford Handbook of Computational Linguistics* (Oxford University Press), Executive Editor of the Journal of Natural Language Engineering (Cambridge University Press), Editor-in-Chief of the NLP book series of John Benjamins publishers, Programme Chair and keynote speaker of many international conferences. Dr Mitkov is Director of the Research Institute in Information and Language Processing (University of Wolverhampton). He was awarded the title of Doctor Honoris Causa in recognition of his outstanding research/professional achievements.

Please cite this article as: F.M. Plaza-del-Arco, M.T. Martín-Valdivia, L.A. Ureña-López et al., Improved emotion recognition in Spanish social media through incorporation of lexical knowledge, Future Generation Computer Systems (2019), https://doi.org/10.1016/j.future.2019.09.034.