SINAI at SemEval-2018 Task 1: Emotion Recognition in Tweets

Flor Miriam Plaza-del-Arco, Salud María Jiménez-Zafra,
M. Teresa Martín-Valdivia, L. Alfonso Ureña-López

Department of Computer Science, Advanced Studies Center in ICT (CEATIC)
Universidad de Jaén, Campus Las Lagunillas, 23071, Jaén, Spain
{fmplaza, sjzafra, maite, laurena}@ujaen.es

Abstract

Emotion classification is a new task that combines several disciplines including Artificial Intelligence and Psychology, although Natural Language Processing is perhaps the most challenging area. In this paper, we describe our participation in SemEval-2018 Task1: Affect in Tweets. In particular, we have participated in EI-oc, EI-reg and E-c subtasks for English and Spanish languages.

1. Introduction

Emotions are playing a significant role in the effective communication of people. In fact, sometimes, emotional intelligence is more important than cognitive intelligence for successful interaction (Pantic et al., 2005). Therefore, affective computing is a key element to the advancement of Artificial Intelligence. The basic task of affective computing is emotion recognition. This task consists of identifying a set of emotions in a document.

The identification of emotions in texts has multiple benefits in different areas, such as psychology to detect some psychological disorder like depression (Cherry et al., 2012), e-learning to improve student motivation (Suero Montero and Suhonen, 2014) or business intelligence to know the preferences of consumers (Cambria, 2016).

Currently, more and more people express their emotions on social media, such as Twitter or Facebook. Therefore, the role of emotion in social media is becoming more important for the researchers in affective computing.

In this paper, we present the different systems we developed as part of our participation in SemEval-2018 Task 1: Affect in Tweets (Mohammad et al., 2018). We have participated in EI-oc, EI-reg and E-c subtasks for English and Spanish. Below, we briefly describe these subtasks:

- **EI-oc** is an emotion intensity ordinal classification task. Given a tweet and an emotion E, it consists of classifying the tweet into one of four ordinal classes of intensity of E that best represents the mental state of the tweeter. Separate datasets are provided for anger, fear, joy, and sadness emotions.

- **EI-reg** is an emotion intensity regression task. Given a tweet and an emotion E, it consists of determining the intensity of E that best represents the mental state of the tweeter. The intensity of E is a real-valued score between 0 (least emotion) and 1 (most emotion). Separate datasets are provided for anger, fear, joy, and sadness emotions.

- **E-c** is an emotion multi-classification task. Given a tweet, it consists of classifying it as 'neutral' or 'no emotion' or as one, or more, of eleven given emotions (anger, anticipation, disgust, fear, joy, love, optimism, pessimism, sadness, surprise, trust) that best represent the mental state of the tweeter.

The rest of the paper is organized as follows. In Section 2 we explain the data used in our methods. Section 3 describes the resources used by our systems. Section 4 presents the details of the proposed systems. Section 5 displays the results and analyses them. We conclude in Section 6 with remarks on future work.

2. Data

To run our experiments, we used the datasets provided by the task organizers (Mohammad et al., 2018) as follows. During pre-evaluation period,

we trained our models on the *train* set, and evaluated our different approaches on the *dev* set. During evaluation period, we trained our models on the *train* and *dev* sets, and tested the model on the *test* set. Table 1 shows the number of tweets for each language and subtask dataset.

Subtask-language	train	dev	test	Total
EI-oc-es	4544	793	2616	7953
EI-oc-en	7102	1465	4070	12637
EI-reg-es	4544	793	2616	7953
EI-reg-en	7102	1464	71816	80382
E-c-es	7561	679	2854	11064
E-c-en	6838	886	3259	10983

Table 1: Number of tweets for each language and subtask dataset

3. Resources

For the development of the task, we used different lexicons that we explain in detail below.

- **Wordnet-Affect (WNA)** (Strapparava et al., 2004). This resource is an extension of Word-Net Domains. WNA provides a set of English emotional words organized in a tree. The leaf nodes represent specific emotions that are grouped into general categories (parent nodes). For example, *anger, hate* and *dislike* belong to the overall emotion *general-dislike*. However, the emotions of WNA are not the same as the emotions of the SemEval subtasks. For this reason, each overall emotion of WNA has been mapped with SemEval subtasks emotions (see Appendix A, Table 8 and Table 9).

  In order to use this resource in Spanish, we have employed the lexical disambiguator Babelfy (Moro et al., 2014) to obtain the corresponding *BalbelNet synset id* of a term. Next, we have used the BabelNet API (Navigli and Ponzetto, 2012) to obtain a correspondence between the *BalbelNet synset id* and the *WordNet synset id*. WNA includes a subset of appropriate synsets of WordNet 1.6 to represent affective concepts. However, the *WordNet synsets id* obtained with BabelNet API corresponds to the 3.0 version of WordNet. Therefore, we have obtained the equivalent synset to the 3.0 version in the 1.6 version. With this, using the synset of the 1.6 version

of WordNet, we can map directly the associated emotion and confident value from WNA.

- **Spanish Emotion Lexicon (SEL)** (Sidorov et al., 2012). It includes 2,036 Spanish words that are associated with the measure of Probability Factor of Affective use (PFA) with respect to at least one basic emotion: joy, anger, fear, sadness, surprise, and disgust. The higher the value of the PFA, the more probable the association of the word with the emotion is.

- **NRC Affect Intensity Lexicon** (Mohammad, 2017). It has almost 6,000 entries in English. Each of them has an intensity score associated to one of the following basic emotions: anger, fear, sadness and joy. The scores range from 0 to 1, where 1 indicates that the word has a high association to the emotion and 0 that the word has a low association to the emotion. However, this resource is not in Spanish. For this reason, we have adapted it to Spanish in the following way. We have translated English terms to Spanish and we have selected the maximum value of intensity if the translation of some terms is the same.

- **NRC Word-Emotion Association Lexicon (EmoLex)** (Mohammad and Turney, 2010). This lexicon has a list of English words associated to one or more of the following emotions: anger, fear, anticipation, trust, surprise, sadness, joy. Moreover, the lexicon is also available for more than one hundred languages (including Spanish). All these versions have been generated by translating the English terms using Google Translate.

4. System description

In this section we describe the systems developed for the subtasks EI-oc, EI-reg and E-c.

In first place, we preprocessed the corpus of tweets provided for each subtask and language (English and Spanish). We applied the following preprocessing steps: the documents were tokenized using NLTK TweetTokenizer[1], stemming was performed using NLTK Snowball stemmer[2], stopwords

---

[1]http://www.nltk.org/api/nltk.tokenize.html
[2]http://www.nltk.org/_modules/nltk/

were removed (only for English), and all letters were converted to lower-case.

In relation to the resources, we have tested several combinations. However, for the final SemEval systems we have used the best systems obtained during the development phase. For EI-oc and EI-reg subtasks in Spanish, we used SEL, NRC Affect Intensity and WNA lexicons adapted to the emotions of these subtasks. On the other hand, for English, we used NRC Affect Intensity and WNA lexicons adapted to the emotions of the EI-oc and EI-reg subtasks. Regarding to subtask E-c, for Spanish, we used SEL, EmoLex Spanish version and WNA lexicons adapted to the emotions of this subtask. However, for English, we used Emolex and WNA lexicon adapted to the emotions of the E-c subtask.

Next, it is described the methodology used for each subtask:

- **Subtask EI-oc**. To perform the classification, we checked the presence of lexicon terms in the tweet and then we added the intensity value of these words grouping them by the emotional category (*anger*, *fear*, *sadness* and *joy*). The result is a vector of four values for each lexicon. Moreover, each tweet is represented as a vector of unigrams using the TF-IDF weighting scheme. The union of the lexicon vectors and the TF-IDF representation of the tweet are used as features for the classification using the SVM algorithm. We selected the SVM formulation, known as C-SVC, the value of the C parameter was 1.0 and the kernel chosen was the linear.

- **Subtask EI-reg**. In this case, we checked the presence of lexicon terms in the tweet and then we computed the sum, the average and the maximum of the intensity value of the words of the tweet grouping them by the emotional category (*anger*, *fear*, *sadness* and *joy*). The result is a vector of twelve values for each lexicon. The union of the lexicon vectors and the TF-IDF representation of the tweet are used as features for the classification using the SVM algorithm with the same configuration as that used in subtask EI-oc.

- **Subtask E-c**. In this subtask, we identified the presence of lexicon terms in the tweet and we assigned 1 as confidence value (CV).

stem/snowball.html

Then, we summed the CV of the words whose emotion is the same obtaining a vector of emotions for each lexicon. The union of these vectors and the TF-IDF representation of the tweet are used as features for the classification using the Random Forest algorithm with 25 as number of trees.

5. Analysis of results

The official competition metric to evaluate the systems in EI-reg and EI-oc subtasks is the Pearson Correlation Coefficient (PCC) between semantic similarity scores of machine assigned and human judgments. In the case of the E-c subtask, systems are evaluated by calculating multi-label accuracy. Since this is a multi-label classification task, each tweet can have one or more gold emotion labels, and one or more predicted emotion labels. Multi-label accuracy is defined as the size of the intersection of the predicted and gold label sets divided by the size of their union. This measure is calculated for each tweet, and then is averaged over all the tweets in the dataset.

The results of our participation in the three subtasks and those of the teams that are in the first and the last position can be seen in Tables 2, 3, 4, 5, 6 and 7. It should be noted that the results of Spanish subtasks are lower than those obtained for English. Another important issue is that the participation in Spanish subtasks is lower than the participation in English subtasks. These facts are due to most of the works and resources for textual emotion mining are in English (Yadollahi et al., 2017).

In relation to our results, in most subtasks we obtained the lowest correlation on *anger* emotion and the best correlation on *joy* emotion. On the contrary, in WASSA-2017 Shared Task on Emotion Intensity (Mohammad and Bravo-Marquez, 2017), most of the systems performed better on *anger* emotion and worse on *fear* and *sadness* emotions. In this competition, it was found that despite using deep learning techniques, training data, and large amounts of unlabeled data, the best systems included features from affect lexicons. Given that, we plan to analyze the recall of the lexicons used in our experiments and to explore new lexicons in order to improve the classification.

On the other hand, it should be noted that we achieved higher ranking positions for Spanish sub-

tasks. In particular, our best participation has been in the E-c subtask. An important difference found between the classification in both languages was that taking stopwords into consideration contributes to the emotion classification for Spanish while the opposite occurs for English. Therefore, we will further study this issue in order to incorporate an specific treatment to those stopwords that can modify the meaning of a sentence, such as negators, intensifiers and diminishers.

(r) Team name	Pearson				
	macro-avg	anger	fear	joy	sadness
(1) SeerNet	0.799	0.827	0.779	0.792	0.798
(39) SINAI	0.342	0.263	0.361	0.444	0.300
(48) TweetGroup	-0.016	-0.043	0.003	-0.011	-0.014

Table 2: Results of subtask EI-reg in English language

(r) Team name	Pearson				
	macro-avg	anger	fear	joy	sadness
(1) AffectThor	0.738	0.676	0.776	0.753	0.746
(10) SINAI	0.321	0.119	0.382	0.360	0.423
(16) AIT2018 Organizers	-0.012	-0.056	0.004	0.018	-0.014

Table 3: Results of subtask EI-reg in Spanish language

(r) Team name	Pearson				
	macro-avg	anger	fear	joy	sadness
(1) SeerNet	0.695	0.706	0.637	0.720	0.717
(24) SINAI	0.449	0.447	0.377	0.519	0.455
(39) TweetGroup	-0.021	0.015	-0.017	-0.029	-0.054

Table 4: Results of subtask EI-oc in English language

6. Conclusions

In this paper, we have presented the systems developed for our participation in 3 subtasks (EI-oc, EI-reg, E-c) of SemEval-2018 Task 1: Affect in Tweets. We have addressed these subtasks in two of the three available languages, English and Spanish. Overall, we have obtained better results in Spanish subtasks than in English subtasks. In future works, we plan to continue working on emotion recognition in Spanish because we have observed that the participation in this language is very low, although it is the second most spoken language. Our next study will focus on exploring more affect lexicons because in WASSA-2017 Shared Task on Emotion Intensity (Mohammad and Bravo-Marquez, 2017), it was demonstrated that using features from affect lexicons is beneficial for this task. Moreover, we will study the use of stopwords in Spanish because in the development phase it was observed that stopwords contribute to the emotion classification.

(r) Team name	Pearson				
	macro-avg	anger	fear	joy	sadness
(1) AffectThor	0.664	0.606	0.706	0.667	0.667
(7) SINAI	0.459	0.378	0.496	0.510	0.453
(16) AIT2018 Organizers	-0.022	0.011	-0.069	-0.005	-0.027

Table 5: Results of subtask EI-oc in Spanish language

(r) Team Name	accuracy
(1) NTU-SLP	0.588
(25) SINAI	0.394
(35) emotion17	0.023

Table 6: Results of subtask E-c in English Language

(r) Team Name	accuracy
(1) MILAB_SNU	0.469
(5) SINAI	0.318
(14) TeamCEN	0.050

Table 7: Results of subtask E-c in Spanish Language

Acknowledgements

References

Erik Cambria. 2016. Affective computing and sentiment analysis. *IEEE Intelligent Systems*, 31(2):102–107.

Colin Cherry, Saif M Mohammad, and Berry De Bruijn. 2012. Binary classifiers and latent sequence models for emotion detection in suicide notes. *Biomedical informatics insights*, 5(Suppl 1):147.

Saif M Mohammad. 2017. Word affect intensities. *arXiv preprint arXiv:1704.08798*.

Saif M Mohammad and Felipe Bravo-Marquez. 2017. Wassa-2017 shared task on emotion intensity. *arXiv preprint arXiv:1708.03700*.

Saif M. Mohammad, Felipe Bravo-Marquez, Mohammad Salameh, and Svetlana Kiritchenko. 2018. Semeval-2018 Task 1: Affect in tweets. In *Proceedings of International Workshop on Semantic Evaluation (SemEval-2018)*, New Orleans, LA, USA.

Saif M Mohammad and Peter D Turney. 2010. 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*, pages 26–34. Association for Computational Linguistics.

Andrea Moro, Alessandro Raganato, and Roberto Na-
vigli. 2014. Entity linking meets word sense disambi-
guation: a unified approach. *Transactions of the Asso-
ciation for Computational Linguistics*, 2:231–244.

Roberto Navigli and Simone Paolo Ponzetto. 2012.
Babelnet: The automatic construction, evaluation and
application of a wide-coverage multilingual semantic
network. *Artificial Intelligence*, 193:217–250.

Maja Pantic, Nicu Sebe, Jeffrey F Cohn, and Thomas
Huang. 2005. Affective multimodal human-computer
interaction. In *Proceedings of the 13th annual ACM in-
ternational conference on Multimedia*, pages 669–676.
ACM.

Grigori Sidorov, Sabino Miranda-Jiménez, Francis-
co Viveros-Jiménez, Alexander Gelbukh, Noé Castro-
Sánchez, Francisco Velásquez, Ismael Díaz-Rangel,
Sergio Suárez-Guerra, Alejandro Treviño, and Juan
Gordon. 2012. Empirical study of machine learning
based approach for opinion mining in tweets. In *Mexi-
can international conference on Artificial intelligence*,
pages 1–14. Springer.

Carlo Strapparava, Alessandro Valitutti, et al. 2004.
Wordnet affect: an affective extension of wordnet. In
Lrec, volume 4, pages 1083–1086. Citeseer.

Calkin Suero Montero and Jarkko Suhonen. 2014.
Emotion analysis meets learning analytics: online lear-
ner profiling beyond numerical data. In *Proceedings of
the 14th Koli calling international conference on com-
puting education research*, pages 165–169. ACM.

Ali Yadollahi, Ameneh Gholipour Shahraki, and Os-
mar R Zaiane. 2017. Current state of text sentiment
analysis from opinion to emotion mining. *ACM Com-
puting Surveys (CSUR)*, 50(2):25.

A. Mapping SemEval emotions

SemEval emotion	WNA emotion
Sadness	apathy, pensiveness, gravity, compassion, sadness, thing
Anger	despair, ingratitude, general-dislike
Fear	ambiguous-fear, ambiguous-expectation, ambiguos-agitation, positive-expectation, daze, shame, anxiety, negative-fear
Joy	humility, surprise, levity, positive-fear, neutral-unconcern, gratitude, fearlessness, affection, self-pride, enthusiasm, positive-hope, calmness, love, liking, joy

Table 8: Mapping of the general emotion of WordNet-Affect to SemEval emotion subtask EI-oc and EI-reg

SemEval emotion	WNA emotion
Sadness	pensiveness, compassion, sadness
Anger	
Fear	ambiguous-fear, gravity, daze, shame, anxiety, negative-fear
Joy	levity, positive-fear, enthusiasm, calmness, joy
Anticipation	ambiguous-expectation, positive-expectation
Disgust	neutral-unconcern, thing, ingratitude, general-dislike
Love	gratitude, affection, love, liking
Optimism	positive-hope
Pessimism	despair
Surprise	ambiguos-agitation, surprise
Trust	fearlessness, self-pride, humility

Table 9: Mapping of the general emotion of WordNet-Affect to SemEval emotion subtask E-c