Procesamiento del Lenguaje Natural, Revista nº 60, marzo de 2018, pp. 37-44 recibido 20-11-2017 revisado 04-01-2018 aceptado 02-02-2018

The democratization of deep learning in TASS 2017

La democratización del aprendizaje profundo en TASS 2017

Manuel C. Díaz-Galiano[1], Eugenio Martínez-Cámara[2], M. Ángel García-Cumbreras[1], Manuel García-Vega[1], Julio Villena-Román[3]
[1]Advanced Studies Center in Information and Communication Technologies (CEATIC), University of Jaén, Jaén, Spain
[2]Department of Computer Science and Artificial Intelligence, University of Granada, Granada, Spain
[3]MeaningCloud, Madrid, Spain
[1]{mcdiaz, magc, mgarcia}@ujaen.es,
[2]emcamara@decsai.ugr.es, [3]jvillena@meaningcloud.com

Abstract: TASS 2017 has brought advances in the state-of-the-art in Sentiment Analysis in Spanish, because most of the systems submitted in 2017 were grounded on Deep Learning methods. Moreover, a new corpus of tweets written in Spanish was released, which is called InterTASS. The corpus is composed of tweets manually annotated at document level. The analysis of the results with InterTASS shows that the main challenge is the classification of tweets with a neutral opinion and those ones that do not express any opinion. Likewise, the organization exposed the project of extending InterTASS with tweets written in different versions of Spanish.
Keywords: TASS, sentiment analysis, deep learning, linguistic resources

Resumen: TASS 2017 ha vuelto a suponer un avance en el estado del arte de análisis de opiniones en español, debido a la exposición de sistemas mayoritariamente fundamentados en métodos de Deep Learning. Además, en esta edición se ha presentado un nueva colección de tuits en español manualmente etiquetados a nivel de documento y que se llama InterTASS. El análisis de los resultados con InterTASS demuestra que en el futuro el esfuerzo investigador se tiene que centrar en la distinción del nivel de intensidad de opinión neutro y la ausencia de opinión. Asimismo, se presentó el proyecto de ampliar el nuevo corpus con tuits escritos en el español que se habla en España y en algunos países de América.
Palabras clave: TASS, análisis de opiniones, aprendizaje profundo, recursos lingüísticos

1 Introduction

After sixth editions, the Workshop Sentiment Analysis at SEPLN (TASS) has become the reference workshop for the research community on Sentiment Analysis (SA) for the Spanish language in microblogs, specifically in Twitter. The main contribution of TASS is the progress of the state-of-the-art as can be read in Villena-Román et al. (2013), Villena-Román et al. (2014), Villena Román et al. (2015) and Martínez Cámara et al. (2016).

The success of TASS may be attributed to: 1) the generation and release of newly annotated corpora in every edition; 2) the organization of competitive evaluations in which the participants submit their systems, which are ranked according to their performance; and 3) the active involvement of the research community in the discussion about the main features of the submitted systems and the state-of-the-art in SA in Spanish, and setting up the challenges for the next edition.

Spanish is the second most widely-spoken language in the world, and it is mainly spoken in Spain and America. Although the language is the same, there exist several varieties with specific lexical and semantic differences among different geographical areas, namely Spain and American countries. Consequently, we set up the project of generating a new corpus of tweets for SA with the novelty of including tweets written in the different varieties of Spanish.

ISSN 1135-5948. DOI 10.26342/2018-60-4

In this paper, the first release of the International TASS Corpus, called InterTASS, is presented. The first version is only composed of tweets written in the Spanish spoken in Spain, but, in contrast to the General Corpus of TASS, InterTASS was manually annotated. Further details about the annotation of the corpus are described in section 2.

TASS 2017 proposed two subtasks: Task 1, polarity classification at document (tweet) level; and Task 2, polarity classification at aspect level (see section 3). Eleven teams from Spain and America submitted several systems and a description paper. Most of the systems are based on the use of Deep Learning methods. Some of them attempted to improve the results using ensemble classifiers. In this paper, we also depict the main features of the best submitted systems and analyse their results (see section 4).

2 Resources

TASS 2017 provided four datasets to the participants for the evaluation of their systems. Three of them were already used in previous editions: General Corpus, Social-TV Corpus, and STOMPOL. A new dataset, InterTASS, was created for Task 1 in TASS 2017.

2.1 InterTASS

The International TASS Corpus (*InterTASS*) is a new corpus released in TASS 2017 for the polarity classification at tweet level in Task 1. This is the first version and includes tweets posted in Spain, all of them are written in the Spanish variety spoken in Spain. The final version of the corpus will be composed of tweets written in the variety of Spanish spoken in different Spanish-speaking countries in America.

In order to prepare this version, over 500,000 tweets were downloaded between July 2016 and January 2017 using some keywords. These tweets were filtered according to the following requirements: 1) tweets should be written in Spanish,[1] 2) each tweet should have at least one adjective, 3) the minimum length of tweets should be four words.

Eight subsets were prepared, sorting the tweets according to their number of words. Using these subsets, the final collection was created by randomly selecting a homogeneous number of tweets from each subset, 3,413 tweets in total.

The annotation process was made by five annotators using a scale of four levels of polarity for the global sentiment of the tweet: positive (P), negative (N), neutral (NEU) and no sentiment (NONE). Tweets were evenly distributed, so that each tweet was annotated by at least three annotators. The annotation guidelines regarding the assignment of the label of each tweet were:

- A label is assigned to a tweet when at least two annotators are agree.
- In case the three annotators are not agree, the other two ones, who are different from the first three, annotate the tweet.
- If the tie persisted, the conflicting annotator decided the label of the tweet.

Each tweet includes its ID (`tweetid`), the creation date (`date`) and the user ID (`user`). Restrictions in the Twitter API Terms of Service,[2] do not allow to release a corpus that includes text contents or information about users. The actual message content of tweets can be obtained by making queries to the Twitter API using the `tweetid`. The corpus is in XML, and Figure 1[3] shows a sample tweet[4].

```
<tweet>
  <tweetid>[ID]</tweetid>
  <user>[USER NAME]</user>
  <content>y lo peor de todo es que
      funcionaba maldita Jaco como te
      quiero </content>
  <date>[DATE]</date>
  <lang>es</lang>
  <sentiment>
    <polarity>
      <value>NEU</value>
    </polarity>
  </sentiment>
</tweet>
```

Figure 1: A tweet in the InterTASS corpus

Finally, the corpus was divided into three datasets: Training, Development and Test. The Training and Development sets were released with the annotations, so the participants could train and validate their models.

[1] `langdetect` Python library was used to check.

[2] `https://dev.twitter.com/terms/api-terms`

[3] The `tweetid`, the `user`, the `date` fields are hidden because of the Twitter term of service.

[4] In English: The worst is that it worked, fucking Jaco I love you too much.

The test corpus was provided without any annotation and was used to evaluate systems. Statistics are shown in Table 1.

	Training	Dev.	Test
P	317	156	642
N	416	219	767
NEU	133	69	216
NONE	138	62	274
Total	1,008	506	1,899

Table 1: Number of tweets per dataset and class in the InterTASS corpus

2.2 General corpus

The General Corpus has been used since the first edition of TASS. It has about 68,000 tweets, written in Spanish by about 150 well-known personalities and celebrities of the world of politics, economy, communication, mass media, and culture, between November 2011 and March 2012. The details of the corpus are described in Villena-Román et al. (2015) and García-Cumbreras et al. (2016).

This corpus was divided into training set (10%) and test set (90%). Each tweet in the training set was annotated with its global polarity in a scale of six intensity levels: strong positive (P+), positive (P), neutral (NEU), negative (N), strong negative (N+) and no sentiment (NONE). The test set was annotated by a meta-classifier based on majority voting, using as base classifiers the submitted systems of previous editions of TASS.

In addition, a selected subset containing 1,000 tweets with a similar class distribution than the training set was extracted in 2015 edition and manually annotated for an additional evaluation of the systems (1k test set).

2.3 Social-TV Corpus

The Social-TV corpus was released in TASS 2014. The tweets were gathered during the 2014 Final of Copa del Rey championship in Spain between Real Madrid and F.C. Barcelona. After filtering out useless tweets, a subset of 2,773 tweets was selected. Further details in Villena-Román et al. (2015), and García-Cumbreras et al. (2016).

The sentiment was manually annotated at aspect level (31 aspects), using only 3 levels of opinion: positive (P), neutral (NEU) and negative (N). The corpus was randomly split into two subsets: training and test (1,773 and 1,000 tweets, respectively).

2.4 STOMPOL

STOMPOL (Spanish Tweets for Opinion Mining about POLitics), released in TASS 2015, is a corpus of Spanish tweets for SA at aspect level. The tweets were gathered from the 23rd to the 24th of April of 2015 during the Spanish political campaign of regional and local elections. Each tweet was manually annotated at aspect level by two annotators, and a third one in case of disagreement. The topics of the tweets are: economics, health system, education, political party, electoral system or environmental policy.

The corpus is composed of 1,284 tweets, and was also divided into training set (784 tweets) and test set (500 tweets). Further details in Villena-Román et al. (2015), García-Cumbreras et al. (2016).

3 Tasks

TASS 2017 proposed two tasks addressing the main challenges of SA in Twitter in Spanish.

3.1 Task 1. Sentiment Analysis at Tweet level

This main task focused on the evaluation of polarity classification systems at tweet level in Spanish. Systems were evaluated on three different datasets: two versions of the General Corpus (the complete test set and the 1k test set), and the new InterTASS corpus.

Participants had to identify the intensity of the opinion expressed in each tweet in any of the 4 intensity levels of polarity in which the datasets were annotated. For the two sets of the General Corpus, which were originally annotated in 6 polarity classes, a direct translation to 4 classes (P+ changed to P and N+ to N) was performed so that the evaluation was consistent with InterTASS.

The three datasets were divided into training, development and test datasets, which were provided to participants in order to train and evaluate their systems. Systems were allowed to use any set of data as training dataset, i.e., the training set of InterTASS, other training sets from the previous editions of TASS or other sets of tweets. However, using the test set of InterTASS and the test set of the datasets of previous editions as training data was obviously forbidden. Apart from that, participants could use any kind of linguistic resource for the development of their classification model.

Participants were expected to submit 3 experiments per each test set, so each participant team could submit a maximum of 9 files of results. Accuracy and the macro-averaged versions of Precision, Recall and F1 were used as evaluation measures. Systems were ranked by the Macro-F1 and Accuracy measures.

3.2 Task 2. Aspect-based Sentiment Analysis

This second task proposed the development of aspect-based polarity classification systems. Two datasets from previous editions were used to evaluate the systems: Social-TV and STOMPOL (see section 2). The aspect, the main category of the aspect, and the opinion in three intensity levels (P, N, textscneu) were annotated in the two datasets.

Participants were expected to submit up to 3 experiments for each corpus. For evaluation, exact match with a single label combining "aspect-polarity" was used. The evaluation measures were the same as in Task 1.

4 Analysis of Submissions

In TASS 2017, the following 11 different groups presented their runs in the tasks:

- ELiRF, Universidad Politécnica de Valencia (Spain)
- RETUYT, Universidad de la República, Montevideo (Uruguay)
- ITAINNOVA, Zaragoza (Spain)
- jacerong, Santiago de Cali (Colombia)
- INGEOTEC, Universidad Panamericana (Mexico)
- tecnolengua, Universidad de Málaga (Spain)
- SINAI, Universidad de Jaén (Spain)
- LexFAR, Universidad Autónoma Metropolitana (Mexico)
- OEG, Universidad Politécnica de Madrid (Spain)
- GSI, Universidad Politécnica de Madrid (Spain)
- C100T-PUCP, Universidad Católica del Perú (Peru)

It must be pointed out that five groups are from countries other than Spain, so the workshop is relevant in other American countries. Table 2 shows the participation of each group in the TASS 2017 tasks: 1I (Task 1, InterTASS corpus), 1G (Task 1, General corpus), 2SO (Task 2, Social-TV corpus) and 2ST (Task 2, STOMPOL corpus).

	1I	1G	2SO	2ST
ELiRF	X	X	X	X
RETUYT	X	X	X	X
ITAINNOVA	X			
jacerong	X	X		
INGEOTEC	X	X		
tecnolengua	X	X		
SINAI	X			
LexFAR	X			
OEG	X	X		
GSI	X	X		
C100T-PUCP				X
Total	10	7	2	2

Table 2: Groups and tasks

Most of the systems were based on Deep Learning techniques, but there were solutions based on traditional machine learning methods and meta-classifiers.

Hurtado, Pla, and González (2017) (ELiRF) created a set of domain-specific word embeddings following the approach of Tang (2015) for tasks 1 and 2. The former word embeddings set is jointly used with a general-domain set of embeddings to represent each token. They evaluated three different neural networks architectures: multilinear perceptron (MLP), convolutional recurrent neural network (CNN) and long-short term memory (LSTM) recurrent networks (RNN).

Cerón-Guzmán (2017) (jacerong) presented an ensemble classifier system for the first task. Their system generated quantitative features from the tweets (the number of words in upper case, the number of words with repeated letters, etc.), and then they used lists of opinion bearing words (iSOL (Molina-González et al., 2013)), as well as the inversion of the polarity of words following a window shifting approach for negation handling. The base classifiers of the ensemble system were Logistic Regression and SVM.

Montañés Salas et al. (2017) (ITAINNOVA) used the FastText classifier (Joulin et al., 2016) for the InterTASS dataset. After a traditional pre-processing to the input tweets, the system substituted the words with an emotional meaning by their synonyms from a list of words with an emotional meaning (Bradley and Lang, 1999).

Rosá et al. (2017) (RETUYT) used three different approaches: an SVM classifier with word embeddings and quantitative linguistic properties as features; a deep neural network grounded on the use of a CNN for encoding the input tweets; and the combination of the two previous classifiers by the selection of the output class with a higher probability mean from the two previous classifiers.

García-Vega et al. (2017) (SINAI) used for InterTASS an SVM classifier that uses word-embeddings as features. They introduced the use of the language of each user in the classification. Other approaches are based on deep neural networks grounded on the use of LSTM RNN for the encoding of the meaning of the input tweets.

The approach by Moctezuma et al. (2017) (INGEOTEC) was an ensemble of SVM classifiers combined into a non-linear model created with genetic programming to tackle the task of global polarity classification at tweet level. They used B4MSA algorithm, a proposed entropy-based term-weighting scheme, which is a baseline supervised learning system based on the SVM classifier, an entropy-based term-weighting scheme. Additionally, they used EvoDAG, a GP system that combines all decision values predicted by B4MSA systems. They also used two external datasets to train the B4MSA algorithm.

Navas-Loro and Rodríguez-Doncel (2017) (OEG) used two classifiers: Multinominal Naïve Bayes and Sequential Minimal Optimization for SVM. Furthermore, they applied morphosyntactic analyses for negation detection, along with the use of lexicons and dedicated preprocessing techniques for detecting and correcting frequent errors and expressions.

Araque et al. (2017) (GSI) applied an RNN architecture composed of LSTM cells followed by a feed-forward network. The architecture makes use of two different types of features: word embeddings and sentiment lexicon values. The recurrent architecture allows to process text sequences of different lengths, while the lexicon inserts directly into the system sentiment information. Two variations of this architecture were used: an LSTM that iterates over the input word vectors, and a combination of the input word vectors and polarity values from a sentiment lexicon.

Tume Fiestas and Sobrevilla Cabezudo (2017) (C100T-PUCP) proposed for Task 2 an approach based on word embeddings for polarity classification at aspect-level. They used vectors of the words to measure their similarity and make a model to classify each polarity of each aspect for each tweet.

Reyes-Ortiz et al. (2017) (LexFAR) used, for Task 1, support vector machines algorithm and lexicons of semantic polarities at the level of lemma for Spanish. Features extracted from lexicons are represented by the bag-of-words model and they are weighted using Term Frequency measure at tweet level.

Moreno-Ortiz and Pérez Hernández (2017) (tecnolengua) proposed a classification model based on the Lingmotif Spanish lexicon with a number of formal text features, both general and CMC-specific, as well as single-word keywords and n-gram keywords. They used logistic regression classifier trained with the optimal set of features, SVM classifier on the same features set.

The fifteen best results reached by systems in Task 1, using the test sets of InterTASS and the General Corpus are showed in Tables 3, 4 and 5.

System	**M-F1**	**Acc.**
ELiRF-UPV-run1	0.493	0.607
RETUYT-svm_cnn	0.471	0.596
ELiRF-UPV-run3	0.466	0.597
ITAINNOVA-model4	0.461	0.576
jacerong-run-2	0.460	0.602
jacerong-run-1	0.459	0.608
INGEOTEC-evodag_001	0.457	0.507
RETUYT-svm	0.457	0.583
tecnolengua-sent_only	0.456	0.582
ELiRF-UPV-run2	0.450	0.436
ITAINNOVA-model3	0.445	0.561
RETUYT-cnn3	0.443	0.558
SINAI-w2v-nouser	0.442	0.575
tecnolengua-run3	0.441	0.576
tecnolengua-sent_only_fixed	0.441	0.595

Table 3: Task 1 InterTASS corpus, fifteen best results

Table 6 and Table 7 show the results reached by the submitted systems in Task 2, using the test sets of Social-TV corpus and STOMPOL corpus respectively.

System	M-F1	Acc.
INGEOTEC-evodag_003	0.577	0.645
jacerong-run-1	0.569	0.706
jacerong-tass_2016-run_3	0.568	0.705
ELiRF-UPV-run2	0.549	0.659
ELiRF-UPV-run3	0.548	0.725
RETUYT-svm_cnn	0.546	0.674
jacerong-run-2	0.545	0.701
ELiRF-UPV-run1	0.542	0.666
RETUYT-cnn	0.541	0.638
RETUYT-cnn3	0.539	0.654
tecnolengua-run3	0.528	0.657
tecnolengua-final	0.517	0.632
tecnolengua-531F1_no_ngrams	0.508	0.652
INGEOTEC-evodag_001	0.447	0.514
OEG-victor2	0.389	0.496

Table 4: Task 1 General Corpus (full test), fifteen best results

System	M-F1	Acc.
RETUYT-svm	0.562	0.700
RETUYT-cnn4	0.557	0.694
RETUYT-cnn2	0.555	0.694
INGEOTEC-evodag_003	0.526	0.595
tecnolengua-run3	0.521	0.638
ELiRF-UPV-run1	0.519	0.630
jacerong-tass_2016-run_3	0.518	0.625
jacerong-run-1	0.508	0.678
jacerong-run-2	0.506	0.673
ELiRF-UPV-run2	0.504	0.596
tecnolengua-final	0.488	0.618
tecnolengua-run4	0.483	0.612
ELiRF-UPV-run3	0.477	0.588
INGEOTEC-evodag_002	0.439	0.431
INGEOTEC-evodag_001	0.388	0.486

Table 5: Task 1 General Corpus (1k), fifteen best results

4.1 InterTASS Analysis

If the test set is grouped by the number of correct labels assigned by one or some of the submitted systems, the obtained results are shown in Figure 2. The test set is balanced according to complexity, and there are more than 10% of the tweets that are not correctly labelled by any system.

Figure 3 analyses the relation with the polarity label and the correct predictions. Most of the rightly predicted tweets are positive or negative, in contrast the submitted systems use to fail in the classification of NONE and NEU tweets.

System	M-F1	Acc.
ELiRF-UPV-run3	0.537	0.615
ELiRF-UPV-run2	0.513	0.600
ELiRF-UPV-run1	0.476	0.625
RETUYT-svm2	0.426	0.595
RETUYT-svm	0.413	0.493

Table 6: Task 2 Social-TV corpus results

System	M-F1	Acc.
ELiRF-UPV-run1	0.537	0.615
RETUYT-svm2	0.508	0.590
ELiRF-UPV-run3	0.486	0.578
ELiRF-UPV-run2	0.486	0.541
C100T-PUCP-run3	0.445	0.528
C100T-PUCP-run1	0.415	0.563
C100T-PUCP-run2	0.414	0.517
RETUYT-svm	0.377	0.514

Table 7: Task 2 STOMPOL corpus results

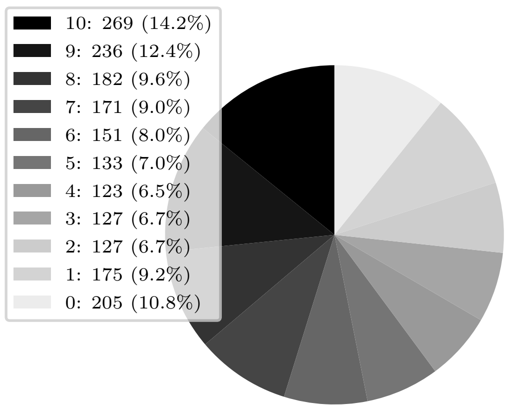

Figure 2: Systems that correctly classified a number of tweets (InterTASS corpus)

Last, we compared the statistics of the correct results with the number of words in tweets, as during the manual labelling, the annotators warned that tweets with a low number of words were noticeably more difficult to annotate. Table 8 shows the statistics. The first column shows different groups with the number of words of the tweets, and the other columns represent the number of systems that have hit the correct label. The percentage is calculated with the total number of tweets regarding the total value of the same column. Statistics are comparable, regardless of the number of words in the tweets, so, apparently, there is not a direct relation between the number of words of the tweets

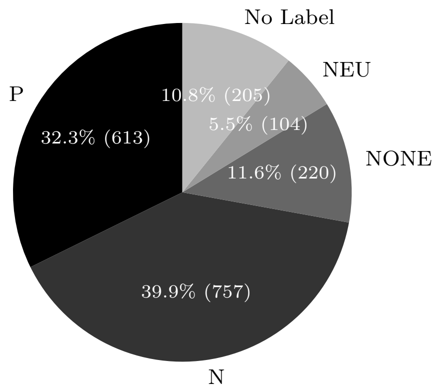

Figure 3: Polarity label and successful results (InterTASS corpus)

and the number of right predictions.

Words	0	1-5	6-9	10
0-4	20 (10%)	98 (14%)	101 (14%)	21 (8%)
5	14 (7%)	101 (15%)	87 (12%)	38 (14%)
6	32 (16%)	79 (12%)	92 (12%)	37 (14%)
7	29 (14%)	86 (13%)	81 (11%)	44 (16%)
8	24 (12%)	84 (12%)	88 (12%)	44 (16%)
9	32 (16%)	67 (10%)	107 (14%)	33 (12%)
11-18	25 (12%)	78 (11%)	97 (13%)	20 (7%)

Table 8: Correct tweet labels vs number of words (InterTASS corpus)

5 Conclusions and future work

The main objectives of TASS 2017 were: 1) to keep the interest of the research community in SA in Spanish; 2) the release of InterTASS, a new corpus for SA in Spanish; and 3) to forward the state-of-the-art through the debate of the features of the systems, most of them based on the use of Deep Learning methods and meta-classifiers.

We analyzed (see section 4) the performance of the submitted systems in the InterTASS corpus, and we conclude that there is room for improvement in the classification of the classes: NEU or NONE. Furthermore, no relation between the length of the tweets and the accuracy of the classification was found.

The work for further editions of TASS will be led by two goals. The first one is to broaden the number of tasks related to SA and semantic analysis with the aim of keep fostering the research in SA tasks in Spanish. The first milestone of the first goal was the update of the name of TASS to Workshop on Semantic Analysis at SEPLN in the edition of 2017. The second milestone will be the invitation to other research groups to organize and generate linguistic resources for SA tasks in Spanish. The second goal is to conclude the development of InterTASS with tweets written in the Spanish varieties of (as many as possible) Spanish speaking countries.

Acknowledgements

This research work is partially supported by REDES project (TIN2015-65136-C2-1-R) and SMART project (TIN2017-89517-P) from the Spanish Government, and a grant from the Fondo Europeo de Desarrollo Regional (FEDER). Eugenio Martínez Cámara was supported by the Juan de la Cierva Formación Programme (FJCI-2016-28353) from the Spanish Government.

References

Araque, O., R. Barbado, J. F. Sánchez-Rada, and C. A. Iglesias. 2017. Applying recurrent neural networks to sentiment analysis of spanish tweets. In *Proceedings of TASS 2017*, volume 1896 of *CEUR Workshop Proceedings*, Murcia, Spain, September. CEUR-WS.

Bradley, M. M. and P. J. Lang. 1999. Affective norms for english words (anew): Stimuli, instruction manual, and affective ratings. Technical report, Center for Research in Psychophysiology, University of Florida.

Cerón-Guzmán, J. A. 2017. Classier ensembles that push the state-of-the-art in sentiment analysis of spanish tweets. In *Proceedings of TASS 2017*.

García-Cumbreras, M. A., J. Villena-Román, E. Martínez-Cámara, M. C. Díaz-Galiano, M. T. Martín-Valdivia, and L. A. Ureña López. 2016. Overview of tass 2016. In *TASS 2016: Workshop on Sentiment Analysis at SEPLN*, pages 13–21.

García-Vega, M., A. Montejo-Ráez, M. C. Díaz-Galiano, and S. M. Jiménez-Zafra.

2017. Sinai en tass 2017: Clasificación de la polaridad de tweets integrando información de usuario. In *Proceedings of TASS 2017*.

Hurtado, L.-F., F. Pla, and J.-A. González. 2017. Elirf-upv en tass 2017: Análisis de sentimientos en twitter basado en aprendizaje profundo. In *Proceedings of TASS 2017*.

Joulin, A., E. Grave, P. Bojanowski, and T. Mikolov. 2016. Bag of tricks for efficient text classification. *arXiv preprint arXiv:1607.01759*.

Martínez Cámara, E., M. A. García Cumbreras, J. Villena Román, and J. García Morera. 2016. TASS 2015-The evolution of the spanish opinion mining systems. *Procesamiento del Lenguaje Natural*, 56(0):33–40.

Moctezuma, D., M. Graff, S. Miranda-Jiménez, E. S. Tellez, A. Coronado, C. N. Sánchez, and J. Ortiz-Bejar. 2017. A genetic programming approach to sentiment analysis for twitter: Tass'17. In *Proceedings of TASS 2017*, volume 1896 of *CEUR Workshop Proceedings*, Murcia, Spain, September. CEUR-WS.

Molina-González, M. D., E. Martínez-Cámara, M.-T. Martí-Valdivia, and J. M. Perea-Ortega. 2013. Semantic orientation for polarity classification in spanish reviews. *Expert Systems with Applications*, 40(18):7250 – 7257.

Montañés Salas, R. M., R. del Hoyo Alonso, J. Vea-Murguía Merck, R. Aznar Gimeno, and F. J. Lacueva-Pérez. 2017. FastText como alternativa a la utilización de deep learning en corpus pequeños. In *Proceedings of TASS 2017*.

Moreno-Ortiz, A. and C. Pérez Hernández. 2017. Tecnolengua lingmotif at tass 2017: Spanish twitter dataset classification combining wide-coverage lexical resources and text features. In *Proceedings of TASS 2017*.

Navas-Loro, M. and V. Rodríguez-Doncel. 2017. Oeg at tass 2017: Spanish sentiment analysis of tweets at document level. In *Proceedings of TASS 2017*, volume 1896 of *CEUR Workshop Proceedings*, Murcia, Spain, September. CEUR-WS.

Reyes-Ortiz, J. A., F. Paniagua-Reyes, B. Priego-Sánchez, and M. Tovar. 2017. Lexfar en la competencia tass 2017: Análisis de sentimientos en twitter basado en lexicones. In *Proceedings of TASS 2017*, volume 1896 of *CEUR Workshop Proceedings*, Murcia, Spain, September. CEUR-WS.

Rosá, A., L. Chiruzzo, M. Etcheverry, and S. Castro. 2017. Retuyt en tass 2017: Análisis de sentimientos de tweets en español utilizando svm y cnn. In *Proceedings of TASS 2017*.

Tang, D. 2015. Sentiment-specific representation learning for document-level sentiment analysis. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, WSDM '15, pages 447–452, New York, NY, USA. ACM.

Tume Fiestas, F. and M. A. Sobrevilla Cabezudo. 2017. C100tpucp at tass 2017: Word embedding experiments for aspect-based sentiment analysis in spanish tweets. In *Proceedings of TASS 2017*, volume 1896 of *CEUR Workshop Proceedings*, Murcia, Spain, September. CEUR-WS.

Villena-Román, J., J. García-Morera, M. A. García-Cumbreras, E. Martínez-Cámara, M. T. Martín-Valdivia, and L. A. Ureña López. 2015. Overview of tass 2015. In *TASS 2015: Workshop on Sentiment Analysis at SEPLN*, pages 13–21.

Villena-Román, J., J. García-Morera, S. Lana-Serrano, and J. C. González-Cristóbal. 2014. Tass 2013 - a second step in reputation analysis in spanish. *Procesamiento del Lenguaje Natural*, 52(0):37–44, March.

Villena-Román, J., S. Lana-Serrano, E. Martínez-Cámara, and J. C. González-Cristóbal. 2013. Tass - workshop on sentiment analysis at sepln. *Procesamiento del Lenguaje Natural*, 50:37–44.

Villena Román, J., E. Martínez Cámara, J. García Morera, and S. M. Jiménez Zafra. 2015. Tass 2014 - the challenge of aspect-based sentiment analysis. *Procesamiento del Lenguaje Natural*, 54(0):61–68.