Procesamiento de Lenguaje Natural, Revista nº 56, marzo de 2016, pp 33-40 recibido 20-11-2015 revisado 14-01-2016 aceptado 03-02-2016

TASS 2015 – The Evolution of the Spanish Opinion Mining Systems

TASS 2015 – La evolución de los sistemas de análisis de opiniones para español

Miguel Ángel García Cumbreras
Eugenio Martínez Cámara
SINAI Research Group. University of Jaén
E-23071 Jaén, Spain
{magc, emcamara}@ujaen.es

Julio Villena Román
Janine García Morera
Sngular Meaning
E-28031 Madrid, Spain
{jvillena, jgarcia}@daedalus.es

Resumen: El análisis de opiniones en microblogging sigue siendo una tarea de actualidad, que permite conocer la orientación de las opiniones que minuto tras minuto se publican en medios sociales en Internet. TASS es un taller de participación que tiene como finalidad promover la investigación y desarrollo de nuevos algoritmos, recursos y técnicas aplicado al análisis de opiniones en español. En este artículo se describe la cuarta edición de TASS, resumiendo las principales aportaciones de los sistemas presentados, analizando los resultados y mostrando la evolución de los mismos. Además de analizar brevemente los sistemas que se presentaron, se presenta un nuevo corpus de tweets etiquetados en el dominio político, que se desarrolló para la tarea de Análisis de Opiniones a nivel de Aspecto.
Palabras clave: TASS 2015, análisis de opiniones, análisis de aspectos, medios sociales.

Abstract: Sentiment Analysis in microblogging continues to be a trendy task, which allows to understand the polarity of the opinions published in social media. TASS is a workshop whose goal is to boost the research on Sentiment Analysis in Spanish. In this paper we describe the fourth edition of TASS, showing a summary of the systems, analyzing the results to check their evolution. In addition to a brief description of the participant systems, a new corpus of tweets is presented, compiled for the Sentiment Analysis at Aspect Level task.
Keywords: TASS 2015, Opinion Mining, Aspect Based Sentiment Analysis, Social TV.

1 Introduction

The Workshop on Sentiment Analysis at SEPLN (TASS, in Spanish) is an experimental evaluation workshop, which is a satellite event of the annual SEPLN Conference, with the aim to promote the research of Sentiment Analysis systems in social media, focused on Spanish language. After successful editions (Villena-Román et al., 2013, Villena-Román et al., 2014), the round corresponding to the year 2015 was held at the University of Alicante.

Twitter is one of the most popular social network, and also the most used microblogging platform. The two main features of Twitter are its simplicity and its real-time nature. Due mainly to those two reasons, people use Twitter to post about what they are doing or what they think. Thus, Twitter is plenty of opinions concerning whatever topic, so that Twitter is a suitable source of opinions.

Sentiment Analysis (SA) is usually defined as the computational treatment of opinion, sentiment and subjectivity in texts, but from our point of view SA is defined in a better way as a series of computational techniques for extracting, classifying, understanding, and assessing the opinions expressed in various online news sources, social media comments, and other user-generated. It is a hard task because even humans often disagree on the sentiment of a given text. SA is also a difficult task because encompasses several Natural Language Processing tasks, and up until now there are several unresolved.

The main characteristic of tweets is their length, 140 characters, which determines the text that the users post in the platform. Furthermore, there are other features that must be taken into account because they make harder the processing of tweets, such as the informal linguistic style utilized by users, the poor grammar and the number of spellings mistakes of the tweets, the lack of context, and the problem related to the data sparsity.

The study of the opinion expressed in a document can be carried out at three different levels of analysis: document level, sentence level and entity or aspect level. Up until now, most of the research conducted by the SA research community is mainly focused on developing polarity classification systems at document level. Polarity classification systems have usually based on two main approaches: a supervised approach, which applies machine learning algorithms in order to train a polarity classifier using a labelled corpus (Pang et al., 2002); an unsupervised approach, known as semantic orientation, which integrates linguistic resources in a model in order to identify the polarity of the opinions (Turney, 2002). The main goal of TASS is to serve as a discussion forum about the progress of SA Analysis research. Work in polarity classification at document level is very active nowadays, so the edition of 2015 included the rerun of the legacy task related to the assessment of polarity classification systems at document level.

Although the processing at document level is a problem still open, the analysis of the opinion at aspect level is more challenging. Furthermore, the industry is demanding polarity classification systems able to identify the opinion valence about specific entities or aspects, or in other words, the industry is demanding the development of polarity classification systems at aspect level. TASS is paying attention to the aspect level analysis since the edition of 2014. Due to the importance of the aspect level analysis, this year was included a rerun of the polarity classification at aspect level, but this year with another corpus of tweets labeled at aspect-level.

The rest of the paper is organized as follows. Section 2 describes the different corpus provided to participants. Section 3 shows the different tasks of TASS 2015. Section 4 describes the participants and the overall results are presented in Section 5. Finally, the last section shows some conclusions and future directions.

2 Corpus

The corpus prepared and provided with the aim of accomplished the tasks defined for the edition of 2015 are described in the subsequent subsections. It must be highlighted the fact that all the corpora compiled by the organization of TASS is available for the research community.

2.1 General corpus

The General Corpus is used in the main legacy task of TASS, which is polarity classification at document level, and it has been used since the first edition of TASS. The General Corpus contains over 68,000 tweets written in Spanish by 150 well-known personalities and celebrities of the world of politics, economy, communication, mass media and culture. It was built between November 2011 and March 2012. It covers some Spanish-speaking world because of the diverse nationality of the authors (from Spain, Mexico, Colombia, Puerto Rico, etc.).

This General Corpus was divided into training set (10%) and test set (90%). As usual, the training set was released to the participants, to train and validate their models, and the test corpus was provided without any annotation to evaluate the results. Each tweet was tagged with its global polarity in a scale of six levels of polarity intensity, which are: strong positive (P+), positive (P), neutral (NEU), negative (N), strong negative (N+) and one additional for no sentiment tweets (NONE). A wider description of the General Corpus is described in (Villena-Román, 2013).

The level of agreement or disagreement of the expressed sentiment within the content was included, with two values: AGREEMENT and DISAGREEMENT. It is useful to make out whether a neutral sentiment comes from neutral keywords or else the text contains positive and negative sentiments at the same time.

Polarity values related to the entities that are mentioned in the text are also included for those cases when applicable. These values are similarly tagged with six possible values and include the level of agreement as related to each entity.

All tagging has been done semi automatically: a baseline machine learning model is first run and then, human experts manually check all tags. In the case of the polarity at entity level, due to the high volume of data to check, this tagging has just been done for the training set.

2.2 Social-TV corpus

The Social-TV Corpus is used in the second task of the edition of 2015, which is focused on polarity classification at aspect level. The Social-TV corpus is a corpus generated in 2014, with tweets collected during the 2014 Final of Copa del Rey championship in Spain between Real Madrid and F.C. Barcelona. This dataset was collected only in one day, on 16 April 2014. Over 1 million of tweets were collected at 15-minute intervals after the match. After filtering useless information a subset of 2,773 was selected.

Three people identified the aspects of the expressed messages and tagged its sentiment manually. Tweets may cover more than one aspect.

Sentiment polarity has been tagged from the point of view of the person who writes the tweet, using 3 levels: P (positive), NEU (neutral) and N (negative). In this case, there is no distinction between no sentiment and neutral sentiment expressed.

The Social-TV corpus was randomly divided into training set (1,773 tweets) and test set (1,000 tweets), with a similar distribution of aspects and sentiments.

2.3 STOMPOL corpus

The STOMPOL Corpus is the new corpus developed for the edition of 2015, and it was used in the task of polarity classification at aspect level. The STOMPOL corpus (corpus of Spanish Tweets for Opinion Mining at aspect level about POLitics) is a corpus of Spanish tweets related to a political aspect that appear in the Spanish political campaign of regional and local elections that were held on 2015, which were gathered from 23rd to 24th of April 2015. These political aspects are the following:

Economics: taxes, infrastructure, markets or labor policy.

- Health System: hospitals, public/private health system, drugs or doctors.
- Education: state school, private school, or scholarships.
- Political party: anything good (speeches, electoral programme...) or bad (corruption, criticism) related to the entity
- Other aspects: electoral system or environmental policy.

Each aspect is related to one or several entities that correspond to one of the main political parties in Spain: Partido Popular (PP), Partido Socialista Obrero Español (PSOE), Izquierda Unida (IU), Podemos, Ciudadanos (Cs) and Unión, Progreso y Democracia (UPyD).

As in previous corpus, two people, and a third one in case of disagreement, manually tagged each tweet. Each tag contains the sentiment polarity from the point of view of the person who writes the tweet, using 3 levels: P (positive), NEU (neutral) and N (negative). Again, no difference is made between no sentiment and a neutral sentiment (neither positive nor negative). Each political aspect is linked to its correspondent political party and its polarity. Figure 1 shows the information of a sample tweet.

```
<tweet id="591267548311769088">@ahorapodemos
@Pablo_Iglesias_ @SextaNocheTV Que alguien pregunte si
habrá cambios en las <sentiment aspect="Educacion" entity=
"Podemos" polarity="NEU">becas</sentiment> MEC para
universitarios, por favor.</tweet>

<tweet id="591192167944736769">#Arroyomolinos lo que le
interesa al ciudadano son Políticos cercanos que se
interesen y preocupen por sus problemas <sentiment aspect=
"Propio_partido" entity="Union_Progreso_y_Democracia"
polarity="P">@UPyD</sentiment> VECINOS COMO TU</tweet>
```

Figure 1 : Sample tweets (STOMPOL corpus)

3 *Description of tasks*

The main goal of TASS is to boost the research on SA in Spanish. In the 2015 edition we analyzed the evolution of the different approaches for SA in Spanish during the last years. So, the traditional SA at global level task was rerun again. Moreover, we wanted to foster the research in the analysis of fine-grained polarity analysis at aspect level (aspect-based SA, one of the new requirements of the market of natural language processing in these areas). So, two legacy tasks were repeated again, to compare results, and a new corpus was created. The proposed tasks are described next.

3.1 Task 1: Sentiment Analysis at Global Level (legacy)

This task consists in performing an automatic polarity classification system to tag the global polarity of each tweet in the test set of the General Corpus. The training set of this General Corpus was provided to the participants.

There were two different evaluations: one based on 6 different polarity labels (P+, P, NEU, N, N+, NONE) and another based on just 4 labels (P, N, NEU, NONE).

Then, the same test corpus of previous years was used to evaluate the results and we compared the evaluation among systems. Two test sets were used: one complete set and set with 1.000 tweets (1k set). It is a subset of the first one, extracted to deal with the problem of the imbalanced distribution of labels between the general training and test set. It is a selected test subset with a similar distribution to the training corpus.

Due to the fact that the task implies the classification in six different classes, for the evaluation of the systems the macro-averaged version of the Precision, Recall and F1 measures were used. Also, the Accuracy measure was taken into account to evaluate the systems.

3.2 Task 2: Aspect-based sentiment analysis

Task 2 consists in performing an automatic polarity classification at aspect level.. Two corpora were provided: Social-TV Corpus and STOMPOL Corpus.

Allowed polarity values were P, N and NEU. For evaluation, a single label combining "aspect-polarity" has been considered. Similarly to the first task, accuracy, and the macro-average versions of Precision, Recall and F1 have been calculated for the global result.

4 Participants

In 2015, 35 groups were registered, and 17 of them sent their submissions and presented their results. The list of active participant groups is shown in Table 1, including the tasks in which they have participated.

The main goal of TASS is not to rank the systems submitted, but it is to compare and discuss the contributions from the different teams to the field of SA in Spanish. Thus, it is prominent to remark the foremost fundamentals of the systems that reached better results in the competition.

LIF team did not submit any paper, so the basics of its system could not be discussed at the workshop. On the other hand, the second best team submitted the description of its system. Hurtado and Pla (2015) (ELiRF team) participated in the two tasks. The polarity classification system is based on a voting system of dissimilar configurations of SVM. However, the key of the successful of Hurtado and Pla (2015) is the compilation of a very informative set of features, which combined the lexical information of tweets (unigrams of tokens and lemmas) and number of positive and negative words according to the lexicons ElhPolar (Saralegi and San Vicente, 2013), iSOL (Molina-González et al., 2013) and AFFIN (Hansen et al., 2011). The polarity classification at aspect-level is based on the determination of the context of each aspect using a fix window size on the left and right side of the aspect.

Group	1	2	Group	1	2
LIF	X		TID-spark	X	X
ELiRF	X	X	BittenPotato	X	
GSI	X	X	SINAI-wd2v	X	
LyS	X	X	DT	X	
DLSI	X		GAS-UCR	X	
GTI-Gradiant	X		UCSP	X	
ITAINNOVA	X		SEDEMO	X	
SINAI-ESMA	X		INGEOTEC	X	
CU	X		**Total groups**	17	4

Table 1: Participant groups

Araque et al., (2015) (GSI team) also participated in the two tasks. For the polarity classification system, the authors applied an approach similar to the one described in (Mohammad et al., 2013), which is based on the use of several lexical, morphosyntactic and sentiment features to represent the information of each tweet. The classification is carried out by a machine learning algorithm, specifically SVM. It must be highlighted that the authors take into account the treatment of negation following the same approach than (Pang et al., 2002). For the polarity classification task, the authors first invest their efforts in the identification of the aspects and their context. In order to detect the aspects, the authors run the Stanford CRF NER (Finkel, Grenager and Manning, 2005) and to identify their context they use a graph-based algorithm (Mukherjee and Bhattacharyya, 2012).

Vilares et al., (2015) (LyS team) propose an approach based on deep learning. Their polarity classifier used the neutral network Long Short-Term Memory (LSTM) with a logistic function at the output layer. The authors also participated in the second task, so they submitted a aspect-level polarity classification system. This system is based on the first one, but it only takes into account the context of each aspect. Regarding the context of each aspect identification, the authors use a fix window size on the left and right side of each aspect, in a similar way than Hurtado and Pla (2015).

The DLSI team (Fernández et al., 2015) attempted again taking advantage from all the lexical information of the tweets. Their system do not use unigrams or bigrams to represent the information of the tweets, they prefer to use skip-grams with the aim of enlarging the covering of the potential vocabulary of the tweets. They measure of the relevance of each skip-gram depends on a sentiment score, which is related to the sentiment class of the training data.

The GTI-Gradiant team (Álvarez-López et al., 2015) present a voting system with two base classifiers, the first on follow a supervised approach and the second one an unsupervised approach. The supervised method is based on a logistic regression classifier, which tries to classify the tweets using as features: unigrams, the POS-tags, the syntactic dependency categories that are in the tweet, and the number of positive and negative words. The unsupervised classifier takes into account the number of positive and negative words, the syntactic dependency structure of the tweet and uses a label propagation method (Caro and Grella, 2013) for obtaining the final sentiment score of the tweet.

The TID-spark team (Park, 2015) proposes an interesting approach for polarity classification based on sociolinguistic information. The author develops a unsupervised classifier that takes into account the information of the users of the tweets of the training data. The author uses this kind of information with the aim of modeling the language of each group of users. For the aspect-level polarity classification task, the author takes into account the possible political affiliation and the likely preference for a football team to build the language model of each group of users.

5 *Results*

The results for each task, in terms of Accuracy, are the following.

5.1 (legacy) Task 1: Sentiment Analysis at Global Level

Table 2 shows the results obtained for Task 1, with the evaluation based on five polarity levels and the whole General test corpus. The best accuracy value achieves 0.67.

Run ID	Acc.	Run ID	Acc.
LIF-Run-3	0.672	TID-spark-1	0.462
LIF-Run-2	0.654	BP-wvoted-v2_1	0.534
ELiRF-run3	0.659	Ensemble exp2_emotions	0.524
ELiRF-run2	0.658	BP-voted-v2	0.535
ELiRF-run1	0.648	SINAI_wd2v_500	0.474
LIF-Run-1	0.628	SINAI_wd2v_300	0.474
GSI-RUN-1	0.618	BP-wvoted-v1	0.522
GSI-RUN-2	0.610	BP-voted-v1	0.522
GSI-RUN-3	0.608	BP-rbf-v2	0.514
LyS-run-1	0.552	Lys-run-3	0.505
DLSI-Run1	0.595	BP-rbf-v1	0.494
Lys-run-2	0.568	CU-Run-2-CompMod	0.362
GTI-GRAD-Run1	0.592	DT-RUN-1	0.560
Ensemble exp1.1	0.535	DT-RUN-3	0.557
SINAI-EMMA-1	0.502	DT-RUN-2	0.545
INGEOTEC-M1	0.488	GAS-UCR-1	0.342
Ensemble exp3_emotions	0.549	UCSP-RUN-1	0.273
CU-Run-1	0.495	BP-wvoted-v2	0.009

Table 2: Results for task 1, 5 polarity levels, whole test corpus

Table 3 shows the results obtained with the 1k test corpus, the selected test subset that contains 1,000 tweets with a similar distribution to the training corpus. In this case the best accuracy value was 0.516, a loss of accuracy of 33% because of a more complex task of labeling.

Run ID	Acc.	Run ID	Acc.
LIF-Run-2	0.516	SINAI-EMMA-1	0.411
GTI-GRAD-Run1	0.509	CU-Run-1-CompMod	0.419
ELiRF-run2	0.488	Ensemble exp3	0.396

		1K	
GSI-RUN-1	0.487	TID	0.400
GSI-RUN-2	0.480	BP-voted-v1	0.408
GSI-RUN-3	0.479	DLSI-Run1	0.385
LIF-Run-1	0.481	CU-Run-2	0.397
ELiRF-run1	0.476	BP-wvoted-v1	0.416
SINAI_wd2v	0.389	BP-rbf-v1	0.418
ELiRF-run3	0.477	SEDEMO-E1	0.397
INGEOTEC-M1	0.431	DT-RUN-1	0.407
Ensemble exp1 1K	0.405	DT-RUN-2	0.408
LyS-run-1	0.428	DT-RUN-3	0.396
Ensemble exp2 1K	0.384	GAS-UCR-1	0.338
Lys-run-3	0.430	INGEOTEC-E1	0.174
Lys-run-2	0.434	INGEOTEC-E2	0.168

Table 3: Results for task 1, 5 polarity levels, selected 1k test corpus

To perform a more in-depth evaluation, previous results were evaluated considering only three polarity levels (positive, negative and neutral) and no sentiment. Tables 4 and 5 show this new evaluation, with the general whole test corpus and the selected 1k test corpus. The accuracy values increase because of a simpler task with three polarity labels. With the whole test corpus the best accuracy value was 0.726, and it was 0.632 with the 1k test corpus. Again, there was a loss of accuracy with the smaller test corpus.

Run ID	Acc.	Run ID	Acc.
LIF-Run-3	0.726	exp1_3_SPARK	0.610
LIF-Run-2	0.725	UCSP-RUN-1-ME	0.600
ELiRF-run3	0.721	BP-wvoted-v1	0.593
LIF-Run-1	0.710	BP-voted-v1 Ensemble	0.593
ELiRF-run1	0.712	exp3_3	0.594
ELiRF-run2	0.722	DT-RUN-2	0.625
GSI-RUN-1	0.690	SINAI_wd2v	0.619
GSI-RUN-2	0.679	SINAI_wd2v_2	0.613
GSI-RUN-3	0.678	BP-rbf-v1	0.602
DLSI-Run1	0.655	Lys-run-2	0.599
LyS-run-1	0.664	DT-RUN-3	0.608
GTI-GRAD-Run1	0.695	UCSP-RUN-1-NB	0.560
TID-spark-1	0.594	SINAI_w2v	0.604
INGEOTEC-M1	0.613	UCSP-RUN-1-DT	0.536
UCSP-RUN-2	0.594	CU-Run2-CompMod	0.481
UCSP-RUN-3 Ensemble	0.613	DT-RUN-1	0.490
exp2_3_SPARK	0.591	UCSP-RUN-2-ME	0.479
UCSP-RUN-1	0.602	SINAI_d2v	0.429
CU-RUN-1 Ensemble	0.597	GAS-UCR-1	0.446

Table 4: Results for task 1, 3 polarity levels, whole test corpus

Run ID	Acc	Run ID	Acc
LIF-Run-1	0.632	INGEOTEC-M1	0.595
ELiRF-run2	0.610	CU-RUN-1	0.600
LIF-Run-2	0.692	SINAI_wd2v_2_500	0.578
BP-wvoted-v1	0.632	UCSP-RUN-1	0.641
GSI-RUN-1	0.658	SINAI_w2v	0.582
GTI-GRAD-Run1	0.674	UCSP-RUN-3	0.627
BP-voted-v1	0.611	SINAI_wd2v	0.626
LyS-run-1	0.634	BP-rbf-v1	0.633
TID-spark-1	0.649	UCSP-RUN-1-NB	0.611
DLSI-Run1	0.637	UCSP-RUN-1-ME	0.636
ELiRF-run1	0.645	Lys-run-2	0.626
DT-RUN-1	0.601	DT-RUN-2	0.605
GSI-RUN-2	0.646	DT-RUN-3	0.583
GSI-RUN-3	0.647	UCSP-RUN-1-DR	0.571
ELiRF-run3	0.595	UCSP-RUN-2-NB	0.495
Ensemble exp3 1K 3	0.614	UCSP-RUN-2-ME	0.559
UCSP-RUN-2	0.586	DT-RUN-1	0.509
Ensemble exp2 1K 3	0.611	GAS-UCR-1	0.514
Ensemble exp1 1K 3	0.503	SINAI_d2v	0.510

Table 5: Results for task 1, 3 polarity levels, selected 1k test corpus

Since 2013 global level systems have developed different variants evolved to the present. Results have also improved, reaching values close to 0.70 of accuracy.

We have analyzed the results obtained with the 1k test corpus, and we try to answer the following questions: a) How many tweets are hard? (The ones not labeled correctly by any system), b) Are the polarities balanced?, and c) Are difficult cases from previous years solved?

Table 6 shows the number of tweets labeled correctly by the 14 groups, task 1, and five levels of polarity. Table 7 shows the statistical distribution of this 1k test set, according to the five levels plus the NONE label.

Correct	Total	%	Correct	Total	%
14	30	3,00%	**6**	59	5,90%
13	53	5,30%	**5**	57	5,70%
12	56	5,60%	**4**	60	6,00%
11	66	6,60%	**3**	74	7,40%
10	53	5,30%	**2**	104	10,40%
9	76	7,60%	**1**	102	10,20%
8	44	4,40%	**0**	109	10,90%
7	57	5,70%	**Total**	**1000**	**100%**

Table 6: Number of tweets labeled correctly, task 1, 5l

Correct	Total	%	Correct	Total	%
P	171	17,1%	**NONE**	121	12,1%
P+	284	28,4%	**NEU**	30	3,0%
N	201	20,1%	**0**	109	10,9%
N+	84	8,4%	**Total**	**1000**	**100%**

Table 7: Statistical distribution of the 1k test set, 5l + NONE

We can conclude that 1) 1k test set is almost balanced, 2) P+ and N are tweets easier to tag, 3) P and N+ are more difficult, 4) NONE values are detected by most systems and 5) NEU values are not detected.

The same analysis was made with three polarity labels, and the conclusions were the same.

We have analyzed the results obtained with hard cases and they are not solved yet. Some of them are hard because it is necessary more information about the user or a complete dialogue, not only an isolated word.

5.2 (legacy) Task 2: Aspect-based sentiment analysis

Tables 8 and 9 show the results obtained for task 2, in terms of Accuracy (Acc).

Run ID	Acc
GSI-RUN-1	0.635
GSI-RUN-2	0.621
GSI-RUN-3	0.557
ELiRF-run1	0.655
LyS-run-1	0.610
TID-spark-1	0.631
GSI-RUN-1	0.533
Lys-run-2	0.522

Table 8: Results for task 2, Social-TV corpus

Run ID	Acc
ELiRF-run1	0.633
LyS-run-1	0.599
Lys-run-2	0.540
TID-spark-1	0.557

Table 9: Results for task 2, STOMPOL corpus

6 *Conclusions and Future Work*

TASS has become a workshop relating to the detection of polarity in Spanish. The Spanish SA research community improves their systems every year, and this area receives great attraction from research groups and companies.

Each year the number of participants increase, as well as the number of different countries and the number of corpora downloads.

Again, the results obtained are comparable to those of the international community. Each year the number of unsupervised increases, and the natural tendency is to incorporate knowledge sources. The other issue is related to the fact that the systems submitted try to use the last methods in the state of the art, like classifiers based on deep learning.

The results obtained in past editions show that the improvement is not relevant, but the systems have checked different methods and resources.

The main purpose for future editions is to continue increasing the number of participants and the visibility of the workshop in international forums, including the participation of Latin American groups.

Acknowledgements

This work has been partially supported by a grant from the Fondo Europeo de Desarrollo Regional (FEDER), REDES project (TIN2015-65136-C2-1-R) and Ciudad2020 (INNPRONTA IPT-20111006) from the Spanish Government.

References

Villena-Román, J., Lana-Serrano, S., Martínez-Cámara, E., González-Cristobal, J. C. 2013. TASS - Workshop on Sentiment Analysis at SEPLN. *Revista de Procesamiento del Lenguaje Natural*, 50, pp. 37-44.

Villena-Román, J, García-Morera, J., Lana-Serrano, S., González-Cristobal, J. C. 2014. TASS 2013 - A Second Step in Reputation Analysis in Spanish. *Revista de Procesamiento del Lenguaje Natural*, 52, pp. 37-44.

Pang, B., Lee, L., and Vaithyanathan, S. 2002. Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing* Volume 10, pp. 79-86. Association for Computational Linguistics.

Turney. P. 2002. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pp. 417–424

Lluís-F. Hurtado, F. Pla and D. Buscaldi. 2015. ELiRF-UPV en TASS 2015: Análisis de Sentimientos en Twitter. In *Proceedings of TASS 2015: Workshop on Sentiment Analysis at SEPLN*, CEUR Series. Volume 1397, pp. 35-40

Saralegi, X., and I. San Vicente. 2013. "Elhuyar at tass 2013." *XXIX Congreso de la Sociedad Española de Procesamiento de lenguaje natural, Workshop on Sentiment Analysis at SEPLN (TASS2013)*

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

Hansen, L. K., Arvidsson, A., Nielsen, F. Å., Colleoni, E., and Etter, M. (2011). Good friends, bad news-affect and virality in twitter. In *Future information technology*, pp. 34-43. Springer Berlin Heidelberg.

Araque, O., I. Corcuera, C. Román, C. A. Iglesias, J. F. Sánchez-Rada. Aspect Based Sentiment Analysis of Spanish Tweets. In *Proceedings of TASS 2015: Workshop on Sentiment Analysis at SEPLN*, CEUR Series. Volume 1397, pp. 35-40

Mohammad, S. M., S. Kiritchenko, and X. Zhu. 2013. Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets. In *Second Joint Conference on Lexical and Computational Semantics*, Volume 2, pp. 321– 327.

Vilares, D., Y. Doval, M. Á. Alonso and C. Gómez-Rodríguez. 2015. LyS at TASS 2015: Deep Learning Experiments for Sentiment Analysis on Spanish Tweets. In *Proceedings of TASS 2015: Workshop on Sentiment Analysis at SEPLN*, CEUR Series. Volume 1397, pp. 47-52

Fernández, J. Gutiérrez, Y. Tomás, D., Gómez, José M., Martínez-Barco, Patricio. 2015. Evaluating a Sentiment Analysis Approach from a Business Point of View. In *Proceedings of TASS 2015: Workshop on Sentiment Analysis at SEPLN*, CEUR Series. Volume 1397, pp. 93-98

Álvarez-López, C.-C., H., T. Juncal-Martínez, J., Celix-Salgado, D., Fernández-Gavilanes, M., Costa-Montenegro, E. and González-Castaño, F. J. 2015. GTI-Gradiant at TASS 2015: A Hybrid Approach for Sentiment Analysis in Twitter. In *Proceedings of TASS 2015: Workshop on Sentiment Analysis at SEPLN*, CEUR Series. Volume 1397, pp. 35-40

Caro, L. Di and M. Grella. 2013. Sentiment analysis via dependency parsing. *Computer Standards and Interfaces*, 35(5): 442– 453.

Park, S. (2015). Sentiment Classification using Sociolinguistic Clusters. In *Proceedings of TASS 2015: Workshop on Sentiment Analysis at SEPLN*, CEUR Series. Volume 1397, pp. 99-104