TASS 2015, septiembre 2015, pp 13-21 recibido 20-07-15 revisado 24-07-15 aceptado 25-07-15

Overview of TASS 2015

Resumen de TASS 2015

Julio Villena Román
Janine García Morera
Daedalus, S.A.
28031 Madrid, Spain
{jvillena, jgarcia, cdepablo}@daedalus.es

Miguel Ángel García Cumbreras
Eugenio Martínez Cámara
M. Teresa Martín Valdivia
L. Alfonso Ureña López
Universidad de Jaén
23071 Jaén, Spain
{magc, emcamara, laurena, maite}@uja.es

Resumen: Este artículo describe la cuarta edición del taller de evaluación experimental TASS 2015, enmarcada dentro del congreso internacional SEPLN 2015. El principal objetivo de TASS es promover la investigación y el desarrollo de nuevos algoritmos, recursos y técnicas para el análisis de sentimientos en medios sociales (concretamente en Twitter), aplicado al idioma español. Este artículo describe las tareas propuestas en TASS 2015, así como el contenido de los corpus utilizados, los participantes en las distintas tareas y los resultados generales obtenidos y el análisis de estos resultados.
Palabras clave: TASS 2015, análisis de opiniones, medios sociales

Abstract: This paper describes TASS 2105, the fourth edition of the Workshop on Sentiment Analysis at SEPLN. The main objective is to promote the research and the development of new algorithms, resources and techniques in the field of sentiment analysis in social media (specifically Twitter), focused on Spanish language. This paper presents the TASS 2015 proposed tasks, the contents of the generated corpora, the participant groups and the results and analysis of them.
Keywords: TASS 2015, sentiment analysis, social media.

1 Introduction

TASS is an experimental evaluation workshop, 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. The fourth edition will be held on September 15th, 2015 at University of Alicante, Spain.

Sentiment analysis (SA) can be defined as the computational treatment of opinion, sentiment and subjectivity in texts (Pang & Lee, 2002). It is a hard task because even humans often disagree on the sentiment of a given text. And it is a harder task when the text has only 140 characters (Twitter messages or tweets).

Text classification techniques, although studied and improved for a longer time, still need more research effort and resources to be able to build better models to improve the current result values. Polarity classification has usually been tackled following two main approaches. The first one applies machine learning algorithms in order to train a polarity classifier using a labelled corpus (Pang et al. 2002). This approach is also known as the supervised approach. The second one is known as semantic orientation, or the unsupervised approach, and it integrates linguistic resources in a model in order to identify the valence of the opinions (Turney 2002).

The aim of TASS is to provide a competitive forum where the newest research works in the field of SA in social media, specifically focused on Spanish tweets, are showed and discussed by scientific and business communities.

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

TASS 2015 experiments are based on three corpus, specifically built for the different editions of the workshop.

2.1 General corpus

The general corpus contains over 68.000 tweets, written in Spanish, about 150 well-known personalities and celebrities of the world of politics, economy, communication, mass media and culture, between November 2011 and March 2012. Although the context of extraction has a Spain-focused bias, the diverse nationality of the authors, including people from Spain, Mexico, Colombia, Puerto Rico, USA and many other countries, makes the corpus reach a global coverage in the Spanish-speaking world.

Eachtweet includes its ID (tweetid), the creation date (date) and the user ID (user). Due to restrictions in the Twitter API Terms of Service (https://dev.twitter.com/terms/api-terms), it is forbidden to redistribute a corpus that includes text contents or information about users. However, it is valid if those fields are removed and instead IDs (including Tweet IDs and user IDs) are provided. The actual message content can be easily obtained by making queries to the Twitter API using the tweetid.

The general corpus has been divided into training set (about 10%) and test set (90%). The training set was released, so the participants could train and validate their models. The test corpus was provided without any tagging and has been used to evaluate the results. Obviously, it was not allowed to use the test data from previous years to train the systems.

Each tweet was tagged with its global polarity (positive, negative or neutral sentiment) or no sentiment at all. A set of 6 labels has been defined: strong positive (P+), positive (P), neutral (NEU), negative (N), strong negative (N+) and one additional no sentiment tag (NONE).

In addition, there is also an indication of the level of agreement or disagreement of the expressed sentiment within the content, with two possible values: AGREEMENT and DISAGREEMENT. This is especially 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.

Moreover, the 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 6 possible values and include the level of agreement as related to each entity.

This corpus is based on a selection of a set of topics. Thematic areas such as "política" ("politics"), "fútbol" ("soccer"), "literatura" ("literature") or "entretenimiento" ("entertainment"). Each tweet in both the training and test set has been assigned to one or several of these topics (most messages are associated to just one topic, due to the short length of the text).

All tagging has been done semiautomatically: a baseline machine learning model is first run and then all tags are manually checked by human experts. 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.

Table 1 shows a summary of the training and test corpora provided to participants.

Attribute	Value
Tweets	68.017
Tweets (test)	60.798 (89%)
Tweets (test)	7.219 (11%)
Topics	10
Users	154
Date start (train)	2011-12-02
Date end (train)	2012-04-10
Date start (test)	2011-12-02
Date end (test)	2012-04-10

Table 1: Corpus statistics

Users were journalists (*periodistas*), politicians (*políticos)* or celebrities (*famosos*). The only language involved this year was Spanish (*es)*.

The list of topics that have been selected is the following:

- Politics (política)
- Entertainment (entretenimiento)
- Economy (economía)
- Music (música)
- Soccer (fútbol)
- Films (películas)
- Technology (tecnología)
- Sports (deportes)
- Literature (literatura)
- Other (otros)

The corpus is encoded in XML. Figure 1 shows the information of two sample tweets. The first tweet is only tagged with the global polarity as the text contains no mentions to any entity, but the second one is tagged with both the global polarity of the message and the polarity associated to each of the entities that appear in the text (UPyD and Foro Asturias).

```
<tweet>
    <tweetid>0000000000</tweetid>
    <user>usuario0</user>
    <content>
        <![CDATA['Conozco a alguien q es adicto al drama! Ja ja ja te suena d algo!]]>
    </content>
    <date>2011-12-02T02:59:03</date>
    <lang>es</lang>
    <sentiments>
        <polarity>
            <value>P+</value>
            <type>AGREEMENT</type>
        </polarity>
    </sentiments>
    <topics>
        <topic>entretenimiento</topic>
    </topics>
</tweet>
<tweet>
    <tweetid>0000000001</tweetid>
    <user>usuario1</user>
    <content>
        <![CDATA['UPyD contará casi seguro con grupo gracias al Foro Asturias.]]>
    </content>
    <date>2011-12-02T00:21:01</date>
    <lang>es</lang>
    <sentiments>
        <polarity>
            <value>P</value>
            <type>AGREEMENT</type>
        </polarity>
        <polarity>
            <entity>UPyD</entity>
            <value>P</value>
            <type>AGREEMENT</type>
        </polarity>
        <polarity>
            <entity>Foro_Asturias</entity>
            <value>P</value>
            <type>AGREEMENT</type>
        </polarity>
    </sentiments>
  - <topics>
        <topic>política</topic>
    </topics>
</tweet>
```

Figure 1: Sample tweets (General corpus)

2.2 Social-TV corpus

The Social-TV corpus was collected during the 2014 Final of Copa del Rey championship in Spain between Real Madrid and F.C. Barcelona, played on 16 April 2014 at Mestalla Stadium in Valencia. After filtering useless information a subset of 2.773 tweets was selected.

All tweets were manually tagged with the aspects and its sentiment polarity. Tweets may cover more than one aspect.

The list of the 31 aspects that have been defined is the following:

- Afición (supporters)
- Árbitro (referee)
- Autoridades (authorities)
- Entrenador (coach)
- Equipo - Atlético de Madrid (Team- Atlético de Madrid)
- Equipo - Barcelona (Team- Barcelona)
- Equipo - Real Madrid (Team - Real Madrid)
- Equipo (any other team)
- Jugador - Alexis Sánchez (Player - Alexis Sánchez)
- Jugador - Álvaro Arbeloa (Player - Álvaro Arbeloa)
- Jugador - Andrés Iniesta (Player - Andrés Iniesta)
- Jugador - Ángel Di María (Player - Ángel Di Maria)
- Jugador - Asier Ilarramendi (Player - Asier Ilarramendi)
- Jugador - Carles Puyol (Player - Carles Puyol)
- Jugador - Cesc Fábregas (Player - Cesc Fábregas)
- Jugador - Cristiano Ronaldo (Player - Cristiano Ronaldo)
- Jugador - Dani Alves (Player - Dani Alves)
- Jugador - Dani Carvajal (Player - Dani Carvajal)
- Jugador - Fábio Coentrão (Player - Fábio Coentrão)
- Jugador - Gareth Bale (Player - Gareth Bale)
- Jugador - Iker Casillas (Player - Iker Casillas)
- Jugador - Isco (Player - Isco)
- Jugador - Javier Mascherano (Player - Javier Mascherano)
- Jugador - Jesé Rodríguez (Player - Jesé Rodríguez)
- Jugador - José Manuel Pinto (Player - José Manuel Pinto)
- Jugador - Karim Benzema (Player - Karim Benzema)
- Jugador - Lionel Messi (Player - Lionel Messi)
- Jugador - Luka Modric (Player - Luka Modric)
- Jugador - Marc Bartra (Player - Marc Bartra)
- Jugador - Neymar Jr. (Player - Neymar Jr.)
- Jugador - Pedro Rodríguez (Player - Pedro Rodríguez)
- Jugador - Pepe (Player - Pepe)
- Jugador - Sergio Busquets (Player - Sergio Busquets)
- Jugador - Sergio Ramos (Player - Sergio Ramos)

- Jugador - Xabi Alonso (Player - Xabi Alonso)
- Jugador - Xavi Hernández (Player - Xavi Hernández)
- Jugador (any other player)
- Partido (Football match)
- Retransmisión (broadcast)

Sentiment polarity has been tagged from the point of view of the person who writes the tweet, using 3 levels: P, NEU and N. No distinction is made in cases when the author does not express any sentiment or when he/she expresses a no-positive no-negative sentiment.

The Social-TV corpus was randomly divided into training set (1.773 tweets) and test set (1.000 tweets), with a similar distribution of both aspects and sentiments. The training set was released previously and the test corpus was provided without any tagging and has been used to evaluate the results provided by the different systems.

The following figure shows the information of three sample tweets in the training set.

```
<tweet id="456544898791907328">
  <sentiment aspect="Equipo-Real_Madrid" polarity="P">#HalaMadrid
  </sentiment> ganamos sin <sentiment aspect="Jugador-Cristiano_Ronaldo"
   polarity="NEU">Cristiano</sentiment>. .perdéis con <sentiment aspect=
  "Jugador-Lionel_Messi" polarity="N">Messi</sentiment>. Hala <sentiment
   aspect="Equipo-Real_Madrid" polarity="P">Madrid</sentiment>! !!!!!
</tweet>
<tweet id="456544898942906369">
  @nevermind2192 <sentiment aspect="Equipo-Barcelona" polarity="P">Barça
  </sentiment> por siempre!!
</tweet>
<tweet id="456544898951282688">
  <sentiment aspect="Partido" polarity="NEU">#FinalCopa</sentiment>
  Hala <sentiment aspect="Equipo-Real_Madrid" polarity="P">Madrid
  </sentiment>, hala <sentiment aspect="Equipo-Real_Madrid" polarity="P"
  >Madrid</sentiment>, campeón de la <sentiment aspect="Partido"
  polarity="P">copa del rey</sentiment>
</tweet>
```

Figure 2: Sample tweets (Social-TV corpus)

2.3 STOMPOL corpus

STOMPOL (corpus of Spanish Tweets for Opinion Mining at aspect level about POLitics) is a corpus of Spanish tweets prepared for the research in the challenging task of opinion mining at aspect level. The tweets were gathered from 23rd to 24th of April 2015, and are related to one of the following political aspects that appear in political campaigns:

- Economics (Economía): taxes, infrastructure, markets, labor policy...
- Health System (Sanidad): hospitals, public/private health system, drugs, doctors...
- Education (Educacion): state school, private school, scholarships...
- Political party (Propio_partido): anything good (speeches, electoral programme...) or bad (corruption, criticism) related to the entity
- Otros_aspectos (Other aspects): electoral system, environmental policy...

Each aspect is related to one or several entities that correspond to one of the main political parties in Spain, which are:

- Partido_Popular (PP)
- Partido_Socialista_Obrero_Español (PSOE)
- Izquierda_Unida (IU)
- Podemos
- Ciudadanos (Cs)
- Unión_Progreso_y_Democracia (UPyD)

Each tweet in the corpus has been manually tagged by two annotators, and a third one in case of disagreement, with the sentiment polarity at aspect level. Sentiment polarity has been tagged from the point of view of the person who writes the tweet, using 3 levels: P, NEU and N. 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 3 shows the information of two sample tweets.

```
<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 3: Sample tweets (STOMPOL corpus)

These three corpora will be made freely available to the community after the workshop. Please send an email to tass@daedalus.es filling in the TASS Corpus License agreement with your email, affiliation (institution, company or any kind of organization) and a brief description of your research objectives, and you will be given a password to download the files in the password protected area. The only requirement is to include a citation to a relevant paper and/or the TASS website.

3 Description of tasks

First of all, we are interested in evaluating the evolution of the different approaches for SA and text classification in Spanish during these years. So, the traditional SA at global level task will be repeated again, reusing the same corpus,

to compare results. Moreover, we want 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 will be repeated again, to compare results, and a new corpus has been created for the second task.

Participants are expected to submit up to 3 results of different experiments for one or both of these tasks, in the appropriate format described below.

Along with the submission of experiments, participants have been invited to submit a paper to the workshop in order to describe their experiments and discussing the results with the audience in a regular workshop session.

The two proposed tasks are described next.

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

This is the same task as previous editions. This task consists on performing an automatic polarity classification to determine the global polarity of each message in the test set of the General corpus. Participants have been provided with the training set of the General corpus so that they may train and validate their models. There will be 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`).

Participants are expected to submit (up to 3) experiments for the 6-labels evaluation, but are also allowed to submit (up to 3) specific experiments for the 4-labels scenario.

Results must be submitted in a plain text file with the following format:

```
tweetid \t polarity
```

where polarity can be:

- `P+`, `P`, `NEU`, `N`, `N+` and `NONE` for the 6-labels case
- `P`, `NEU`, `N` and `NONE` for the 4-labels case.

The same test corpus of previous years will be used for the evaluation, to allow for comparison among systems. Accuracy, macroaveraged precision, macroaveraged recall and macroaveraged F1-measure have been used to evaluate each run.

Notice that there are two test sets: the complete set and 1k set, a subset of the first one. The reason is that, to deal with the problem of the imbalanced distribution of labels between the training and test set, a selected test subset containing 1.000 tweets with a similar distribution to the training corpus was extracted to be used for an alternate evaluation of the performance of systems.

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

Participants have been provided with a corpus tagged with a series of aspects, and systems must identify the polarity at the aspect-level. Two corpora have been provided: the Social-TV corpus, used in TASS 2014, and the new STOMPOL corpus, collected in 2015 (described above). Both corpora have been splitted into training and test set, the first one for building and validating the systems, and the second for evaluation.

Participants are expected to submit up to 3 experiments for each corpus, each in a plain text file with the following format:

```
tweetid \t aspect \t polarity
```

[for the Social-TV corpus]

```
tweetid \t aspect-entity \t polarity
```

[for the STOMPOL corpus]

Allowed polarity values are P, N and NEU.

For evaluation, a single label combining "aspect-polarity" has been considered. Similarly to the first task, accuracy, macroaveraged precision, macroaveraged recall and macroaveraged F1-measure have been calculated for the global result.

4 Participants and Results

This year 35 groups registered (as compared to 31 groups last year) but unfortunately only 7 groups (14 last year) sent their submissions. The list of active participant groups is shown in Table 2, including the tasks in which they have participated.

Fourteen of the seventeen participant groups sent a report describing their experiments and results achieved. Papers were reviewed and included in the workshop proceedings. References are listed in Table 3.

Group	1	2
LIF	X	
ELiRF	X	X
GSI	X	X
LyS	X	X
DLSI	X	
GTI-GRAD	X	
ITAINNOVA	X	
SINAI-ESMA	X	
CU	X	
TID-spark	X	X
BittenPotato	X	
SINAI_wd2v	X	
DT	X	
GAS-UCR	X	
UCSP	X	
SEDEMO	X	
INGEOTEC	X	
Total groups	17	4

Table 2: Participant groups

Group	Report
ELiRF	ELiRF-UPV en TASS 2015: Análisis de Sentimientos en Twitter
GSI	Aspect based Sentiment Analysis of Spanish Tweets
LyS	LyS at TASS 2015: Deep Learning Experiments for Sentiment Analysis on Spanish Tweets
DLSI	Evaluating a Sentiment Analysis Approach from a Business Point of View
GTI-GRAD	GTI-Gradiant at TASS 2015: A Hybrid Approach for Sentiment Analysis in Twitter
ITAINNOVA	Ensemble algorithm with syntactical tree features to improve the opinion analysis
SINAI-EMMA	SINAI-EMMA: Vectores de Palabras para el Análisis de Opiniones en Twitter
CU	Spanish Twitter Messages Polarized through the Lens of an English System
TID-spark	Sentiment Classification using Sociolinguistic Clusters
BittenPotato	BittenPotato: Tweet sentiment analysis by combining multiple classifiers
SINAI_wd2v	Participación de SINAI DW2Vec en TASS 2015
DT	DeustoTech Internet at TASS 2015: Sentiment analysis and polarity classification in Spanish tweets
UCSP	Comparing Supervised Learning Methods for Classifying Spanish Tweets
INGEOTEC	Sentiment Analysis for Twitter: TASS 2015

Table 3: Participant reports

5 Results

Results for each task are described next.

5.1 Task 1: Sentiment Analysis at Global Level

Submitted runs and results for Task 1, evaluation based on 5 polarity levels with the whole General test corpus, are shown in Table 4. Accuracy, macroaveraged precision, macroaveraged recall and macroaveraged F1-measure have been used to evaluate each individual label and ranking the systems.

Run Id	Acc
LIF-Run-3	0.672
LIF-Run-2	0.654
ELiRF-run3	0.659
LIF-Run-1	0.628
ELiRF-run1	0.648
ELiRF-run2	0.658
GSI-RUN-1	0.618
run_out_of_date	0.673
GSI-RUN-2	0.610
GSI-RUN-3	0.608
LyS-run-1	0.552
DLSI-Run1	0.595
Lys-run-2	0.568
GTI-GRAD-Run1	0.592
Ensemble exp1.1	0.535
SINAI-EMMA-1	0.502
INGEOTEC-M1	0.488
Ensemble exp3_emotions	0.549
CU-Run-1	0.495
TID-spark-1	0.462
BP-wvoted-v2_1	0.534
Ensemble exp2_emotions	0.524

BP-voted-v2	0.535
SINAI_wd2v_500	0.474
SINAI_wd2v_300	0.474
BP-wvoted-v1	0.522
BP-voted-v1	0.522
BP-rbf-v2	0.514
Lys-run-3	0.505
BP-rbf-v1	0.494
CU-Run-2-CompMod	0.362
DT-RUN-1	0.560
DT-RUN-3	0.557
DT-RUN-2	0.545
GAS-UCR-1	0.342
UCSP-RUN-1	0.273
BP-wvoted-v2	0.009

Table 4: Results for Task 1, 5 levels, whole test corpus

As previously described, an alternate evaluation of the performance of systems was done using a new selected test subset containing 1.000 tweets with a similar distribution to the training corpus. Results are shown in Table 5.

In order to perform a more in-depth evaluation, results are calculated considering the classification only in 3 levels (POS, NEU, NEG) and no sentiment (NONE) merging P and P+ in only one category, as well as N and N+ in another one. The same double evaluation using the whole test corpus and a new selected corpus have been carried out, shown Tables 8 and 9.

Run Id	Acc
ELiRF-run2	0.488
GTI-GRAD-Run1	0.509
LIF-Run-2	0.516
GSI-RUN-1	0.487
GSI-RUN-2	0.48
GSI-RUN-3	0.479
LIF-Run-1	0.481
ELiRF-run1	0.476
SINAI_wd2v	0.389
ELiRF-run3	0.477
INGEOTEC-M1	0.431
Ensemble exp1 1K	0.405
LyS-run-1	0.428
Ensemble exp2 1K	0.384
Lys-run-3	0.430
Lys-run-2	0.434
SINAI-EMMA-1	0.411
CU-Run-1-CompMod	0.419
Ensemble exp3 1K	0.396
TID	0.400
BP-voted-v1	0.408
DLSI-Run1	0.385
CU-Run-2	0.397
BP-wvoted-v1	0.416
BP-rbf-v1	0.418
SEDEMO-E1	0.397
DT-RUN-1	0.407
DT-RUN-2	0.408
DT-RUN-3	0.396
GAS-UCR-1	0.338
INGEOTEC-E1	0.174
INGEOTEC-E2	0.168

Table 5: Results for Task 1, 5 levels, selected 1k corpus

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

Table 6: Results for Task 1, 3 levels, whole test corpus

Run Id	Acc

LIF-Run-1	0.632
ELiRF-run2	0.610
LIF-Run-2	0.692
BP-wvoted-v1	0.632
GSI-RUN-1	0.658
GTI-GRAD-Run1	0.674
BP-voted-v1	0.611
LyS-run-1	0.634
TID-spark-1	0.649
DLSI-Run1	0.637
ELiRF-run1	0.645
DT-RUN-1	0.601
GSI-RUN-2	0.646
GSI-RUN-3	0.647
ELiRF-run3	0.595
Ensemble exp3 1K 3	0.614
UCSP-RUN-2	0.586
Ensemble exp2 1K 3	0.611
Ensemble exp1 1K 3	0.503
INGEOTEC-M1	0.595
CU-Run-2-CompMod	0.600
CU-RUN-1	0.578
SINAI_wd2v_2_500	0.641
UCSP-RUN-1	0.582
SINAI_w2v	0.627
UCSP-RUN-3	0.626
SINAI_wd2v	0.633
BP-rbf-v1	0.611
UCSP-RUN-1-NB	0.636
UCSP-RUN-1-ME	0.626
Lys-run-2	0.605
DT-RUN-2	0.583
DT-RUN-3	0.571
UCSP-RUN-1-DR	0.495
UCSP-RUN-2-NB	0.559
UCSP-RUN-2-ME	0.509
DT-RUN-1	0.514
GAS-UCR-1	0.556
SINAI_d2v	0.510

Table 7: Results for Task 1, 3 levels, selected 1k corpus

5.2 Task 2: Aspect-based Sentiment Analysis

Submitted runs and results for Task 2, with the Social-TV and STOMPOL corpus, are shown in Tables 10 and 11. Accuracy, macroaveraged precision, macroaveraged recall and macroaveraged F1-measure have been used to evaluate each individual label and ranking the systems.

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 10: 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 11: Results for Task 2, STOMPOL corpus

6 Conclusions and Future Work

TASS was the first workshop about SA focused on the processing of texts written in Spanish. Clearly this area receives great attraction from research groups and companies, as this fourth edition has had a greater impact in terms of registered groups, and the number of participants that submitted experiments in 2015 tasks has increased.

Anyway, the developed corpus and gold standards, and the reports from participants will for sure be helpful for other research groups approaching these tasks.

TASS corpora will be released after the workshop for free use by the research community. In 2014 the corpora had been downloaded up to date by more than 60 research groups, 25 out of Spain, by groups coming from academia and also from private companies to use the corpus as part of their product development. We expect to reach a similar impact with this year's corpus.

Acknowledgements

This work has been partially supported by a grant from the Fondo Europeo of Desarrollo Regional (FEDER), ATTOS (TIN2012-38536-C03-0) and Ciudad2020 (INNPRONTA IPT-20111006) projects from the Spanish Government, and AORESCU project (P11-TIC-7684 MO).

References

Villena-Román, Julio; Lana-Serrano, Sara; Martínez-Cámara, Eugenio; González-Cristobal, José Carlos. 2013. *TASS - Workshop on Sentiment Analysis at SEPLN*. Revista de Procesamiento del Lenguaje Natural, 50, pp 37-44. http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/4657.

Villena-Román, Julio; García-Morera, Janine; Lana-Serrano, Sara; González-Cristobal, José Carlos. 2014. *TASS 2013 - A Second Step in Reputation Analysis in Spanish*. Revista de Procesamiento del Lenguaje Natural, 52, pp 37-44. http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/4901.

Vilares, David; Doval, Yerai; Alonso, Miguel A.; Gómez-Rodríguez, Carlos. *LyS at TASS 2014: A Prototype for Extracting and Analysing Aspects from Spanish tweets*. In Proc. of the TASS workshop at SEPLN 2014. 16-19 September 2014, Girona, Spain.

Perea-Ortega, José M. Balahur, Alexandra. *Experiments on feature replacements for polarity classification of Spanish tweets*. In Proc. of the TASS workshop at SEPLN 2014. 16-19 September 2014, Girona, Spain.

Hernández Petlachi, Roberto; Li, Xiaoou. *Análisis de sentimiento sobre textos en Español basado en aproximaciones semánticas con reglas lingüísticas*. In Proc. of the TASS workshop at SEPLN 2014. 16-19 September 2014, Girona, Spain.

Montejo-Ráez, A.; García-Cumbreras, M.A.; Díaz-Galiano, M.C. *Participación de SINAI Word2Vec en TASS 2014*. In Proc. of the TASS workshop at SEPLN 2014. 16-19 September 2014, Girona, Spain.

Hurtado, Lluís F.; Pla, Ferran. *ELiRF-UPV en TASS 2014: Análisis de Sentimientos, Detección de Tópicos y Análisis de Sentimientos de Aspectos en Twitter*. In Proc. of the TASS workshop at SEPLN 2014. 16-19 September 2014, Girona, Spain.

Jiménez Zafra, Salud María; Martínez Cámara, Eugenio; Martín Valdivia, M. Teresa.; Ureña López, L. Alfonso. *SINAI-ESMA: An unsupervised approach for Sentiment Analysis in Twitter*. In Proc. of the TASS workshop at SEPLN 2014. 16-19 September 2014, Girona, Spain.

San Vicente Roncal, Iñaki; Saralegi Urizar, Xabier. *Looking for Features for Supervised Tweet Polarity Classification*. In Proc. of the TASS workshop at SEPLN 2014. 16-19 September 2014, Girona, Spain.