Procesamiento del Lenguaje Natural, Revista nº 50 marzo de 2013, pp 77-84 recibido 01-12-12 revisado 14-02-13 aceptado 19-02-13

Sentiment Analysis and Topic Classification based on Binary Maximum Entropy Classifiers

Análisis de Sentimiento y Categorización de texto basado en clasificadores binarios de máxima entropía

Fernando Batista, Ricardo Ribeiro

Laboratório de Sistemas de Língua Falada (L2F) - INESC-ID Lisboa
R. Alves Redol, 9, 1000-029 Lisboa, Portugal

Instituto Universitário de Lisboa (ISCTE-IUL)
Av. Forças Armadas, 1649-026 Lisboa, Portugal

{Fernando.Batista, Ricardo.Ribeiro}@inesc-id.pt

Resumen: En este trabajo se presenta una estrategia basada en clasificadores binarios de máxima entropía para el análisis de sentimiento y categorización de textos de Twitter enfocados al español. El sistema desarrollado consigue los mejores resultados para la categorización temática, y el segundo lugar para el análisis de sentimiento, en un esfuerzo de evaluación conjunta (Villena-Román et al., 2012). Se han explorado diferentes configuraciones para ambas tareas. Esto llevó a la utilización de una cascada de clasificadores binarios para el análisis de sentimiento y una estrategia de tipo uno-vs-todo para la clasificación de tema, donde los temas más probables para cada tweet fueron seleccionados.
Palabras clave: Análisis de sentimiento, categorización de texto en temas de interés, medios sociales, regresión logística, máxima entropía.

Abstract: This paper presents a strategy based on binary maximum entropy classifiers for automatic sentiment analysis and topic classification over Spanish Twitter data. The developed system achieved the best results for topic classification, and the second place for sentiment analysis in a joint evaluation effort – the TASS challenge (Villena-Román et al., 2012). Different configurations have been explored for both tasks, leading to the use of a cascade of binary classifiers for sentiment analysis and a one-vs-all strategy for topic classification, where the most probable topics for each tweet were selected.
Keywords: Sentiment analysis, topic detection, social media, logistic regression, maximum entropy.

1 *Introduction*

Social Networks take part in the nowadays life of a large number of people, providing revolutionary means for people to communicate and interact. Each social network targets different audiences, offering a unique range of services that people find useful in the course of their lives. Twitter offers a simple way for people to express themselves, by means of small text messages of at most 140 characters that can be freely used.

Twitter can be accessed in numerous ways, ranging from computers to mobile phones or other mobile devices. That is particularly important because accessing and producing content becomes a trivial task, therefore assuming an important part of people's lives. One relevant aspect that differentiates Twitter from other communication means is its ability to rapidly propagate such content and make it available to specific communities, selected based on their interests. Twitter data is a powerful source of information for assessing and predicting large-scale facts. For example, (O'Connor et al., 2010) capture large-scale trends on consumer confidence and political opinion in tweets, strengthening the potential of such data as a supplement for traditional polling. In what concerns stock markets, (Bollen, Mao, and Zeng, 2010) found that Twitter data can be used to significantly improve stock market predictions accuracy.

The huge amount of data, constantly being produced in a daily basis, makes it impracticable to manually process such content. For that reason, it becomes urgent to apply automatic

ISSN 1135-5948

processing strategies that can handle, and take advantage, of such amount of data. However, processing Twitter is all but an easy task, not only because of specific phenomena that can be found in the data, but also because it may require to process a continuous stream of data, and possibly to store some of the data in a way that it can be accessed in the future.

This paper tackles two well-known Natural Language Processing (NLP) tasks, commonly applied both to written and speech corpora: sentiment analysis and topic detection. The two tasks have been performed over Spanish Twitter data provided in the context of a contest proposed by "TASS – workshop on Sentiment Analysis" (Villena-Román et al., 2012), a satellite event of the SEPLN 2012 conference.

The paper is organized as follows: Section 2 overviews the related work, previously done for each task. Section 3 presents a brief description of the data. Section 4 describes the most relevant strategies that have been considered for tackling the problem. Section 5 presents and analyses a number of experiments, and reports the results for each one of the approaches. Section 6 presents some conclusions and discusses the future work.

2 *Related work*

Sentiment analysis and topic detection are two well-known NLP (Natural Language Processing) tasks. Sentiment analysis is often referred by other names (e.g. sentiment mining) and consists of assigning a sentiment, from a set of possible values, to a given portion of text. Topic detection consists of assigning a class (or topic) from a set of possible predefined classes to a given document. Often, these two tasks are viewed as two classification problems that, despite being characterized by their specificities, can be tackled using similar strategies. The remainder of this section overviews the related work previously done concerning these two tasks.

2.1 Sentiment analysis

Sentiment analysis can be performed at different complexity levels, where the most basic one consists just on deciding whether a portion of text contains a positive or a negative sentiment. However, it can be performed at more complex levels, like ranking the attitude into a set of more than two classes or, even further, it can be performed in a way that different complex attitude types can be determined, as well as finding the source and the target of such attitudes.

Dealing with the huge amounts of data available on Twitter demand clever strategies. One interesting idea, explored by (Go, Bhayani, and Huang, 2009) consists of using emoticons, abundantly available on tweets, to automatically label the data and then use such data to train machine learning algorithms. The paper shows that machine learning algorithms trained with such approach achieve above 80% accuracy, when classifying messages as positive or negative. A similar idea was previously explored by (Pang, Lee, and Vaithyanathan, 2002) for movie reviews, by using star ratings as polarity signals in their training data. This latter paper analyses the performance of different classifiers on movie reviews, and presents a number of techniques that were used by many authors and served as baseline for posterior studies. As an example, they have adapted a technique, introduced by (Das and Chen, 2001), for modeling the contextual effect of negation, adding the prefix `NOT_` to every word between a "negation word" and the first punctuation mark following the negation word.

Common approaches to sentiment analysis involve the use of sentiment lexicons of positive and negative words or expressions. The General Inquirer (Stone et al., 1966) was one of the first available sentiment lexicons freely available for research, which includes several categories of words, such as: positive vs. negative, strong vs. week. Two other examples include (Hu and Liu, 2004), an opinion lexicon containing about 7000 words, and the MPQA Subjectivity Cues Lexicon (Wilson, Wiebe, and Hoffmann, 2005), where words are annotated not only as positive vs. negative, but also with intensity. Finally, (Baccianella, Esuli, and Sebastiani, 2010) is another available resource that assigns sentiment scores to each synset of the wordnet.

Learning polarity lexicons is another research approach that can be specially useful for dealing with large corpora. The process starts with a seed set of words and the idea is to increasingly find words or phrases with similar polarity, in semi-supervised fashion (Turney, 2002). The final lexicon contains much more words, possibly learning domain-specific information, and therefore is more prone to be robust. The work reported by (Kim and Hovy, 2004) is another example of learning algorithm that uses WordNet synonyms and antonyms to learn polarity.

2.2 Topic Detection

Work on Topic Detection has its origins in 1996 with the Topic Detection and Tracking (TDT) initiative sponsored by the US government (Allan, 2002). The main motivation for this initiative was the processing of the large amounts of information coming from newswire and broadcast news. The main goal was to organize the information in terms of events and stories that discussed them. The concept of topic was defined as the set of stories about a particular event. Five tasks were defined: story segmentation, first story detection, cluster detection, tracking, and story link detection. The current impact and the amount information generated by social media led to a state of affairs similar to the one that fostered the pioneer work on TDT. Social media is now the context for research tasks like topic (cluster) detection (Lee et al., 2011; Lin et al., 2012) or emerging topic (first story) detection (Kasiviswanathan et al., 2011).

In that sense, closer to our work are the approaches described by (Sriram et al., 2010) and (Lee et al., 2011), where tweets are classified into previously defined sets of generic topics. In the former, a conventional bag-of-words (BOW) strategy is compared to a specific set of features (authorship and the presence of several types of twitter-related phenomena) using a Naïve Bayes (NB) classifier to classify tweets into the following generic categories: News, Events, Opinions, Deals, and Private Messages. Findings show that authorship is a quite important feature. In the latter, two strategies, BOW and network-based classification, are explored to classify clusters of tweets into 18 general categories, like Sports, Politics, or Technology. In the BOW approach, the clusters of tweets are represented by tf-idf vectors and NB, NB Multinomial, and Support Vector Machines (SVM) classifiers are used to perform classification. The network-based classification approach is based on the links between users and C5.0 decision tree, k-Nearest Neighbor, SVM, and Logistic Regression classifiers were used. Network-based classification was shown to achieve a better performance, but being link-based, it cannot be used for all situations.

3 Data

Experiments described in this paper use Spanish Twitter data provided in the context of the TASS contest (Villena-Román et al., 2012). The provided training data consists of an XML file containing about 7200 tweets, each one labelled with sentiment polarity and the corresponding topics. We decided to consider the first 80% of the data for training our models (5755 tweets) and the remaining 20% for development (1444 tweets). The provided test data is also available in XML and contains about 60800 unlabeled tweets. The goal consists in providing automatic sentiment and topic classification for that data.

Each tweet in the labelled data is annotated in terms of polarity, using one of six possible values: *NONE*, *N*, *N+*, *NEU*, *P*, *P+* (Section 4.2 contains information about their meaning). Moreover, each annotation is also marked as *AGREEMENT* or *DISAGREEMENT*, indicating whether all the annotators performed the annotation coherently. In what concerns topic detection, each tweet was annotated with one or more topics, from a list of 10 possible topics: *política* (politics), *otros* (others), *entretenimiento* (entertainment), *economía* (economics), *música* (music), *fútbol* (football), *cine* (movies), *tecnología* (technology), *deportes* (sports), and *literatura* (literature).

It is also important to mention that, besides the tweets, an extra XML file is also available, containing information about each one of the users that authored at least one of the tweets in the data. In particular, the information includes the type of user, assuming one of three possible values – *periodista* (journalist), *famoso* (famous person), and *politico* (politician) – which may provide valuable information for these tasks.

Apart from the provided data, some experiments described in this paper also made use of Sentiment Lexicons in Spanish[1], a resource created at the University of North Texas (Perez-Rosas, Banea, and Mihalcea, 2012). From this resource, only the most robust part was used, known as *fullStrengthLexicon*, and containing 1346 words automatically labelled with sentiment polarity.

4 Approach

We have decided to consider both tasks as classification tasks, thus sharing the same method. The most successful and recent experiments cast the problem as a binary classification problem, which aims at discriminating between two possible classes. Binary classifiers are easier to develop, offer faster convergence ratios, and can be executed in parallel. The final results are then produced by combining all the different binary classifiers.

The remainder of this section describes the method and the architecture of the system when applied to each one of the tasks.

[1] http://lit.csci.unt.edu/

4.1 Maximum Entropy models

We have adopted an approach based on logistic regression classification models, which corresponds to the maximum entropy classification for independent events, firstly applied to natural language problems in (Berger, Pietra, and Pietra, 1996). This approach provides a clean way of expressing and combining different aspects of the information, and naturally implements feature selection. That is specially useful for twitter data, in which a large number of sparse features are used. A ME model estimates the conditional probability of the events given the corresponding features. Let us consider the random variable $y \in C$ that can take k different values, corresponding to the classes c_1, c_2, ... ,c_k. The ME model is given by the following equation:

$$P(c|d) = \frac{1}{Z_\lambda(F)} \times exp\left(\sum_i \lambda_{ci} f_i(c,d)\right) \quad (1)$$

determined by the requirement that $\sum_{c \in C} P(c|d)$=1. $Z_\lambda(F)$ is a normalizing term, used just to make the exponential a true probability, and is given by:

$$Z_\lambda(F) = \sum_{c' \in C} exp\left(\sum_i \lambda_{c'i} f_i(c',d)\right) \quad (2)$$

f_i are feature functions corresponding to features defined over events, and $f_i(c,d)$ is the feature defined for a class c and a given observation d. The index i indicates different features, each of which has associated weights λ_{ci}, one for each class. The ME model is estimated by finding the parameters λ_{ci} with the constraint that the expected values of the various feature functions match the averages in the training data. These parameters ensure the maximum entropy of the distribution and also maximize the conditional likelihood $\prod_i P(y^{(i)}|d^{(i)})$ of the training samples. Decoding is conducted for each sample individually and the classification is straightforward, making it interesting for on-the-fly usage. ME is a probabilistic classifier, a generalization of Boolean classification, that provides probability distributions over the classes. The single-best class corresponds to the class with the highest probability, and is given by:

$$\hat{c} = \underset{c \in C}{argmax} P(c|d) \quad (3)$$

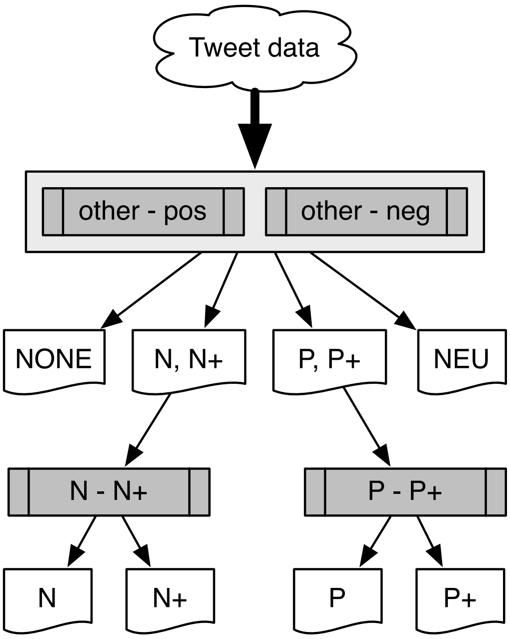

Figure 1: Approach for sentiment analysis.

The ME models used in this study are trained using the MegaM tool (Daumé III, 2004), which uses an efficient implementation of conjugate gradient (for binary problems).

4.2 Sentiment analysis

As previously mentioned, the sentiment classification considers 6 possible classes: *N*, *N*+ → negative polarity; *P*, *P*+ → positive polarity; *NEU* → contains both positive and negative sentiments; *NONE* → without polarity information. The plus sign (+) signals the sentiment intensity.

The first interesting results were achieved by combining 5 different binary classifiers, one for each class. A first classifier <`NONE,other`> was used to discriminate between NONE and all the other classes. Two other classifiers <`other,neg`> and <`other,pos`> were applied after the first classifier for detecting negative and positive sentiments, respectively. These two latest classifiers make it possible to distinguish between three classes: *Positive*, *Negative*, and *Neutral*. These three classifiers, one can now discriminate between four classes: *NONE*, *Negative*, *Positive* and *Neutral*. Finally, two other classifiers: <`N,N+`> and <`P,P+`>, allow perceiving the sentiment intensity. Only tweets annotated as *N* and *N*+ were used for training the <`N,N+`> classifier, and only tweets marked as *P* or *P*+ were used for training the second. That is different from the first three classifiers, which have used all the available data for training.

After some other experiments, we observed

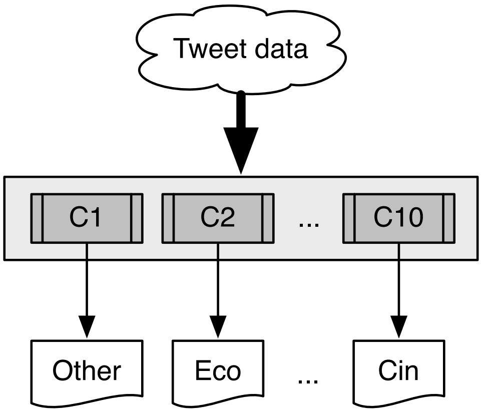

Figure 2: Approach for topic classification.

that similar results can be achieved by using the second and third classifiers to also indicate if no sentiment was present and then eliminating the need of the first classifier. The idea is that the classifiers <`other,N`> and <`other,P`> can, in fact, discriminate between four classes, by considering the class NONE whenever both return "other". Figure 1 illustrates the resulting configuration, where only four binary classifiers are used in a cascade fashion.

4.3 Topic classification

Figure 2 illustrates the classification process, where 10 distinct binary classifiers have been used, one for each topic. Each classifier selects its corresponding topic, which may lead to zero, one, or several topics. The number of selected topics have not been limited to a maximum, but when no topic is selected, the most probable topic is chosen based on the available classification probabilities.

5 Experiments

This section describes the steps taken, the features that have been used, and experiments that have been conducted using the previously described approaches.

5.1 Tweet content pre-processing

The content of each tweet was firstly tokenized using *twokenize*, a tokenization tool for English tweets[2], with some minor modifications for dealing with Spanish data instead of English.

5.2 Features

The following features, concerning the tweet text, were used for each tweet:

- Punctuation marks.
- Words occurring after the words "*nunca*" (never) or "*no*" (no) were prefixed by "`NO_`" until reaching some punctuation mark or until reaching the end of the tweet content (Pang, Lee, and Vaithyanathan, 2002).
- Each token starting with "http:" was converted into the token "HTTP", and it's weight as a feature was reduced.
- All tokens starting with "#" were expanded into two features: one with "#", and other without it. A lesser weight was given to the stripped version of the token.
- All tokens starting with "@" were used as feature, but the feature "@USER" was introduced as well, with a smaller weight.
- All words containing more than 3 repeating letters were also used. Whenever such words occur, two more features are produced: "LONG_WORD" with a lower weight, and the corresponding word without repetitions with a high weight (3 times the standard weight).
- All cased words were used, but the corresponding lowercase words were used as well. Uppercase words were assigned also to a higher weight, since they are often used for emphasis.

Apart from the features extracted from the text, two more features were used:

- *Username* of the author of the tweet.
- *Usertype*, corresponding to the user classification, according to `users-info.xml`.

Most of the previously described features were used both for sentiment analysis and for topic detection. Some of them were combined as bigrams for some experiments. Feature bigrams involve the following tokens: *HTTP*, words starting with # without the diacritic #, *@USER*, *LONG_WORD*, all other words converted to lowercase.

5.3 Results for sentiment analysis

Our experiments for sentiment analysis consider 6 possible classes, as described in Section 4.2. The Initial experiments achieved 52.5 Acc (Accuracy) in the development set, using all previously described features except punctuation, tweet's author name, and the *user type*. This baseline result was then further improved to 53.6 Acc [+1.1] by using the tweet's author name, and by adding the *user type* it was further improved to

[2] By Brendan O'Connor (brenocon@gmail.com)

	development	test
Unigrams only	55.2	63.4
Unigrams, Bigrams	53.8	62.2
Sentiment lexicon	54.8	63.2

Table 1: Submitted runs (Accuracy).

54.2 [+0.6]. The best results in our development set were achieved by also providing punctuation marks as features: 55.2 Acc [+1].

After establishing the feature set, bigrams and a sentiment lexicon were also tested as additional resources. Table 1 summarizes the obtained results for the development and test sets, revealing that results over the development set are consistent with results over the test set. However, we have concluded that sentiment lexicons and bigram-based features turned out not to be helpful the way they have been used. Nevertheless, differences were not statistical significant using Wilcoxon signed-rank test.

5.4 Results for topic classification

The evaluation performed in the scope of the TASS challenge assumes that the set of topics manually labeled for each tweet must be matched (Villena-Román et al., 2012). For example, if a tweet was previously marked with topic t_1 and t_2, then the system must also suggest the same set of two topics.

Differences across experiments are always subtle, because improvements in one classifier may worsen results in another classifier. In terms of feature usage, experiments revealed that adding the author's name produced slightly better results but, contrarily to what was expected, providing the *user type* as a feature did not improve results. Adding punctuation marks decreased the overall performance. The best combination of features, using unigrams, led to 43.2 Acc in the development set and 64.9 Acc in the test set.

Apart from the previous evaluation, we have also performed evaluations for each topic individually. Table 2 shows the corresponding results for the test set, sorted by SER (Slot Error Rate) (Makhoul et al., 1999) performance. The first column shows the number of correct classifications, and the other columns show the corresponding *Precision*, *Recall*, *F1-measure*, and SER, respectively.

Similarly to what has been done for sentiment analysis, we have also performed experiments that combined features as bigrams. That strategy proved to be a good solution for the test set (65.4 Acc [+0.5]), but not so good for the development set (42.5 Acc [-0.7]). However, a deeper analysis on the test set, considering each topic individually have revealed that such strategy increases the recall but decreases the precision, leading to lower F1-measure [-0.5%].

Topic	**Cor**	**Prec**	**Rec**	**F1**	**SER**
política	26830	0.89	0.87	0.88	0.237
literatura	45	0.48	0.44	0.46	0.239
música	1345	0.90	0.47	0.62	0.268
deportes	70	0.52	0.63	0.57	0.271
tecnología	205	0.71	0.63	0.67	0.274
cine	418	0.70	0.44	0.54	0.287
fútbol	444	0.54	0.65	0.59	0.299
entreten.	5055	0.93	0.48	0.63	0.357
economía	2212	0.87	0.47	0.61	0.379
otros	18039	0.64	0.91	0.75	0.442
Total	54663	0.79	0.77	0.78	0.442

Table 2: Separate results per topic.

Figure 3 shows the confusion matrix for topic detection (each topic is represented by its first letter, except for "entertainment" which is represented by ET). As expected the highest values appear in the diagonal of the matrix. However, it is possible to observe that topic "others" is frequently assigned to tweets classified in the reference with more than one topic, being "others" one of them. It is also possible to observe that "others" is also incorrectly predicted for tweets classified in the reference as "movies", "economics", "entertainment", and "music", something that it is not very surprising. Also expected are the misclassifications of "economics" as "politics". For the construction of the confusion matrix, we have used evaluation criterion of the TASS challenge, but it also possible to perceive that for classification with more than one topic, in general, our approach correctly predicts at least one of the topics (usually by predicting only one of the reference topics).

6 *Conclusions*

The paper describes a shared classification approach that has been applied to automatic sentiment analysis and topic classification over Spanish Twitter data. The strategy, based on binary maximum entropy classifiers, is easy to develop, offer fast convergence ratios, can be executed in parallel, and is language independent, except for the detection of the negation. A cascade of binary classifiers was used for discriminating between six possible sentiment classes, and a one-vs-all strategy was used for topic classification, where the most probable topics for each tweet were

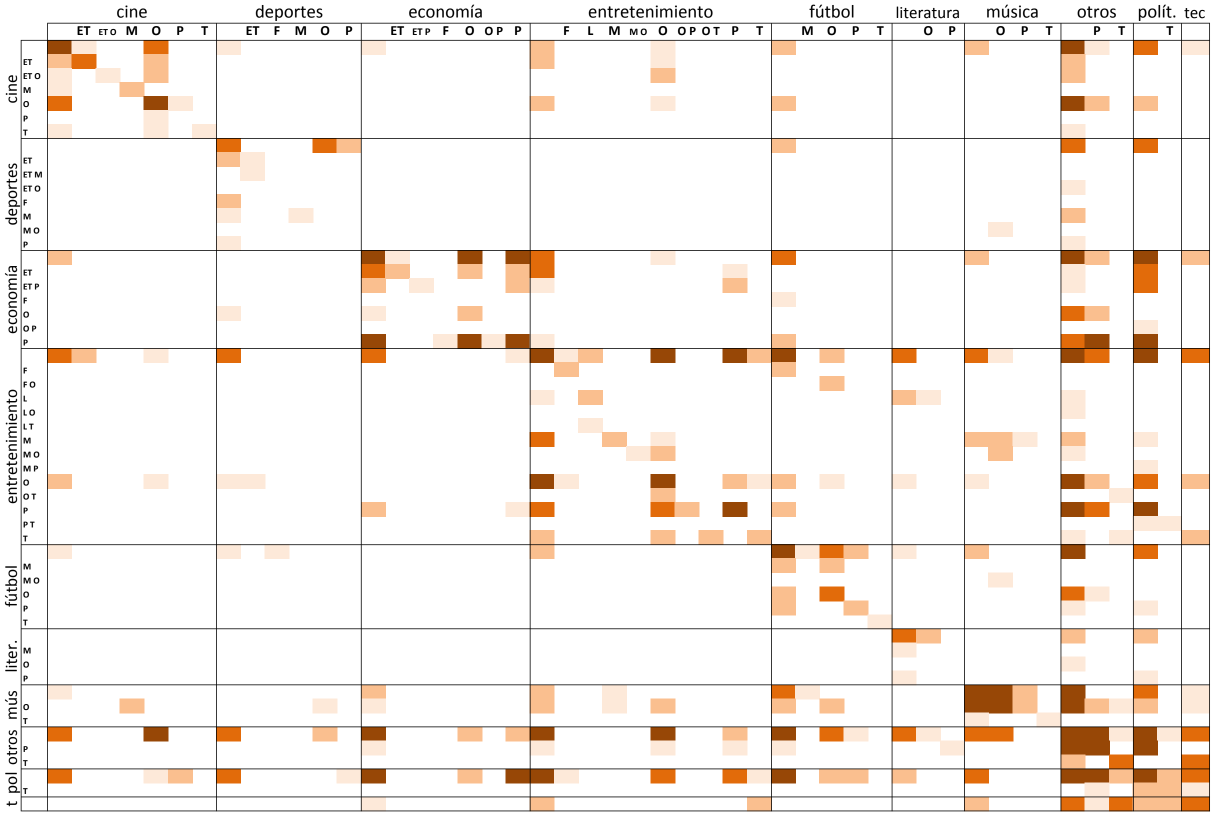

Figure 3: Topics confusion matrix (higher values darker; lower values lighter).

selected. The developed system achieved the best results for topic classification (+5.2 Acc, with statistical significance using the Wilcoxon signed-rank test: $W = 9425$, $p < 0.001$), and the second place for sentiment analysis (-1.9 Acc, without statistical significance, also using the Wilcoxon signed-rank test) in a joint evaluation effort (Villena-Román et al., 2012). In what concerns sentiment analysis, our experiments have shown that knowledge about the author and punctuation marks contribute to improved results. However, using bigram-based features and sentiment lexicons did not show a positive contribution with our setup. In what concerns the topic classification, the author type did not show a strong contribution, contrarily to what was expected.

Future experiments will make use of the remainder information available. The sentiment polarity type (AGREEMENT, DISAGREEMENT), together with other information about the user (e.g. number of tweets, followers, and following), will probably have impact on the results. Another possible direction is to automatically learn lexicons from the data and use them as an additional source of information.

Acknowledgements

This work was partially supported by national funds through FCT – Fundação para a Ciência e Tecnologia, under project PEst-OE/EEI/LA0021/2011, and by DCTI - ISCTE-IUL – Lisbon University Institute.

References

Allan, James, editor. 2002. *TOPIC DETECTION AND TRACKING: Event-based Information Organization*. Kluwer Academic Publishers.

Baccianella, S., A. Esuli, and F. Sebastiani. 2010. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta. ELRA.

Berger, A. L., S. Della Pietra, and V. Della Pietra. 1996. A maximum entropy approach to natural language processing. *Computational Linguistics*, 22(1):39–71.

Bollen, Johan, Huina Mao, and Xiao-Jun Zeng. 2010. Twitter mood predicts the stock market. *CoRR*, abs/1010.3003.

Das, Sanjiv and Mike Chen. 2001. Yahoo! for Amazon: Extracting market sentiment from stock message boards. In *Proceedings of the Asia Pacific Finance Association Annual Conference (APFA)*.

Daumé III, Hal. 2004. Notes on CG and LM-BFGS optimization of logistic regression. http://hal3.name/megam/.

Go, Alec, Richa Bhayani, and Lei Huang. 2009. Twitter Sentiment Classification using Distant Supervision. Technical report, Stanford University.

Hu, Minqing and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the* 10^th *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '04, pages 168–177. ACM.

Kasiviswanathan, S. P., P. Melville, A. Banerjee, and V. Sindhwani. 2011. Emerging Topic Detection using Dictionary Learning. In *CIKM'11: Proceedings of the 20th ACM international conference on Information and Knowledge Management*, pages 745–754. ACM.

Kim, Soo-Min and Eduard Hovy. 2004. Determining the sentiment of opinions. In *Proceedings of the 20th International Conference on Computational Linguistics (COLING '04)*. ACL.

Lee, K., D. Palsetia, R. Narayanan, M. Patwary, A. Agrawal, and A. Choudhary. 2011. Twitter trending topic classification. In *International Conference on Data Mining Workshops (ICDMW)*, pages 251–258. IEEE.

Lin, C., Y. He, R. Everson, and S. Rüger. 2012. Weakly Supervised Joint Sentiment-Topic Detection from Text. *IEEE Transactions On Knowledge And Data Engineering*, 24(6):1134–1145.

Makhoul, J., F. Kubala, R. Schwartz, and R. Weischedel. 1999. Performance measures for information extraction. In *Proceedings of the DARPA Broadcast News Workshop*.

O'Connor, B., R. Balasubramanyan, B. R. Routledge, and N. A. Smith. 2010. From tweets to polls: Linking text sentiment to public opinion time series. In *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media (ICWSM'10)*.

Pang, Bo, L. Lee, and S. Vaithyanathan. 2002. Thumbs up? sentiment classification using machine learning techniques. In *Proceedings of Empirical Methods in Natural Language Processing (EMNLP)*, pages 79–86.

Perez-Rosas, V., C. Banea, and R. Mihalcea. 2012. Learning Sentiment Lexicons in Spanish. In N. Calzolari, K. Choukri, T. Declerck, M. U. Doğan, B. Maegaard, J. Mariani, J. Odijk, and S. Piperidis, editors, *Proceedings of the International Conference on Language Resources and Evaluation (LREC'12)*. ELRA.

Sriram, Bharath, David Fuhry, Engin Demir, Hakan Ferhatosmanoglu, and Murat Demirbas. 2010. Short Text Classification in Twitter to Improve Information Filtering. In *SIGIR'10: Proceedings of the 33rd international ACM SIGIR conference on Research and Development in Information Retrieval*, pages 841–842. ACM.

Stone, P., D. Dunphy, M. Smith, and D. Ogilvie. 1966. *The General Inquirer: A Computer Approach to Content Analysis*. MIT Press.

Turney, Peter D. 2002. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pages 417–424. ACL.

Villena-Román, J., J. García-Morera, C. Moreno-Garcia, L. Ferrer-Ureña, S. Lana-Serrano, J. C. González-Cristobal, A. Westerski, E. Martínez-Cámara, M. A. García-Cumbreras, M. T. Martín-Valdivia, and Ureña-López L. A. 2012. TASS-Workshop on Sentiment Analysis at SEPLN. *Procesamiento de Lenguaje Natural*, 50.

Wilson, T., J. Wiebe, and P. Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP)*, pages 347–354. ACL.