Domain Adaptation of Polarity Lexicon combining Term Frequency and Bootstrapping

Salud M. Jiménez-Zafra, M. Teresa Martín-Valdivia,
M. Dolores Molina-Gonzalez and L. Alfonso Ureña-López
Department of Computer Science, Universidad de Jaén (Spain)
Campus Las Lagunillas, E-23071, Jaén, Spain
{sjzafra, maite, mdmolina, laurena}@ujaen.es

Abstract

In this paper we study several approaches to adapting a polarity lexicon to a specific domain. On the one hand, the domain adaptation using Term Frequency (TF) and on the other hand, the domain adaptation using pattern matching with a BootStrapping algorithm (BS). Both methods are corpus based and start with the same polarity lexicon, but the first one requires an annotated collection of documents while the second one only needs a corpus where it looks for linguistic patterns. The performance of both methods overcomes the baseline system using the general polarity lexicon iSOL. However, although the TF approach achieves very promising results, the BS strategy does not give as much improvement as we expected. For this reason, we have combined both methods in order to take advantage of the positive aspects of each one. With this new approach the results obtained are even better that those with the systems applied individually. Actually, we have achieved a significant improvement of 11.50% (in terms of accuracy) in the polarity classification of the movie reviews with respect to the results achieved with the general purpose lexicon iSOL.

1 Introduction

Sentiment Analysis (SA) is a discipline that combines Natural Language Processing (NLP) and data mining techniques to deal with the subjectivity in textual information. Several tasks have been studied but perhaps Polarity Classification is the most well known that focuses on determining the semantic orientation of a document: positive, negative or neutral.

Although different approaches have been applied to the field of polarity classification, the mainstream basically consists of two major methodologies. On the one hand, the Machine Learning (ML) approach that is based on using a collection of data to train the classifiers (Pang et al., 2002). On the other hand, the approach based on Semantic Orientation (SO) that does not need prior training but takes into account the orientation of words, positive or negative (Turney, 2002). In this paper we focus on semantic orientation in order to tackle one of the open issues related to polarity classification: domain adaptation.

Our main goal is to propose a method for automatically adapting a general polarity lexicon to a specific domain. Specifically, we are going to work with the movie domain because, as several papers demonstrate, this is a very difficult domain to deal with and adapt in SA (Turney, 2002; Taboada et al., 2009; Molina-González et al., 2015b). In addition, we will focus on Spanish since we consider that a language other than English is a more challenging task in the NLP area in general, and in SA in particular, due to the scarcity of resources.

Different methods have been proposed for tackling the domain adaptation problem by automatically generating polarity lexicons. One of the primary studies related to SA is (Blitzer et al., 2007). They note that the polarity of a particular word can carry opposing sentiments depending on the domain, so the general purpose lexicon should be adapted to the specific domain in order to improve

137

the effectiveness. Two main approaches to creating polarity lexicons automatically have been studied: dictionary-based and corpus-based. Dictionary-based approaches use external resources such as thesaurus or dictionaries in order to enrich a set of polar terms by aggregating new subjective words (Esuli and Sebastiani, 2006; Hu and Liu, 2004; Molina-González et al., 2013). On the other hand, corpus-based approaches also start from a set of polar terms but instead of using dictionaries, which are domain dependant and usually difficult to find in some languages, they try to integrate external knowledge from document collections (Hatzivassiloglou and McKeown, 1997; Turney, 2002; Molina-González et al., 2015b).

In this paper we investigate two corpus-based methods used to adapt a polarity lexicon to a specific domain. On the one hand, domain adaptation using Term Frequency (TF) and on the other hand, domain adaptation using pattern matching with a BootStrapping algorithm (BS). Both methods are corpus based and start with the same polarity lexicon, but the first one requires an annotated collection of documents while the second one only needs a corpus where it seeks the specific patterns. The performance obtained with both methods significantly overcomes the baseline system using the general polarity lexicon. Finally, we propose an approach that combines both domain adaptation methods and the results obtained are even better that those with the systems applied individually.

Most studies usually start with a very small set of polar terms and then apply some methods to extract and append new subjective words to the original list. However, in this paper we use a very large general lexicon as our starting point. Specifically, in the two approaches presented in the paper, we have used as seed the Spanish polarity lexicon iSOL (Molina-González et al., 2013). This lexicon has been successfully applied in several studies, showing a very good performance in Spanish SA (González et al., 2015a; Martínez-Cámara et al., 2014; Cruz et al., 2014). In addition, we have carried out our experiments with the Spanish movie corpus MuchoCine (Cruz et al., 2008).

Regarding the domain approaches applied, the first one is based on the Term Frequency (TF) in an annotated corpus. This approach has already been applied in previous studies using different domains (Molina-González et al., 2015b; Molina-González et al., 2014; González et al., 2015c). However, although in all the cases the results are improved, we have noted that some polar terms are wrongly added and sometimes some noise is introduced into the systems. Thus, in this paper we investigate a new method that not only adds new words but also, if some terms are detected to be highly subjective, they are eliminated from the adapted lexicon. This is the second approach used in this paper and it is based on detecting patterns in a corpus and applying a bootstrapping algorithm in order to enrich and clean the polarity lexicon. We will call this approach Boot-Strapping (BS). One of the best points of this second method is that we do not need any annotated corpus to adapt the system. We only need a polarity lexicon and a corpus of documents. From this corpus, we extract patterns using the information in the lexicon and we apply the bootstrapping algorithm in order to append or eliminate polar terms from the original lexicon. In this paper we only look for patterns including adjectives, because according to several studies this kind of word is the best clue to express subjectivity in documents (Wiegand et al., 2013).

As we will show in Section 5, the results obtained with the TF approach are very promising although the method requires an annotated corpus. On the other hand, although the results with the BS strategy also surpass the baseline system with the general opinion lexicon, the improvement is lower than we would hope. For this reason, we have decided to combine both methods in order to take advantage of them. Thus, we first apply the TF approach obtaining a new adapted polarity lexicon. This new list is used as the seed of polar terms for the BS algorithm. In this way, our method not only appends terms that are adapted to the domain but it also eliminates polar terms that can be considered highly subjective in this specific domain. The results achieved with this combined method improve on the performance of both approaches when they are applied individually.

The rest of the paper is organised as follows: The following section presents a review of the main methods of domain adaptation, focusing mainly on the corpus-based approaches and commenting on some studies that deal with Spanish. Section 3 introduces the methodology of adaptation based on

138

term frequency and Section 4 presents domain adaptation using bootstrapping. Section 5 exhibits the resources employed and shows the results obtained. In addition, we propose the combination of both methods in order to achieve an improvement in the final system. Section 6 discusses the different results obtained and analyses the systems proposed showing their advantages and disadvantages. Finally, in Section 7 conclusions and future work are presented.

2 Background

In this paper we focus on Spanish domain adaptation. We follow two different corpus based methods and then we combine both of them. Our work is based on the papers briefly described below.

2.1 Methods using Bootstrapping (BS)

One of the first learning studies to extract linguistic patterns for determining the polarity of sentences was (Hatzivassiloglou and McKeown, 1997). They consider only adjectives and using a large corpus classify the new extracted terms as positive or negative. Turney (2002) follows the same approach but includes adverbs along with adjectives, and uses Pointwise Mutual Information and Information Retrieval to estimate the semantic orientation. Riloff et al. (2003) present a bootstrapping algorithm that learns subjective nouns from an unannotated corpus. The approach uses two high precision classifiers (HP-Subj and HP-Obj) to automatically identify subjective and objective sentences. These classifiers give very high precision but low recall. The extracted sentences are then added to the training data to learn patterns. The learned patterns are then used to automatically identify more subjective and objective sentences. The bootstrapping algorithm increases the recall of the final system while high precision is maintained.

Regarding the studies dealing with Spanish documents, (Vázquez et al., 2012) is one of the first works applying a bootstrapping algorithm to discover new subjective adjectives in a Spanish corpus, following the same approach as Hatzivassiloglou and McKeown (1997). However, they use a small set of polar seeds that they have previously shown to be domain independent (Vázquez and Bel, 2012). Actually, they introduce the concept of "highly sub-

jective adjectives" as those adjectives which can be considered not only positive for one domain and negative for another ("predictable steering" in car domain vs "predictable plot" in movie domain), but which could also change their prior polarity even within the same domain ("antique car" can be positive for a classic person and negative for a modern person).

Our method using a bootstrapping algorithm is based on this idea of "highly subjective adjectives". We first identify and extract the adjectives learned by the linguistic patterns, but before being included in the original lexicon, we first check whether the adjective must be eliminated or included in a set of "highly subjective adjectives".

2.2 Methods using Term Frequency (TF)

The method based on term frequency proposed in this paper has already been applied, although the domain and corpora used are different. For example, the proposed approach in (Du et al., 2010) is based on the idea that a word must be negative (or positive) if it appears in many negative (or positive) documents, among other assumptions. The authors select three datasets of different domains, and from the relationships between them, they generate two labelled sentiment lexicons (domain independent and domain dependent) for each domain. In (Dehkharghani et al., 2012) a method is proposed for building a domain-dependent polarity classification system. The hotel and movie domains are selected by the authors. Each review is represented by a set of domain-independent features and a set of domain-dependent ones. The domain independent features are extracted from SentiWordNet (Baccianella et al., 2010). To build the set of domain-dependent features the authors propose taking the lexicon built by Hu and Liu (2004) and choosing those positive/negative words that occur in a significant number of positive/negative reviews of the training corpus used for the experimentation.

For Spanish domain adaptation, García et al. (2012) develop a polarity classification system based on the use of a list of opinion words generated by the authors. The corpus used for the evaluation is a set of hotel reviews written in Spanish and gathered from TripAdvisor. The method consists of counting the number of positive and negative words that ap-

pear in the text. Molina-González et al. (2014) propose a method taking as a base the iSOL lexicon and then enrich it with the most frequent words in positive/negative reviews from a Spanish corpus in the tourism domain (subset of SFU corpus[1]). They implement an automatic method to determine the best ratio between positive and negative words in order to integrate them into the new adapted lexicon. The system is tested on a different corpus of Spanish hotel reviews composed of more than 32,000 opinions. The results obtained improve greatly on the result with other general purpose lexicon. Afterward, the same approach is extended to the 8 different domains present in the Spanish SFU corpus (Molina-González et al., 2015b). In this paper we have applied the same approach as Molina-González et al. (2014), but instead of focusing on the tourism domain we have applied the method to the movie domain. We have also carried out several experiments in order to determine the best ratio between positive and negative terms to consider polar terms.

3 Domain adaptation based on Bootstrapping

The bootstrapping algorithm implemented in this paper follows the approach described in (Vázquez et al., 2012) that proposes the use of a bootstrapping method to automatically create lists of polar adjectives relevant for a domain and to detect "highly subjective adjectives" (that is, adjectives that could change their polarity even in the same domain). It is a corpus-based method that only needs a set of linguistic patterns extracted from a corpus and a seed polarity lexicon. The hypothesis of this algorithm is that there are some linguistic patterns that provide evidence of the semantic orientation of the words, and therefore this information can be iteratively used to identify the polarity of new words. The patterns selected by Vázquez et al. (2012) are the ones corresponding in Spanish to those presented in Hatzivassiloglou and McKeown (1997) where the authors hypothesize that the conjunction "and" ("y"/"e") joins adjectives of the same orientation while the conjunction "but" ("pero"/"aunque") joins adjectives of different orientation. As seed words, they use a set of

28 positive and 7 negative adjectives that five human annotators manually labeled as domain independent (Vázquez and Bel, 2012). They test the effectiveness of the proposed approach over a set of 200 Spanish documents manually tagged with the polar adjectives that should be in the final polarity lexicon. From all the adjectives labeled by the bootstrapping algorithm (67% of the total are identified), 97.6% of the positive adjectives and 71.5% of the negatives adjectives are correctly tagged. The authors validate the method and obtain promising results, but they do not test the generated sentiment lexicon in the task of polarity classification, therefore we have decided to use it in our experimentation to check how it works.

The bootstrapping algorithm operates as follows. In first place, all the pairs of adjectives that match with any of the patterns defined (adj1 **"y"/"e"** adj2, adj1 **"pero"/"aunque"** adj2) are extracted from the training corpus. After this, the following process is iteratively repeated until there are no changes (insertion or removal) in the polarity lexicon.

For each pattern found, if any of the adjectives is in the seed polarity lexicon we proceed as follows:

- If the adjectives are joined by "y"/"e" and the polarity of one of them is unknown (that is, the adjective is not in the polarity lexicon yet), then the unknown adjective has the **same** semantic orientation as the other adjective and consequently it is added to the polarity lexicon with its corresponding polarity.

- If the adjectives are connected by "y"/"e" and the polarity of the two is known; if both have the same polarity all is well, but if they have opposite polarity (positive adj "y"/"e" negative adj, negative adj "y"/"e" positive adj), the two adjectives will be added to the list of highly subjective adjectives and both will be removed from the polarity lexicon.

- If the adjectives are joined by the conjunction "pero"/"aunque" and the polarity of one of them is unknown, then the unknown adjective has the **opposite** semantic orientation as the other adjective and consequently it is added to the polarity lexicon with its corresponding polarity.

- If the adjectives are joined by "pero"/"aunque" and the polarity of the two is known; if both have opposite polarity all is well, but if they have the same polarity (positive adj "pero"/"aunque" positive adj, negative adj "pero"/"aunque" negative adj), the two adjectives will be added to the list of highly subjective adjectives and both will be removed from the polarity lexicon.

We first applied this approach directly using the 35 adjectives proposed by Vázquez and Bel (2012) over the MuchoCine Corpus. However, the results were very poor and, thus, we decided to start from the iSOL lexicon as seed for the BS algorithm.

The main advantages of this approach are that it does not need a corpus previously tagged with the semantic orientation of each text, it can be applied to any domain and it not only adds adjectives to the seed sentiment lexicon but also cleans it by removing the highly subjective adjectives. However, we also find some disadvantages. It only takes into account one part of speech, the adjective. Furthermore, it removes an adjective if it appears in a contradictory construction[2] without considering that the adjective could appear in more correct constructions than in those that are contradictory. For example, a user can make a mistake and write "clean and dirty", then the algorithm will remove both adjectives from the positive and negative lists respectively, but if it took into account that there are a great quantity of correct constructions for the adjective "dirty" (for example, "dirty and ugly", "dirty and dusty"...) it should not be removed from the negative list. Thus, we should consider other parameters before removing a specific polar term.

4 Domain adaptation based on Term Frequency

In order to implement the Term Frequency approach, we have followed the same assumptions as Du et al. (2010). According to this strategy, a word must be negative (or positive) if it appears in many negative (or positive) documents. Therefore, we have implemented an automatic method to determine the

---

[2]Positive adj + "y"/"e" + negative adj; negative adj + "y"/"e" + positive adj; positive adj "pero"/"aunque" + positive adj; negative adj + "pero"/"aunque" + negative adj

groups of terms in order to integrate them into the new adapted lexicons. The groups of words are selected using the following equation:

$$
polarity(word) = \begin{cases} pos & \text{if } (f^- = 0 \wedge f^+ \geq n) \vee (\frac{f^+}{f^-} \geq n) \\ \\ neg & \text{if } (f^+ = 0 \wedge f^- \geq n) \vee (\frac{f^-}{f^+} \geq n) \end{cases}
$$

(1)

Where f^+ is the absolute frequency of the occurrences of a given word in positive reviews and f^- is the absolute frequency of the occurrences of a given word in negative reviews. Therefore, n is the ratio between the amount of positive and negative words.

The main advantage of this method is its simplicity and quick implementation. Nevertheless, we find several disadvantages. On the one hand, the task of finding available corpora labelled with polarity at the document level is sometimes difficult, particularly for certain domains and languages. On the other hand, the inclusion in the adapted lexicon of all types of words without any discrimination, depending only on a ratio, sometimes introduces noise that does not improve the result in polarity classification.

5 Experimental framework and results

We tested two corpus-based approaches and the combination of them for the domain adaptation of a polarity lexicon in Spanish. The list of opinion words taken as a starting point was iSOL (Molina-González et al., 2013). iSOL is a Spanish polarity lexicon generated from the automatic translation of the Bing Liu Lexicon (Hu and Liu, 2004) and the manual revision of it. It is composed of 2,509 positive and 5,626 negative words.

For the adaptation of this lexicon and for testing the TF and BS approaches we used the Spanish MuchoCine corpus (MC) (Cruz et al., 2008). This dataset consists of 3,878 movie reviews collected from the MuchoCine website. The reviews are written by web users, therefore the sentences found in the reviews may include spelling mistakes or informal expressions and they may not always be grammatically correct. The dataset contains about 2 million words and an average of 546 words per review. The opinions in the corpus are rated on a scale from 1 to 5. A rank of 1 means that the opinion is very

bad and 5 means very good. Reviews with a rating of 3 can be categorized as "neutral", which means that the user considers the movie is neither bad nor good. In our experiments the neutral reviews were not taken into account, the opinions with ratings of 1 or 2 were considered as negative and those with ratings of 4 or 5 as positive (in total 1,351 positive and 1,274 negative reviews). The 60% of these reviews (781 positive and 794 negative reviews) were employed for the domain adaptation of iSOL and the remaining 40% (570 positive and 480 negative reviews) were used for testing the resultant lists in the task of polarity classification.

In order to apply the method based on frequency (TF), the punctuations and the stopwords of the documents were first removed. After this, the absolute frequency of each word in the positive and negative reviews was determined. Subsequently, several experiments were carried out to add to iSOL those words of the corpus that verify Equation 1, considering different ratios in order to fix the best ratio between positive and negative terms to consider polar terms.

In the case of the approach based on bootstrapping (BS), the documents were first tokenized and splitted into sentences and each token was tagged with its pertinent part of speech, using Freeling[3] (Carreras et al., 2004). Afterwards, the pairs of adjectives that matched with any of the defined patterns were extracted using regular expressions. Finally, two experiments were performed. In the first one, the bootstrapping algorithm was applied using as seed the opinion lexicon iSOL, and in the second the opinion lexicon resultant from applying the TF method with the best ratio (eSOLMovie) was employed as seed.

In order to evaluate the experiments we used the traditional measures employed in text classification: precision (P), recall (R), F1 and Accuracy. On the other hand, to calculate the polarity (p) of a review (r) with each lexicon, we take into account the total number of positive words (#positive) and the total number of negative words (#negative) within the review, according to the following strategy:

$$polarity(review) = \begin{cases} 1 & \text{if } (\#positive > \#negative) \\ -1 & \text{if } (\#positive \leq \#negative) \end{cases} \qquad (2)$$

As the baseline of our experimentation we took the general purpose lexicon iSOL, in order to adapt it to the movie domain with the proposed approaches and with the combination of them. The result of the polarity classification of the documents following the strategy defined previously and using iSOL is of 62.95%, in terms of accuracy.

Regarding the Term Frequency methodology, we first carried out different experiments in order to determine the best ratio to use in our final experiment combining both strategies. Thus, after testing several ratios we determined that the best one is obtained using the ratio n=4 (Figure 1). Therefore, this lexicon was taken as the seed list for the combined experiment with bootstrapping. Table 1 shows the results obtained over the MC corpus using iSOL (domain independent) and eSOLMovie^n lexicons (adapted to the movie domain with the ratios n=3,4,5,6).
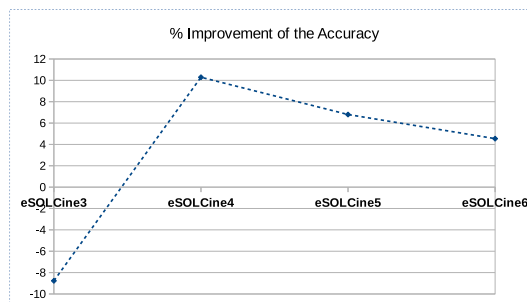
Figure 1: Improvement of the Accuracy respect to iSOL with the different eSOLMovie^n lexicons.

In relation to the Bootstraping strategy, it found 1,841 "y"/"e" patterns and 39 "pero"/"aunque" patterns and it was tested using iSOL and eSOLMovie^4 as seeds. The experiment with iSOL detected 292 highly subjective adjectives, inserted 659 adjectives, removed 110 adjectives and converged in 5 iterations, achieving 63.71% of accuracy in the polarity classification of the corpus. On the other hand, the experiment with eSOLMovie4 detected 296 highly subjective adjectives, appended 626 adjectives, deleted 228 adjectives and converged in 4 iterations, achieving 70.19% of accuracy. Table 2

Resource	Macro-P	Macro-R	Macro-F1	Accuracy	Improvement Macro-F1	Improvement Accuracy
iSOL	62.59%	62.28%	62.43%	62.95%	-	-
eSOLMovie[3]	70.51%	60.51%	65.13%	*57.43%	4.32%	-8.77%
eSOLMovie[4]	**69.74%**	**68.39%**	**69.06%**	***69.43%**	**10.62%**	**10.29%**
eSOLMovie[5]	67.40%	66.16%	66.77%	*67.24%	6.95%	6.80%
eSOLMovie[6]	65.56%	65.09%	65.33%	65.81%	4.65%	4.54%

Table 1: Results obtained using the different lexicons eSOLMovie^n.
An asterisk (*) means significantly different according to the Chi-square test ($\alpha = 0.05$) compared to the result with the baseline approach (iSOL).

Resource	Macro-P	Macro-R	Macro-F1	Accuracy	Improvement Macro-F1	Improvement Accuracy
iSOL	62.59%	62.28%	62.43%	62.95%	-	-
iSOL+Bootstrapping	63.63%	62.55%	63.09%	63.71%	1.06%	1.21%
eSOLMovie[4]	69.74%	68.39%	69.06%	*69.43%	10.62%	10.29%
eSOLMovie[4]+Bootstrapping	**70.45%**	**69.22%**	**69.83%**	***70.19%**	**11.85%**	**11.50%**

Table 2: Results obtained with the different approaches.

shows the results achieved in the classification at the document level of the MC corpus using iSOL (domain independent) adapted to the movie domain with the BS approach, with the TF method (using the best ratio) and with the combination of both strategies (the bootstrapping algorithm over eSOL-Movie4).

6 Results analysis

Table 3 shows a summary of the results obtained and Table 4 presents the total number of positive and negative polar terms inserted and eliminated from the original lexicon after applying each method. As we can see, the results obtained with the TF method (eSOLMovie^n) are very promising. It improves the accuracy of the classification with respect to the general purpose lexicon (iSOL) by 10.29%, inserting only 132 positive words and 126 negative words (Table 3 and 4). However, the restriction of this approach is that we need a corpus previously tagged with the polarity of the documents. Moreover, this strategy only appends new words to the original lexicon and sometimes the new terms introduce noise (for example, we consider that the words "fácil" (easy) and "rápido" (fast) could not be indicators of negative opinion in the movie domain, see Table 5).

Therefore, we also decided to conduct experiments with a technique based on bootstrapping that does not require an annotated corpus and that not only appends words but also removes some of them. The application of the BS approach for the adap-

tation of iSOL to the movie domain also achieves an improvement in the classification (iSOL + bootstrapping) over the baseline (iSOL), although this improvement is not as great as we expected (it is only about 1.21%, see Table 3). We think that one of the reasons could be that in this approach we have only used patterns for the extraction of adjectives, while the TF method appends a word independently of its PoS. Furthermore, we think that the fact of not only inserting but also removing adjectives would be promising but if we take a look at some of the words removed (Table 5), there are adjectives that a priori it seems they should not have been eliminated (for example, the adjectives "agradable" (pleasant) and "espectacular" (spectacular) of the positive list).

Due to this fact, we decided to combine both methods in order to take advantage of them. Thereby, the list generated with the TF method (eSOLMovie[4]) was used as input in the BS algorithm and the resultant list (eSOLMovie[4] + Bootstrapping) was employed for the classification of the movie reviews, achieving an improvement with respect to the baseline (iSOL) of 11.50%, in terms of accuracy (Table 3). Moreover, the results obtained with this combined method improve on the performance of both approaches (iSOL + bootstrapping and eSOLMovie[4]) when applied individually.

7 Conclusions and future work

In this paper we have presented two corpus-based approaches for the domain adaptation of a polarity

Resource	#positive words	#negative words	Accuracy	Improvement Accuracy
iSOL	2509	5626	62.95%	-
iSOL+Bootstrapping	2743	5839	63.71%	1.21%
eSOLMovie[4]	2641	5752	*69.43%	10.29%
eSOLMovie[4]+Bootstrapping	2857	5934	*70.19%	11.50%

Table 3: Results summary.

Resource	#positive words inserted	#negative words inserted	#positive words removed	#negative words removed
iSOL+Bootstrapping	339	320	105	105%
eSOLMovie[4]	132	126	-	-%
eSOLMovie[4]+Bootstrapping	326	300	110	118%

Table 4: Total of positive/negative words inserted into iSOL and removed from iSOL with the different approaches.

	iSOL+Bootstrapping	eSOLMovie[4]	eSOLMovie[4]+Bootstrapping
Example of positive words inserted	Relevante(relevant), amena(pleasant), azul(blue), humilde(humble), última(last), expectante(expectant), destacable(remarkable), talentosa(talented)...	Debut(debut), imprescindible(essential), peliculón(great movie), cruda(crude), sorprendente(surprising), impactante(impressive), cruel(cruel), catelera(billboard)...	Orgulloso(proud), llamativa(striking), anecdótica(anecdotal), original(original), atrapadora(trapping), breve(short), familiar(familiar), desgarrador(heartbreaking)...
Example of negative words inserted	Maquiavélica(machiavellian), asquerosilla(disgusting), simplona(simpleton), peculiar(peculiar), repelente(repellent), extraordinaria(axtraordinary), predecible(predictable), horroroso(horrifying)...	Fallida(failed), previsible(foreseeable), montón(be an average), secuela(sequel), exceso(excess), experimento(experiment), fácil(easy), rápido(fast)...	Cansino(tiresome), aburridísimo(very boring), gigantes(giants), claustrofóbica(claustrophobic), intrascendente(trivial), asfixiante(stifling), bochornosa(embarrassing), tétrica(gloomy)...
Example of positive words removed	Sensacional(sensational), poderoso(powerful), interesante(interesting), agradable(pleasant), espectacular(spectacular), consistente(consistent), primera(first), realista(realistic)...		Cruel(cruel), brillante(brilliant), aceptable(acceptable), recomendable(recommendable), rico(rich), espectacular(spectacular), accesible(accessible), vital(vital)...
Example of negative words removed	Violentos(violents), extraña(strange), amenazante(threatening), sangrienta(bloody), despiadado(ruthless), terrible(terrible), enfermo(sick), doloroso(painful)...		Rápido(fast), fácil(easy), perverso(perverse), gris(grey), trágica(tragic), misteriosa(mysterious), última(last), inesperado(unexpected)...

Table 5: Examples of the words inserted into iSOL and removed from iSOL in the different experiments.

lexicon. Both methods are language independent and can be applied to any domain. One of them, the based on term frequency (TF), needs a corpus previously tagged with the polarity of the documents and the other one, the based on a bootstrapping algorithm (BS), does not require an annotated corpus, it only needs as input a set of patterns and a seed sentiment lexicon. The TF approach achieves very promising results while the BS strategy, although it improves on the baseline system with the general purpose lexicon, does not improve as much as we expected. Due to this fact we have combined both methods, in order to take advantage of the positive aspects of each of them. With this new lexicon we have achieved an improvement of 11.50% (in terms of accuracy) in the polarity classification of the movie reviews with respect to the results achieved with the general purpose lexicon iSOL.

In future work, we consider adding some improvements to the domain adaptation approach based on bootstrapping in order to solve one of the main disadvantages detected and to strengthen its main advantage. Thus, we plan to consider a different approach to remove an adjective from the sentiment lexicon, for example, if two adjectives are in contradictory constructions before removing them it could be a good idea to check the number of correct and contradictory constructions in which each of them appear, and remove them only if the number of contradictory constructions exceeds the number of correct constructions in a threshold. We also plan to check, before adding a new adjective to the sentiment lexicon, its polarity according to other important lexical resources (such as SentiWordNet using Multilingual Central Repository (Gonzalez-Agirre et al., 2012) to map sentiment labels to Spanish). Moreover, we will incorporate new patterns to the algorithm in order to extract polar words with another PoS (not only adjectives) such as nouns, verbs and adverbs.

References

Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *LREC*, volume 10, pages 2200–2204.

John Blitzer, Mark Dredze, Fernando Pereira, et al. 2007. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In *Proceedings of the 45th annual meeting of the association of computational linguistics*, volume 7, pages 440–447, Prague, Czech Republic. Association for Computational Linguistics.

Xavier Carreras, Isaac Chao, Lluis Padró, and Muntsa Padró. 2004. Freeling: An open-source suite of language analyzers. In *LREC*.

Fermín L Cruz, José A Troyano, Fernando Enriquez, and Javier Ortega. 2008. Clasificación de documentos basada en la opinión: experimentos con un corpus de críticas de cine en español. *Procesamiento de Lenguaje Natural*, 41.

Fermín L Cruz, José A Troyano, Beatriz Pontes, and F Javier Ortega. 2014. Building layered, multilingual sentiment lexicons at synset and lemma levels. *Expert Systems with Applications*, 41(13):5984–5994.

Rahim Dehkharghani, Benin Yanikoglu, Dilek Tapucu, and Yucel Saygin. 2012. Adaptation and use of subjectivity lexicons for domain dependent sentiment classification. In *Data Mining Workshops (ICDMW), 2012 IEEE 12th International Conference on*, pages 669–673. IEEE.

Weifu Du, Songbo Tan, Xueqi Cheng, and Xiaochun Yun. 2010. Adapting information bottleneck method for automatic construction of domain-oriented sentiment lexicon. In *Proceedings of the third ACM international conference on Web search and data mining*, pages 111–120. ACM.

Andrea Esuli and Fabrizio Sebastiani. 2006. Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of LREC*, volume 6, pages 417–422. Citeseer.

Aitor García, Sean Gaines, Maria Teresa Linaza, et al. 2012. A lexicon based sentiment analysis retrieval system for tourism domain. *Expert Syst Appl Int J*, 39(10):9166–9180.

Aitor Gonzalez-Agirre, Egoitz Laparra, and German Rigau. 2012. Multilingual central repository version 3.0. In *LREC*, pages 2525–2529.

M Dolores Molina González, Eugenio Martínez Cámara, and M Teresa Martín Valdivia. 2015a. Crisol: Base de conocimiento de opiniones para el español. *Procesamiento del Lenguaje Natural*, 55:143–150.

M Dolores Molina González, Eugenio Martínez Cámara, M Teresa Martín Valdivia, and Salud M Jiménez Zafra. 2015c. esolhotel: Generación de un lexicón de opinión en español adaptado al dominio turístico. *Procesamiento del Lenguaje Natural*, 54:21–28.

Vasileios Hatzivassiloglou and Kathleen R McKeown. 1997. Predicting the semantic orientation of adjectives. In *Proceedings of the 35th annual meeting of the association for computational linguistics and eighth conference of the european chapter of the association for computational linguistics*, pages 174–181. Association for Computational Linguistics.

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

Eugenio Martínez-Cámara, M Teresa Martín-Valdivia, M Dolores Molina-González, and José M Perea-Ortega. 2014. Integrating Spanish lexical resources by meta-classifiers for polarity classification. *Journal of Information Science*, page 0165551514535710.

M Dolores Molina-González, Eugenio Martínez-Cámara, María-Teresa Martín-Valdivia, and José M Perea-Ortega. 2013. Semantic orientation for polarity classification in Spanish reviews. *Expert Systems with Applications*, 40(18):7250–7257.

M Dolores Molina-González, Eugenio Martínez-Cámara, M Teresa Martín-Valdivia, and L Alfonso Urena-López. 2014. Cross-domain sentiment analysis using Spanish opinionated words. In *Natural Language Processing and Information Systems*, pages 214–219. Springer.

M Dolores Molina-González, Eugenio Martínez-Cámara, M Teresa Martín-Valdivia, and L Alfonso Ureña-López. 2015b. A Spanish semantic orientation approach to domain adaptation for polarity classification. *Information Processing & Management*, 51(4):520–531.

Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 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*, EMNLP '02, pages 79–86, Stroudsburg, PA, USA. Association for Computational Linguistics.

Ellen Riloff, Janyce Wiebe, and Theresa Wilson. 2003. Learning subjective nouns using extraction pattern bootstrapping. In *Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003-Volume 4*, pages 25–32. Association for Computational Linguistics.

Maite Taboada, Julian Brooke, and Manfred Stede. 2009. Genre-based paragraph classification for sentiment analysis. In *Proceedings of the SIGDIAL 2009 Conference: The 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 62–70. Association for Computational Linguistics.

Peter D. Turney. 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*, ACL '02, pages 417–424, Stroudsburg, PA, USA. Association for Computational Linguistics.

Silvia Vázquez and Núria Bel. 2012. A classification of adjectives for polarity lexicons enhancement. In *Calzolari N, Choukri K, Declerck T (et al.), editors. Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, pages 3557–3561.

Silvia Vázquez, Muntsa Padró, Núria Bel, and Julio Gonzalo. 2012. Automatic extraction of polar adjectives for the creation of polarity lexicons. In *Kay M, Boitet C, editors. Proceedings of COLING 2012: Posters: 24th International Conference on Computational Linguistics COLING 2012; 2012 December 8-15; Mumbai, India. Mumbai: The COLING 2012 Organizing Committee; 2012. p. 1271-1280.*, pages 1271–1280. ACL (Association for Computational Linguistics).

Michael Wiegand, Manfred Klenner, and Dietrich Klakow. 2013. Bootstrapping polarity classifiers with rule-based classification. *Language resources and evaluation*, 47(4):1049–1088.