Expert Systems With Applications 58 (2016) 57–75

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Unsupervised method for sentiment analysis in online texts

Milagros Fernández-Gavilanes*, Tamara Álvarez-López, Jonathan Juncal-Martínez, Enrique Costa-Montenegro, Francisco Javier González-Castaño

AtlantTIC, University of Vigo, Campus, 36310 Vigo, Spain

ARTICLE INFO

Article history:
Received 26 October 2015
Revised 17 March 2016
Accepted 18 March 2016
Available online 1 April 2016

MSC:
68Q55
68T50

Keywords:
Sentiment analysis
Opinion mining
NLP
Artificial intelligence

ABSTRACT

In recent years, the explosive growth of online media, such as blogs and social networking sites, has enabled individuals and organizations to write about their personal experiences and express opinions. Classifying these documents using a polarity metric is an arduous task. We propose a novel approach to predicting sentiment in online textual messages such as tweets and reviews, based on an unsupervised dependency parsing-based text classification method that leverages a variety of natural language processing techniques and sentiment features primarily derived from sentiment lexicons. These lexicons were created by means of a semiautomatic polarity expansion algorithm in order to improve accuracy in specific application domains. The results obtained for the Cornell Movie Review, Obama-McCain Debate and SemEval-2015 datasets confirm the competitive performance and the robustness of the system.

1. Introduction

The field of *sentiment analysis* (SA) has received increasing attention in recent years (Liu, 2012), particularly due to the explosive growth of social media, blogs and forums, which has enabled individuals and organizations to write about experiences and express opinions using colloquial and compact language. This new form of expression is potentially a source of extremely valuable information. For example, Twitter, one of the most popular social media networks, grew from 5000 new tweets per day in 2007 to 500 million tweets per day in 2013[1] by over 240 million users[2] and today has over 500 million users. Consequently, an increasing number of companies are focusing their marketing campaigns on the analysis of online comments from these potential customers, for instance, to predict the acceptance level of certain products (Jansen, Zhang, Sobel, & Chowdury, 2009).

However, it is difficult and costly to manually extract relevant knowledge from such large volumes of data, which is why automated machine prediction is so attractive (Bothos, Apostolou, & Mentzas, 2010). SA represents an interdisciplinary challenge that leverages a variety of *natural language processing* (NLP) techniques in order to determine the sentiment expressed in texts and decide whether they are positive, negative or neutral.

Of the different approaches applied to polarity classification, we can basically distinguish between *supervised machine learning* (ML) and *unsupervised lexicon-based* approaches. Although ML has proven to be extremely useful in the field of SA, an obvious disadvantage is its limited applicability to subject domains other than the domain it was designed for. Moreover, training of the classifier requires labeled datasets (Moreno Ortiz & Pérez Hernández, 2013), which are often difficult or even impossible to obtain. This is because their generation require people labeling data which is too labor-intensive and time-consuming.

We describe an unsupervised method for SA in English that used dependency parsing to determine the polarity of tweets and a previously created sentiment lexicon that took into consideration the special structure and linguistic content of messages. We analyzed the linguistic peculiarities of the texts and used a new SA algorithm based on sentiment propagation for dependency parsing that does not need prior or specific training.

While some of these unsupervised studies tried to compare their results with "*ad-hoc*" supervised methods, developed directly by the same author, in our case we preferred to compare our results with existing ones. As a testbed, we evaluated the performance of our approach for the movie review domain,

* Corresponding author. Tel.: +34 986 814081.
E-mail addresses: milagros.fernandez@gti.uvigo.es (M. Fernández-Gavilanes), talvarez@gti.uvigo.es (T. Álvarez-López), jonijm@gti.uvigo.es (J. Juncal-Martínez), kike@gti.uvigo.es (E. Costa-Montenegro), javier@det.uvigo.es (F. Javier González-Castaño).

[1] http://www.telegraph.co.uk/technology/twitter/9945505/Twitter-in-numbers.html.

[2] http://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/.

http://dx.doi.org/10.1016/j.eswa.2016.03.031
0957-4174/

represented by the *Cornell Movie Review*. Our system was able to correctly classify 74.80% of the test set, a result that improves on results obtained for this dataset by the methods in other works: (Annett & Kondrak, 2008; Zhou, Zhao, & Zeng, 2014) (unsupervised), (Li, Zhang, & Sindhwani, 2009) (semisupervised), (Annett & Kondrak, 2008) (supervised) and (Carrillo de Albornoz, Plaza, & Gervás, 2010) (hybrid), all without using training data.

We also evaluated our approach on the Obama-McCain Debate dataset (Shamma, Kennedy, & Churchill, 2009) in comparison with other unsupervised methods (Hu, Tang, Gao, & Liu, 2013; Zhou et al., 2014); as well as on the dataset provided for SemEval-2015 Task 10[3] on SA in Twitter (Rosenthal et al., 2015). For this message-level task (40 submissions by 40 teams), in which we participated for the first time, we achieved satisfactory results. The results for both datasets therefore confirm the competitive performance and robustness of our system.

The paper is organized as follows. Section 2 discusses related work on polarity classification. Section 3 describes the system proposed for this task and the semiautomatic domain adaptation of the sentiment lexicon. Section 4 describes experimental results for the Cornell Movie Review Data, Obama-McCain Debate and SemEval-2015 datasets. Finally, Section 5 summarizes the main findings and conclusions.

2. Related work

SA systems development follows two basic steps: identification and classification (Medhat, Hassan, & Korashy, 2014; Pang & Lee, 2008).

The first step identifies subjective features in texts. These can be selected using (which is not always trivial) methods that can use some paradigm words and word similarities in order to obtain words expressing similar opinions. According to the way the similarities are obtained, these methods may be divided into semantic thesaurus-based and domain corpus-based approaches.

Semantic thesaurus-based approaches rely on the existence of semantic thesaurus created by human annotators, like WordNet[4] or General Inquirer[5]. The approaches in this category depend on the kind of relationships, synonyms or antonyms, between sentiment terms and the glosses in the thesaurus, expanding the polarity lexicon from a small set of seed words with known polarity. In (Hu & Liu, 2004; Kim & Hovy, 2004) two positive and negative verb and adjective seed lists were bootstrapped using WordNet in order to produce a larger lexicon. Similarly, in (Kamps, Marx, Mokken, & de Rijke, 2004) a lexical network was built by linking synonyms provided by the thesaurus, and the sentiment polarity was defined by the distance from the seed words "*good*" and "*bad*" in the network. Furthermore, in (Esuli & Sebastiani, 2007) an inverse and bidirectional model of random walking algorithm was proposed. These methods commonly rely on the assumption that adjectives share the same polarities with their synonyms and opposite polarities with their antonyms. It could be argued that they rely on prior semantic thesaurus resources without considering the domain-dependent characteristic of the sentiment lexicon.

In recent years, domain corpus-based approaches have been more widely studied. They are built on the basic assumption that polar terms conveying the same polarities co-occur with each other in domain corpuses, with context-specific orientations, usually relying on syntactic or statistical techniques like co-occurrence of a word with another word of known polarity. For example, in (Hatzivassiloglou & McKeown, 1997), the orientation of adjectives was predicted using other adjectives linked to the first ones by "*and*" (the same orientation) and "*but*" (the opposite orientation). Another example can be found in (Turney, 2002) where semantic orientation was assigned by means of association relationships between an unknown word and a set of selected seeds (like "*excellent*" and "*poor*"). In other studies, such as (Read & Carroll, 2009), the polarity of a word was identified by studying its frequency in a large annotated corpus of texts. If the word occurred more frequently among positive (negative) texts, then polarity was assumed to be positive (negative). If neither positive nor negative texts were dominant, polarity was assumed to be neutral. Qiu, Liu, Bu, and Chen (2011) analyzed manually and summarized eight dependency rules between opinion words and opinionated targets, and proposed a double propagation algorithm to expand the opinionated targets and sentiment lexicon iteratively. Other studies treated the problem of detecting polarities of words by means of graph propagation algorithms, such as Rao and Ravichandran (2009) (with a label propagation algorithm) and Huang, Niu, and Shi (2014) (with a constrained label propagation one using chunk dependency information and prior generic lexicon).

More recently, many studies have also tried to exploit prior sentiment knowledge in source domains to assist sentiment lexicon construction in the target domain. This was the case in (Tan & Wu, 2011), where the lexicon construction was modeled as a random walking process over four types of relationships between documents and words from both the source and target domains. In (Liu˜K., 2015), a method for co-extracting opinion targets and opinion words by using a word alignment model is described. They detected opinion relations between them. In (Zhang & Singh, 2014), a semisupervised framework was proposed. Instead of using sentences, they used segments of them and their dependency relation pairs in order to capture the contextual sentiment words for sentiment lexicon construction. Other works, such as (Lu, Castellanos, Dayal, & Zhai, 2011), focus on the problem of learning a sentiment lexicon that is not only domain specific but also dependent on aspects in some context given an unlabeled opinionated text collection. Finally, in (Tang, Wei, Qin, Zhou, & Liu, 2014a), they have applied a seed expansion algorithm to enlarge a small list of sentiment seeds using prior web knowledge.

The second step is correct classification of the overall sentiment of a given text. As already commented, methods can be broadly divided into two categories: supervised ML and unsupervised lexicon-based approaches (Maynard & Funk, 2012). The former are often classifiers built from linguistic features that use two sets of documents: a labeled training set to learn the differentiating characteristics of texts and a test set to check classifier performance.

The most widely used supervised ML techniques applied for SA described in the literature include naive Bayes, maximum entropy and support vector machines (SVM) (Vohra & Teraiya, 2013). Once a supervised classification technique is selected, it is important to determine how the documents are represented, as the selection of adequate features is crucial for classification success. Many authors treat the documents as *bags of words* (Hu & Liu, 2004; Pak & Paroubek, 2010) comprising unigrams or *n*-grams with their frequencies, because of the resulting simplicity of classification (Pang, Lee, & Vaithyanathan, 2002). Other authors propose including linguistic information such as part-of-speech (POS) tagging in order to disambiguate text sense according to lexical category (Pang & Lee, 2008), for instance, identifying adjectives and adverbs used as sentiment indicators. Negation words reverse the sentiment (König & Brill, 2006; Pang & Lee, 2008) with the polarity of opinion words and phrases determined using WordNet (Hu & Liu, 2004). In this direction, most studies focus on selecting effective features to improve performance

[3] SemEval is an international forum for natural-language shared tasks and dataset is available at http://alt.qcri.org/semeval2015/.

[4] Available at http://wordnet.princeton.edu/

[5] Available at http://www.wjh.harvard.edu/~inquirer/

(Mohammad, Kiritchenko, & Zhu, 2013). For this reason, nowadays the goal is to attain an effective representation of a sentence (or document) from the representations of the words in order to use them as features (Tang et al., 2014b). However, the major disadvantages of this kind of approaches are that: classifiers trained for a domain-specific problem do not perform well in other domains (Aue & Gamon, 2005); feature engineering is very important but hard to implement (Bengio, Courville, & Vincent, 2013); and the performance of these methods usually relies on manually labeled training data. This motivates the problem of achieving sentiment classification via unsupervised paradigms (Severyn & Moschitti, 2015; Zhou et al., 2014).

A traditional way to perform unsupervised SA is that of lexicon-based approaches (Taboada, Brooke, Tofiloski, Voll, & Stede, 2011). The aim is to employ a sentiment lexicon composed of a collection of known and precompiled sentiment terms tagged with their semantic orientation to determine the overall sentiment of a given text. For example, Turney (2002) determined the polarity of a message by means of search results returning hits for the target word conjoined with the words "*good*" and "*bad*", while Kamps et al. (2004) used the WordNet database to estimate the minimum path distance between a word and pivot words ("*good*" and "*bad*"). Other works have shown that it is possible to apply polarity scores directly, aggregating them from a sentence or a document and computing the resulting sentiment on a continuous scale (Fahrni & Klenner, 2008; Missen & Boughanem, 2009; Tsytsarau, Palpanas, & Denecke, 2010). More sophisticated methods employ strategies involving lexis, syntax and semantics (Quinn, Monroe, Colaresi, Crespin, & Radev, 2010) and then aggregate their values.

Other works like Li et al. (2009) did not use polarity scores and proposed a constrained non-negative matrix trifactorization approach to SA, with a domain-independent sentiment lexicon as prior knowledge. The idea here is that words could be represented as data points from a geometric perspective, meaning that two words sufficiently close to each other tend to share the same sentiment polarity, and should be preserved even after the matrix factorization. Later, this idea was enhanced by Zhou et al. (2014) through a novel algorithm, called graph co-regularized non-negative matrix trifactorization. Other authors exploit emoticons and product ratings as emotional signals that are associated with sentiments in order to employ them as prior knowledge to guide the SA process (Hu et al., 2013).

It can be concluded that there is a huge interest in SA across a variety of domains, so that the datasets used are an important issue. The main data sources are taken from commerce, such as product reviews (Jansen et al., 2009; Pang & Lee, 2004). Nevertheless, SA is also applied in fields such as health (Salathé & Khandelwal, 2011), disaster management (Mandel etÂ al., 2012), stock markets (Nguyen, Shirai, & Velcin, 2015) and social network and micro-blogging sites. The latter are considered a valuable source of information because people share and discuss their opinions about a certain topic freely.

3. System overview

The main objective of this research was to predict whether an online text expresses positive, negative or neutral sentiments without the need of supervision. The review of the state of the art reveals that many of the existing learning or lexicon-based systems only take into account isolate words and not the relationships between them (Hu & Liu, 2004; Kamps et al., 2004; Pak & Paroubek, 2010; Turney, 2002). Some try to simulate the comprehension of certain linguistic constructions, such as negation, but fail due to the complexity of human language (König & Brill, 2006; Pang & Lee, 2008; Quinn et al., 2010; Taboada et al., 2011). Others employ dependencies between words but give up the interpretation of the complex human language to supervised systems (Nakagawa, Inui, & Kurohashi, 2010; Zhang & Singh, 2014). For this reason, we propose an alternative unsupervised approach to correctly detecting the sentiment expressed in texts that exploited the linguistic information present in dependencies retrieved from a parsing analysis. NLP techniques were used to capture linguistic peculiarities that were later used to improve sentiment detection performance. Our approach consists of four stages: preprocessing, lexical and syntactic analysis, the creation of a sentiment lexicon and sentiment analysis through propagation.

3.1. Preprocessing

Several challenges for NLP are posed by the language used in online media, such as social media sites, blogs and forums. The main problem is that words can be used with particular orthographic and typographical characteristics, such as letter or word duplication, and those are not found in a dictionary. Hence, before applying our approach, it was necessary to preprocess data to normalize the language, remove noisy elements and generalize the vocabulary used to express sentiment. The aim was to restore, as much as possible, message language to natural language by eliminating atypical expressions so as to minimize noise in later stages. Preprocessing involved four types of changes:

- URL links (such as "*http://url*"), hashtag links (such as "*#hashtag*") and username links (such as "*@username*") were replaced with the placeholders "URL", "HASHTAG" and "USERNAME" respectively.
- Replicated characters were removed to restore words to their standard spelling, e.g., *sweeeeet* → *sweet*.
- Emoticons[6] were replaced by one of nine labels: `e_laugh`, `e_happy`, `e_surprise`, `e_positive_state`, `e_neutral_state`, `e_inexpressive`, `e_negative_state`, `e_sad` and `e_sick`. For instance, :-(was replaced with `e_sad`.
- Abbreviations[7] were spelled out in full, e.g., *h8* → *hate*.

3.2. Lexical and syntactic analysis

In order to derive the syntactic context, each preprocessed message was split into tokens and then into sentences. To ensure that all inflected forms of a word were included, FreeLing (Atserias et al., 2006; Padró & Stanilovsky, 2012) was applied to lemmatization and POS tagging using hidden Markov (HMM) implementation (Brants, 2000). The FreeLing library, which includes language analysis services for several languages, including English, Spanish and French, probabilistically predicts categories for unknown words. POS tagging enables identification of lexical items that can contribute to correct recognition of the sentiment in a message. The accuracy rate is around 90%, but we conservatively assumed this accuracy rate to drop for social media, blog and forum texts. The lemmatized and POS-tagged messages were fed to the FreeLing Parser (Padró & Stanilovsky, 2012) which transformed output into a dependency tree.

As a result, for a given text T, we obtained a dependency parse for each sentence and a tree structure with nodes and edges. Each node represents the lexical information associated with a term in position i, denoted as t_i, including form, lemma and tag. Each edge represents the binary syntactical interaction between two terms t_i and t_j, denoted by $dep(t_i;\ t_j)$, with the former considered to be the head of t_j, and the latter considered to be dependent on t_i.

[6] List available at http://www.datagenetics.com/blog/october52012/index.html

[7] Lists available at http://chatslang.com/terms/abbreviations.

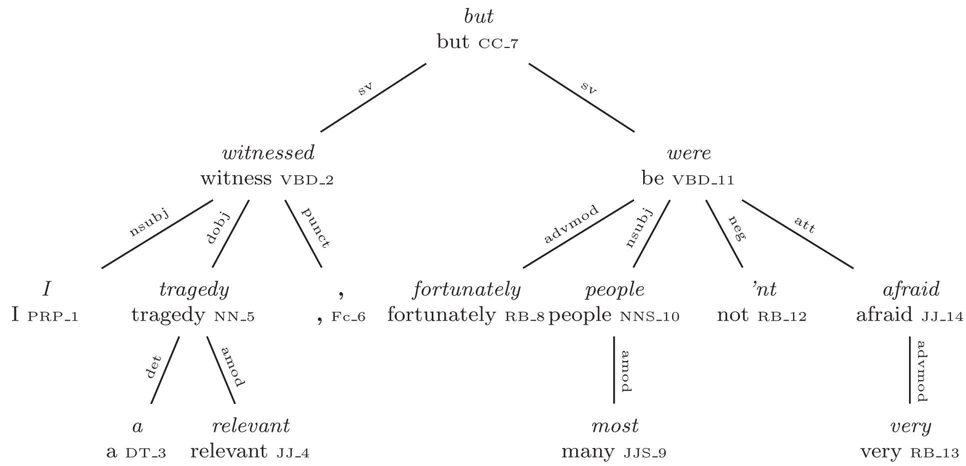

Fig. 1. Dependency parse for the running example.

As an example, Fig. 1 shows the dependency parse for the English sentence "*I witnessed a relevant tragedy, but fortunately most people weren't very afraid*" (used throughout this paper as a running example). The node "*witnessed*" is a verb in the second position (*VBD_2*), whose lemma is "*witness*". It is also the head term for three dependencies, *dep*(witness; I), *dep*(witness; tragedy) and *dep*(witness; ,).

3.3. Sentiment lexicon creation

A sentiment lexicon, also called a *polarity* or *opinion lexicon*, consists of a list of words with associated values representing their sentiments. These values are usually integers expressing polarity and polarity intensity as increasing or decreasing absolute values. The literature describes several sentiment lexicon creation methods, mostly complex and time-consuming manual approaches, but also automatic or semiautomatic dictionary generation methods. One such method is based on correlation measures that take into account the co-occurrence of a word with other positive or negative predefined words (Turney & Littman, 2002). Another approach is dictionary-based (Hu & Liu, 2004), whereby a group of initial words or seeds are expanded from synonym and antonym relationships given by dictionaries like WordNet. SentiWordNet (Esuli & Sebastiani, 2006), for instance, applies the concept that different words contained in the same categories in WordNet will have similar polarities.

Some examples of polarity lexicons available, mostly manually created as general purpose lexicons, include:

- SO-CAL (Taboada et al., 2011) is composed of lists of words according to lexical category. Polarity is expressed on a scale between –5 and +5.
- AFINN (Nielsen, 2011) consists of language used for microblogging. Polarity is expressed on a scale between –5 and –1 for negative items and between +1 and +5 for positive items.
- SentiWordNet (Esuli & Sebastiani, 2006) is an extension to WordNet created with semisupervised algorithms. It adds two values to each word, expressing negativity and positivity on a scale between -1 and +1.
- NRC Word-Emotion Association Lexicon (Mohammad & Turney, 2013) is, as its name indicates, a list of word-emotion associations. It is not exactly a polarity lexicon, as words are classified according to a set of basic emotion categories, such as *joy, trust, fear*, etc.
- MPQA Lexicon (Wiebe, Wilson, & Cardie, 2005) is a list of words classified in different ways, e.g., in terms of positivity/negativity, strength, etc. As in the previous case, polarity is not expressed as a numerical value.

However, these lexical resources are intrinsically non-contextualized, so it is necessary to improve their coverage. To do this, we need to acquire polarities for subjective words not present in generic dictionaries and adapt their scores using the available data. The advantage of this approach is that, since a word can have different connotations in different contexts, sentiment is obtained for words adapted to a specific context.

We used a non-supervised method that is an adaptation of the PolarityRank algorithm (Cruz, Troyano, Ortega, & Enríquez, 2011) based on PageRank (Page, Brin, Motwani, & Winograd, 1998). The starting point was a set of positive and negative words used as seeds, with polarities expanded through a graph constructed using dependencies between words.

The first step was to build the graph from a set of texts from a particular domain analyzed lexically and syntactically as in Fig. 1. Specifically, we created an undirected graph $\mathcal{G}=(N, E)$, where $N=\{v_i\}$ is a set of nodes, E is a set of undirected edges between pairs of nodes and where each edge joining nodes i and j has an associated weight denoted by w_{ij}. Nodes were selected by extracting all the resulting lemmatized nouns, adjectives and verbs, discarding adverbs because they do not carry an inherent sentiment polarity, but merely alter the degree of the polarity of the words they modify (Dragut & Fellbaum, 2014). Edges were created between pairs of nodes according to syntactic dependencies and between all nodes contained in children descendant branches, where each w_{ij} was the number of constructions observed between the linked terms. These links were bidirectional so polarity was propagated in both directions.

The resulting graph for the running example, shown in Fig. 2, was grown by the addition of new nodes and edges for all the other sentences in the dataset. Note that the verb "*to be*" was not included because it has no polarity and so introduces noise into the expanded lexicon. For this reason, we decided to establish an edge between its two siblings in order to relate the nodes "*people*" and "*afraid*". The same occurred with connectors. We established an edge between its two siblings in order to relate both descendants through the graph. The frequency of a word in the whole dataset was taken into account, as well as the frequency of edges established between two words/nodes. Frequencies for both nodes and edges were needed for the propagation algorithm to assess the relevance of each node in the expansion method.

Once the big graph was created, polarity propagation could commence. The starting point was the selection of positive and

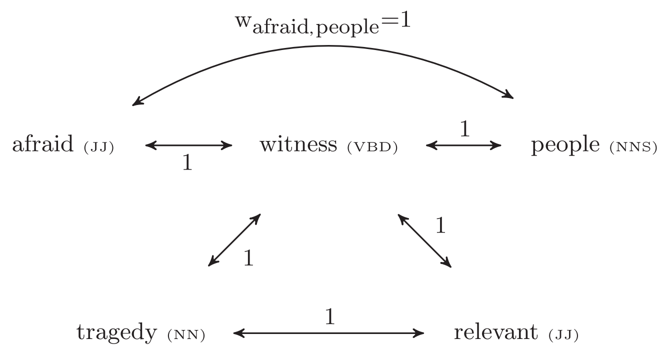

Fig. 2. Graph built for lexical expansion of the running example.

negative seed nodes as two lists of words related to strong positive and negative sentiments, respectively. These seeds were extracted from among the most negative and positive words in the general lexicons according to their frequency in the dataset. Since they participated in many edges, more nodes were reached during the first iteration of the algorithm. The PolarityRank algorithm calculated two scores for each node, positive (PR^+) and negative (PR^-), that depend on the relevance of each node and on the relevance of the nodes connected to the first nodes by an undirected edge. Let $E(v_i)$ be a set of indices j of the nodes for which an edge is connected to node v_i and let e_i^+ and e_i^- be two nonzero positive or negative values, respectively, if node v_i contains a seed. Finally, the parameter d is a damping factor used to ensure convergence, which we set to 0.85 (as recommended in the original definition of PageRank (Cruz et al., 2011; Page et al., 1998)). PR^+ and PR^- were estimated as follows:

$$PR^+(v_i) = (1-d)e_i^+ + d\sum_{j\in E(v_i)} \frac{w_{ij}}{\sum_{k\in E(v_j)}|w_{jk}|} PR^+(v_j)\ , \tag{1}$$

$$PR^-(v_i) = (1-d)e_i^- + d\sum_{j\in E(v_i)} \frac{w_{ij}}{\sum_{k\in E(v_j)}|w_{jk}|} PR^-(v_j) \tag{2}$$

We computed several iterations of the algorithm, obtaining two final values for each node PR^+ and PR^-. The process was be iterated until convergence (Cruz, Vallejo, Enríquez, & Troyano, 2012) to a fixed point or until a fixed approximation threshold is reached. In our case, we also established a stopping criteria setting a maximum of 200 iterations, which was decided after testing this process on different datasets. In practice, according to the experimental results on the datasets used in this paper, estimations were estabilished before executing all those iterations. We then computed the final *semantic orientation* (SO) of the word represented by the node as

$$SO(v_i) = \frac{PR^+(v_i) - PR^-(v_i)}{PR^+(v_i) + PR^-(v_i)} \tag{3}$$

3.4. Sentiment analysis through propagation

Lexicon-based approaches to SA are based on the idea that the polarity of a text can be obtained from the polarities of the words that compose that text. Therefore, given the dependency parse of T, we defined its sentiment, represented by SO_T, as a real number that was the sum of the polarities conveyed by each sentence making up T. In turn, the polarity of each sentence, which depends on the sentiment of each term, denoted by SO_{t_i}, was positive or negative according to the sentiment lexicon used, or neutral when t_i did not express sentiment (the case of determinants or punctuation marks), which corresponds to $SO_{t_i} = 0$. The result of assigning SO to each term in a parsed sentence is called *dependency parsed sentence sentiment* (DPSS), exemplified in Fig. 3, where words like "*I*" and "*people*" are neutral, while "*relevant*" is positive.

However, this approach based only on lexical information is over-simplistic, given the complexity of natural language (Musto, Semeraro, & Polignano, 2014). We therefore centered our efforts on the concept of *sentiment propagation* (Moilanen & Pulman, 2007), which assumes that the syntactic structure of a DPSS enables the intrinsic sentiment value of each linguistic element to be propagated. The basic idea is that these values influence each other at a syntactic level (Caro & Grella, 2013), so the result will be a propagation of the sentiment from the leaves to the parent nodes of the tree represented by the DPSS until the root is reached, as shown in Eq. (4)

$$SO_{t_i}^{new} = SO_{t_i} + \sum_{\forall t_j/\exists\ dep(t_i;t_j)} SO_{t_j}^{new} \tag{4}$$

To improve sentiment propagation it is very important to identify linguistic phenomena such as intensification or negation from the outset, as these can affect words, branches and even whole sentences and require specific propagation treatment. Once propagation is complete, many terms will have changed their sentiment values, and it is also possible that terms without a sentiment value before propagation will be assigned one during the process. When propagation is applied in a case where no linguistic phenomena are taken into account, we have what is called *basic propagation*.

Since the value computed by the PolarityRank algorithm lies in $[-1, 1]$, we transform each SO_{t_i} to its corresponding value in

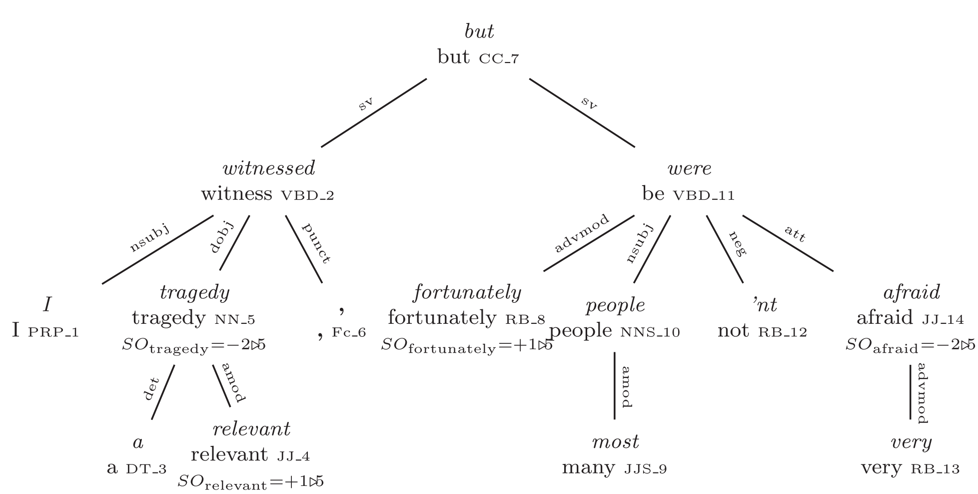

Fig. 3. Dependency parsed sentence sentiment for the running example.

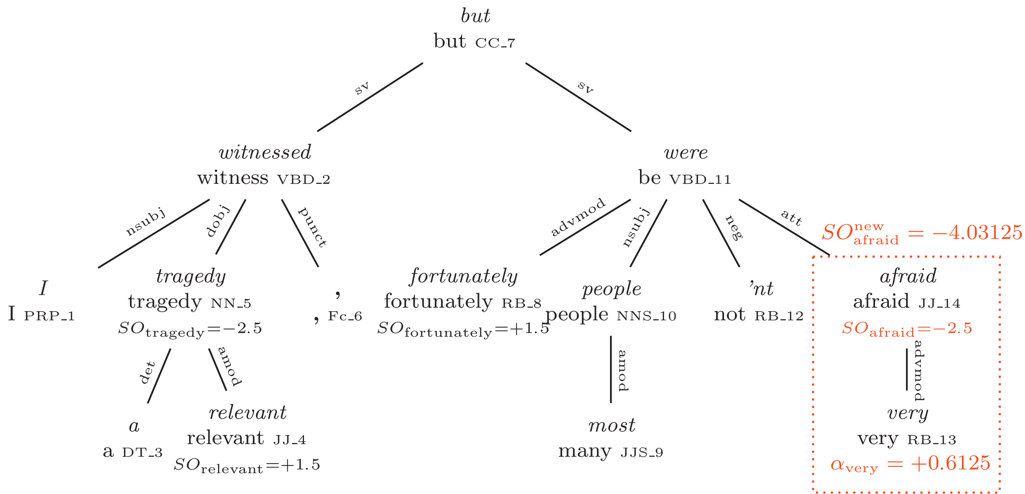

Fig. 4. Intensification propagation for the running example.

[−5, 5]. Considering a 10-point scale (excluding zero), it is easier to distinguish the neutral interval with a naked eye when a word frequency distribution of the sentiment lexicon is created. Furthermore, applying this scale, it should not be necessary to take into account an increase in treatment of the number of decimals during the estimations of the sentiment propagation to be more precise. The use of this interval reflects a tradeoff with the desire to differentiate clearly word meanings Brooke (2009).

next we describe how to propagate sentiment for intensification, modification, negation and adversative/concessive relations.

3.4.1. Intensification

Intensifiers and diminishers are linguistic terms that refer to elements that do not contribute to the propositional meaning of a sentence but serve to emphasize or attenuate, respectively, the strength of a word or expression and provide an emotional context. As semantically vacuous fillers, they are grammatical expletives. For example, in the sentence, "*People were very afraid*", the phrase "*very afraid*" is more negative than "*afraid*" alone. In contrast, if the word "*very*" is replaced by "*little*", the intensity of "*afraid*" is reduced.

The most common way of identifying these valence[8] shifters is using lists of words, such as adverbs and adjectives, associated with fixed values for intensifiers and diminishers, as was done in (Kennedy & Inkpen, 2006) and (Polanyi & Zaenen, 2004), where effects were calculated by adding/subtracting a point to/from the base valence of a term without distinguishing between intensifier or diminisher types. Additionally, Kennedy and Inkpen (2006) included the added constraint that all positive and negative sentiment terms could only have values of +2 and −2 respectively. Thus, the intensifier "*very*" in "*very afraid*" enhanced the negative sentiment orientation of "*afraid*" from −2 to −3, whereas the diminisher "*barely*" in "*barely courageous*" weakened the force of "*courageous*" from +2 to +1. In appraisal-based approaches, in contrast, intensifiers and diminishers modify the force of an attribute to varying degrees. Whitelaw, Garg, and Argamon (2005), for instance, considered three values for this force: low, neutral and high. Thus, "*courageous*" was neutral, whereas the inclusion of "*barely*" decreased force to low.

However, these approaches have one main limitation in that they assign a fixed value to all intensifiers and diminishers, regardless of the fact that they vary in the extent to which they intensify or diminish emotions. For example, the use of "*very*" and "*extremely*" in the sentences "*People were very afraid*" and "*People were extremely afraid*" clearly represents different shades of negativity, which is why an intensity scale is necessary.

We used a list of intensifiers and diminishers, adapted from Brooke (2009), where each element is a modifier that emphasizes or attenuates words. For each one, a mean average value was estimated depending on common sense values set by two annotators. The value of this *intensification coefficient* (IC) enabled propagation of the intrinsic intensification sentiment through the syntactic structure. For this reason, given an intensifier or diminisher x with an IC denoted by $\alpha_x \in [-1, 1]$ and dependency denoted by $dep(t_i; x)$, where t_i is the head term of x, the propagation is represented as

$$SO^{new}_{t_i} = SO_{t_i} * (1 + \alpha_x) \quad (5)$$

where α_x will be positive if x is an intensifier and negative if x is a diminisher. Our system detected these structures and used parsing to identify the extent of intensification whose polarity was altered.

For instance, in Fig. 4, there is a dependency between "*very*", which is an intensifier with $\alpha_{\text{very}} = +0.6125$, and "*afraid*", denoted by $dep(\text{afraid}; \text{very})$. Applying Eq. (5), the negativity of "*afraid*" is enhanced, yielding the new sentiment $SO^{\text{new}}_{\text{afraid}} = -4.03125$.

In the same way, superlative adjectives (JJS) were also taken into account, it being assumed that they would behave like a word accompanied by an intensifier, e.g., "*greatest*" was considered to intensify the word "*great*". No distinction was made between superlatives: all superlatives were considered to apply the same IC. But, in contrast to (Neviarouskaya, Prendinger, & Ishizuka, 2010), we obtain better results taking +0.25.

3.4.2. Modification

In grammar, a modifier is an optional element (Huddleston & Pullum, 2002) that changes the meaning of another dependent element. Unlike intensifiers, modifiers contribute to the meaning of words or expressions. For example, consider the phrase "*she got a beautiful place in heaven*", where "*beautiful*" denotes the sentiment towards noun it modifies, "*place*". The noun "*place*" thus becomes a sentiment expression itself, with the same positive polarity as the sentiment adjective "*beautiful*" (Nasukawa & Yi, 2003). In other

[8] Valence is a term from psychology that refers to the intrinsic attractiveness (positive valence) or aversiveness (negative valence) of an event, object, or situation (Frijda, 1986). It is also used to characterize and categorize negative and positive emotions, such as "*anger*" or "*happiness*".

words, the polarity of the modifier is propagated to the modified noun which itself has no inherent polarity (Min & Park, 2011).

The most common approach to determining sentiment is by incrementally combining word polarities from a sentiment lexicon. Simpler approaches calculate the ratio of positive and negative words in a sentence (Musto et al., 2014). Take, for example, the following sentences:

$$\text{“That is a } \underline{\textit{sensitive}}^{(+1.0)}\ \underline{\textit{love}}^{(+3.0)}\text{”, and} \tag{6}$$

$$\text{“I witnessed an } \underline{\textit{unacceptable}}^{(-1.0)}\ \underline{\textit{tragedy}}^{(-2.5)}\text{”} \tag{7}$$

Both “*sensitive*” and “*love*” are positive (with polarities of $+1.0$ and $+3.0$, respectively), and both “*unacceptable*” and “*tragedy*” are negative (with values of -1.0 and -2.5, respectively), so we can conclude that in (6) “*sensitive love*” is more positive than just “*love*”, while in (7) “*unacceptable tragedy*” is more negative than just “*tragedy*”. However, if we calculate the ratio of positive and negative words, those assumptions are not satisfied, because the total values are $+2.0$ and -1.75, which are lower than would be expected.

Another problem arises when a modifier is combined with its modified expression but they have different polarity signs. Using the above examples, the modifiers are now replaced with modifiers expressing the opposite sentiment

$$\text{“That is a } \underline{\textit{heartless}}^{(-2.0)}\underline{\textit{love}}^{(+3.0)}\text{”, and} \tag{8}$$

$$\text{“I witnessed a } \underline{\textit{relevant}}^{(+1.5)}\underline{\textit{tragedy}}^{(-2.5)}\text{”} \tag{9}$$

We now have “*heartless*” with negative polarity (-2.0) and “*relevant*” with positive polarity ($+1.5$), so “*heartless love*” and “*relevant tragedy*” should both be quite negative.

In such cases, it may be supposed that negative polarities are more dominant than positive ones. More specifically, positive modifiers tend to affirm or intensify, whereas negative modifiers tend to negate, the negative or positive polarity, respectively, of the modified expressions (Klenner, Tron, Amsler, & Hollenstein, 2014; Tron, 2013). To illustrate, if we apply the same simple approach as used above, the total values in examples (8) and (9) will be now $+0.5$ and -0.5, which are barely positive and barely negative, respectively. As can be seen in examples (6)–(9), the estimated values are not consistent with real sentiments.

We designed a mechanism to assign relatively accurate quantitative values in such cases. Our approach detected words accompanied by adjectives that were used attributively and not predicatively by locating dependencies that involved adjectives as modifiers of head terms where both had positive and/or negative polarities. We then propagated the modifier polarity to the upper level node in the dependency tree structure. Thus, given dependency $dep(t_i;\ x)$, where t_i is the head term of x and x is a modifier of t_i, we distinguished between the following situations:

- The term t_i has positive polarity, while the modifier x has negative polarity. This means that the new value propagated to t_i will change to negative. Here, modifiers negate the positive polarity of the head term and simultaneously attenuate it. The value of this attenuation, called a *diminished modifier coefficient* (DMC), enables the sentiment expressed by the modifier to be propagated. For this reason, given dependency $dep(t_i;\ x)$ and a DMC associated with x, denoted by $\beta_x^D \in [-1, 0]$, propagation is represented as follows:

  $$SO_{t_i}^{new} = SO_{t_i} * \beta_x^D \tag{10}$$

  where $\beta_x^D = \frac{SO_x}{5}$ represents conversion of the polarity of x (polarity values lie in $[-5, 5]$). Applying our approach to the example in (8), the polarity of “*heartless love*” will now be -1.2 rather than $+0.5$, with the initial value of “*love*” reduced by 160%.
- The term t_i and the modifier x both have positive polarities. This means that the new value propagated to t_i should be more positive. The same occurs when the head term t_i is negative and the modifier x is either positive or negative. Here, modifiers are used to emphasize the head term. The value of this intensification, called, in this case, the *intensified modifier coefficient* (IMC), enables the sentiment expressed by the modifier to be propagated through the syntactic structure. For this reason, given a dependency $dep(t_i;\ x)$ and an IMC associated with x denoted by $\beta_x^I \in [0, 1]$, the propagation is represented as follows:

  $$SO_{t_i}^{new} = SO_{t_i} * (1 + \beta_x^I) \tag{11}$$

  where $\beta_x^I = \frac{|SO_x|}{5}$ represents conversion of the polarity of x into an intensification. Applying our approach to example (6), the polarity of “*sensitive love*” will now be $+3.6$ rather than $+2.0$. The same occurs with the example in (7), where the polarity of “*unacceptable tragedy*” will now be -3.0 rather than -1.75. Finally, in example (9), the value of “*relevant tragedy*” will be -3.25 rather than -0.5, which corresponds to the polarity conflict reflected in Fig. 5. That means that the values for “*love*” and “*tragedy*” increase by 20% and 30%, respectively.

In this way, each modifier with positive polarity has an IMC assigned to it, whereas each modifier with negative polarity has both an IMC and a DMC assigned to it.

3.4.3. Negation

Negation permits an affirmative statement to be converted into a negative statement (Horn, 2010). Basically, an affirmative statement is used to express the validity or truth of an assertion, while a negative statement expresses its falsity (Bosque & Gutiérrez-Rexach, 2009; Morante & Sporleder, 2012).

In SA, the most common way to model negation is found in early supervised approaches to polarity classification, where texts are considered as a *bag of words* and vectors are built without any explicit knowledge of polarity, with each entry indicating the presence or frequency of individual words (Pang et al., 2002) or *n*-grams (Mejova & Srinivasan, 2011). Negation is usually detected from words like “*not*” and then modeled by simply attaching these terms to the nearest words; e.g., in “*You don't like his football skills*”, the term “*like*” is converted into “*like-NOT*”. However, the scope of the negation is usually not properly modeled. In some studies it is assumed that the negation spans from the negation token to the next punctuation mark (Pang et al., 2002) or a maximum number of words is set as being affected by negation (Hu & Liu, 2004).

Polar lexicon-based approaches go a step further in handling negation. Negation terms are considered as polarity shifters of polar expressions that produce the opposite polarity. For example, for Polanyi and Zaenen (2004), if a polar expression fell within the negation scope, its polarity value was simply inverted; thus, if the word “*perfect*” in the sentiment lexicon had a positive polarity of $+4.0$, then “*not perfect*” would score -4.0.

Such approaches have some limitations, mainly that negation is always considered to invert the polarity of the opinion expressed. However, we are of the opinion that it would sometimes be preferable to decrease it rather than directly invert it. For instance, the sentence, “*The work is not perfect*” does not mean that the work is imperfect and ridden with mistakes. Existing approaches thus try to simply approximate negation without accurately determining its scope.

Our approach was to first detect the negator term using the list of terms shown in Table 1 extracted from previous works (Carrillo de Albornoz & Plaza, 2013; Councill, McDonald, & Velikovich, 2010; Zhang, Ferrari, & Enjalbert, 2012). We ruled out negated modal auxiliaries because in using FreeLing, which lemmatizes words,

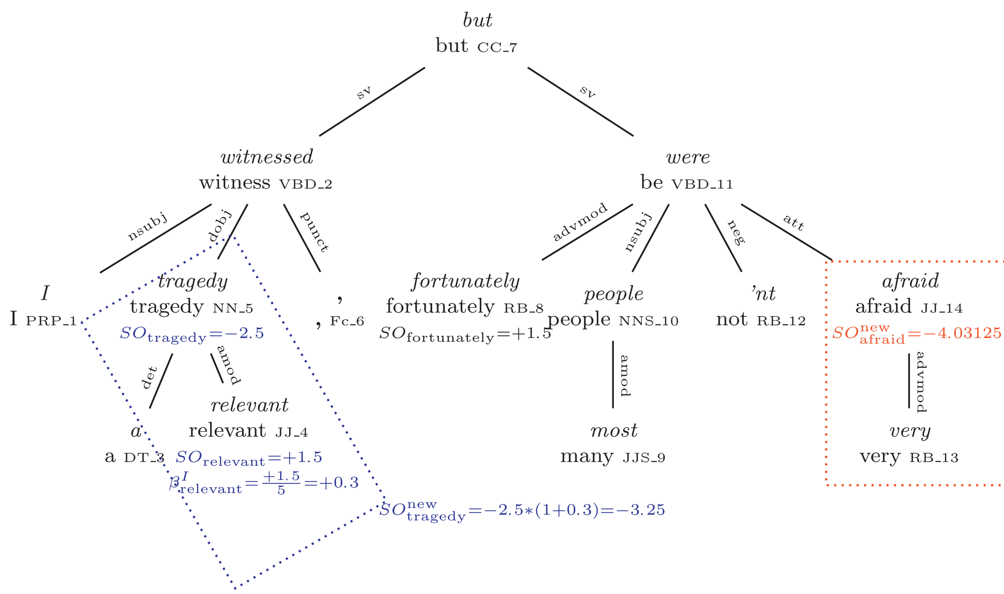

Fig. 5. Modification propagation for the running example.

Table 1
List of negator terms.

No	No one	Non	Not	Nothing	None of	Nowhere
Never	Nobody	Nor	None	Neither	Less	Without

modal auxiliaries are separated out into two words, e.g., "*can't*" and "*cannot*" into "*can*" "*not*".

We next estimated the negation scope of the negator term using the dependency tree for each sentence and key linguistic constructions. In particular, the syntactic structure created by the FreeLing parser to represent negation represents the negator term as a leaf node, with the candidate scope embracing both the head term and descendants of a sibling node. In order to detect which sibling node to choose, we applied a syntactic procedure based on dependency types (Vilares, 2013), selecting the first node occurring after the negator term according to any of the following criteria, in this order:

- Predicative complement/direct object: the first sibling node labeled with a dependency of the predicative complement type, such as in "*The sky was not outstandingly beautiful*" or as a direct object, such as in "*The sky does not have nice clouds*".
- Adjunct: the first sibling node labeled with a dependency of the adjunct type, such as in "*He is not on vacation until next week*".
- Default (when neither of the previous criteria is selected): the scope considered, besides the negator head term, is the descendant of the first sibling node of the negator node.

However, the syntactic structure sometimes has a negator term that is not a leaf node and so has children. In this special case, the candidate scope is the negator head term and its descendants. Fig. 6 illustrates the dependency structures obtained for each case.

Once scope has been identified, the effect of negation on it was considered. We applied a shift negation to both the head term and the selected descendant nodes, so each negator term had two kinds of associated values, one referring to the negation shift applied to the head term, called the *head negation coefficient* (HNC), and the other referring to the negation shifts applied to the selected descendant nodes, called the *selected negation coefficient* (SNC). Both coefficients enabled the intrinsic sentiment of the negation to be propagated through the syntactic structure. For this reason, given a negator x, with a HNC denoted by $\gamma_x^H \in (0, 5.0]$ and a SNC denoted by $\gamma_x^S \in (0, 5.0]$ and given:

- a dependency $dep(t_i; x)$, where t_i is the head term of x, and a dependency $dep(t_i; t_j)$, where t_j is a sibling node of x, which is the parent of the selected scope of the negation
  or
- a dependency $dep(t_i; x)$, where t_i is the head term of x, and a dependency $dep(x; t_j)$, where t_j is a child node of x, which is the parent of the selected scope of the negation,

the propagation is represented as follows:

$$SO_{t_i}^{new} = SO_{t_i} - H_{t_i} * \gamma_x^H + SO_{t_j}^{new} - S_{t_j} * \gamma_x^S \qquad (12)$$

where:

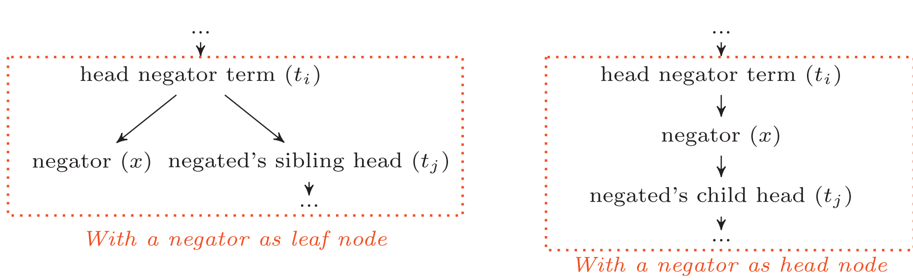

Fig. 6. Possible structures for detecting the scope of negation.

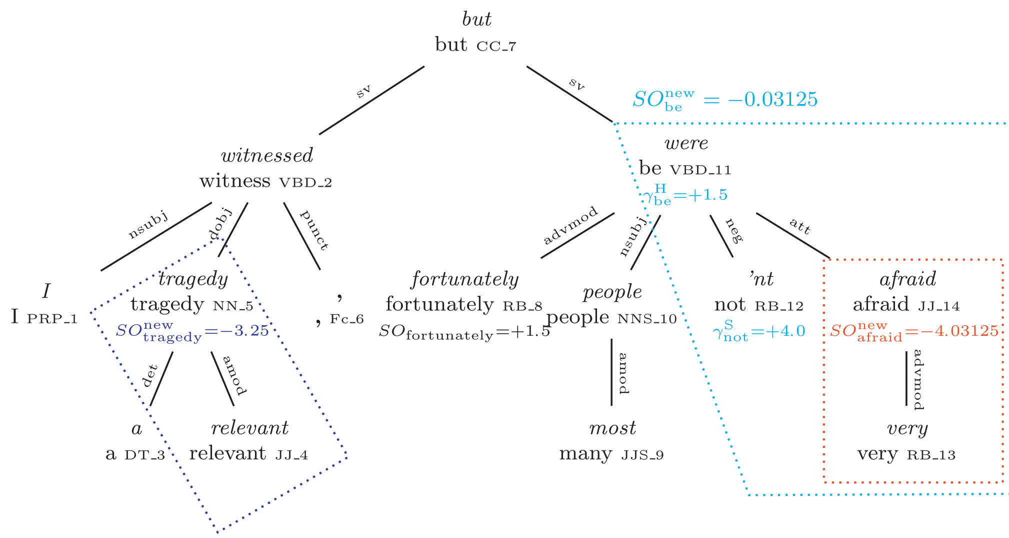

Fig. 7. Negation propagation for the running example.

- $SO^{new}_{t_j}$ represents the mean value of the sentiment polarities present in t_j descendants, expressed, in another way, as

$$SO^{new}_{t_j} = \frac{\sum_{t_k \in Q_{t_j}} SO_{t_k}}{|\{\forall t_k \; / \; t_k \in Q_{t_j} \wedge SO_{t_k} \,!= 0\}|}, \tag{13}$$

  where Q_{t_j} is the set of nodes that are descendants of t_j, including the latter.
- H_{t_i} and S_{t_j} are two constants that permit the values of $SO^{new}_{t_i}$ and $SO^{new}_{t_j}$ to be shifted through the fixed values expressed by γ^H_x and γ^S_x, respectively. These constants are

$$H_{t_i} = \begin{cases} 1 & \text{if } SO_{t_i} \geqslant 0 \\ 0 & \text{if } SO^{new}_{t_j} \,!= 0 \\ -1 & \text{otherwise} \end{cases} \qquad S_{t_j} = \begin{cases} 1 & \text{if } SO^{new}_{t_j} \geqslant 0 \\ -1 & \text{otherwise} \end{cases} \tag{14}$$

- γ^H_x and γ^S_x are both fixed values.

Thus, the new value of SO_{t_i} is the sum of the shifted value of the head term t_i and the shifted value obtained by the mean average of the descendants of t_j (which form the negation scope). In both cases, the shifted values are obtained by applying the coefficients γ^H_x and γ^S_x, with values of +1.5 and +4.0 for any negators. Experimental results showed an improvement in accuracy when these values were used.

Referring to the running example, there is a negator term "*'nt*" whose lemma is "*not*" and whose head term is "*were*". The scope is that of the descendants of the first sibling node of "*not*", which has a dependency label tagged with "*att*", as shown in Fig. 7. Applying Eq. (12), the mean value of the sentiment polarities present in the descendants of "*afraid*" is -4.03125, and the new value of SO_{be} is:
$SO^{\text{new}}_{\text{be}} = SO_{\text{be}} - (1.5 * H_{\text{be}}) + SO^{\text{new}}_{\text{afraid}} - (4 * S_{\text{afraid}}) = 0 - (1.5 * 0) + (-4.03125) - (4 * -1) = -0.03125$

3.4.4. Adversative/concessive relations

Adversative and concessive clauses are constructions that express antithetical circumstances (Crystal, 2011), which can be represented in several grammatical ways, such as through conjunctions ("*but*"), adverbs ("*however*") or prepositions ("*despite*") (Allan & Brown, 2010). If we focus on the communication itself, what is said in the main clause is always the relevant part, at least for the author. However when an adversative or a concessive clause is introduced by one of the connectors mentioned, the fact or event commented is usually in clear opposition to the main clause.

The main difference between adversative and concessive clauses is how the author's opinion is positioned within the whole sentence. For instance, it is not the same to say:

"*Bill Maher may be a little out there, but he does make some points*"

versus

"*Bill Maher may be a little out there, despite making some points*".

In both sentences, the opinion of the author concerning the related facts is found in the second dependent clause, but in the first sentence, this is an adversative clause, whereas in the second sentence, it is a concessive clause.

In SA, few existing approaches consider adversative clauses, while, to the best of our knowledge, no approach takes account of concessive ones. For example, the limitation of works that only detect structures that use an overt coordination cue and limit themselves to adversative markers like "*but*", "*however*", etc. is evident, as the polarity calculation of the overall sentence is only determined by the polarity of the adversative clause, i.e., the right member of the construction. If this is unspecified, the polarity of the left member is inverted (Poria, Cambria, Winterstein, & Huang, 2014).

Distinguishing between adversative and concessive clauses is crucial nonetheless. Adversative clauses are always found in postposition to the main clause, whereas concessive clauses can be found in both anteposition or postposition. The restriction does not always mean a limitation on the propositional content but may reflect a more precise formulation of what the speaker has in mind. The affirmation of concessive clause is then added to the reality of the preceding information in the main clause. The concessive clause remains in a position of secondary importance; it is the main clause which expresses the main facts (Rudolph, 1996), yet this does not mean that the other is discarded.

We treated the constructions as an extension of intensification propagation, where the sentiment formulated could be diminished or intensified, depending on both adversative/concessive and main clauses. In this context the entire clause and not just a term was intensified or diminished. We used a list of adversative and concessive connectors, as shown in Table 2, to identify clauses in dependency structures.

Unfortunately, the FreeLing dependency tree representing a sentence containing an adversative/concessive clause can be shown in different ways depending on whether the connector is a conjunction, an adverb or a preposition. To apply the algorithm correctly, we had to reorganize and homogenize all syntactic representations of these clauses, as shown in Fig. 8. The idea is that the connector

Table 2
Connectors for adversative and concessive clause detection.

Type	Connector	Tag	MIC	SIC	Type	Connector	Tag	MIC	SIC
Adversative	but	CC	-0.6	0.4	**Concessive**	although	CC	0.3	-0.3
	however	RB				despite	IN		
	nevertheless	RB				even if	CC		
	only	RB				even though	CC		
	still	RB				in spite of	IN		
	yet	CC				regardless of	RB		
	whereas	IN				though	IN		

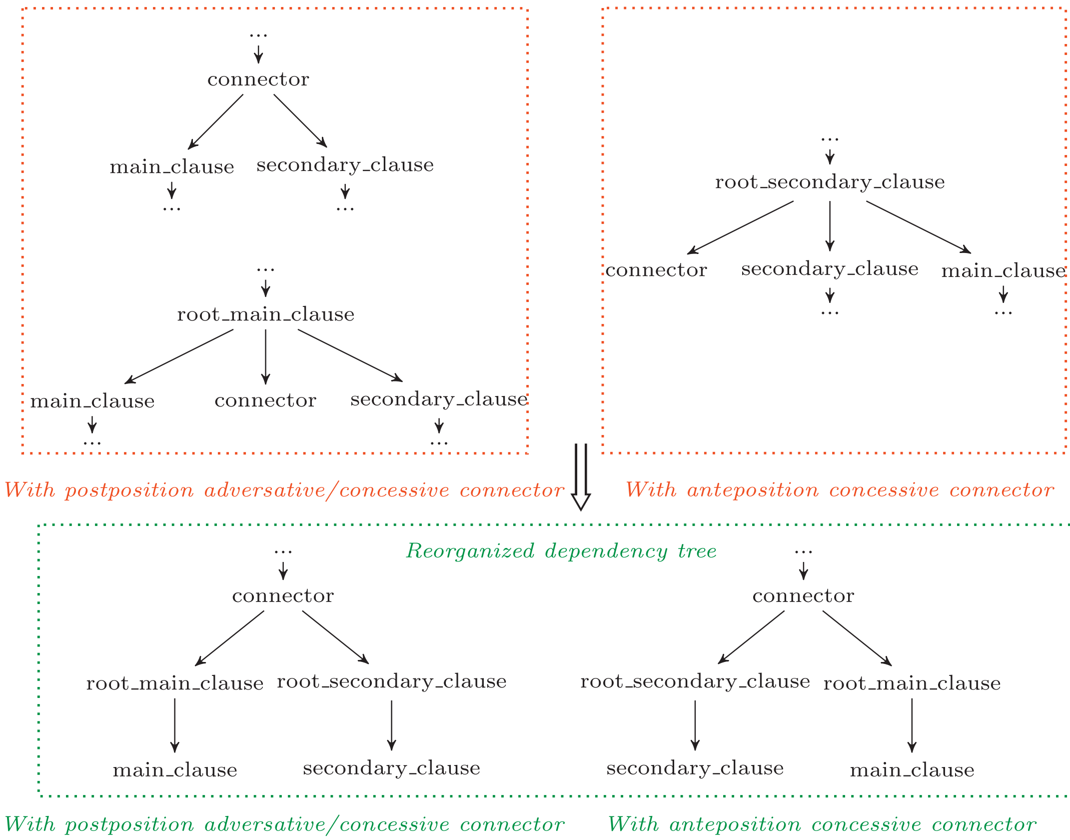

Fig. 8. Reorganization of dependencies with adversative/concessive connectors.

had to be found as head of both roots of the main and adversative/concessive clauses so that each clause (main and secondary) could propagate independently and easily.

The top part of Fig. 8 shows how the connector can be placed in different locations depending on the sentence, while the bottom part shows the result after reorganization according to the kind of connector. The connector will therefore always be the root node with two children branches: the main clause and the secondary clause in that order for adversative sentences, and the secondary clause and the main clause in that order for concessive sentences.

Once the dependencies were reorganized, two values were considered per connector, each emphasizing or attenuating the main or adversative/concessive clauses, depending on type. The first value, called the *main intensification coefficient* (MIC), enabled the intrinsic sentiment of the intensification to be propagated through the syntactic structure of the main parsed clause. The second value, called the *secondary intensification coefficient* (SIC), propagated through the syntactic structure of the adversative/concessive parsed clause.

Therefore, given an adversative connector x (concessive connector x, respectively) with a MIC denoted by $\delta_x^M \in [-1, 0]$ ($\delta_x^M \in [0, 1]$, respectively) and a SIC denoted by $\delta_x^S \in [0, 1]$ ($\delta_x^S \in [-1, 0]$, respectively) and given two dependencies $dep(x;\, t_i)$ and $dep(x;\, t_j)$, where x is the head term of t_i (which is the head of a main clause) and t_j (which is the head of an adversative/concessive clause), the propagation is represented as follows:

$$SO_x{}^{\text{new}} = SO_{t_i} * (1 + \delta_x^M) + SO_{t_j} * (1 + \delta_x^S) \quad (15)$$

where δ_x^M (δ_x^S) will be positive or negative if the semantic orientation of the main clause (the concessive/adversative clause) is intensified or diminished, respectively. In this sense, the values shown in Table 2 were obtained after testing different values between the corresponding intervals for both main and secondary clauses using steps of 0.1.

Fig. 9 shows adversative/concessive propagation applied to our running example. The connector term, "*but*", was first detected and then the roots of the main and secondary clauses. Since "*but*" is an adversative connector, the main root is "*witness*" and the secondary root is "*be*". Once located, the new sentiment value was propagated through these roots. For example, $SO_{witness}{}^{\text{new}}$ is -3.25 and $SO_{be}{}^{\text{new2}}$ is $+1.46875$. Applying the propagation, the final result is:

$SO_{but}{}^{\text{new}} = SO_{witness}{}^{\text{new}} * (1 + \gamma_{but}^M) + SO_{be}{}^{\text{new2}} * (1 + \gamma_{but}^S) = -3.25 * (1 - 0.6) + 1.46875 * (1 + 0.4) = -1.3 + 2.05625 = +0.75625$

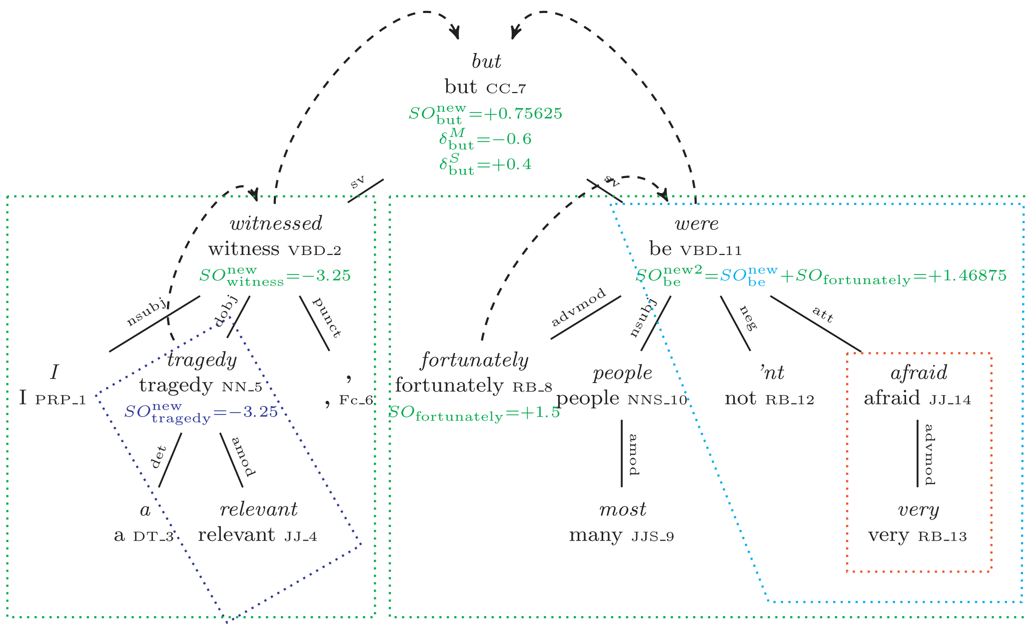

Fig. 9. Adversative/concessive propagation for the running example.

Table 3
Notation components of the sentiment propagation on DPSS.

Notation	Explanation
t_i	Term at position i.
$dep(t_i; t_j)$	Dependency between two terms t_i and t_j, where t_i is the head term of t_j.
SO_{t_i}	Initial sentiment polarity of t_i.
$SO^{new}_{t_i}$	Sentiment polarity of t_i after applying sentiment propagation on DPSS.
α_x	*Intensification coefficient* (IC) of a given intensifier or diminisher x, where $\alpha_x \in [-1, 1]$.
β^D_x	*Diminished modifier coefficient* (DMC) of a given negative modifier x, where $\beta^D_x \in [-1, 0]$.
β^I_x	*Intensified modifier coefficient* (IMC) of a given positive modifier x, where $\beta^I_x \in [0, 1]$.
γ^H_x	*Head negation coefficient* (HNC) of a given negator x, where $\gamma^H_x \in (0, 5.0]$.
γ^S_x	*Selected negation coefficient* (SNC) of a given negator x, where $\gamma^S_x \in (0, 5.0]$.
H_{t_i}	Constant that permits the value of $SO^{new}_{t_i}$ to be shifted through a fixed value expressed as a function of γ^H_x, where $H_{t_i} \in \{1\} \cup \{0\} \cup \{-1\}$.
S_{t_j}	Constant that permits the value of $SO^{new}_{t_j}$ to be shifted through a fixed value expressed as a function of γ^S_x, where $S_{t_j} \in \{1\} \cup \{-1\}$.
δ^M_x	*Main intensification coefficient* (MIC) of a given connector x applied on the main parsed clause, where $\delta^M_x \in [-1, 1]$.
δ^S_x	*Secondary intensification coefficient* (SIC) of a given connector x applied on an adversative/concessive parsed clause, where $\delta^S_x \in [-1, 1]$.

In order to facilitate the understanding of the formulas, Table 3 summarizes the notation of all elements involved in our SA approach. Finally, Table 4 shows the notation of sentiment propagation.

4. Evaluation and experimental results

4.1. Datasets

To test the performance of our SA system, we used three annotated datasets:

- The Cornell Movie Review dataset of Tagged Blogs v2.0 [9], a collection of 1,000 positive and 1,000 negative film reviews in individual text files, introduced in (Pang & Lee, 2004).
- The Obama-McCain Debate dataset [10], a collection of 3,268 tweets crawled during the presidential TV debate on September 26, 2008 between Barack Obama and John McCain (Shamma et al., 2009). Sentiment labels were acquired using Amazon Mechanical Turk, where each tweet was rated by at least three annotators as either positive, negative, mixed, or other. We considered those sentiment labels, on which two-third of the voters agreed, as final labels of the tweets. This resulted in a set with 1,196 negative and 710 positive tweets. The remainder were mixed or other tweets.
- The SemEval-2015 Task 10 dataset[11](Rosenthal et al., 2015), a collection of tweets constructed for the Twitter Sentiment Analysis Task (Task 10) of the 2015 International Workshop on Semantic Evaluation (SemEval-2015) and consisting of an official 2015 test - a set of Twitter messages (Rosenthal et al., 2015)- and a progress test - a rerun of SemEval-2014 Task 9 (Rosenthal, Ritter, Nakov, & Stoyanov, 2014) which includes Twitter messages and other kinds of texts from different domains. All the tweets were manually annotated with negative, positive and neutral labels by Amazon Mechanical Turk workers. Table 5 shows the distribution for the labels.

[9] Available at http://www.cs.cornell.edu/people/pabo/movie-review-data/.
[10] Available at https://bitbucket.org/speriosu/updown/downloads.
[11] Available at http://alt.qcri.org/semeval2015/task10/.

Table 4
Notation for representing formulas for sentiment propagation.

Propagation	Notation	Explanation
Basic	$SO_{t_i}^{new} = SO_{t_i} + \sum_{\forall t_j / \exists\ dep(t_i;t_j)} SO_{t_j}^{new}$	Propagation of the sentiment of each term from the leaves to the parent nodes of the tree represented by the DPSS until the root is reached.
Intensification	$SO_{t_i}^{new} = SO_{t_i} * (1 + \alpha_x)$	Propagation of an IC value of a given intensifier or diminisher x on a dependency, where t_i is the head term of x.
Modifier	$SO_{t_i}^{new} = SO_{t_i} * \beta_x^D$	Propagation of a DMC value from a given modifier x with a negative polarity on a dependency, t_i, with positive polarity, is the head term of x.
	$SO_{t_i}^{new} = SO_{t_i} * (1 + \beta_x^I)$	Propagation of an IMC value from a given modifier x on a dependency, where t_i is the head term of x.
Negation	$SO_{t_i}^{new} = SO_{t_i} - H_{t_i} * \gamma_x^H + SO_{t_j}^{new} - S_{t_j} * \gamma_x^S$	Propagation of an HNC and an SNC value of a given negator x applied both to the head term t_i and to the selected descendant nodes that determine the negation scope.
Adversative/Concessive	$SO_x{}^{\text{new}} = SO_{t_i} * (1 + \delta_x^M) + SO_{t_j} * (1 + \delta_x^S)$	Propagation of a MIC and an SIC value of a given connector x on main and secondary clauses, where t_i is the head term of the main clause of x and t_j is the head term of the secondary one.

Table 5
SemEval-2015 official test and progress test statistics: distribution.

Type		Positive	Negative	Neutral
Official Test	Twitter2015	1038	365	987
	TwitterSarcasm2015	18	26	16
Progress Test	Twitter2013	1572	601	1640
	SMS2013	492	394	1207
	Twitter2014	982	202	669
	LiveJournal2014	427	304	411
	TwitterSarcasm2014	33	40	13

Participants in the SemEval-2015 Task 10 used this dataset to evaluate their systems for context polarity detection and message polarity detection, among other challenges.

In this sense, the parameters of the system, such as HNC or SNC were adjusted using Regarding the question about cross validation: no. We have not based our results on cross validation. We have adjusted some parameters of our system, such as HNC or SNC , depending on experiments performed on the SemEval-2015 training dataset.

4.2. Results for the Cornell Movie Review dataset

We applied our system propagating sentiment values across dependencies. Table 6 compares our approach using just basic propagation with five variants of the lexical approach described in (Annett & Kondrak, 2008): (1) a blog post labeled as positive or negative from the dominant presence of positive or negative tokens compared to two General Inquirer[12] lists of adjectives (Stone, Dunphy, Smith, & Ogilvie, 1966); (2) a stemmer applied to each word in the dictionary (increasing its size) and to each token; (3) the addition of more words to the dictionary using a POS tagger and a search for slang adjectives and adverbs in Yahoo! Web Search API; (4) weights assigned to dictionary words using the minimum path distance from pivot words in WordNet; and (5) combination of variants (2), (3) and (4) using the baseline. We also compared our approach with a lexicon-based method described in (Zhou et al., 2014) that incorporated intensification and negation as proposed by Taboada et al. (2011), using a sentiment lexicon generated by Hu and Liu (2004). Then, we applied two variants of our approach: one employing the sentiment lexicon PolarityRank 40 (PR40), created completely with our method using a total of 40 positive and negative seeds; and the other employing the sentiment lexicons SO-CAL together with PR40, for which we calculated an average of polarities when a word existed in both lexicons.

The two variants of our method, using only basic propagation, improved on the accuracy of the best Annett result by about 7% and 9%, and by about 5% and 7% on the Zhou result. The PR40 lexicon created using our method had a vocabulary size of 24, 072 unique words, 8546 with a negative sign and 15, 526 with a positive sign. It was easy to obtain an overview of the general distribution of the words as a function of the polarities in the form of a histogram of polarity scores as shown in Fig. 10. As would be expected, most scores were rather neutral, as was the case also for the SO-CAL+PR40 lexicon, which had a vocabulary size of 25, 961 unique words, 10, 423 with a negative sign and 15, 538 with a pos-

[12] Available at http://www.wjh.harvard.edu/~inquirer/.

Table 6
Accuracy rates for seven lexicon-based propagation methods.

Approach		Accuracy
	baseline	50.00%
	baseline + Stemming	50.20%
Annett *et al.*, 2008	baseline + Yahoo! Words	57.70%
	baseline + WordNet	60.40%
	baseline + Stemming + WordNet + Yahoo! Words	55.70%
Zhou *et al.*, 2014	baseline + intensification + negation	63.20%
Our approach	our baseline + PR40	68.60%
	our baseline + SO-CAL+PR40	69.95%

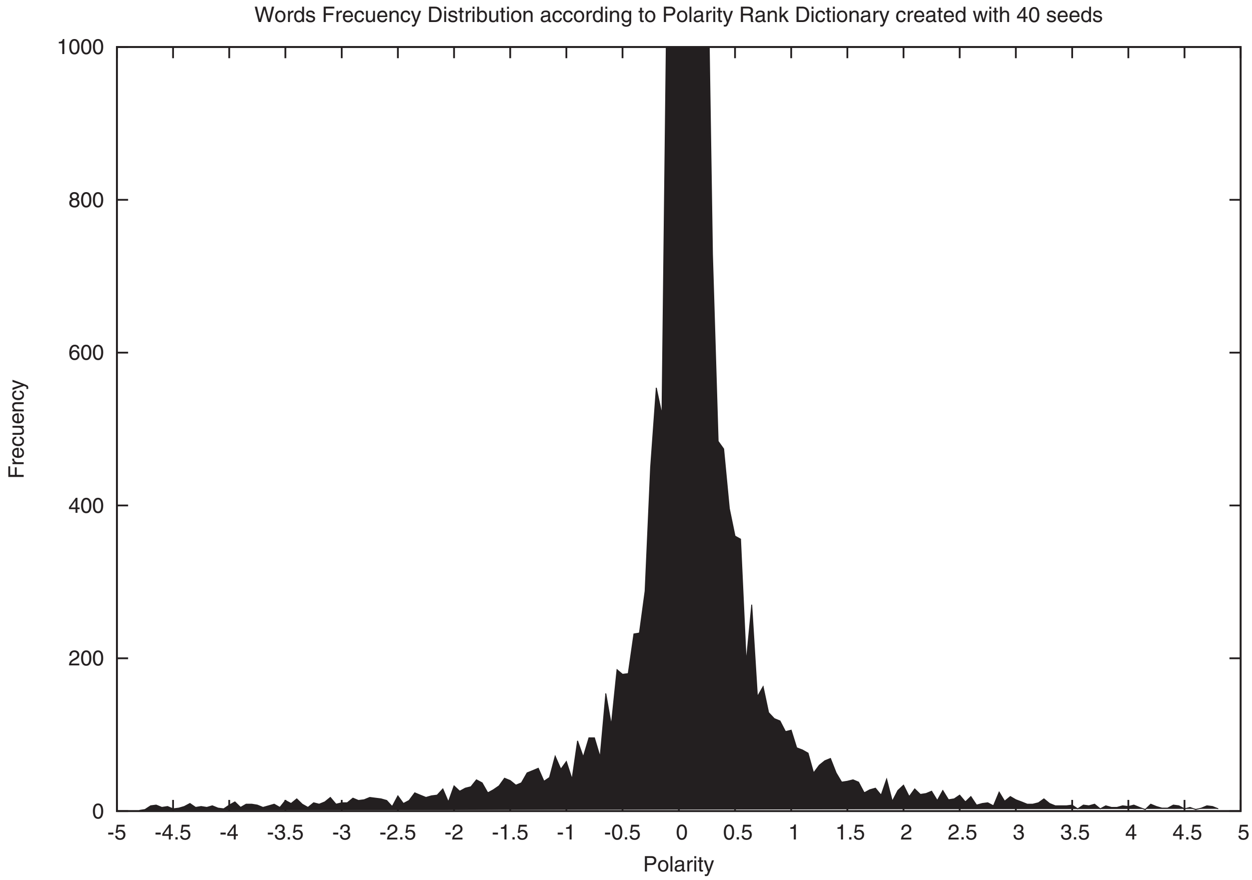

Fig. 10. Word frequency distribution for PR40.

itive sign. In the histogram in Fig. 11, integer polarity values show an increase in the number of different words due to SO-CAL.

Consequently, selecting words that are included in the dictionary was very important for the lexical approach. As mentioned by Annett and Kondrak (2008), it is important to obtain a balanced dictionary (originally the seed words were 50%-50%). Otherwise, accuracy might descend, because the distribution of the dictionary might skew the results in one direction, thereby reducing the number of correct classifications. If the dictionary is too sparse or too exhaustive, because it is imbalanced, there is the risk of over- or under-analyzing results. After performing several tests, we set the neutral sentiment intervals to $[-2.0, 2.0]$ for both lexicons. Our results demonstrate that it is possible to exceed a 65% accuracy level using a purely lexical approach (Annett & Kondrak, 2008).

We also checked the performance of our method applying different incremental propagations to the Cornell Movie Review dataset. Table 7 shows accuracy rates for the PR40 and SO-CAL+PR40 lexicons, along with precision, recall and F-measure values for positive (**P**) and negative (**N**) results. Total accuracy improved by around 2% for the PR40 lexicon and by around 5% for the SO-CAL+PR40 lexicon.

Table 8 compares results for our approach based on propagation with PR40 and SO-CAL+PR40 with results for four variants of the ML approaches described in Annett and Kondrak (2008), two variants of the non-negative matrix trifactorization approach (NMTF) for semisupervised SA described in (Li et al., 2009), a hybrid approach based on ML techniques and lexical rules described in (Carrillo de Albornoz et al., 2010) and six different unsupervised approaches described in (Zhou et al., 2014), all applied to the Cornell Movie Review dataset.

The four variants of the Annett and Kondrak (2008) approach applied POS tagging to each blog post and created a word set of the most popular adjectives and adverbs to produce different kinds of features, such as the number of positive/negative/negating words and the presence or absence of words in that word set. The blog posts were classified as positive or negative with differences in classification methods as follows (1) a fixed subset of the most frequent popular words, called unigrams, was used as the features set, considering integer frequencies; (2) only word presence or absence was taken into account; (3) three additional features were included, namely, the numbers of positive, negative and neutral words; and (4) a binary representation was assigned to the unigrams. For each variant, three ML methods were considered: SVMLight, naive Bayes and Alternating Decision Tree (ADTree).

The approach described in (Li et al., 2009) was based on NMTF employing domain-independent lexical prior knowledge in conjunction with domain-dependent unlabeled data and some labeled data. The two variants incorporated 10% and 40% of labeled data for training.

The approach described in (Carrillo de Albornoz et al., 2010) was based on ML techniques and lexical rules to classify sentences according to their polarity and intensity, working with concepts rather than terms and using an affective lexicon to label these concepts with emotional categories. The system described tackled the effect of (1) negations, (2) quantifiers and (3) both in SA evaluating polarity classification comparing the results using a logistic regression model (Logistic), a C4.5 decision tree (J48Graph) and a support vector machine (LibSVM).

The six unsupervised approaches described in (Zhou et al., 2014) were divided in three categories as follows (1) four document clustering methods without making use of a sentiment lex-

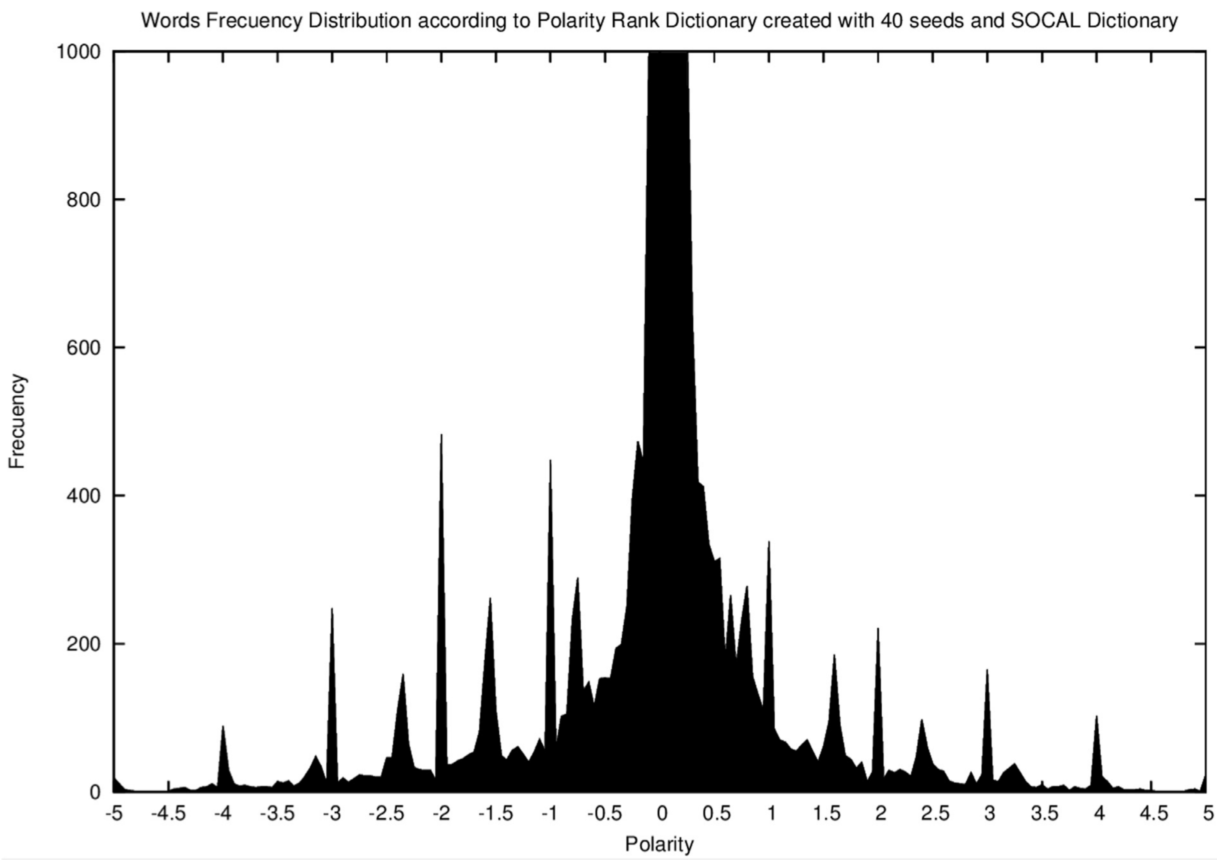

Fig. 11. Word frequency distribution for SO-CAL + PR40.

Table 7
Performance for different configurations applied to the Cornell Movie Review dataset.

		Precision		Recall		F-measure		
Type	Lexicon	P	N	P	N	P	N	Accuracy
Baseline	PR40	65.47	73.28	78.50	58.70	71.40	65.19	68.60
	SO-CAL+PR40	65.17	79.12	85.70	54.20	74.04	64.33	69.95
+intensification	PR40	65.94	73.37	78.40	59.50	71.63	65.71	68.95
	SO-CAL+PR40	65.43	79.68	86.10	54.50	74.35	64.73	70.30
+modifier	PR40	66.24	73.18	77.90	60.30	71.60	66.12	69.10
	SO-CAL+PR40	65.90	80.06	86.20	55.40	74.70	65.48	70.80
+negation	PR40	73.42	67.99	63.80	76.90	68.27	72.17	70.35
	SO-CAL+PR40	75.34	72.88	71.50	76.60	73.37	74.70	74.05
+adversative/concessive	PR40	74.27	68.12	63.50	78.00	68.46	72.73	70.75
	SO-CAL+PR40	76.38	73.40	71.80	77.80	74.02	75.53	74.80

icon, namely K-means, NMTF, information-theoretic co-clustering (ITCC) (Dhillon, Mallela, & Modha, 2003) and Euclidean co-clustering ECC (Cho, Dhillon, Guan, & Sra, 2004); (2) the constrained NMTF proposed by Li et al. (2009) incorporating the sentiment lexicon generated by Hu and Liu (2004) that contains 2,006 positive and 4,783 negative words and incorporates 10% of labeled data for training; (3) a graph co-regularized non-negative matrix trifactorization (GNMTF) method, which the assumption that if two words are sufficiently close to each other, they tend to share the sentiment polarity, encoding the geometric information by constructing the nearest neighbor graphs with a NMTF.

Comparing the results, the great advantage of our system is that it achieved reasonable accuracy without training (which requires creating a large corpus of manually tagged data). The approaches applied by Carrillo de Albornoz et al. (2010), the ML algorithm ADTree used by Annett and Kondrak (2008) were the least effective. The NMTF approach with lexical knowledge by Li et al. (2009) with 10% of labeled documents also performed poorly. All of them achieved accuracy rates worse than our PR40 and SO-CAL+PR40 approaches. Applying the latter with 40% of labeled documents together with SO-CAL+PR40 improved results. The naive Bayes algorithm performed quite well in all tests, outperforming our SO-CAL+PR40 lexicon, although only by between 0.70% and 2.70%. SVMLight also outperformed our best approach when used without aggregation, but only by between 2.20% and 2.60%. The document clustering methods K-means, NMTF and ECC, as well as NMTF with lexical knowledge used by Zhou et al. (2014), were also less effective than our approach. Even though ITCC and GNMTF achieved better results than our PR40 approach, that was not the case for SO-CAL+PR40. In comparison with the remaining ML cases, our approach obtained better results.

4.3. Results for the Obama-McCain Debate dataset

The sentiment lexicon was created in the same way as for the Cornell Movie Review dataset, from a total of 15 positive and negative seeds, employing SO-CAL and AFINN. Thus, when a word existed

Table 8
Accuracy rates for various ML, semisupervised and unsupervised (not lexicon-based) methods.

Approach			Accuracy
Annett et al., 2008	Unigram Integer	SVM-Light	77.40%
		Naive Bayes	77.10%
		ADTree	69.30%
	Unigram Binary	SVM-Light	77.00%
		Naive Bayes	75.50%
		ADTree	69.30%
	Unigram Integer + Aggregation	SVM-Light	68.20%
		Naive Bayes	77.30%
		ADTree	67.40%
	Unigram Binary + Aggregation	SVM-Light	65.40%
		Naive Bayes	77.50%
		ADTree	67.40%
Li et al., 2009	NMTF + lexical knowledge	10% doc label	60.00%
		40% doc label	73.50%
Carrillo et al., 2010	Negations	Logistic	61.60%
		J48	62.60%
		LibSVM	60.10%
	Quantifiers	Logistic	61.90%
		J48	59.50%
		LibSVM	59.00%
	Negations+Quantifiers	Logistic	62.40%
		J48	62.10%
		LibSVM	60.50%
Zhou et al., 2014	Document clustering method	K-means	54.30%
		NMTF	56.10%
		ECC	67.80%
		ITCC	71.40%
	NMTF + lexical knowledge	10% doc label	69.50%
	GNMTF	10% doc label	73.60%
Our approach	Sentiment propagation	PR40	70.75%
		SO-CAL+PR40	74.80%

Table 9
Performance of different configurations applied to the Obama-McCain Debate dataset.

	Precision		Recall		F-measure		
Type	P	N	P	N	P	N	Accuracy
Baseline	54.91	55.99	55.45	73.58	72.72	73.15	66.49%
+intensification	55.05	56.14	55.59	73.67	72.80	73.23	66.60%
+modifier	55.22	55.99	55.60	73.67	73.05	73.36	66.70%
+negation	59.53	56.84	58.15	75.06	77.07	76.05	69.54%
+adversative/concessive	59.64	56.70	58.13	75.04	77.24	76.12	69.59%

Table 10
Accuracy rates for various machine learning methods.

Approach			Accuracy
Hu et al., 2013	Lexicon-based method	GI-label	60.20%
		MPQA-label	62.90%
		intensification+negation	63.60%
	Document clustering method	K-means	56.80%
		NMTF	56.50%
	Method with emotional signals	MOODLENS	58.30%
		CSMF	65.70%
		ESSA	69.20%
Saif et al., 2014	Lexicon-based method	original	66.79%
		update	69.29%
		update+expand	69.20%
Our approach	Sentiment propagation	SO-CAL+AFINN+PR15	69.59%

in SO-CAL or AFINN and in our lexicon, its polarity was calculated as the mean of the polarities. After performing several tests on the dataset, we set the neutral sentiment interval to $[-0.25, 0.25]$.

Table 9 shows the performance of our system for the Obama-McCain Debate dataset, with different propagations enabled incrementally, for the SO-CAL+AFINN+PR15 lexicon, presenting the precision, recall and F-measure values for positive (**P**) and negative (**N**) results. Total accuracy improved by around 4% regarding the initial baseline for this lexicon.

Table 10 compares the results of our approach, based on propagation with SO-CAL+AFINN+PR15, with results of three categories of unsupervised methods described in Hu et al. (2013) and three variants described in (Saif, He, Fernandez, & Alani, 2014). We attained the best performance in terms of accuracy and F-measure. We also

Table 11
Performance for different configurations applied to the SemEval-2015 dataset.

Type		Precision			Recall			F-measure			Overall
		P	N	NEU	P	N	NEU	P	N	**NEU**	
Baseline	T'15	70.80	44.31	59.48	64.93	49.04	62.31	67.74	46.55	60.86	**57.15**
	TS'15	38.10	65.22	37.50	44.44	57.69	37.50	41.03	61.22	37.50	**51.13**
	T'13	68.32	56.34	65.97	73.66	45.09	65.85	70.89	50.09	65.91	**60.49**
	SMS'13	53.68	62.50	77.47	71.14	46.95	73.49	61.19	53.62	75.43	**57.41**
	T'14	76.58	53.47	57.31	67.92	53.47	66.82	71.99	53.47	61.70	**62.73**
	LJ'14	71.43	75.29	56.04	69.09	43.09	75.67	70.24	54.81	64.39	**62.52**
	TS'14	50.00	80.00	26.67	69.70	20.00	61.54	58.23	32.00	37.21	**45.11**
+intensification	T'15	70.85	45.12	60.04	65.80	50.68	61.80	68.23	47.74	60.91	**57.99**
	TS'15	40.91	68.18	37.50	50.00	57.69	37.50	45.00	62.50	37.50	**53.75**
	T'13	67.93	55.94	66.33	74.11	46.26	64.76	70.89	50.64	65.54	**60.76**
	SMS'13	53.21	62.20	78.92	72.36	51.78	71.67	61.33	56.51	75.12	**58.92**
	T'14	75.93	53.77	58.33	68.43	56.44	65.92	71.99	55.07	61.89	**63.53**
	LJ'14	70.71	73.37	56.69	69.56	44.41	74.21	70.13	55.33	64.28	**62.73**
	TS'14	50.00	80.00	28.57	72.73	20.00	61.54	59.26	32.00	39.02	**45.63**
+modifier	T'15	71.01	44.31	60.36	65.61	51.23	61.70	68.20	47.52	61.02	**57.86**
	TS'15	40.91	64.00	38.46	50.00	61.54	31.25	45.00	62.75	34.48	**53.87**
	T'13	68.09	56.18	66.44	74.11	46.92	64.82	70.97	51.13	65.62	**61.05**
	SMS'13	53.14	61.82	78.90	72.15	51.78	71.58	61.21	56.35	75.07	**58.78**
	T'14	76.05	53.95	58.26	68.23	57.43	65.92	71.93	55.64	61.85	**63.78**
	LJ'14	70.98	72.04	56.40	69.32	44.08	73.97	70.14	54.69	64.00	**62.42**
	TS'14	50.00	80.00	28.57	72.73	20.00	61.54	59.26	32.00	39.02	**45.63**
+negation	T'15	71.95	42.17	61.24	64.74	59.73	58.26	68.15	49.43	59.71	**58.79**
	TS'15	42.86	68.00	42.86	50.00	65.38	37.50	46.15	66.67	40.00	**56.41**
	T'13	69.23	54.87	67.21	72.26	59.07	62.50	70.71	56.89	64.77	**63.80**
	SMS'13	56.31	58.09	80.90	70.73	69.29	67.36	62.70	63.19	73.51	**62.95**
	T'14	76.62	53.85	58.98	68.43	65.84	64.28	72.30	59.24	61.52	**65.77**
	LJ'14	74.87	74.13	61.15	69.09	63.16	72.75	71.86	68.21	66.44	**70.03**
	TS'14	52.17	81.25	33.33	72.73	32.50	61.54	60.76	46.43	43.24	**53.59**
+adversative/ concessive	T'15	72.13	41.57	61.72	66.09	59.45	57.35	68.98	48.93	59.45	**58.95**
	TS'15	45.00	68.00	40.00	50.00	65.38	37.50	47.37	66.67	38.71	**57.02**
	T'13	68.54	54.56	67.69	73.73	59.73	60.43	71.04	57.03	63.85	**64.03**
	SMS'13	56.57	58.47	81.64	71.75	70.05	67.44	63.26	63.74	73.87	**63.50**
	T'14	76.27	51.94	60.03	70.06	66.34	62.18	73.04	58.26	61.09	**65.65**
	LJ'14	73.53	74.52	62.00	70.26	64.47	71.05	71.86	69.14	66.21	**70.50**
	TS'14	52.17	87.50	37.50	72.73	35.00	69.23	60.76	50.00	48.65	**55.38**

observed that performance for negative sentiment detection was always higher than for positive sentiment detection.

The first category described in (Hu et al., 2013) is a traditional lexicon-based method, which employs a word-matching scheme to perform unsupervised sentiment classification, by means of a sentiment lexicon. In this way, overall sentiment score is computed as the sum of sentiment scores of the words in the target text. The author distinguished three variants: (1) using the General Inquirer lexicon; (2) using the MPQA lexicon; and (3) incorporating intensification and negation to refine the sentiment score, as proposed by Taboada et al. (2011). The second category described by Hu et al. (2013) corresponded to methods based on document clustering, choosing as first variant K-means as baseline and then testing clustering performance on a NMTF method. Finally, the last category was further divided into three methods that incorporated emotional signals: (1) MOODLENS, which used available emoticons in a corpus as noisy label information to train a naive Bayes classifier, and applied the trained classifier to infer sentiment polarity; (2) CSMF, learning from lexical prior knowledge from domain-independent sentiment terms and domain-dependent unlabeled data; and (3) ESSA, a variant of NMTF with emotional signals.

The approach described in (Saif et al., 2014) was based on a contextual semantic approach using three adaptation settings with Thelwall-lexicon Thelwall, Buckley, Paltoglou, Cai, and Kappas (2010): using (1) the original lexicon; (2) the lexicon resulting from updating the prior sentiment of its terms; and (3) the lexicon resulting from combining the aforementioned setting with the expansion of Thelwall-lexicon with new opinionated terms.

Comparing the results of binary sentiment classification, our approach achieved an average improvement in accuracy between 5.99% and 9.39% for the lexicon-based method of Hu et al. (2013). The same occurred versus document clustering methods, where the improvement was around 13%, whereas, for methods with emotional signals , the improvement was up to 11.29%. Finally, the results are slightly better than those reported for the lexicon-based method described by Saif et al. (2014).

4.4. Results for the SemEval-2015 dataset

The sentiment lexicon was created, in the same way as for the Obama-McCain Debate dataset, from a total of 15 positive and negative seeds, and employing SO-CAL and AFINN. Thus, when a word existed in SO-CAL or AFINN and in our lexicon, its polarity was calculated as the mean of the polarities. After performing several tests on the training dataset provided by the SemEval-2015 organizers, we set the neutral sentiment interval to $[-1.0, 1.0]$.

Table 11 shows the performance of our system for the SemEval-2015 message-level task (subtask B), with different propagations enabled incrementally, presenting the overall score, as well as precision, recall and F-measure values for positive (**P**), negative (**N**) and neutral (**NEU**) results for both the official and progress tests.

All the features contributed to improving overall performance for each SemEval-2015 subset, with differences in performance depending on the kind of text. However, since our results were adjusted, we can confirm that our approach was not biased for any particular result, performing quite well for all three types of polarity. In general, our approach showed itself to be quite stable, as it generated similar results for the different kinds of texts evaluated.

Also notable were the good results obtained for sarcastic tweets by our system in the SemEval-2015 shared task, which obtained us

Table 12
Ranking provided by the SemEval organization of our approach for each test.

System	T'15	TS'15	T'13	SMS'13	T'14	LJ'14	TS'14
Webis	**64.84**$_1$	53.59$_{22}$	68.49$_{10}$	63.92$_{14}$	70.86$_7$	71.64$_{14}$	49.33$_{12}$
IOA	62.62$_7$	**65.77**$_1$	71.32$_5$	68.14$_3$	71.86$_4$	74.52$_2$	51.48$_9$
Splusplus	63.73$_5$	60.99$_7$	**72.80**$_1$	67.16$_5$	**74.42**$_1$	**75.34**$_1$	42.36$_{31}$
ECNU	59.72$_{18}$	52.67$_{23}$	65.25$_{23}$	**68.49**$_1$	66.37$_{20}$	74.40$_3$	45.87$_{25}$
Gradiant	60.62$_{16}$	56.45$_{16}$	65.29$_{22}$	61.97$_{21}$	66.87$_{17}$	72.63$_{11}$	**59.11**$_1$
Our approach	58.95$_{21}$	57.02$_{13}$	64.03$_{25}$	63.50$_{16}$	65.65$_{22}$	70.50$_{17}$	55.38$_6$
Submission means	57.07	52.12	63.22	60.19	64.88	68.13	47.02

Table 13
Matching matrix for results of SemEval-2015.

		Detected					Detected		
		P	Neu	N			P	Neu	N
T'15	P	66.09%	24.76%	9.15%	**TS'15**	P	50.00%	27.78%	22.22%
	Neu	21.38%	57.35%	21.28%		Neu	37.50%	37.50%	25.00%
	N	14.79%	25.75%	59.45%		N	14.36%	19.31%	66.34%
T'13	P	73.73%	19.72%	6.55%	**SMS'13**	P	71.75%	19.31%	8.94%
	Neu	27.62%	60.43%	11.95%		Neu	19.97%	67.44%	12.59%
	N	13.14%	27.12%	59.73%		N	7.61%	22.34%	70.05%
T'14	P	70.06%	24.24%	5.7%	**LJ'14**	P	70.26%	23.65%	6.09%
	Neu	27.65%	62.18%	10.16%		Neu	18.98%	71.05%	9.98%
	N	14.36%	19.31%	66.34%		N	9.87%	25.66%	64.47%
TS'14	P	72.73%	27.27%	0%					
	Neu	15.38%	69.23%	15.38%					
	N	50.00%	15.00%	35.00%					

the thirteenth (test dataset - TS'15) and sixth positions (progress dataset - TS'14) in subtask B, out of submissions by 40 participants, as shown in Table 12. For comparison purposes, this table also includes the top-ranked systems in each category as well as the mean scores across all submissions. The SA methods for the other approaches, unlike our approach, required problem-dependent supervision based on ML algorithms. Taking into account that no training was performed, the results achieved by our approach can be considered as satisfactory, since they were consistently better than the mean of the results of the submissions.

Table 13 , which shows the matching matrix of results for SemEval-2015, shows how our approach behaved for individual labels. It can be observed that when a tweet was wrongly classified, for instance, as positive in T'15, we usually labeled it as neutral, rarely negative. The same occurred with errors for negative labels, which were almost always labeled as neutral.

5. Conclusions and future work

We created an approach to SA that detects the sentiment both in short messages such as tweets and SMS and also reviews. Its main characteristic is that it is unsupervised and can be applied to subject domains other than the domain it was designed for. It also has the advantage that it does not need labeled text for prior training. Our approach performed in the SemEval-2015 Sentiment Analysis in Twitter Task 10 and demonstrated its performance on the SemEval-2014 progress, Obama-McCain Debate and Cornell Movie Review datasets.

Our approach was based on determining dependencies between lemmatized tagged words using a sentiment propagation algorithm that took into account and distinguished between key linguistic phenomena, namely, intensification, modification, negation and adversative and concessive relations. All of these, but especially negation, play an important role in SA. Our approach was more sophisticated (and therefore more precise) than the simplistic approach of reversing word polarity, as has been done in other studies. In our approach, we first detected the negation scope and then negated this scope. In this way we were able to significantly improve, by between 3 and 7 percentage points, the overall performance of our SA system for all the datasets.

To the best of our knowledge, this is the first time that such linguistic aspects are all taken into account in sentiment propagation and the first time that a distinction is drawn between adversative and concessive clauses. Another advantage of our approach derives from the use of sentiment dictionaries that yield polarities for words. Although some generic lexicons can be found on the Internet, they are not context-based, which is why we applied a context-based algorithm that automatically created a dictionary for each particular context. Our experiments showed that newly created lexicons are superior in sentiment prediction using our unsupervised approach.

Experiments using various datasets including the dataset from the SemEval-2015 workshop where this system was presented (Fernández-Gavilanes, Álvarez-López, Juncal-Martínez, Costa-Montenegro, & González-Castaño, 2015) demonstrate quite satisfactory results for our algorithm that compare favorably with the results obtained by supervised algorithms. These supervised algorithms achieve somewhat better accuracy rates, but have the drawback that they require corpora to be specifically annotated for each domain of interest, which in turn requires human judgment for the labeling of documents for training. They also have the disadvantages of taking more time for algorithm training and being highly dependent on the quantity and quality of the features set.

In view of the results and conclusions drawn regarding our approach to SA, we plan to further develop this line of research by including new sentiment propagation modules. We also are of the opinion that our unsupervised method could be used to train supervised ML methods or could be combined with ML methods to develop a hybrid system that would improve SA results for tweets and other opinion texts. We also plan to adapt our sentiment analysis system to languages other than English.

Another future research line is sentiment detection regarding different aspects in the same domain. With the current system we only detect the global sentiment of a text, but it would also be interesting to carry on a deeper study, in order to provide the sentiment about the different targets that can be found within the same domain. For doing this, dictionary generation also needs to be reconsidered in order to provide different polarities for the same word, depending on the entities referred.

Acknowledgments

This work was supported by the Spanish Government, co-financed by the European Regional Development Fund (ERDF) under the TACTICA project.

References

Allan, K., & Brown, K. (2010). *Concise encyclopedia of semantics*. *Concise Encyclopedias of Language and Linguistics*. Elsevier Science.

Annett, M., & Kondrak, G. (2008). A comparison of sentiment analysis techniques: Polarizing movie blogs. In *Proc. of the canadian society for computational studies of intelligence, 21st conference on advances in artificial intelligence*. In *Canadian AI'08* (pp. 25–35). Berlin, Heidelberg: Springer-Verlag.

Atserias, J., Casas, B., Comelles, E., González, M., Padró, L., & Padró. , M. (2006). FreeLing 1.3: Syntactic and semantic services in an open-source NLP library. In *Proc. of the 5th international conference on language resources and evaluation (LREC 2006)*. Genoa, Italy: ELRA.

Aue, A., & Gamon, M. (2005). Customizing sentiment classifiers to new domains: A case study. In *Proc. of recent advances in natural language processing (ranlp)*.

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transaction on Pattern Analysis and Machine Intelligence, 35*(8), 1798–1828. doi:10.1109/TPAMI.2013.50.

Bosque, I., & Gutiérrez-Rexach, J. (2009). *Fundamentos de sintaxis formal*. *Akal universitaria: Serie Lingüística*. Ed. Akal. Madrid.

Bothos, E., Apostolou, D., & Mentzas, G. (2010). Using social media to predict future events with agent-based markets.. *IEEE Intelligent Systems, 25*(6), 50–58. URL: http://dblp.uni-trier.de/db/journals/expert/expert25.html#BothosAM10.

Brants, T. (2000). Tnt: A statistical part-of-speech tagger. In *Proc. of the 6th conference on applied natural language processing*. In *ANLC '00* (pp. 224–231). Stroudsburg, USA: ACL.

Brooke, J. (2009). A semantic approach to automated text sentiment analysis. Simon Fraser University Ph.D. thesis.

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

Carrillo de Albornoz, J., & Plaza, L. (2013). An emotion-based model of negation, intensifiers, and modality for polarity and intensity classification. *JASIST, 64*(8), 1618–1633. doi:10.1002/asi.22859.

Carrillo de Albornoz, J., Plaza, L., & Gervás, P. (2010). A hybrid approach to emotional sentence polarity and intensity classification. In *Proc. of the 14th Conference on Computational Natural Language Learning (CoNLL 2010)* (pp. 153–161).

Cho, H., Dhillon, I. S., Guan, Y., & Sra, S. (2004). Minimum sum-squared residue co-clustering of gene expression data. In M. W. Berry, U. Dayal, C. Kamath, & D. B. Skillicorn (Eds.), *Sdm* (pp. 114–125). Lake Buena Vista, Florida, USA: SIAM.

Councill, I., McDonald, R., & Velikovich, L. (2010). What's great and what's not: learning to classify the scope of negation for improved sentiment analysis. In *Proc. of the workshop on negation and speculation in natural language processing* (pp. 51–59). Uppsala, Sweden: University of Antwerp. URL: http://www.aclweb.org/anthology/W/W10/W10-3110.

Cruz, F., Vallejo, C., Enríquez, F., & Troyano, J. (2012). Polarityrank: Finding an equilibrium between followers and contraries in a network. *Information Processing & Management, 48*(2), 271–282.

Cruz, F. L., Troyano, J. A., Ortega, F. J., & Enríquez, F. (2011). Automatic expansion of feature-level opinion lexicons. In *Proc. of the 2nd workshop on computational approaches to subjectivity and sentiment analysis*. In *WASSA '11* (pp. 125–131). Stroudsburg, USA: ACL.

Crystal, D. (2011). *Dictionary of linguistics and phonetics*. *The Language Library*. Malden, MA: Blackwell Pub.

Dhillon, I. S., Mallela, S., & Modha, D. S. (2003). Information-theoretic co-clustering. In *Proc. of the 9th acm sigkdd international conference on knowledge discovery and data mining*. In *KDD '03* (pp. 89–98). New York, USA: ACM.

Dragut, E., & Fellbaum, C. (2014). The role of adverbs in sentiment analysis. In *Proc. of frame semantics in nlp: A workshop in honor of chuck fillmore (1929-2014)* (pp. 38–41). Baltimore, MD, USA: ACL.

Esuli, A., & Sebastiani, F. (2006). Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proc. of the 5th conference on language resources and evaluation (lrec'06)* (pp. 417–422).

Esuli, A., & Sebastiani, F. (2007). Random-walk models of term semantics: An application to opinion-related properties. In *Proc. of ltc 2007*.

Fahrni, A., & Klenner, M. (2008). Old wine or warm beer: target-specific sentiment analysis of adjectives. In *Symposium on affective language in human and machine, aisb 2008 convention* (pp. 60–63). Affective Language in Human and Machine, Proc. Vol. 2; AISB 2008 Convention: Communication, Interaction and Social Intelligence; 1st-4th April 2008; University of Aberdeen

Fernández-Gavilanes, M., Álvarez-López, T., Juncal-Martínez, J., Costa-Montenegro, E., & González-Castaño, F. J. (2015). GTI: An unsupervised approach for sentiment analysis in Twitter. In *Proc. of the 9th international workshop on semantic evaluation (semeval 2015)* (pp. 533–538). Denver, Colorado: ACL. URL: http://www.aclweb.org/anthology/S15-2089.

Frijda, N. (1986). *The emotions*. *Studies in Emotion and Social Interaction*. New York, USA: Cambridge University Press.

Hatzivassiloglou, V., & McKeown, K. R. (1997). Predicting the semantic orientation of adjectives. In *Proc. of the 35th annual meeting of the acl and 8th conference of the european chapter of the acl*. In *ACL '98* (pp. 174–181). Stroudsburg, USA: ACL.

Horn, L. R. (2010). *The expression of negation*. *Expression of cognitive categories*. Berlin, Boston: De Gruyter Mouton.

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In *Proc. of the 10th acm sigkdd international conference on knowledge discovery and data mining*. In *KDD '04* (pp. 168–177). New York, USA: ACM.

Hu, X., Tang, J., Gao, H., & Liu, H. (2013). Unsupervised sentiment analysis with emotional signals. In *Proc. of the 22nd international conference on world wide web*. In *WWW '13* (pp. 607–618). Canton of Geneva, Switzerland: Int. World Wide Web Conf. Steering.

Huang, S., Niu, Z., & Shi, C. (2014). Automatic construction of domain-specific sentiment lexicon based on constrained label propagation. *Knowledge-Based Systems, 56*, 191–200.

Huddleston, R., & Pullum, G. (2002). *The Cambridge grammar of the english language*. *Cambridge textbooks in linguistics*. New York, USA: Cambridge University Press. URL: http://books.google.es/books?id=2yoQhHikxE8C.

Jansen, B. J., Zhang, M., Sobel, K., & Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth. *Journal of the American Society for Information Science and Technology, 60*(11), 2169–2188.

Kamps, J., Marx, M., Mokken, R. J., & de Rijke, M. (2004). Using WordNet to measure semantic orientation of adjectives. In *Lrec 2004: 4* (pp. 1115–1118). URL: http://citeseer.ist.psu.edu/kamps04using.html.

Kennedy, A., & Inkpen, D. (2006). Sentiment classification of movie reviews using contextual valence shifters. *Computational Intelligence, 22*(2), 110–125. (Special Issue on Sentiment Analysis)

Kim, S.-M., & Hovy, E. (2004). Determining the sentiment of opinions. In *Proc. of the 20th international conference on computational linguistics*. In *COLING '04*. Stroudsburg, USA: ACL.

Klenner, M., Tron, S., Amsler, M., & Hollenstein, N. (2014). The detection and analysis of bi-polar phrases and polarity conflicts. In *Proc. of 11th international workshop on natural language processing and cognitive science*. Venezia, Italy: De Gruyter. URL: http://dx.doi.org/10.5167/uzh-99629.

König, A. C., & Brill, E. (2006). Reducing the human overhead in text categorization. In *Proc. of the 12th acm sigkdd international conference on knowledge discovery and data mining*. In *KDD '06* (pp. 598–603). New York, USA: ACM.

Li, T., Zhang, Y., & Sindhwani, V. (2009). A non-negative matrix tri-factorization approach to sentiment classification with lexical prior knowledge. In *Proc. of the joint conference of the 47th annual meeting of the acl and the 4th international joint conference on natural language processing of the afnlp: Volume 1 - volume 1*. In *ACL '09* (pp. 244–252). Stroudsburg, USA: ACL.

Liu, B. (2012). *Sentiment analysis and opinion mining*. *Synthesis digital library of engineering and computer science*. Morgan & Claypool. URL: http://books.google.es/books?id=Gt8g72e6MuEC.

Liu, K., Xu, L., & Zhao, J. (2015). Co-extracting opinion targets and opinion words from online reviews based on the word alignment model. *IEEE Transactions on Knowledge & Data Engineering, 27*(3), 636–650.

Lu, Y., Castellanos, M., Dayal, U., & Zhai, C. (2011). Automatic construction of a context-aware sentiment lexicon: An optimization approach. In *Proceedings of the 20th international conference on world wide web*. In *WWW '11* (pp. 347–356). New York, NY, USA: ACM. doi:10.1145/1963405.1963456.

Mandel, B., Culotta, A., Boulahanis, J., Stark, D., Lewis, B., & Rodrigue, J. (2012). A demographic analysis of online sentiment during hurricane irene. In *Proc. of the second workshop on language in social media*. In *LSM '12* (pp. 27–36). Stroudsburg, USA: ACL. URL: http://dl.acm.org/citation.cfm?id=2390374.2390378.

Maynard, D., & Funk, A. (2012). Automatic detection of political opinions in tweets. In *Proc. of the 8th international conference on the semantic web*. In *ESWC'11* (pp. 88–99). Berlin, Heidelberg: Springer-Verlag.

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal, 5*(4), 1093–1113. URL: http://www.sciencedirect.com/science/article/pii/S2090447914000550.

Mejova, Y., & Srinivasan, P. (2011). Exploring feature definition and selection for sentiment classifiers. . In L. A. Adamic, R. A. Baeza-Yates, & S. Counts (Eds.), *Proc. of the 5th International Conference on Weblogs and Social Media, Barcelona, Catalonia, Spain, July 17–21*. San Francisco, California USA: The AAAI Press.

Min, H.-J., & Park, J. C. (2011). Detecting and blocking false sentiment propagation. In *Proc. of 5th international joint conference on natural language processing* (pp. 354–362). Chiang Mai, Thailand: Asian Federation of Natural Language Processing.

Missen, M. M. S., & Boughanem, M. (2009). Using wordnet's semantic relations for opinion detection in blogs. In M. Boughanem, C. Berrut, J. Moth, & C. Soulé-Dupuy (Eds.), *Advances in Information Retrieval: 31th European Conference on IR Research, ECIR 2009, Toulouse, France, April 6–9, LNCS 5478* (pp. 729–733). Berlin, Heidelberg: Springer.

Mohammad, S., & Turney, P. D. (2013). Crowdsourcing a word-emotion association lexicon. *Computational Intelligence, 29*(3), 436–465. URL: http://dblp.uni-trier.de/db/journals/ci/ci29.html#MohammadT13.

Mohammad, S. M., Kiritchenko, S., & Zhu, X. (2013). Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets. *CoRR, abs/1308.6242*. URL: http://dblp.uni-trier.de/db/journals/corr/corr1308.html#MohammadKZ13.

Moilanen, K., & Pulman, S. (2007). Sentiment composition. In *Proc. of ranlp 2007*.Borovets, Bulgaria.

Morante, R., & Sporleder, C. (2012). Modality and negation: An introduction to the special issue. *Computational Linguistics, 38*(2), 223–260. doi:10.1162/COLI_a_00095.

Moreno Ortiz, A., & Pérez Hernández, C. (2013). Lexicon-based sentiment analysis of twitter messages in spanish.. *Procesamiento del Lenguaje Natural, 50*, 93–100. URL: http://dblp.uni-trier.de/db/journals/pdln/pdln50.html#OrtizH13.

Musto, C., Semeraro, G., & Polignano, M. (2014). A comparison of lexicon-based approaches for sentiment analysis of microblog posts. *Proc. of the 8th International Workshop on Information Filtering and Retrieval, 1314*, 59.

Nakagawa, T., Inui, K., & Kurohashi, S. (2010). Dependency tree-based sentiment classification using crfs with hidden variables. In *Human language technologies: The 2010 annual conference of the north american chapter of the association for computational linguistics*. In *HLT '10* (pp. 786–794). Stroudsburg, PA, USA: Association for Computational Linguistics. URL: http://dl.acm.org/citation.cfm?id=1857999.1858119.

Nasukawa, T., & Yi, J. (2003). Sentiment analysis: Capturing favorability using natural language processing. In *Proc. of the 2nd international conference on knowledge capture*. In *K-CAP '03* (pp. 70–77). New York, USA: ACM.

Neviarouskaya, A., Prendinger, H., & Ishizuka, M. (2010). Recognition of affect, judgment, and appreciation in text. In *Proc. of the 23rd international conference on computational linguistics*. In *COLING '10* (pp. 806–814). Stroudsburg, USA: ACL.

Nguyen, T. H., Shirai, K., & Velcin, J. (2015). Sentiment analysis on social media for stock movement prediction. *Expert Systems with Applications, 42*(24), 9603–9611.

Nielsen, F. Å. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. In *Making sense of microposts (#MSM2011)* (pp. 93–98). URL: http://ceur-ws.org/Vol-718/paper_16.pdf.

Padró, L., & Stanilovsky, E. (2012). Freeling 3. 0: Towards wider multilinguality. In N. C. C. Chair), K. Choukri, T. Declerck, M. U. Doğan, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, & S. Piperidis (Eds.), *Proc. of the 8th international conference on language resources and evaluation (lrec'12)*. Istanbul, Turkey: ELRA.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1998). *The PageRank citation ranking: Bringing order to the web*. *Technical Report;*. Stanford University.

Pak, A., & Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. In N. C. C. Chair), K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, & D. Tapias (Eds.), *Proc. of the 7th international conference on language resources and evaluation (lrec'10)*. Valletta, Malta: ELRA.

Pang, B., & Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proc. of the 42nd annual meeting on acl*. In *ACL '04*. Stroudsburg, USA: ACL.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundation and Trends in Information Retrieval, 2*(1-2), 1–135.

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up?: Sentiment classification using machine learning techniques. In *Proc. of the acl-02 conference on empirical methods in natural language processing - volume 10*. In *EMNLP '02* (pp. 79–86). Stroudsburg, USA: ACL. doi:10.3115/1118693.1118704.

Polanyi, L., & Zaenen, A. (2004). Contextual valence shifters. In *Working notes — exploring attitude and affect in text: Theories and applications (aaai spring symposium series)*.

Poria, S., Cambria, E., Winterstein, G., & Huang, G.-B. (2014). Sentic patterns: Dependency-based rules for concept-level sentiment analysis. *Knowledge-Based Systems, 69*(0), 45–63.

Qiu, G., Liu, B., Bu, J., & Chen, C. (2011). Opinion word expansion and target extraction through double propagation. *Computational Linguistics, 37*(1), 9–27. doi:10.1162/coli_a_00034.

Quinn, K. M., Monroe, B. L., Colaresi, M., Crespin, M. H., & Radev, D. R. (2010). How to Analyze Political Attention with Minimal Assumptions and Costs. *American Journal of Political Science, 54*(1), 209–228.

Rao, D., & Ravichandran, D. (2009). Semi-supervised polarity lexicon induction. In *Proc. of the 12th conference of the european chapter of the acl*. In *EACL '09* (pp. 675–682). Stroudsburg, USA: ACL.

Read, J., & Carroll, J. (2009). Weakly supervised techniques for domain-independent sentiment classification. In *Proc. of the 1st international cikm workshop on topic-sentiment analysis for mass opinion*. In *TSA '09* (pp. 45–52). New York, USA: ACM.

Rosenthal, S., Nakov, P., Kiritchenko, S., Mohammad, S. M., Ritter, A., & Stoyanov, V. (2015). Semeval-2015 task 10: Sentiment analysis in twitter. In *Proc. of the 9th international workshop on semantic evaluation*. In *SemEval '2015*. Denver, Colorado

Rosenthal, S., Ritter, A., Nakov, P., & Stoyanov, V. (2014). Semeval-2014 task 9: Sentiment analysis in twitter. In *Proc. of the 8th international workshop on semantic evaluation (semeval 2014)* (pp. 73–80). Dublin, Ireland: ACL and Dublin City University.

Rudolph, E. (1996). *Contrast: Adversative and concessive relations and their expressions in English, German, Spanish, Portuguese on sentence and text level*. *Research in text theory*. Walter de Gruyter.

Saif, H., He, Y., Fernandez, M., & Alani, H. (2014). Adapting sentiment lexicons using contextual semantics for sentiment analysis of twitter. In V. Presutti, E. Blomqvist, R. Troncy, H. Sack, I. Papadakis, & A. Tordai (Eds.), *The semantic web: Eswc 2014 satellite events*. In *LNCS: 8798* (pp. 54–63). Berlin, Heidelberg: Springer Int. Publishing.

Salathé, M., & Khandelwal, S. (2011). Assessing vaccination sentiments with online social media: Implications for infectious disease dynamics and control.. *PLoS Computational Biology, 7*(10).

Severyn, A., & Moschitti, A. (2015). Unitn: Training deep convolutional neural network for twitter sentiment classification. In *Proc. of the 9th international workshop on semantic evaluation (semeval 2015)* (pp. 464–469). Denver, Colorado: ACL.

Shamma, D. A., Kennedy, L., & Churchill, E. F. (2009). Tweet the debates: Understanding community annotation of uncollected sources. In *Proc. of the first sigmm workshop on social media*. In *WSM '09* (pp. 3–10). New York, USA: ACM.

Stone, P. J., Dunphy, D. C., Smith, M. S., & Ogilvie, D. M. (1966). The general inquirer: A Computer Approach to Content Analysis. MIT Press. URL: http://www.webuse.umd.edu:9090/.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics, 37*(2), 267–307. doi:10.1162/COLI_a_00049.

Tan, S., & Wu, Q. (2011). A random walk algorithm for automatic construction of domain-oriented sentiment lexicon.. *Expert Systems with Application, 38*(10), 12094–12100. URL: http://dblp.uni-trier.de/db/journals/eswa/eswa38.html#TanW11a.

Tang, D., Wei, F., Qin, B., Zhou, M., & Liu, T. (2014a). Building large-scale twitter-specific sentiment lexicon : A representation learning approach. In *Proc. of coling 2014, the 25th international conference on computational linguistics: Technical papers* (pp. 172–182). Dublin, Ireland: Dublin City University and ACL. URL: http://www.aclweb.org/anthology/C14-1018.

Tang, D., Wei, F., Yang, N., Zhou, M., Liu, T., & Qin, B. (2014b). Learning sentiment-specific word embedding for twitter sentiment classification. In *Proc. of the 52nd annual meeting of the acl 2014, baltimore, md, usa, volume 1: Long papers* (pp. 1555–1565). URL: http://aclweb.org/anthology/P/P14/P14-1146.pdf.

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas, A. (2010). Sentiment in short strength detection informal text. *Journal of the American Society for Information Science and Technology, 61*(12), 2544–2558. doi:10.1002/asi.v61:12.

Tron, S. (2013). A verb-centered sentiment analysis for french. Zürich University Ph.D. thesis..

Tsytsarau, M., Palpanas, T., & Denecke, K. (2010). Scalable discovery of contradictions on the web. In *Proc. of the 19th international conference on world wide web, WWW 2010, raleigh, north carolina, usa, april 26-30, 2010* (pp. 1195–1196).

Turney, P., & Littman, M. (2002). *Unsupervised learning of semantic orientation from a hundred-billion-word corpus*. *Technical Report NRC Technical Report ERB-1094*. Institute for Information Technology, National Research Council Canada.

Turney, P. D. (2002). Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews. In *Proc. of the 40th annual meeting on acl*. In *ACL '02* (pp. 417–424). Stroudsburg, USA: ACL.

Vilares, D. (2013). Sentiment analysis for reviews and microtexts based on lexico-syntactic knowledge. In *Proc. of 5th bcs-irsg symposium on future directions in information access*. In *FDIA '13* (pp 38–43).

Vohra, S., & Teraiya, J. (2013). A comparative study of sentiment analysis techniques. *Information, Knowledge and Research in Computer Engineering, 2*(2), 313–317.

Whitelaw, C., Garg, N., & Argamon, S. (2005). Using appraisal groups for sentiment analysis. In *Proc. of the 14th acm international conference on information and knowledge management*. In *CIKM '05* (pp. 625–631). New York, USA: ACM.

Wiebe, J., Wilson, T., & Cardie, C. (2005). Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation, 1*(2), 0. URL: http://www.cs.pitt.edu/~wiebe/pubs/papers/lre05withappendix.pdf.

Zhang, L., Ferrari, S., & Enjalbert, P. (2012). Opinion analysis: The effect of negation on polarity and intensity. In J. Jancsary (Ed.), *Proc. of konvens 2012* (pp. 282–290). ÖGAI.

Zhang, Z., & Singh, M. P. (2014). Renew: A semi-supervised framework for generating domain-specific lexicons and sentiment analysis. In *Proceedings of the 52nd annual meeting of the association for computational linguistics (volume 1: Long papers)* (pp. 542–551). Baltimore, Maryland: Association for Computational Linguistics. URL: http://www.aclweb.org/anthology/P/P14/P14-1051.

Zhou, G., Zhao, J., & Zeng, D. (2014). Sentiment classification with graph co-regularization. In *COLING 2014, 25th international conference on computational linguistics, proc. of the conference: Technical papers, august 23-29, 2014, dublin, ireland* (pp. 1331–1340).

Update

Expert Systems With Applications

Volume 61, Issue , 1 November 2016, Page 394

DOI: https://doi.org/10.1016/j.eswa.2016.06.007

Expert Systems With Applications 61 (2016) 394

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Corrigendum

Corrigendum to "Unsupervised method for sentiment analysis in online texts" Expert Systems with Applications 58 (2016) 57–75

Milagros Fernández-Gavilanes*, Tamara Álvarez-López, Jonathan Juncal-Martínez, Enrique Costa-Montenegro, Francisco Javier González-Castaño

AtlantTIC, *University of Vigo*, Campus, 36310 Vigo, *Spain*

The authors regret that the printed version of the above article contained a number of errors. The correct and final version follows. The authors would like to apologize for any inconvenience caused.

The corrections of the article are:

- In p. 61, Section 3.4.1, paragraph 4th: the sentence *"We used a list of intensifiers and diminishers, adapted from (Brooke, 2009), where each element is a modifier that emphasizes or attenuates words."* was replaced by *"We used a list of intensifiers and diminishers, adapted from (Brooke, 2009), where each element is a modifier that emphasizes or attenuates words* ***(Taboada et al., 2011)****."*.
- In p. 64, Section 3.4.3, paragraph 6th: the sentence *"In order to detect which sibling node to choose, we applied a syntactic procedure based on dependency types (Vilares, 2013) (...)"* was replaced by *"In order to detect which sibling node to choose, we applied a syntactic procedure based on dependency types* ***(Vilares et al., 2015)*** *(...)"*.
- In p. 65, Section 3.4.4, paragrah 4th: the sentence *"In SA, few existing approaches consider adversatives clauses, while to the best of our knowledge, no approach takes account of concessive ones."* was replaced by *"In SA, few existing approaches consider adversatives clauses* ***(Poria et al., 2014; Vilares et al., 2015)****, while to the best of our knowledge, no approach takes account of concessive ones."*.
- In p. 75, the reference *"Vilares, D. (2013). Sentiment analysis for reviews and microtexts based on lexico-syntactic knowledge. In Proc. of 5th bcs-irsg symposium on future directions in information access. In FDIA '13 (pp. 38–43)."* was replaced by *"**Vilares, D., Alonso, M.A., Gómez-Rodríguez, C. (2015). A syntactic approach for opinion mining on spanish reviews. Natural Language Engineering, 21, 139-163. URL:*** **http://journals.cambridge.org/article_S1351324913000181.** ***doi:10.1017/S1351324913000181****."*.

DOI of original article: 10.1016/j.eswa.2016.03.031

* Corresponding author Tel.: +34986814081.
E-mail address: milagros.fernandez@gti.uvigo.es (M. Fernández-Gavilanes).

http://dx.doi.org/10.1016/j.eswa.2016.06.007