Sentiment Analysis for Reviews in Spanish: Algorithm for Negation Handling

Samara Gretel Villalba-Osornio, Juan Andrés Pérez-Celis,
Luis Villaseñor-Pineda, Manuel Montes-y-Gómez

Instituto Nacional de Astrofísica Óptica y Electrónica,
Puebla, Mexico

{svillalba, celis, villasen, mmontesg}@inaoep.mx

Abstract. Sentiment Analysis (SA) is a challenging area where several problems still are open, one of them is treating the negation. Most of the research focusses on reviews written in English. Therefore, algorithms to deal with the negation in other languages are needed. This work proposes an algorithm to determine the scope of negative particles and to treat the affected words on reviews in Spanish. Our algorithm includes a novel weight scheme which combines the relative frequency of a word with the vocabulary of the class. The experiments show that handling the negation improves the accuracy in Spanish reviews in most cases. Likewise, the proposed weight scheme reaches a maximum accuracy of 67% before treating the negation. Combining our algorithm with the proposed weight scheme yields an accuracy greater than the one reported in the state of the art.

Keywords: Sentiment analysis, negation treatment, Spanish, reviews, polarity classification, negative particles scope.

1 Introduction

Nowadays, the social web provides efficient tools to share opinions regarding social events, politics, products, and services. However, this information is not hasty to process automatically due to the lack of structure and the subjectivity included. The scientific community has shown a growing interest in the matter for the challenges it represents. Likewise, the business world share some interest in the matter for the possible market prediction. In order to cope with the subjectivity the field of Opinion Minning (OM), also known as Sentiment Analysis (SA), emerged [15].

The OM field tries to classify the documents by the polarity of the opinion expressed [10, 15]. Currently, there is a fair amount of research ongoing on that deal with several aspects of the SA [3]. This work focusses on one aspect: the negation; determining its scope and performing a suitable treatment.

The negation handling has a wide importance as people express opinions with negated words. For example, "La película no me gustó" (I did not like the

movie) is a negative opinion with a negated positive word. For the other case, "La historia no es mala" (The story is not bad) expresses a positive opinion with a negated negative word.

Even though, most of the work regarding the negation treatment focuses on English opinions, the increased presence of documents in other languages creates the need for developing research in languages different than English. This paper deals with the problem of negation in Spanish documents. The main contribution is an algorithm that establishes the reach of negative particles for opinions written in Spanish. The goal is to improve the correct prediction of the polarity in opinion texts.

In addition to the algorithm, a novel weight scheme based on a supervised approach is proposed. The scheme takes into account the relative frequency weight of terms in a class' document and the vocabulary of the classes. The novel weight technique was tested in two corpus of different domains with successful results.

The rest of the paper is organized as follows: Section 2 gives an overview of the related work: the negation treatment. Section 3 describes the classification method and presents the proposed algorithm. Section 4 shows a brief analysis of the two corpus used and details the experiments performed. Section 5 presents the results and a thorough analysis of them, highlighting the findings. Section 6 gives concluding remarks and suggests future work.

2 Related Work

There is previous work on the negation handling. Mostly the works are for English texts [1, 3, 8, 14]. There is also research ongoing on other languages such as Chinese, French and Spanish [2]. The most relevant works are detailed bellow, starting with the works for English reviews.

The first research dealing with the negation treatment, to the best of our knowledge, was performed for the English language in 2001. A simple approach is taken by adding a marker to the words following some negative particles until the end of the sentence [6]. The same approach is taken in [11] but only considering the negative particle *not* and its contraction. The authors consider that the effect of the negation is negligible. In [1] the algorithm adds a flag which changes its state when *not* is found in a document. Then the words after *not* are marked, if another *not* is found or a punctuation mark, the marker stops tagging words.

In [8] the model used modifies the polarity of a word next to a negative particle, an intensifier or a mitigator. The polarity of a word gets inverted for the case of a negative particle. The work states the treating the negation has a significant effect on classifying opinions. This model can be seen as if it uses a window of size 1, .i.e, the effect of negative particles presents only to the word next to them.

The window size is modified in some works. In [16] a window of size 4 with effect to the right is proposed. The windows can have different size and can have

an effect to the left too. In [3] the author used a window of size 3 to the right and 1 to the left.

To stop the effect of a negative particle not only a size of a window has been used. Nexus or words of a certain category can stop the effect [4, 14]. More sophisticated information from dependency trees has been used to stop the effect of negative particles [5, 7, 14].

For the Spanish language, the number of works is limited. In [14] English opinions where translated into Spanish for their further classification with a method proposed for reviews in English [4]. A more recent work considers the syntactic structure of the text in matter [12].

Lastly, in [2] the authors utilize the syntactic structure with grounds on the information provided by dependency trees for each sentence. Moreover, negative particles are included, such as "tampoco" (neither), "nadie" (none), "jamás" (never), "ni" (nor) y "nada" (nothing). This work reports a relative improvement of 3.02% when handling the negation. However, there is no analysis of the results to define the reasons for the improvement.

The work herein uses the same corpus as [2]. The difference with our work is that no dependency trees where used and the approach taken in this work is supervised; i.e., the list of words from eSOL, iSOL, SentiWordNet where not used. We inspired in the techniques previously reported for the English language and propose a much simpler approach. To establish the reach of the negative particles windows, nexus and punctuation marks were used.

3   Methodology

In this paper a Negation Algorithm (NA) that determine the scope of negative particles is proposed. The NA is incorporated into a classification method with a supervised approach. The method is described in the following subsections.

3.1   Negation Particles

The first step in the negation analysis is the identification of the negative particles in Spanish. According to the Real Academia Española (RAE)[1] several negative particles exist. In this work the particles considered are: "no", "sin" (whitout), "ni"(nor), "nada"(nothing), "nunca"(never) and "tampoco"(neither).

3.2   Negation Scope

The second step is to design the algorithm that will determine the words in a sentence affected by the negative particles. A fair number of approaches have been reported for the English language. This work takes ideas of the previous research in English and combine them to get an appropriate algorithm for negation handling in Spanish.

Namely, the algorithm takes into account the following aspects:

---

[1] "Nueva gramática de la lengua española: Fonética y fonología. Espasa Libros. (2011)"

- A window of size N affecting a N number of words following a negative particle. In this work windows of 2, 4, 6 and 8 were tested.
- The negative particle affects words between itself and a punctuation mark.
- When a negative particle is found a flag activates and so the effect of it. If another negative particle is found within the reach of the first one, then the flag is turn off and so the effect. Consider the sentence "No había visto nada tan bueno" (I have not seen anything that good), the particle *no* will affect *había* and *visto*, the particle *nada* will turn off the effect so the remaining words will be left without the negation effect.
- The effect of a negative particle is ended when an adversative nexus is within the reach of the negation. For example in "No estuvo romántica pero me gustó" (It was not romantic but I like it), the effect will be stopped by the word *pero* (but). The adversative nexus used in this work are: "pero"(but), "aunque"(even though), "sino" (if not, but), "sin embargo"(however), "no obstante"(nevertheless), "al contrario"(on the contrary) and "a pesar de" (despite).

After the algorithm is executed, the words that are affected by a negative particle are replaced with the negated forms in the document. That modified document is classified. For example if the original document contains "La película no me aburrió, está divertida." (The movie did not bore me, it is fun.). The modified document will contain "la película no_me no_aburrió está buena" (the movie not_bore not_me it is fun).

The algorithm is shown on Fig. 1.

Algorithm 1 Determinig the Negation Scope

1: Read Document.
2: **for** Each word in the document **do**
3:    **if** The word is a negative particle **then**
4:       **for** $i = 0$ until $i =$ window size **do**
5:          **if** The next word is not a negative particle, and is not a puntuation mark, and is not an adversative nexus **then**
6:             Add "no_"+ Word.
7:          **else**
8:             Break and continue with the next word in the document.
9:          **end if**
10:       **end for**
11:    **end if**
12: **end for**

Fig. 1. Algorithm for the scope of negative particles.

3.3 Words' Weight Computation

The third step is to compute the weight of each word or term for the negative and positive class. For this step three different weights were calculated.

In the first weight scheme the weight of a term is related with the number of documents where the term appears. The weight scheme is shown in eq. 1. It will be referred as relative frequency scheme.

$$W(t_i) = \frac{D_c\,t}{|D_c|} \; ; \tag{1}$$

where:

- $D_c\,t$ are the documents of class c that contain the term t.
- $|D_c|$ are the total number of documents in the class c.

The second weight is described in eq. 2. It will be referred as Frequency scheme.

$$W(t_i) = \frac{F_c\,t}{\sum^d F_c\,t'} \; ; \tag{2}$$

where:

- $F_c\,t$ is the frequency of term t in documents of the class c.
- $F_c\,t'$ are the frequencies of all terms in the class c.

Lastly, the third weight scheme, shown in eq. 3 takes a novel approach. The core is the same as eq. 1, the relative frequency weight. The modification lies in $\frac{1}{Voc_c}$, this term was added with the aim of making a compensation as people use more positive words to express themselves in English and in other languages [13], including Spanish.

$$W(t_i) = \frac{D_c t}{|D_c|} \times \frac{1}{Voc_c} \; ; \tag{3}$$

where:

- $\frac{D_c t}{|D_c|}$ is the relative Frequency weight.
- Voc_c is the vocabulary of class c.

3.4   Polarity Modification

Considering that the weights are computed and the negation scope is defined, the next step is modifying the polarity of the terms affected by a negative particle. Starting from the fact that there are two lists, one with positive weights and one with negative weights for each word, the modification is as follows. For each word or term added to the vocabulary the negative form is also added, i.e., a word comprised by *"no_" + word* is added. This *no_* prefix acts as a mirror for the weights. The weights of a term in its negated form are the opposite weights of that same term in its normal state. For example: if the word *good* has a positive weight of 0.7 and a negative weight of 0.3, the word *no_good* will have a positive weight of 0.3 and a negative weight of 0.7.

With the weights for negative and positive words computed, the last step is the classification of the documents.

3.5 Classification Method

The decision whether a document D belongs to the positive class or negative class is taken based on the sum of the weights for all the words in D. It is noteworthy to recall that the document to be classified is the modified document, which is the output after executing the proposed algorithm for the scope of negative particles. The decision is based on comparing the result of eq. 4 for each of the two classes. The document will be classified as positive if the sum of the positive weights for all words in the document under test is greater than the sum of the negative weights for all the words in the document, and will be negative otherwise. In case the result of eq. 4 is equal for both classes, the document will be classified as negative.

$$Weigth(D_x) = \sum_{i=1}^{n} W(t_i) + \sum_{j=1}^{m} W(\overline{t_j}) \; ; \tag{4}$$

where:

- n is the total of non-negated terms in the document x.
- m is the total of negated terms in the document x.

The summation on the left expresses the sum of the normal weights for all non-negated terms. The summation on the right adds the mirrored weights for all negated terms.

4 Experimental Framework

Tests where conducted on two corpus in Spanish. The first corpus is of movie reviews [5] and the second corpus is the SFU [9]. The description of both corpus is detailed on the next subsections.

4.1 Corpus Details

The Corpus of Movie Reviews (CMR) is a Spanish corpus containing 3878 movie reviews obtained from the website MuchoCine [2]. The reviews are rated from 1 to 5, where 1 is a bad movie and 5 an excellent movie. For this work the reviews rated with 3 are considered neutral and are disregarded from the work. The opinions rated with 1 or 2 are negative opinions. The opinions with 4 or 5 are positive opinions.

The corpus SFU [3] is comprised by 400 reviews regarding eight topics (books, cars, computers, electro domestics, hotels, movies, music and cellphones). Each topic has 50 reviews, 25 negative and 25 positives. For the two corpus, the opinions were written by users of the web, there are no modifications on the writing.The opinions preserve spelling mistakes, grammar mistakes and colloquial expressions. The Table 1 shows details of the documents in the corpus.

---

[2] "http://www.lsi.us.es/ fermin/corpusCine.zip."
[3] "https://www.sfu.ca/ mtaboada/research/SFU_Review_Corpus.html."

Table 1. Details of CMR and SFU corpus.

Classes in the CMR.				
Class.	**Documents.**	**Vocabulary.**	**#of Words.**	**Ave. # of Words.**
Positive	1352	42245	736172	545
Negative	1274	37542	560639	440
Classes in the SFU corpus.				
Positive	200	14578	120227	601
Negative	200	10714	77159	386

The number of negative particles in each class was counted for this corpus. The results shown in Table 2 suggest that treating the negative particles will have a effect in the classification as the number of appearances is high. It is interesting to point out that the most used particle *no* is used more times in the positive class than in the negative class. This fact may be a consequence of having longer reviews and greater number of words in the positive class.

Table 2. Appearance of negative particles in both Corpus.

	CRM		**SFU Corpus.**	
	Positive	**Negative**	**Positive**	**Negative**
NO	7932	7606	4797	1528
SIN	2217	1600	256	169
NI	876	1233	161	199
NADA	719	866	183	211
NUNCA	387	284	81	39
TAMPOCO	201	315	33	45

4.2 Experiments Considerations

Several experiments were conducted with both corpus, however the main corpus is CMR due to the greater extension in the number of reviews and to the fact that it is only focused on one domain. For the CMR, the classes were evened by disregarding the last reviews. Both classes remain with 1270 reviews. The experiments on the SFU corpus were conducted mainly to test the novel weight scheme proposed in this paper in various corpus.

The experiments detailed in this paper were performed in Python using the Natural Language ToolKit (NLTK). For the tests in both corpus a preprocessing was performed. This step began by changing characters containing accents of tilde. Characters such as á, é, were changed to their forms without accents. The special character ñ was changed by *n*. Symbols were changed for a space. Punctuation marks were substituted by special markers for the sake of the algorithm. These markers have no influence in the classification step.

During the experiments stop words have a weight of zero in both classes. This rule allows stop words to be counted for the window size of negative particles but they have no effect in the total weight. The list of stop words used can be found in the RANK NL website [4].

Finally, all the tests use cross validation. The results report the macro-average and the standard deviation. For the CMR ten folds were used. For the SFU corpus five folds were used. Folds have even classes. For each fold first 80% of the reviews was used to compute the weights and the remaining 20% was classified. Folds were chosen to ensure that all reviews were classified; once for the SFU corpus and twice for the CMR.

4.3 Experiments

The first experiment consists in classifying the documents in the CMR with the three schemes mentioned in the paper. The experiment was performed with the negation algorithm disabled and enabled. The results of Table 3 show that the greatest accuracy is obtained with the proposed weight scheme and the negation algorithm enabled. The standard deviation is shown bellow each macro-average. For the three weight scheme enabling the negation algorithm modifies the accuracy. In the case of the Relative Frequency Scheme (RFS) the accuracy drops. In the Proposed Weight Scheme (PWS) and in the Frequency Weight Scheme (FWS) the accuracy shows an improvement.

Table 3. Classification of CMR with the different weight schemes.

	Proposed Weight Scheme							
	Without Negation				With Negation			
	Prec.	Recall.	F1.	Acc.	Prec.	Recall.	F1.	Acc.
Positive	0.9177	0.3774	0.5276	0.6727	0.7947	0.5872	0.7096	**0.7373**
	(±0.0299)	(±0.1111)	(±0.1116)	(±0.0549)	(±0.0372)	(±0.0645)	(±0.0280)	(±0.0145)
Negative	0.6114	0.9680	0.7486		0.7029	0.8288	0.7589	
	(±0.0442)	(±0.0083)	(±0.0323)		(±0.0279)	(±0.0546)	(±0.0151)	
	Frequency Weight Scheme							
Positive	0.9884	0.0247	0.0481	0.5116	0.8780	0.3095	0.4893	**0.6470**
	(±0.0880)	(±0.0094)	(±0.0179)	(±0.0048)	(±0.0433)	(±0.0508)	(±0.0580)	(±0.0289)
Negative	0.5058	0.9984	0.6715		0.5916	0.9535	0.7300	
	(±0.0024)	(±0.0020)	(±0.0023)		(±0.0202)	(±0.2878)	(±0.0181)	
	Relative Frequency Scheme							
Positive	0.5249	0.9952	0.6871	**0.5465**	0.5109	0.9083	0.6760	0.5210
	(±0.0149)	(±0.0061)	(±0.0118)	(±0.0265)	(±0.0096)	(±0.0020)	(±0.0081)	(±0.0182)
Negative	0.9652	0.0977	0.1726		0.9928	0.0429	0.0800	
	(±0.0358)	(±0.0571)	(±0.0927)		(±0.0225)	(±0.0381)	(±0.0678)	

The next experiment consists on changing the size of the window. Sizes of 2, 4, 6 and 8 were chosen. The results are shown in Table 4. The best accuracy is reached with a window size of 4. Likewise, Table 4 reports the results of

---

[4] http://www.ranks.nl/stopwords/spanish.

another experiment conducted with the goal of identifying the impact of the adversative nexus as disablers of the negation effect. The necessary changes were performed in the negation algorithm to ignore the effect of adversative nexus. The experiment is based in the PWS and was conducted with 2, 4, 6 and 8 as the size of the window.

Table 4. Windows size variations.

Size	Class	With Adversative Nexus.				Without Adversative Nexus.			
		Prec.	Recall.	F1.	Acc.	Precc.	Recall.	F1.	Acc.
2	Pos	0.8008 (±0.0323)	0.5775 (±0.0698)	0.7054 (±0.0341)	0.7369 (±0.0174)	0.8625 (±0.0363)	0.4740 (±0.0948)	0.6439 (±0.1255)	0.7174 (±0.0324)
	Neg	0.6994 (±0.0320)	0.8385 (±0.2568)	0.7610 (±0.0120)		0.6593 (±0.0383)	0.9133 (±0.2777)	0.7642 (±0.0165)	
4	Pos	0.7945 (±0.0372)	0.5872 (±0.0645)	0.7096 (±0.0280)	**0.7373** (±0.0145)	0.8516 (±0.0271)	0.4880 (±0.0922)	0.6535 (±0.0644)	**0.7203** (±0.0339)
	Neg	0.7029 (±0.0279)	0.8288 (±0.0546)	0.7589 (±0.0151)		0.6640 (±0.0404)	0.9042 (±0.2742)	0.7643 (±0.0186)	
6	Pos	0.7862 (±0.0290)	0.5862 (±0.0602)	0.7063 (±0.0297)	0.7334 (±0.0169)	0.8456 (±0.0261)	0.4848 (±0.0863)	0.6498 (±0.0614)	0.7172 (±0.0325)
	Neg	0.7002 (±0.0276)	0.8220 (±0.2514)	0.7549 (±0.0144)		0.6614 (±0.0374)	0.9011 (±0.2730)	0.7817 (±0.0179)	
8	Pos	0.7821 (±0.0279)	0.5882 (±0.0609)	0.7060 (±0.0301)	0.7321 (±0.0171)	0.8417 (±0.0308)	0.4787 (±0.0802)	0.6437 (±0.0566)	0.7126 (±0.0287)
	Neg	0.7004 (±0.0286)	0.8172 (±0.2499)	0.7530 (±0.0140)		0.6572 (±0.0330)	0.8987 (±0.2727)	0.7581 (±0.0159)	

The following test was performed over the SFU corpus. The test consist in using the three weight schemes (PWS, FWS and FRS) to classify the reviews. The window size is 4 and the adversative nexus are taken into account. The test was conducted without the negation algorithm and with the algorithm enabled. The results are shown in Table 5.

5   Analysis of Results and Discussion

The PWS got the greatest accuracy for the CMR and a better accuracy than BWS in the SFU corpus. The fact suggest that considering the vocabulary length for the weight schemes can improve the accuracy when classifying opinions. As mentioned before the PWS was inspired by the work in [13] which states that the vocabulary of a positive class is greater then the negative one. In an attempt to verify this statement, an analysis of the vocabulary for each of the ten folds was performed. The results in Table 6 show that the vocabulary for each fold of the the positive class is greater than the negative one.

The results of the PWS with a window size of 4 applying cross fold validation, as in all the experiments, were compared to those in [2]. The same corpus was used. The reviews with three stars were also disregarded in [2]. The approach taken by the authors is unsupervised. The Table 7 shows the comparisson. Our

Table 5. Results with the different weight schemes in the SFU corpus.

	Prec.	Recall.	F1.	Acc.	Prec.	Recall.	F1.	Acc.
Proposed Weight Scheme								
	Without Negation				**With Negation**			
Positive	0.7333	0.0400	0.0846	0.5162	0.7850	0.0916	0.1907	**0.5375**
	(±0.4346)	(±0.0285)	(±0.0385)	(±0.0162)	(±0.2233)	(±0.0418)	(±0.0674)	(±0.0318)
Negative	0.5077	0.9900	0.6712		0.5202	0.9650	0.6759	
	(±0.0095)	(±0.0136)	(±0.0108)		(±0.0174)	(±0.3954)	(±0.0092)	
Frequency Scheme Weight								
Positive	0.8000	0.0297	0.0571	0.5087	0.6520	0.1541	0.2783	**0.5400**
	(±0.2738)	(±0.0113)	(±0.0213)	(±0.0121)	(±0.1307)	(±0.0928)	(±0.0990)	(±0.0298)
Negative	0.5038	0.9901	0.6678		0.5242	0.8950	0.6601	
	(±0.0072)	(±0.0135)	(±0.0092)		(±0.0196)	(±0.3694)	(±0.0199)	
Relative Frequency Weight								
Positive	0.5395	1	0.7003	**0.5700**	0.5240	0.8333	0.6875	0.5450
	(±0.0367)	(±0)	(±0.0304)	(±0.0603)	(±0.0187)	(±0)	(±0.0160)	(±0.0387)
Negative	1	0.1400	0.2307		1	0.0900	0.1595	
	(±0)	(±0.1206)	(±0.1754)		(±0)	(±0.0707)	(±0.1117)	

Table 6. Average length of the vocabularies.

Vocabularies		
Positive	**Negative**	**Total**
36269	33216	49379
(±276)	(±103)	(±222)

approach is almost 4% more accurate without the Negation Algorithm (NA) and almost 9% more accurate with the NA enabled.

Table 7. Comparisson of PWS with [2].

Jiménez, Martínez, Martín & Molina.							
Without Negation				**With Negation.**			
Prec.	**Recall.**	**F1.**	**Acc.**	**Prec.**	**Recall.**	**F1.**	**Acc.**
0.6365	0.6276	0.6320	0.6312	0.6519	0.6430	0.6474	0.6475

This Work							
Prec.	**Recall.**	**F1.**	**Acc.**	**Prec.**	**Recall.**	**F1.**	**Acc.**
0.7645	0.6727	0.7156	0.6727	0.7488	0.7080	0.7278	0.7373

The difference of the accuracy between the results of the experiments where the adversative nexus were considered and then disregarded for several size of windows is rather small 0.0195. To the end of understanding the reason of this effect, the Table 8 counts the condition which made the negative particles lose their effect.

The number of times the negation is stopped by punctuation marks increases as the windows size do so. The trend for number of times the negation is stopped

by the window size decreases as the size increased. It is an interesting finding that the number of times the negation effect is stopped by an adversative nexus is only 146 times for a window size of 8.

Table 8. Determining the negation scope.

Window Size	Positive Class.				Negative Class.			
	Size 2	Size 4	Size 6	Size 8	Size 2	Size 4	Size 6	Size 8
Punctuation Mark	1511	3740	5592	6795	1392	3494	5132	6242
Window	9503	6505	4320	2901	8866	5834	3871	2587
Nexus	30	83	120	146	25	71	102	122
Negative Particle	481	839	987	1082	649	1091	1238	1320
NO	31	83	120	146	34	92	158	210
SIN	6	32	45	58	8	28	35	46
NI	182	329	383	409	264	448	501	520
NADA	216	335	345	350	301	457	474	474
NUNCA	23	35	36	39	22	39	41	41
TAMPOCO	23	26	27	27	20	27	29	29

Another remarkable finding is that after analyzing the content of the reviews, the impact of considering double negative particles to end the effect of negation is not reflecting the actual case of the nested negative particles. Plenty sentences where nested negative particles present seek to produce an effect of intensification. This is commonly found in negative reviews. For the following examples the structure of the sentences is shown and an English translation that may not be gramatically correct as the point is to show the effect of Negative Particles (NP) in the sentences. In the negative class we found expressions such as:

- "No tiene ni subtramas ni similares ..." (NP + Verb + NP + Noun + NP + Adjective).
- "Ni engancha el guión ni interesa la historia ..." (NP + Verb + Article + Noun + NP + Verb + Article + Noun).
- "La trama no me parece ni original ni demasiado interesante ..." (Article + Noun + NP + Pronoun + Verb + NP + Adjective + NP + Adverb + Adjective).

In the possitive class nested negative particles are to be found in coloquial expressions or neutral phrases such as:

- "No tiene ni ton ni son ..." (NP + Verb + NP + Noun + NP + Noun). The meaning of this expression is that it lacks of interesting moments.
- "Ni más ni menos ..." (NP + Adjective + NP + Adjective).
- "No es ni mala ni buena ..." (NP + Verb + NP + Adjective + NP + Adjective).

An analysis of the output of the classification was conducted. The cases where the Negation Algorithm (NA) was enable and disable are compared. Two interesting situations were noticed:

1. All the negative reviews classified as positive without the NA remain positive after enabling the NA.
2. All the positive reviews classified correctly without the NA remain positive after enabling the NA.

For the first case, the analysis shows that the length of those reviews (of point 1) are above the average of positive and negative reviews. Likewise, negative particles seldom appear on those reviews making the negation handling effect negligible. Review #8 is a clear example. It has 985 words and says:

"...y es que cuando los paisajes te parecen más bonitos e interesantes que la historia de amor que se está contanto, es algo que no funciona, falla la pasión, falla la emoción, falla lo conmovedor, ...". (...and when the landscapes seem prettier and more interesting to you than the love story that is being told, there is something not working right, the passion fails, the emotion fails, the touching fails, ...).

It is clear that this is a negative review, however, the word "fails" is not used by the NA as a negative particle. Using verbs for negating the meaning of words was not considered in the NA. Moreover, combining those verbs with strongly positive weighted words makes the NA to err in the classification.

The second case have the same characteristics. The length of those reviews are above the average of positive and negative reviews. Likewise, the reviews lack of negative particles making the effect of the NA negligible.

The case where the reviews were misclassified before the NA and after the NA was also presented. This reviews have a fair amount of negative particles too close to each other, breaking the effect of the negation and causing a small change in the total weight summation. The difference in the total sum with the NA and without the NA is not enough to correctly classify the review in matter.

The reviews that where misclassified before the NA but where correctly classified after applying the NA have a particular characteristic. Representative words of a certain class are negated, i.e., words with a weight significantly greater in a certain class than in the other are negated. After applying the NA the weights of those words are mirrored yielding a correct classification. The following fragment of a positive review serves as an example:

– "...no es que sea mala ..." (...it is not that it is bad ...).

In the classification state the only word to consider is *mala* (bad) as the other words are stop words. The word *mala* has a weight of 0.0008231 and 0.0054293 respectively. When classifying without the NA the review will be placed on the negative class. Nonetheless, after applying the NA the weights are mirrored and the opinion is classified correctly as a positive review.

6 Conclusions

The algorithm proposed for handling the negation has a direct impact in the classification of Spanish reviews. In the Proposed Weight Scheme (PWS) the

effect increases the accuracy from 0.6727 to 0.7373 and from 0.5162 to 0.5375 for the Corpus of Movie Reviews (CMR) and the SFU corpus respectively. The Frequency Weight Scheme (FWS) improves from 0.5116 to 0.6470 and from 0.5087 to 0.54 for the CMR and the SFU corpus respectively. The relative frequency Weight Scheme have a counterproductive effect the accuracy drops from 0.5465 to 0.5210 and from 0.57 to 0.5450 for the CMR and the SFU corpus respectively. The greatest accuracy is achieved with the PWS and applying the Negation Algorithm (NA).

The tests performed for different size of windows suggest that the best performance comes when using a size of 4, at least for the CMR. The result agrees with the ones reported for English reviews. More tests should be conducted to verify if the behavior is the same in other domains of reviews.

The use of nexus in the NA increases the classification accuracy in 1.9% for the best case: with a window size of 2. The analysis showed in 8 suggest that the increase is not greater because only a few times the effect of negative particles is stopped by adversative nexus. 90% of the time the negation loses its effect due to the window size or the punctuation marks.

The PWS has a satisfying performance for classifying Spanish reviews. Giving importance to the vocabulary of the classes translate into a maximum improvement of 21% over the BWS, which only differs in the factor $\frac{1}{Voc_c}$. The PWS outperforms in the CMR the FWS by 12% without handling the negation an by 9% when the NA is used.

For the SFU corpus, the PWS has a similar behavior as the FWS. The PWS is 1% better when the NA is off and 0.25% worse then the NA is on. However there is still an improvement in both weights when applying the NA if compared to the result when the NA is off. The highest accuracy is achieved by the BWS. Compared to our PWS the BWS have a 6% higher accuracy without the NA and 0.8% with the NA enabled. The results of the PWS may have dropped in the SFU corpus due to the fact that the corpus is small (400 reviews) and have 8 different categories. Further tests will be performed to identify the exact reason. Our guess is that the vocabulary for each category varies dramatically in number and in the words used, making our PWS not able to compensate the greater use of vocabulary in positive reviews competently.

The supervised approach has a better performance for classifying Spanish reviews. Even without using our NA, with the PWS the accuracy achieved is 4% higher than the accuracy reported in [2]. When enabling the NA our approach outperforms by 9% the same work.

In the future work the algorithm presented for negation handling will be applied in the training set to obtain a vocabulary that includes words with the prefix "no_" indicating that was affected by a negation. The words with the mark will be treated in the same way of the rest of the features, calculating their probabilities according to their frequencies in the documents.

References

1. Narayanan, V., Arora, I., Bhatia, A.: Fast and accurate sentiment classification using an enhanced Naive Bayes model. In: Intelligent Data Engineering and Automated Learning–IDEAL 2013, pp. 194–201, Springer, Heidelberg (2013)
2. Jiménez Zafra, S. M., Martínez Cámara, E., Martín Valdivia, M. T., Molina González, M. D.: Tratamiento de la Negación en el Análisis de Opiniones en español (2015)
3. Bakliwal, A., Arora, P., Patil, A., Varma, V.: Towards enhanced opinion classification using NLP techniques. In: Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology, pp. 101–107, IJCNLP (2011)
4. Brooke, J., Tofiloski, M., Taboada, M.: Cross-Linguistic Sentiment Analysis: From English to Spanish. In: RANLP, pp. 50–54 (2009)
5. Cruz, F. L., Troyano, J. A., Enriquez, F., Ortega, J.: Clasificación de documentos basada en la opinión: experimentos con un corpus de criticas de cine en espanol. Procesamiento de Lenguaje Natural, vol. 41. (2008)
6. Das, S., Chen, M.: Yahoo! for Amazon: Extracting market sentiment from stock message boards. In: Proceedings of the Asia Pacific finance association annual conference, Vol. 35, p. 43 (2001)
7. Jia, L., Yu, C., Meng, W.: The effect of negation on sentiment analysis and retrieval effectiveness. In: Proceedings of the 18th ACM conference on Information and knowledge management, pp. 1827–1830, ACM (2009)
8. Kennedy, A., Inkpen, D.: Sentiment classification of movie reviews using contextual valence shifters. Computational intelligence, vol. 22, pp. 110–125 (2006)
9. Konstantinova, N., de Sousa, S. C., Díaz, N. P. C., López, M. J. M., Taboada, M., Mitkov, R.: A review corpus annotated for negation, speculation and their scope. In: LREC, pp. 3190–3195 (2012)
10. Molina-González, M. D., Martínez-Cámara, E., Martín-Valdivia, M. T., Perea-Ortega, J. M.: Semantic orientation for polarity classification in Spanish reviews. Expert Systems with Applications, vol. 18, pp. 7250–7257 (2013)
11. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up?: sentiment classification using machine learning techniques. In: Proceedings of the ACL-02 conference on Empirical methods in natural language processing, vol. 10, pp. 79–86, Association for Computational Linguistics (2002)
12. Vilares, D., Alonso, M. A., Gómez-Rodríguez, C.: Clasificación de polaridad en textos con opiniones en espanol mediante análisis sintáctico de dependencias. Procesamiento del lenguaje natural, vol. 50, pp. 13–20 (2013)
13. Dodds, P. S., Clark, E. M., Desu, S., Frank, M. R., Reagan, A. J., Williams, J. R., Danforth, C. M.: Human language reveals a universal positivity bias. In: Proceedings of the National Academy of Sciences, vol. 8, pp. 2389–2394 (2015)
14. Taboada, M., Brooke, J., Tofiloski, M., Voll, K., Stede, M.: Lexicon-based methods for sentiment analysis. Computational linguistics, vol. 37, pp. 267–307 (2011)
15. Pang, B., Lee, L.: Opinion mining and sentiment analysis. Foundations and trends in information retrieval, vol. 2, pp. 1–135 (2008)
16. Wilson, T., Wiebe, J., Hoffmann, P.: Recognizing contextual polarity in phrase-level sentiment analysis. In: Proceedings of the conference on human language technology and empirical methods in natural language processing, Association for Computational Linguistics, pp. 347–354 (2005)