

A Concise Representation for Detailed Sentiment Analysis

Víctor M. Morales¹, Saúl León Silverio¹, Maya Carrillo¹, Aurelio López-López², Luis Enrique Colmenares-Guillen¹

¹ Benemérita Universidad Autónoma de Puebla,
Faculty of Computer Science, Puebla,
Mexico

² INAOE, Computational Science Department, Tonantzintla, Puebla,
Mexico

{victor.morales, cmaya, lecolme}@cs.buap.mx
saul.ls@live.com, allopez@inaoep.mx

Abstract. This paper presents initial results in sentiment analysis classification, as an attempt to go beyond categorizing texts only by ‘positive’ or ‘negative’ orientation, using fine-grained features for this purpose. We present a method for sentiment classification based on a concise representation built from analyzing appraisal groups such as “very good” or “not terrible”. An appraisal group is represented as a set of attribute values anteceding an appraisal word (adjective). An appraisal lexicon is used to identify adjectives guiding the analysis. We performed experiments classifying movie reviews in Spanish using features based upon attitude taxonomy information, and report improvements on precision with eight dimensional vectors and a support vector machine algorithm.

Keywords: Concise representation, appraisal group, sentiment analysis.

1 Introduction

Many traditional data mining tasks in natural language processing focus on extracting data from documents and mining it according to topic. In recent years, the natural language processing community has recognized the value of analyzing opinions and emotions expressed in free text. Sentiment analysis is the task of having computers automatically extract and classify the opinions in a text. Sentiment analysis has become a growing field for commercial applications, with at least a dozen of companies offering products and services for sentiment analysis, with very different sets of goals and capabilities.

Some applications of sentiment analysis focused on classifying movie or product reviews as positive or negative, or identifying positive and negative sentences, but more complicated applications involve opinion mining in ways that require a more detailed analysis of the sentiment expressed in texts. One such application

is the use of opinion mining to determine areas of a product that need to be improved by summarizing product reviews to identify what parts of the product are generally considered good or bad by users. For instance, if the product is considered of good quality but its presentation is not very pleasant. Another application requiring a more detailed analysis of sentiment is to understand where political writers fall on a political spectrum, something that can only be done by looking at support or opposition to specific issues. There are others applications [1], which allow politicians to have a better understanding of how their constituents view different issues, or predicting stock prices based on opinions that people have about the companies.

These applications can be tackled with a better understanding of the expressed opinions. Sentiment analysis researchers are currently working on creating the techniques to handle these more complicated problems, defining the structure of opinions and the techniques to extract such structure. For instance, the typology described by Appraisal Theory is reviewed in [2], presenting a methodology for annotating appraisal expressions, and its use to annotate a corpus of book reviews. The authors discuss an inter-annotator agreement study, and consider instances of common disagreement that indicate areas in which appraisal can be refined.

EmotiNet is presented in [3], a knowledge base for representing and storing affective reaction to real-life contexts. The authors detail the methodology employed in designing, populating, and evaluating such knowledge base. The basic model was further extended from additional resources, such as VerbOcean, ConceptNet, and SentiWordNet, with the aim of generalizing the knowledge, and later evaluated.

Appraisal theory was applied in news articles, specifically in the domain of stock markets [4]. They implemented an approach of Chinese new words detection by using N-gram model and applied the result for Chinese word segmentation and sentiment classification, employing Naïve Bayes, K-nearest Neighbor and Support Vector Machine algorithms. The best accuracy result reaches 82.9% in their experiments.

A different kind of domain specific feature-based heuristic for aspect-level sentiment analysis of movie reviews is reported in [5]. In particular, they developed an aspect oriented scheme that analyses the reviews of a movie and assign it a sentiment label on each aspect. The scores on the aspects from multiple reviews are then aggregated and a net sentiment profile of the movie is generated on all parameters. The scheme used is based on SentiWordNet, with two different linguistic feature selections comprising of adjectives, adverbs and verbs and n-gram feature extraction. Such scheme is also employed to compute the document-level sentiment for each movie reviewed and compared the results with those obtained with an API. The sentiment profile of a movie is also compared with the document-level sentiment result. They claim that the scheme produces a more accurate and focused sentiment profile than a simple document-level sentiment analysis.

Text classification with fine-grained attitude labels is reported in [6]. The

system developed relies on the compositionality principle and a novel approach based on rules elaborated for semantically distinct verb classes. The evaluation of the method on 1000 sentences, describing personal experiences, showed promising results in terms of accuracy for 14, 7 and 3 labels.

In the direction of these effort to have fine-grained sentiment analysis, and specifically taking advantage of Appraisal Theory, the goal of our research is to analyze opinions at a fine grain level that will allow an improved understanding of the opinion expressed in the text.

We developed a method for sentiment classification based on a concise representation built from analyzing appraisal groups such as “very good” or “not terrible”, where an appraisal group is represented as a set of attribute values antecedent an appraisal word. An appraisal lexicon is used to identify adjectives guiding the analysis. We performed experiments classifying movie reviews in Spanish using features based upon attitude taxonomy features.

This paper is organized as follows: Section 2 provides a brief overview of related works, Section 3 describes our proposed approach in detail, Section 4 reports experiments and results, and finally section 5 includes conclusions and further work.

2 Related Work

Appraisal Theory is a particular approach for exploring, describing and explaining the way language is used to evaluate, to adopt stances, to construct textual personas and to manage interpersonal positioning and relationships. Thus, this theory explores how speakers and writers transmit judgments on people and object, generally. The Appraisal Theory has emerged over a period of almost 20 years as a result of work conducted by a group of researchers lead by Prof. James Martin of the University of Sydney ([7], [8], [9]). The taxonomy of Appraisal Theory has four systems: affect, graduation, orientation and polarity, and three subsystems: affect, judgment and appreciation, as follows:

- Attitude provides the type of appraisal being expressed as affect, appreciation, or judgment.
- Affect is concerned with emotional response and disposition and is typically realized through mental processes of reaction (“This pleases me”, “I hate chocolate”, etc).
- Judgment encompasses meanings which serve to evaluate human behavior positively and negatively by reference to a set of institutionalized norms.
- Appreciation system by which evaluations are made of products and processes. It encompasses values which fall under the general heading of aesthetics, as well as a non-aesthetic category of ‘social valuation’ which includes meanings such as significant and harmful.
- Graduation describes the intensity of appraisal in terms of two independent dimensions of force (or ‘intensity’) and focus (‘prototypicality’)
- Orientation is whether the appraisal is positive or negative (often simply termed ‘sentiment’).

- Polarity of an appraisal marked if it is scoped in a polarity marker (such as ‘not’), or unmarked otherwise.

The research to map this theory to automatic sentiment analysis is recent and has been made for English language. In this context, we have the work of [10] where they apply the appraisal linguistic theory and find that it could be helpful in distinguishing different types of subjective texts (e.g., movie reviews from consumer product reviews).

In [11], two representations of text are compared, within the same experimental setting for sentiment orientation analysis. The two representations are bag-of-words (BoW) and a nine dimensional vector (9Dim), i.e. a compact representation, as we are experimenting. The former represents text with a high dimensional feature vector as we also tested. But, the short representation encodes grammatical knowledge of clauses in sentences, which is lexicon-independent. The aim of their paper was to demonstrate how text sentiment orientation classifiers’ performance is improved when the length of the sentence comprising a training vector is varied. Their results indicated that the classification accuracy benefits from increasing the text length, and the results also indicated that the compact method can provide comparable results to BoW under the same sentiment classification algorithm. They also employed the support vector machines (SVM) algorithm.

3 Proposed Approach

We now describe the method for sentiment classification that basically consist on getting a concise representation built from analyzing appraisal groups, using an appraisal lexicon to identify adjectives revealing opinions. We test two different representations that are described. The lexical resource needed for the analysis is described as well as the way the appraisal groups are handled.

3.1 Method

The proposed method counts the number of positive, negative, judgment, appreciation and affects words present in an opinion. Each opinion is represented as a five entry vector, each one corresponding to the entries that a word has in the appraisal lexicon.

To evaluate an opinion, each word is looked up in the appraisal lexicon, if the word is found, the opinion vector is increased according to the values that the word has in the lexicon. However, the presence of appraisal modifiers plays a key role in this process, because depending on the type of modifier; values may increase, decrease or even be inverted, so as a result the values of the opinion vector are affected.

Two approaches (models) are used to evaluate the opinions under analysis, which are described below:

Model-1 Weighted Terms: In this approach, each opinion is represented as a vector of five entries, each one corresponds to the registered values in the attitude lexicon (i.e. positive, negative, judgment, appreciation and affect). The opinion is reviewed word by word, when a term is found in the lexicon; all its five values are stored in a temporary vector. Therefore, a window of three words before the term is analyzed, if appraisal modifiers are detected, then following steps are followed:

- If the appraisal modifier is an inversion, temporary positive and negative values are exchanged.
- If the appraisal modifier is considered of high impact, all the values of the temporary vector are duplicated.
- If the appraisal modifier is of low impact, all the values of the temporal vector are halved.

Finally, the temporary vector is added to the opinion vector.

Model-2 Weighted Terms by Polarity: In the previous model, vector entries for affect, judgment and appreciation, do not help a lot to determine the opinion polarity. With the aim that these aspects contribute to clarify the opinion polarity, in the second model, each of these attributes was represented by two vector entries. Then each opinion is represented by a vector of eight entries (i.e. positive, negative, positive judgment, negative judgment, positive appreciation, negative appreciation, positive affect and negative affect).

To determine the polarity of each lexicon term, as there are words that are valued both positively and negatively, the positive and negative values are subtracted, if the remainder is greater than zero, the term is considered positive; if subtraction is less than zero, the term is considered negative, and if the remainder is zero we say that the word is neutral.

As in the previous model, this approach processes the opinion word by word, when a term is found in the lexicon, the actual polarity of the term is calculated, i.e. the negative value is subtracted from the positive value and then, in accordance to the criteria shown in Table 1, the value of the eight temporary vector entries are established.

Afterwards, a window of three preceding words, as in the previous approach, is considered, if attitude modifiers are found, the following steps are done:

- If the attitude modifier is an inversion, temporary positive and negative values are exchanged, i.e. positive for negative, positive judgment for negative judgment, and so on.
- If the attitude modifier is of high impact, all the values of the temporal entries are duplicated.
- If the attitude modifier is considered of low impact, all the values of the temporary entries are halved. Finally, temporary entries are added to the final opinion vector.

Table 1. Vector entries.

Vector Entry	Description
Positive	Correspond to the positive term value in the aptitude lexicon.
Negative	Correspond to the negative term value in the aptitude lexicon.
Positive Judgment	If the term actual polarity is positive, the term judgment value in the lexicon is assigned to this entry, otherwise 0 is assigned.
Negative Judgment	If the term actual polarity is negative, the term judgment value in the lexicon is assigned to this entry, otherwise 0 is assigned.
Positive Appreciation	If the term actual polarity is positive, the term appreciation value in the lexicon is assigned to this entry, otherwise 0 is assigned.
Negative Appreciation	If the term actual polarity is negative, the term appreciation value in the lexicon is assigned to this entry, otherwise 0 is assigned.
Positive Affect	If the term actual polarity is positive, the term affect value in the lexicon is assigned to this entry, otherwise 0 is assigned.
Negative Affect	If the term actual polarity is negative, the term affect value in the lexicon is assigned to this entry, otherwise 0 is assigned.

3.2 Attitude Lexicons

To identify the attitude words in the opinions, the attitude lexicon build by [12] was used. It has 3,005 word entries, where each word was manually evaluated, considering three values: 0, 1, and 2, to establish its polarity (positive, negative) and its correspondence to an attitude subsystem (i.e. judgment, appreciation, affect). Here, 0 indicates the lower level and 2 the highest level. Some examples are showed in Table 2, along the values for each entry:

Table 2. Examples of words in attitude dictionary.

Word	Attitude Feature Value				
	POS	NEG	AFF	JUD	APP
Audaz (<i>bold</i>)	2	0	0	2	0
Suicidio (<i>suicide</i>)	0	2	0	2	0
Fértil (<i>fertile</i>)	2	0	0	2	2
Añejo (<i>old</i>)	2	1	0	0	2

Analyzing the word *suicide* in Table 2, one can notice that it has a negative connotation, but at the same time denotes a judgment term with a high level. On the other hand, *old* does not have a clear orientation (more positive than negative) but has a high value in terms of appreciation (i.e. social or aesthetics valuation). This lexicon is further referred as HLM.

The original attitude lexicon was enlarged adding the words in two different lexicons gathered by [13] to have as many words as possible, however the new words only are categorized as positive or negative, leaving judgment, appreciation and affect without values. This enlarged lexicon is here on referred as PBM.

3.3 Appraisal Groups

An appraisal group is defined as a group of words with a main adjective that belongs to an attitude subsystem type and a list of preceding modifiers [14]. Each group denotes the transformation of one or more attitude attributes. For example sentences: “It is a very boring movie”, “It is a little entertaining film” and “It is not an interesting movie.” In this ongoing work, to analyze opinions we consider just three kind of modifiers to define an appraisal group:

- Inversion group are words denoting a polarity inversion. (e.g. “not”)
- Low Impact words that decrement the polarity value. (e.g. “little”)
- High Impact words that increase the value of the polarity. (e.g. “very”)

The appraisal modifiers considered are depicted in Table 3 (Spanish and English translation):

Table 3. Modifiers in appraisal phrases.

Effect	Spanish Terms
Inversion	No
Low Impact	Poco (little), escaso (little), escasamente (barely), exiguo (meager), falta (lacking), insuficiente (insufficient), insuficientemente (insufficiently), falta (missing)
High Impact	Muy (very), demasiado (too), bastante (quite), bastantemente (sufficiently), mucho (much), sumo (most), sumamente (most), excesivo (excessive), asaz (exceedingly), excesivamente (excessively)

4 Experiments and Results

The results obtained by the proposed approach are described in this section. The learning process was carried out with Weka, using the support vector machine (SVM) algorithm with default parameters and 10-fold cross validation.

4.1 Corpus

Sentimental Analysis has been a topic of interest in the last years. However the available corpora in Spanish are few. Fortunately, [15] made a contribution to the community by creating a corpus for sentiment analysis in Spanish. Their corpus is about movie reviews. The authors prepared the corpus according to the following criteria:

- A high number of available reviews (more than two thousand).
- When the content is generated by users, a minimum texts quality has to be ensured.

- Each user who reviews a movie has to assign a review score to the movie, which allows distinguishing whether a critics contains a favorable or unfavorable opinion.
- The web publishing license should allow to use freely the contents.

According with the previous criteria, the authors selected the mucho-cine web site. They chose 3,878 reviews (opinions). Each one has associated a reviewer, a review summary, a body review and a score given by the reviewer to the movie. The movie score is a number between 1 and 5, where 1 means an unfavorable review (negative) and 5 a favorable one (positive).

In our experiments the reviews ranked as 3 were eliminated, because they are considered neutral reviews, then the total reviews considered for our experiments were 2,625. The reviews ranked 1 and 2 where considered negative and those ranked 4 and 5 as positive. Therefore, the experiments were carried out with a two class corpus: 1,351 positive review and 1,274 negative reviews.

4.2 Experiment Outcomes

To show the importance of attitude lexicon, three experiments were performed, the first using HLM lexicon, the second using the PBM lexicon, and the third combining both lexicons. Table 4 shows the results obtained where MD1 refers to the first model and MD2 to the second.

Table 4. Results with different lexicons.

Experiment	Measure Value		
	Precision	Recall	F-Measure
MD1-HLM Lex	0.634	0.621	0.615
MD1-PBM Lex	0.575	0.564	0.556
MD1-Both	0.625	0.614	0.608
MD2-HLM Lex	0.652	0.643	0.641
MD2-PBM Lex	0.575	0.564	0.556
MD2-Both	0.649	0.641	0.639

We can notice that combining polarity with the information of appraisal subsystems contributed to improve classification (MD1-HLM versus MD2-HLM or MD1-Both versus MD2-Both), in terms of recall and precision, using the HLM lexicon. However, increasing the lexicon did not help to improve classification (in fact, affecting the classification, see for instance MD2-HLM versus MD2-Both), given that the new entries did not include additional information about the appraisal subsystems.

Establishing the utility of appraisal groups has been difficult mainly because no all the movie reviews have an appraisal group. To evaluate their utility, the reviews where at least one appraisal group appears, were extracted. Then two experiments were executed: the first makes use of appraisal groups and the

second is limited to perform the count of lexicon values, without the use of appraisal groups. HLM lexicon is used, because is the only one built according to the Appraisal Theory (i.e. it has values in all the entries). Table 5 shows the results of both experiments.

Table 5. Results when using appraisal groups.

Experiment	Measure Value		
	Precision	Recall	F-Measure
MD1	0.668	0.642	0.605
MD1 w/AG	0.655	0.634	0.599
MD2	0.689	0.681	0.667
MD2 w/AG	0.719	0.706	0.693

First, we observe that in the first model (MD1, more limited, since it has less information of polarity) is affected by the addition of the appraisal information. However, we also notice that when representing separately the appraisal components (MD2), the classification improves by taking advantage of the polarity identified in the analysis, that seems to be a consistent significant improvements, but we still have to test such significance.

From results obtained by MD2 w/AG, it is relevant to analyze some cases where the classification was done correctly and incorrectly, as showed below:

In the movie reviews:

- Movie review 1) “*El guion fue un tanto **aburrido**, además de una **muy mala** interpretación del papel por parte de los actores principales, lo **aceptable** solo fue su **excelente** musicalización*”.

The attitude terms were detected, and the modifier “*muy*” preceding to the term “*mala*” was processed by duplicating the values that the word has in the attitude lexicon and then added to the corresponding vector entries, which increases finally the negative polarity review. Therefore, the review is correctly classified as negative

- Movie review 2) “*El mayor **fallo** de *Fin*, protagonizada por Maribel Verdú, Clara Lago y Daniel Grao, es que es casi **imposible sentir** algún tipo de **empatía** por sus personajes*.”

Similarly, the attitude terms are detected and their values are taken from the attitude lexicon. However, the phrase “*casi imposible sentir algún tipo de empatía*”, turns out to be a complex appraisal group that modifies the “*empatía*” word intention, but it can not be detected. So, the values that the word has in the attitude lexicon are not modified and they are directly added to the respective entries in the final vector. This problem causes that the classification model incorrectly assigns a positive polarity to the review.

still making improvements on This situation shows that the proposed method has a good performance with reviews that contain basic appraisal groups. However,

those groups that have a more complex grammatical structure can affect partially the classification of the review polarity. Therefore, we are working on improving the detection of more complex appraisal groups.

Despite the work done, the proposed method could not overcome the baseline, which used a vector representation of the reviews; this had a precision of 0.77. However, it uses vectors of 58,474 entries in contrast with our approaches that used vectors of 5 and 8 entries, and provides a good approximation in quite a short time.

5 Conclusions and Future Work

We have described a method to evaluate the polarity of opinions using the attitude subsystem of the Appraisal Theory, together with simple appraisal groups. As it was shown in experiments, when at least one appraisal group is present in a review (opinion) and separating the positive and negative values for each attitude attribute, the precision improves, obtaining a precision value quite close to that obtained by a traditional vector representation, so we would expect that in a corpus where opinions have an adequate number of appraisal groups, the advantage of our proposal would be more evident.

On the other hand, our method depends on the attitude lexicon as also showed in experiments, and then if we enrich the lexicon, we would expect a precision improvement, also.

Our results are also consistent with previous explorations of the advantages of using adverbs in sentiment analysis [16]. In our case, the contribution of adverbs in the appraisal groups was noticeable. This also opens the opportunity to consider a similar detailed attention to adverbs during our analysis of appraisal groups.

An advantage of our method is the vector dimension reduction from 58,474 to 8, which has a significant impact on storage and time processing. This can be advantageous when dealing with big data collections.

Our ongoing work consists on completing the lexicon (that unfortunately requires expert knowledge), considering more elaborated appraisal groups, and using or building other corpus appropriate to test our proposed approach and execute a qualitative analysis of the results.

We also plan to experiment adding grammatical knowledge of clauses and on different sentences lengths, as previous related research. We are also considering taking into account other syntactic class information as features for classification.

References

1. Chen H., Zimbra D.: AI and opinion mining. *Intelligent Systems, IEEE*, 25(3), pp. 74–80 (2010)
2. Read J., Carroll J.: Annotating expressions of Appraisal in English. *Lang Resources & Evaluation*, vol. 46, pp. 421–447 (2012)

3. Balahur A., Hermida J. M., Montoyo A.: Building and exploiting emotinet, a knowledge base for emotion detection based on the appraisal theory model. *IEEE Transactions on Affective Computing* 3(1), pp. 88–101 (2012)
4. Gao Y., Zhou L., Zhang Y., Xing C., Sun Y., Zhu X.: Sentiment classification for stock news. In: *5th International Conference on Pervasive Computing and Applications (ICPCA)*, pp. 99–104, IEEE Press (2010)
5. Singh V. K., Piryani R., Uddin A., Waila P.: Sentiment analysis of movie reviews: A new feature-based heuristic for aspect-level sentiment classification. In: *2013 International Multi-Conference on Automation, Computing, Communication, Control and Compressed Sensing (iMac4s)*, pp. 712–717, IEEE Press (2013)
6. Neviarouskaya A., Prendinger H., Ishizuka M.: Recognition of affect, judgment, and appreciation in text. In: *Proceedings of the 23rd International Conference on Computational Linguistics, Association for Computational Linguistics*, pp. 806–814 (2010)
7. Martin J. R.: Reading Positions/Positioning Readers: JUDGEMENT in English. *Prospect: a Journal of Australian TESOL*, 10(2), pp. 27–37 (1995)
8. Martin J. R.: Beyond Exchange: APPRAISAL Systems in English. In: *Evaluation in Text*, S. Hunston & G. Thompson (Eds.), pp. 142–175, Oxford: Oxford University Press (2000)
9. Martin J. R., White P. R. R.: *The Language of Evaluation: Appraisal in English*. London, U.K.: Palgrave (2005)
10. Taboada M., Grieve J.: Analyzing Appraisal Automatically. In: *Proceedings of AAAI Spring Symposium on Exploring Attitude and Affect in Text*, pp. 158–161, AAAI Technical Report SS-04-07, Stanford University, CA, AAAI Press (2004)
11. Wang J., Dong A.: A comparison of two text representations for sentiment analysis. In: *2010 International Conference on Computer Application and System Modeling (ICCASM)*, vol. 11, pp. V11–35, IEEE Press (2010)
12. Hernández L., López-López A., Medina J. E.: Classification of Attitude Words for Opinion Mining. *International Journal of Computational Linguistics and Applications* 2(1-2), pp. 267–283 (2011)
13. Pérez-Rosas V., Banea C., Mihalcea R.: Learning Sentiment Lexicons in Spanish. In: *Proceedings of the International Conference on Language Resources and Evaluations (LREC 2012)*, pp. 3077–3081 (2012)
14. Whitelaw C., Garg N., Argamon S.: Using appraisal groups for sentiment analysis. In: *Proceedings of the 14th ACM international conference on Information and knowledge management*, pp. 625–631, ACM (2005)
15. Cruz F. L., Troyano J. A., Enríquez F., Ortega J.: Clasificación de documentos basada en la opinión: experimentos con un corpus de críticas de cine en español. *Procesamiento del Lenguaje Natural*, vol. 41, pp. 73–80 (2008)
16. Benamara F., Cesarano C., Picariello A., Reforgiato D., Subrahmanian V. S.: Sentiment analysis: Adjectives and adverbs are better than adjectives alone. In: *Proceedings of the International Conference on Weblogs and Social Media (ICWSM)* (2007)