Polarity Classification of Tourism Reviews in Spanish

Clasificación de polaridad de revisiones de turismo en español

Iñaki San Vicente Roncal
Elhuyar Fundazioa
Zelai Haundi 3, 20170 Usurbil
i.sanvicente@elhuyar.com

Xabier Saralegi Urizar
Elhuyar Fundazioa
Zelai Haundi 3, 20170 Usurbil
x.saralegi@elhuyar.com

Resumen: Este artículo presenta un clasificador de polaridad para críticas de recursos turísticos en español. Se ha creado una nueva colección de datos compuesta por críticas de recursos turísticos (hospedajes, restaurantes y actividades) del País Vasco en Español, extraídos de la web de críticas TripAdvisor. Adoptamos una estrategia supervisada y analizamos varios modelos configurados según diferentes atributos: un modelo de unigramas y otros basados en la información léxica proporcionada por un lexicón de polaridad adaptado al dominio del turismo. El sistema basado en el léxico obtiene un 83% de precisión para la tarea de clasificación de 3 categorías, y un 57% a la hora de clasificar 5 categorías. La mejora respecto al modelo de unigramas no es significativa, pero el número de atributos se reduce a la mitad, redundando en una mejora de la eficiencia. Asimismo, se ha evaluado el sistema para diferentes sub-dominios del turismo, que incluyen alojamientos, restaurantes y actividades.
Palabras clave: Análisis de sentimiento, Minería de opiniones, Detección de polaridad

Abstract: This article describes a polarity classifier for Spanish tourism reviews. We created a new data-set comprised by reviews of tourism resources (accommodations, restaurants, and activities) from the Basque Country in Spanish, by crawling the TripAdvisor review website. We adopt a supervised approach, and analyze various feature sets: an unigram model and various models that rely on the lexical information provided by a polarity lexicon, adapted to the tourism domain. The lexicon-based system achieves 83% accuracy for a 3-category classification task, and a 57% accuracy for a 5-category classification. Although the improvement over the unigram model is not significant it uses the half number of features which is more efficient. On top of that, evaluation is carried out for tourism resources sub-domains, including accommodation, restaurants and activities.
Keywords: Sentiment Analysis, Opinion-mining, Polarity detection

1 Introduction

The tourism sector has seen during the last years how traditional selling systems based on intermediary agents have lost ground in favor of the Internet which offer greater direct sale capabilities. The rise of the Web 2.0 and the expansion of resources involving user-/client participation has brought down the barriers between the off-line and the on-line worlds. Consumers opinion are now public and accessible to everyone. Those opinions are becoming more and more relevant in decision processes such as the election of a holiday destiny, or the hotel we will stay. According to (Comscore/the K. Group, 2007; Horrigan, 2008), and based on studies carried out over a sample of 2,000 adults, the 81% of the Internet users have done research on a resource online, at least once, and 73% to 87% of tourism review site users (including restaurants hotels and other services) admit that the reviews they read had great influence on the final decision. Moreover, according to those studies, consumers are ready to pay more for resources with higher assessments. Users do contribute to this new system, 32% of them have submitted at least one review online. The Travel Industry Association of America estimates that 67% of travelers/tourists with Internet access look for information about their possible destinies on the net (Travel Industry Association and others, 2005).

Being able to identify and extract the opinions of users about topics or resources would enable many organizations to obtain global feedback on their activities. Some studies (O'Connor et al., 2010) have pointed out that such systems could perform as well as traditional polling systems, but at a much lower cost. In this context, social media like microblogs and user review sites constitute a very valuable source when seeking opinions and sentiments. Our final objective is to provide a tool for classifying and summarizing opinions referring to a certain resource and its characteristics. The work in this paper focuses on classifying a review as positive, neutral or negative in a scale from very negative to very positive. The review can be only a few lines long or it can be up to some paragraphs.

We compare an unigram model with models that rely on the lexical information provided by a polarity lexicon. The polarity lexicon is built by translating an existing lexicon and is adapted to the tourism domain by extending it with polarity words automatically extracted from corpora. We created a new data-set comprised by reviews of tourism resources (accommodations, restaurants, and activities) from the Basque Country area in Spanish, by collecting reviews from the TripAdvisor[1] website. The data-set is divided in order to separate testing examples from examples used for training purposes and for extracting polarity words as well. Evaluation is done over the whole test-set, and also over various sub-sets composed of the different resources of in our data-set.

The rest of the paper is organized as follows. Section 2 reviews the state of the art in the polarity detection field, placing special interest on paragraph level detection, and on tourism reviews, in particular. The third section describes the data-sets we built for the experiments. Next the method for constructing the polarity lexicon is explained. Following we describe the classifiers we developed, and the features we included in our supervised system. The next section presents the evaluation we performed and results obtained over the test-sets. The last section draws some conclusions and future directions.

2 Related Work

Regarding the algorithms used in sentiment classification, although there are approaches based on averaging the polarity of the words appearing in the text (Choi and Cardie, 2009; Hu and Liu, 2004; Kim and Hovy, 2004; Turney, 2002), machine learning methods have become the more widely used approach. Pang et al. (Pang, Lee, and Vaithyanathan, 2002) proposed a unigram model using Support Vector machines which does not need any prior lexicon to classify movie reviews. Read (Read, 2005) confirmed the necessity to adapt the models to the application domain, and Choi and Cardie (Choi and Cardie, 2009) address the same problem for polarity lexicons.

As for the polarity classification of Spanish texts, it must be mentioned the task of classifying Spanish tweets organised within the TASS 2012 workshop (Villena-Román et al., 2013) where both rule-based and supervised systems took part. Another interesting work dealing with polarity over Spanish texts is (Vilares, Alonso, and Gómez-Rodrıguez, 2013) which introduces a classification based on syntactic information.

Regarding the tourism domain, different tasks have been addressed by various authors. Waldhör (Waldhör and Rind, 2008) proposes to use Opinion Mining for monitoring the satisfaction level with respect to tourism resources in blogs and travel forums. The system includes a web crawler and a sentiment classifier. The final prototype sends alerts to a company (e.g., a hotel) that has received a number of negative comments in the analyzed sources. (Shimada et al., 2011) present a system that does a complete analysis of a target destination, including service demand, opinion mining and visitors timeline patterns. Polarity classification is done using a rule-based system which relies on a polarity lexicon. Although they work with tweets, they point out the necessity to include other source such as blogs, especially for less known destinations with little presence in Twitter.

[1]http://www.tripadvisor.es

(Ye, Zhang, and Law, 2009) compares various supervised methods, for classifying the polarity of reviews on 7 popular destinations of USA and Europe. SVM and n-gram based classifiers perform best, achieving an 80% accuracy. Zheng and Ye (Zheng and Ye, 2009) classify hotel reviews in chinese from the www.ctrip.com site, obtaining 80% accuracy with an SVM classifier.

Lin and Chao (Chao and Lin, 2010) detect and classify opinions in tourism blog articles. They propose differents strategies to process the target opinions: keywords, correference expressions and machine learning. Their best precision and recall results are 51.30% and 54.21% respectively. (Gräbner et al., 2012) classify separately hotel reviews as positive, negative or neutral. Their approach consist on extracting polarity lexicons adapted each of the categories. 84%, 80% and 92% accuracies are obtained for positive, negative and neutral categories respectively.

3 Data-sets

In order to develop a supervised system, we need polarity annotated corpora. In our case, and since we are dealing with tourism reviews, we have resorted to the web in order to construct such a corpus. This is a suitable strategy for us, because on the one hand, it is a cheap way to get annotated data, and on the other, because the system's final version will work in that environment.

The data-set C used in this work is composed of Spanish reviews crawled from the TripAdvisor review site. The crawling is limited to the geographical area of the Basque Country. Reviews for three types of tourism resources were extracted:

- Accommodations: Here we include hotels, hostels, bed&breakfasts and cottages. House rentals were discarded.
- Restaurants.
- Activities: TripAdvisor differentiates beween activities, attractions and shopping. We only extracted information related to the activities category.

A total amount of 30672 reviews are included in C, corresponding to 2419 different resources (see Table 1). Initially more than 62,000 reviews were collected, but only those in Spanish were selected. Some of the reviews that were not written in Spanish were machine translated. After a manual evaluation of a sample of 50 such reviews, we concluded that the quality of the machine translated texts varies greatly from one text to the other. Thus, we discarded such reviews. The corpus is divided in two parts created by choosing reviews randomly: the first one, C_{train}, contains the 75% of the reviews in C and it is used both for training the classifier and for inferring polarity words for the lexicon. The second part, C_{test}, contains the other 25% of the reviews and is used for evaluating our system.

Apart from C, we created three sub-sets corresponding to each of the aforementioned resource types, C_acc (Accommodations), C_res (Restaurants) and C_Act (Activities). Our objective is to measure if creating a general tourism classifier is enough to successfully classify reviews of its sub-domains or if a more refined domain adaptation is still necessary. Again, those sub-sets are divided in two parts for training and testing, following the same methodology as before.

Corpus	**# Resources**	**# Reviews**
Accommodations C_acc	1,127	14,530
Restaurants C_res	1,655	11,706
Activities C_act	195	4,439
Total C	2,419	30,672

Table 1: Statistics for the collected reviews.

For each review, apart from some other metadata, the corpus includes the global rating the user gave to the resource and the text of the review. We do not require a minimum length to include a review in the corpus C. Reviews can be a single sentence or a few paragraphs long. The average review length in C is 91 words with a very high standard deviation ($\sigma = 84.1$). Other information is also stored in the corpus, such as the rates given to specific features, but that information is not relevant for this work.

The global rating used by TripAdvisor is ranged from 1.0 to 5.0. We map those punctuations as Table 2 shows. As we can see in the same table, positive reviews are far more than the negative ones.

4 Polarity Lexicon

Our initial lexicon P_{es} is a general domain Spanish polarity lexicon (Saralegi and San Vicente, 2012) created for the TASS

Trip Advisor Rating	Our Annotation scheme	#Reviews in C_{train}	#Reviews in C_{test}
1.0	Strong negative (N+)	1,003 (4.36%)	349 (4.55%)
2.0	Negative (N)	1,125 (4.89%)	415 (5.41%)
3.0	Neutral (NEU)	3,355 (14.58%)	1,098 (14.31%)
4.0	Positive (P)	7,756 (33.72%)	2,561 (33.4%)
5.0	Strong Positive (P+)	9,665 (42.01%)	3,245 (42.32%)

Table 2: Rating distribution of the collected reviews.

2012 Workshop. The process of building the lexicon consisted of translating the OpinionFinder English polarity lexicon (Wilson et al., 2005) automatically and then manually correcting both translations and polarities.

In order to adapt P_{es} to the tourism domain, we extended the initial lexicon with words automatically extracted from the training corpus C_{train}. In order to extract the words most associated with a certain polarity; let us say positive, we divided the corpus into two parts: positive reviews and the rest of the corpus. Using the Loglikelihood ratio (LLR) we obtained the ranking of the most salient words in the positive part with respect to the rest of the corpus. The same process was conducted to obtain negative candidates. The top 2,000 negative and top 2,000 positive words (and also multiword units, e.g., *"visita obligada", "hoja de reclamaciones"*) were manually checked. Among them, 309 negative and 271 positive words were selected for the polarity lexicon PT_{es} (see sixth column in Table 3).

polarity	#words in the initial lexicon P_{es}	#words manually selected from C_{train}	#words in final lexicon PT_{es}
negative	2,435	309	2,744
positive	1,518	271	1,799
Total	3,953	580	4,543

Table 3: Statistics of the polarity lexicons.

5 *Supervised System*

We chose to build a supervised classifier, because it allowed us to combine the various features more effectively. We used the SMO implementation of the Support Vector Machine algorithm included in the Weka (Hall et al., 2009) data mining software. Default configuration was used. All the classifiers built over the training data C_{train} were evaluated against the test-set C_{test}.

We apply some heuristics in order to preprocess the reviews and solve the main problems detected in user generated contents:

- Replication of characters (e.g., *"Sueñooo"*): Sequences of the same characters are replaced by a single character when the pre-edited word is not included in Freeling's (Padró et al., 2010) dictionary and the post-edited word appears in Freeling's dictionary.
- Abbreviations (e.g., *"q"*, *"dl"*, ...): A list of abbreviations is created from the training corpus. These abbreviations are extended before the lemmatisation process.
- Overuse of upper case (e.g., *"MIRA QUE BUENO"*). Upper case is used to give more intensity to the review. If we detect a sequence of two words all the characters of which are upper case and which are included in Freeling's dictionary as common, we change them to lower case.

5.1 Unigram model

We implemented an unigram representation using all lemmas in the training corpus as features (10,549 altogether). Lemmatisation was done by using Freeling (Padró et al., 2010). Contrary to (Pang, Lee, and Vaithyanathan, 2002) who used the presence of the lemmas, we stored the frequency of the lemmas in a review. We tested the unigram model using either frequency or presence, and frequency obtained better results for both 5-category (1.8% improvement) and 3-category (1.5% improvement) classification. Thus, for the sake of simplicity, all the experiments reported in this paper make use of the frequency.

5.2 Lexicon-based models

Emoticons and interjections are very strong expressions of sentiments. A list of emoticons is collected from a Wikipedia article about emoticons and all of them are classified as positive (e.g., ":)", ":D" ...) or negative (e.g., ":(" , "u_u" ...). 23 emoticons were classified as positive and 35 as negative. A list of 54 negative (e.g., *"mecachis"*, *"sniff"*, ...) and 28 positive (e.g., *"hurra"*, *"jeje"*, ...) interjections including variants modelled by regular expressions were also collected from different webs as well as from the training corpora. The frequency of each emoticon and interjection type (positive or negative) is included as a feature of the classifier.

These clues did not provide significant improvement. Although in other domains such features indeed help to detect the polarity (Koulompis, 2011) the low density of such elements in review texts (only 394 were found in 30,672 reviews) explains the low impact of these features in our case.

5.2.1 Selection of Polarity Words (SP)

Only lemmas corresponding to words included in the polarity lexicon PT_{es} (see section 4) were selected as features. This allows the system to focus on features that express the polarity, without further noise. Another effect is that the number of features decreases significantly (from 10,549 to 4,543), thus reducing the computational costs of the model.

5.2.2 Frequency of Polarity Words (FP)

The SP classifier does not interpret the polarity information included on the lexicon. We explicitly provide that information as a feature to the classifier. Furthermore, without the polarity information, the classifier will be built taking into account only those polarity words appearing in the corpus. Including the polarity frequency information explicitly, the polarity words included in the PT_{es} but not in the corpus will be used by the classifier.

Two new features are created to be included in the polarity information: a score of the positivity and a score of the negativity of a review. In principle, positive words in PT_{es} add 1 to the positivity score and negative words add 1 to the negativity score. However, depending on various phenomena, the score of a word can be altered. These phenomena are explained below.

Treatment of Negations and Adverbs

The polarity of a word changes if it is included in a negative clause. Syntactic information provided by Freeling is used for detecting those cases. The polarity of a word increases or decreases depending on the adverb which modifies it. We created a list of increasing (e.g., *"mucho"*, *"absolutamente"*, ...) and decreasing (e.g., *"apenas"*, *"poco"*, ...) adverbs. If an increasing adverb modifying a polarity word is detected, the polarity is increased (+1). If it is a decreasing adverb, the polarity of the words is decreased (−1). Syntactic information provided by Freeling is used for detecting these cases.

Text Position Information

The importance of the word in the review determines the influence it can have on the polarity of the whole review. Important sentences can be located in certain fixed positions in the text, such as first and last sentences (Edmundson, 1969). After testing different weighting strategies (words in the first and last positions in text are weighted higher, only words in last positions, and only words in first positions), we obtained the best results by rewarding words in the first positions of the text ($weight(w) = 1/position(w)$).

6 Evaluation and Results

Two tasks were set up in order to evaluate the feature sets we designed: a fine grained polarity classification task and a coarse grained one.

In the first one reviews have to be classified in five categories (P+, P, NEU, N and N+), using the annotation scheme presented in section 3. This task aims to give us insight on whether our system is capable of identifying the intensity of a review.

The second task consist on classifying the reviews between positive (P), negative (N) and neutral (NEU). For that task all reviews regarded as positive (P and P+) are grouped into a single category (P), and the same is done for negative reviews (both N and N+ reviews become N). This task has the objective of testing whether the classifier is able to determine the polarity of a review, without taking into account its intensity.

All the evaluation experiments have been done by testing the models trained on the C_{train} corpus against the C_{test} test-set.

The results show that the overall performance of the systems is higher on the 3-category classification than on 5-category classification. If we analyze the performance category by category on the 5-category task (see Table 4) we see that the system performs best on the P+ category, while NEU and specially N are the worst classified categories.

Metric/ System	Acc. (5 cat.)	P+	P	NEU	N	N+
Baseline	.575	.694	.517	.452	.31	.497
SP	.575	.701	.52	.398	.255	.485
SP+FP	.576	.701	.52	.408	.256	.505

Table 4: Accuracy results obtained for the 5-category classification, and f-score results for each category.

Looking at the confusion matrix (see Table 5) we can see that NEU examples tend to be classified as P and N examples tend to be classified as NEU. These results rises doubts about the annotation scheme used in TripAdvisor. Our hypothesis is that users tend to give not too negative reviews, and thus, they do not regard a middle rating as neutral, but as fair. Therefore, only clearly negative reviews have enough "clues" for the classifier to label them correctly.

classified as →	P+	P	NEU	N	N+
P+	2426	769	44	6	0
P	1029	1369	150	11	2
NEU	158	482	395	42	21
N	42	59	172	82	60
N+	18	25	75	85	146

Table 5: Confusion Matrix of the SP+FP classifier for the 5-category classification.

Regarding to the performance when identifying the intensity of the reviews, if we look at the confusion matrix of the 5-category classifier (see Table 5) we realize that many classification errors are because the system incorrectly classifies P and P+ classes, and also N and N+ to some extent. Thus, we conclude that our configuration is not adequate at the moment to identify the intensity of opinions. When we turn our eyes to the 3-category classification task (see Table 7), where P and P+ are grouped, and so are N and N+, the aforementioned errors are not present and the overall performance of the classifier improves greatly.

Otherwise, 3-category classification task results remain similar to the 5-category ones. Best classification performance is achieved over positive reviews. NEU reviews are again the most difficult ones for classifying and tend to be classified as P, according to the confusion matrix (Table 7).

As for different classifiers, results show

Metric/ System	Acc. (3 cat.)	P	NEU	N
Baseline	.831	.915	.445	.672
SP	.83	.914	.335	.66
SP+FP	.836	.918	.358	.676

Table 6: Accuracy results obtained for the 3-category classification, and f-score results for each category.

classified as →	P	NEU	N
P	5665	119	22
NEU	709	299	90
N	164	153	447

Table 7: Confusion Matrix of the SP+FP classifier for the 3-category classification.

that there are only slight differences between them. Even if the polarity lexicon-based model with statistical information (SP+FP in tables 6 and 4) outperforms the unigram model, the improvement is not statistically significant. However the lexicon-based classifier is computationally much more efficient, since it uses half the number of features.

Resource type evaluation

As a second part of the evaluation, we analyzed the performance of the best system (SP+FP) with regard to the resource type. Since the results for the 5-category classification task showed that the classifier does not adequately differentiate the intensity of the reviews, we only provide results for this experiments on a 3-category classification basis.

We tested each of the resource type r test-sets against the SP+FP classifier, trained over three different sets of examples:

- Training based on examples of the other two resource types (C_{train}-C_r_{train}): this set is built by taking out the example in C_{train} that belong to the same resource type as r. We can consider this as an out-of-domain training-set.
- Trained on C_{train} examples: this set represents a training-set of the general tourism domain, including both in-domain and out-of-domain examples.
- Trained only on examples of the same resource type (C_r_{train}): this set can be regarded as an in-domain training-set.

The examples in this set are those examples in C_{train} that correspond to the same resource type as r.

Test-set / Training-set	Restaurants	Accommodations	Activities
C_{train}-C_r_{train} Examples of the other resource types	.812	.799	.858
C_{train}	.837	.828	.859
C_r_{train} resource type examples	.839	.83	.859

Table 8: Accuracy results for each resource type over the various training-sets, in the 3-category classification task.

Results in Table 8 show that, as expected, training over examples of the same type of resources provides the best results (C_r_{train}). However, including examples of other resource types C_{train} trains a classifier which performs almost as good as the best one and with a broader coverage. Finally, not including in-domain examples in the training-set (C_{train}-C_r_{train}) leads to a decrease in the accuracy, which is notable in the case of accommodations and Restaurants. In the case of activities, there is almost no decrease. Overall, accommodations seem to be the most sensitive to the lack of adequate examples, and thus, they achieve the lowest results.

7 Conclusions

This paper presents SVM classifiers for assigning the polarity to Spanish tourism reviews. One system relies on a semi-automatically built polarity lexicon, but it also includes some statistical features. That system has a very good performance, comparable with state of the art results, achieving 83% accuracy for the 3-category classification task, and 57% accuracy for the 5-category classification. The little improvement obtained over the unigram model leads to the conclusion that lexicon-based word selection and lexicon-based statistics are not more useful than a minimum frequency based unigram selection in terms of accuracy, at least in the tourism domain. Nevertheless, even if the improvement obtained with the lexicon-based models is not statistically significant, the fact that it has the half number of features is an advantage in terms of computational efficiency.

Regarding the classification of the different touristic resource types, we have showed that a classifier trained over all resource types performs as well as independent in-domain classifiers. However, results also show that only using examples of other resource types leads to a performance decrease, which surfaces the need to adapt the system not only to the tourism domain but also to the resource type, at least to some extent.

Lexicon-based approaches have been indeed successful in other domains, so we would like to analyze deeply the reasons behind our results. Other short term goals include identifying the features of a target resource, and detecting the polarity of those features, which would give us the possibility to summarize opinions according to those features.

Acknowledgments

This work has been partially founded by the Industry Department of the Basque Government under grant IE11-305 (KnowTOUR project).

References

Chao, Chuan-Jie and Pin-Hsien Lin. 2010. Tourism-related opinion mining. In *Proceedings of the 22nd Conference on Computational Linguistics and Speech Processing (ROCLING 2010)*.

Choi, Yejin and Claire Cardie. 2009. Adapting a polarity lexicon using integer linear programming for domain-specific sentiment classification. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2 - Volume 2*, EMNLP '09, pages 590–598.

Comscore/the K. Group. 2007. Online Consumer-Generated Reviews Have Significant Impact on Offline Purchase Behavior. November.

Edmundson, H. P. 1969. New methods in automatic extracting. *J. ACM*, 16(2):264–285, April.

Gräbner, Dietmar, Markus Zanker, Günther Fliedl, and Matthias Fuchs. 2012. Classification of customer reviews based on sentiment analysis. In *19th Conference on Information and Communication Technologies in Tourism (ENTER)*. Springer.

Hall, Mark, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. 2009. The WEKA data mining software: an update. *SIGKDD Explor. Newsl.*, 11(1):10–18, november.

Horrigan, J. A. 2008. Online shopping. Technical report.

Hu, M. and B. Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the 10th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177.

Kim, Soo-Min and Eduard Hovy. 2004. Determining the sentiment of opinions. In *Proceedings of the 20th international conference on Computational Linguistics*, COLING '04.

O'Connor, Brendan, Ramnath Balasubramanyan, Bryan R. Routledge, and Noah A. Smith. 2010. From tweets to polls: Linking text sentiment to public opinion time series. In *4th International AAAI Conference on Weblogs and Social Media*, May.

Padró, Lluís, Miquel Collado, Samuel Reese, Marina Lloberes, and Irene Castellón. 2010. Freeling 2.1: Five years of open-source language processing tools. In *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC'10).*

Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing*, EMNLP '02, pages 79–86.

Read, Jonathan. 2005. Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In *Proceedings of the ACL Student Research Workshop*, ACLstudent '05, pages 43–48.

Saralegi, Xabier and Iñaki San Vicente. 2012. Tass: Detecting sentiments in spanish tweets. In *Proceedings of the TASS Workshop at SEPLN.*

Shimada, Kazutaka, Shunsuke Inoue, Hiroshi Maeda, and Tsutomu Endo. 2011. Analyzing tourism information on twitter for a local city. In *Software and Network Engineering (SSNE), 2011 First ACIS International Symposium on*, pages 61–66. IEEE.

Travel Industry Association et al. 2005. Executive summaries-travelers' use of the internet, 2004 edition. *Washington, DC: Author.*

Turney, Peter D. 2002. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics - ACL '02*, page 417, Philadelphia, Pennsylvania.

Vilares, David, Miguel A Alonso, and Carlos Gómez-Rodrıguez. 2013. Clasificación de polaridad en textos con opiniones en español mediante análisis sintáctico de dependencias. *SEPLN*, 50:13–20.

Villena-Román, Julio, Sara Lana-Serrano, Eugenio Martínez-Cámara, and José Carlos González-Cristóbal. 2013. Tass-workshop on sentiment analysis at sepln. *SEPLN*, 50:37–44.

Waldhör, Klemens and Alexander Rind. 2008. etbloganalysis —mining virtual communities using statistical and linguistic methods for quality control in tourism. In *Information and Communication Technologies in Tourism 2008.* pages 453–462.

Wilson, Theresa, Paul Hoffmann, Swapna Somasundaran, Jason Kessler, Janyce Wiebe, Yejin Choi, Claire Cardie, Ellen Riloff, and Siddharth Patwardhan. 2005. OpinionFinder. In *Proceedings of HLT/EMNLP on Interactive Demonstrations*, pages 34–35, Vancouver, British Columbia, Canada.

Ye, Qiang, Ziqiong Zhang, and Rob Law. 2009. Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. *Expert Systems with Applications*, 36(3):6527–6535.

Zheng, Wenying and Qiang Ye. 2009. Sentiment classification of chinese traveler reviews by support vector machine algorithm. In *3d International Symposium on Intelligent Information Technology Application - IITA 2009.*, volume 3, pages 335–338. IEEE.