

17th International Conference in Knowledge Based and Intelligent Information and Engineering Systems - KES2013

Identifying Customer Preferences about Tourism Products using an Aspect-Based Opinion Mining Approach

Edison Marrese-Taylor^a, Juan D. Velásquez^a, Felipe Bravo-Marquez^b, Yutaka Matsuo^c

^aDepartment of Industrial Engineering, University of Chile, República #701, Santiago, Chile

^bDepartment of Computer Science, University of Chile, Blanco Encalada #2120, Santiago, Chile

^cInstitute of Engineering Innovation, Graduate School of Engineering, The University of Tokyo 2-11-16 Yayoi, Bunkyo-ku, Tokyo 113-8656, Japan

Abstract

In this study we extend Bing Liu's aspect-based opinion mining technique to apply it to the tourism domain. Using this extension, we also offer an approach for considering a new alternative to discover consumer preferences about tourism products, particularly hotels and restaurants, using opinions available on the Web as reviews. An experiment is also conducted, using hotel and restaurant reviews obtained from TripAdvisor, to evaluate our proposals. Results showed that tourism product reviews available on web sites contain valuable information about customer preferences that can be extracted using an aspect-based opinion mining approach. The proposed approach proved to be very effective in determining the sentiment orientation of opinions, achieving a precision and recall of 90%. However, on average, the algorithms were only capable of extracting 35% of the explicit *aspect expressions*.

© 2013 The Authors. Published by Elsevier B.V.

Selection and peer-review under responsibility of KES International

Keywords: opinion mining, aspect-based, tourism, customer preferences, natural language processing, web mining

1. Introduction

With the inception of the Web 2.0 and the explosive growth of social networks, enterprises and individuals are increasingly using the content in these media to make better decisions [1], [2], [3]. More people are checking the opinions of other shoppers before buying a product when trying to make a good choice. On the other hand, for organizations, the vast amount of information available publicly on the Web could make polls, focus groups and some similar techniques an unnecessary requirement in market research. In particular, results provided by aspect-based opinion mining techniques could represent a real alternative in finding customer preferences about a product. An aspect-based opinion mining approach permits analyzing opinions about product features such as product components and attributes. As established in Lancaster's *new theory of consumer demand*, customer preferences about a product are intrinsically related to its features, i.e. *aspects*. He states that consumer behavior is a process of choosing bundles of product characteristics or attributes inherent in goods and services, rather than simply choosing bundles of goods or services themselves [4]. Thus, discovering what these features are and

E-mail address: emarrese@wi.dii.uchile.cl.

defining how customers feel about these features will undoubtedly lead to a better comprehension of consumer preferences.

In this study, we propose an extension of Liu's aspect-based opinion mining methodology in order to apply it to the tourism domain. This extension is concerned with the fact that users refer differently to different kinds of products when writing reviews on the web. Consider a *generic product*, which refers to the conceptual commodity produced by an industry. This product can take a wide variety of real forms, each one of them providing the same function [5]. In the literature, most of the authors, including Kotler [6], tend to classify these generic products using two categories: physical goods and intangible services. However, these are not discrete categories but rather a continuum with pure service on one terminal point and pure commodity on the other [7]. Most of the existing work in product reviews mining, including Liu's, is focused on physical product reviews. In these kinds of reviews, users generally go straight to the point and talk directly about the product features they liked or did not like. Furthermore, few people will care about issues like who has designed or manufactured the product. However, for other kinds of products, different phenomena occur. For instance, when a person writes a movie review, he probably comments not only on movie elements, but also on movie-related people [8]. If we focus on the other terminal point, considering tourism products like restaurants, which provide a physical good (the food) but also services in the form of ambience and the setting, we see some other special features appear. Consider the following example, taken from a real review in TripAdvisor:

“When we arrived to the hotel, it looked really good and only after trying several rooms we discovered the whole hotel was really mouldy in the interior. I barely had enough room to move around the 2 very small/short twin beds and the bathroom was smaller than most standard closets.”

We see that users tend to *tell stories* about their experiences when writing reviews about tourism products. These stories are more likely to have longer and more complex sentences, which often include features in them that are mentioned multiple times. Reviewers also usually mention objects that do not correspond to attributes or components of the reviewed product and use many different and complex expressions to refer to the things that we actually consider as aspects. Finally, a considerable number of sentences do not contain opinions.

To the best of our knowledge, existing approaches do not address these special issues. Here, the contribution of our extension consisted in the development of new rules to cover the appearance of those aspects that are composed of more than one term and to also cover those aspects that appear more than once in a sentence. In addition, we include the formalization of some concepts and the creation of special corpora or datasets for the evaluation of our proposals. Ultimately, we also state that a system implementing our extension would permit the discovery of customer preferences when applied to tourism product reviews. These preferences are conceived as an individual's attitude towards a set of objects, typically reflected in an explicit decision-making process [9] or as an evaluative judgment in the sense of liking or disliking an object [10]. In the case of the tourism industry, the study of customer preferences is usually implemented using traditional tools that fail to cover a significantly representative number of participants because they are applied to specific groups of people. In this context, aspect-based opinion mining offers a larger scope method to understand aggregated consumer preferences.

The rest of this paper is structured in the following manner. In the first place, we briefly present most of the important ideas about aspect-based opinion mining in section 2, with a special mention of the approach developed by Liu et al. After that, in section 3, we extend Liu's ideas and propose our own approach. Later, in section 4, we present an experimental setup to evaluate how the proposed approach performs for tourism product reviews in the X Región de Los Lagos, Chile, also showing the most important results. This intends to encompass the current situation in the region, where tourism operators try to understand customer preferences using studies with limited scope. Finally, in section 5, the main conclusions of this paper and future work are detailed.

2. Previous Work

Aspect-based opinion mining techniques divide input texts into *aspects*, also called features or subtopics in literature, that usually correspond to arbitrary topics considered important or representative of the text that is being analyzed. The aspect-based approach is very popular and many authors have developed their own perspectives and

models. Examples of them are the works of Lu et al. [11], Popescu and Etzioni [12], Archak et al. [13], Decker and Trusov [14], Ku et al. [15], Titov and McDonald [16], Zhuang et al. [8] and Zhao and Li [17]. However, research showed that the work of Liu et al. is probably the most comprehensive one in this context and that is why it was used here by us as inspiration. Reviews of the state-of-the-art opinion mining techniques can be found in [18], [19] and [20].

2.1. Initial Definitions

Liu's approach in [19] proposes that opinions are 5-tuples, composed of (1) An *entity*: Proposed to denote the opinion objective or, in other words, what is being evaluated by the opinion. An *entity* can contain a set of components and attributes and, similarly, each *entity* component can have its own subcomponents and attributes. Finally, an *entity* can be decomposed into a tree or hierarchy of subattributes and subcomponents. (2) An *aspect*: Because it is difficult to study an *entity* at an arbitrary hierarchy level, this hierarchy is simplified to one or two levels, denoting as *aspect* every component or attribute of the *entity*. In this way, the root of the hierarchy or tree becomes the *entity* itself, each leaf is an *aspect* and links are part-of relationships. (3) The *Sentiment orientation*, considering that opinions express a *positive* or *negative* sentiment about what they evaluate. (4) The *Opinion holder*, which corresponds to the user (a person, an enterprise, etc.) that gives the opinion. (5) *Time*: Time and date when the opinion was given. Opinions are then considered to be a positive or negative view, attitude, emotion or appraisal about an *entity* or an *aspect* of that *entity* from an *opinion holder* in a specific time. The following concepts are also introduced:

- *Entity expression*: Corresponds to the actual word or phrase written by the user to denote or indicate an *entity*. As a result, *entities* are then generalizations of every *entity expression* used in the analyzed documents, or a particular occurrence of an *entity expression*. In [19] this concept is called *entity name*.
- *Aspect expression*: As for an *entity expression*, the *aspect expression* is the actual word or phrase written by the user to denote or indicate an *aspect*. Thus, aspects are also general concepts that comprise every *aspect expression*. They are called *aspect names* by Bing Liu.

It is then possible to define a model of an *entity* and a model of an opinionated document. In this manner, an *entity* e_i is represented by itself as a whole and a finite set of *aspects*, $A_i = \{a_{i1}, a_{i2}, \dots, a_{im}\}$. The *entity* can be expressed with any one of a final set of *entity expressions* $EE_i = \{ee_{i1}, ee_{i2}, \dots, ee_{is}\}$. Each *aspect* a_{ij} of A_i of *entity* e_i can be expressed by any one of a finite set of *aspect expressions* $AE_{ij} = \{ae_{ij1}, ae_{ij2}, \dots, ae_{ijm}\}$. On the other hand, an opinionated document $d_k \in D$ contains opinions on a set of *entities* e_1, e_2, \dots, e_r from a set of opinion holders h_1, h_2, \dots, h_p . The opinions on each *entity* e_i are expressed on the *entity* itself and a subset A_{ik} of its *aspects*.

2.2. Process Steps

Kim et al. gives a good review of historical and state-of-the-art aspect-based developments in [18]. The authors indicate that the process is commonly made up of three distinct steps, which are also considered by Liu.

1. Aspect identification, to find and extract important topics in the text that will then be used to summarize. In [21], Hu and Liu present a technique based in NLP and statistics. In their proposal, part-of-speech (POS) tagging and syntax tree parsing (or chunking) are used to find nouns and noun phrases or NPs. Then, using frequent itemset mining, the most frequent nouns and NPs are extracted. The extracted sets of nouns and NPs are then filtered using special linguistic rules. These rules ensure that the terms inside those aspects that are composed of more than one word are likely to represent real objects together and also eliminate redundant aspects.
2. Sentiment Prediction, to determine the sentiment orientation on each *aspect*. Ding, Liu and Yu offer a lexicon and rule-based approach in [22]. This method relies on a sentiment word dictionary that contains a list of positive and negative words (called *opinion words*) that are used to match terms in the opinionated text. Also, since other special words might also change the orientation, special linguistic rules are proposed. These rules consider cases like negations words *no* or *not* and also some common negation patterns. However, despite how simple these rules might appear, it is important to handle them with care, because not all

occurrences of such rules or word apparitions will always have the same meaning. In this context, rules developed by Ding, Liu and Yu include an aggregation score function to determine the orientation of an aspect in a sentence combining multiple *opinion words*. This function will be explained in detail in the next section, since it will be used and extended by us.

3. Summary Generation, to present processed results in a simple manner. In this context, defined opinion quintuples are a good source of information for generating quantitative summaries. In particular, Liu defines a kind of summary called aspect-based opinion summary [23] [24], that consists of bar charts that show the number of positive and negative opinions about every *aspect* of one *entity*. In [25], Liu also proposes that the bar charts could be used to compare a set of selected products, showing the set of all aspects of the chosen products in the chart. In this case, each bar above or below the x-axis can be displayed in two scales: (1) the actual number of positive or negative opinions normalized with the maximal number of opinions on any feature of any product and (2) the percent of positive or negative opinions, showing the comparison in terms of percentages of positive and negative reviews.

3. Proposed Extension

Our extension takes Liu's methods as a basis and considers the same set of structured steps mentioned in the previous section. Here, we discuss issues on each one of the tree steps and propose our own approach in the context of tourism product reviews.

3.1. Aspect expression extraction

As defined by Liu, *aspects* do not directly appear in a text but they exist as *aspect expressions*. Accordingly, when trying to apply Liu's opinion model to extract opinions from real data, concepts can be somewhat confusing or unclear. It is also unclear how *aspects* that appear more than once in a document are managed. Having noticed these issues, a model to build opinion tuples from an opinionated document has been developed here.

To make things simpler, consider a set of opinionated documents $D_i = \{d_{i1}, d_{i2}, \dots, d_{im}\}$ about only one *entity*, e_i . This seems a realistic assumption since opinions are usually available in the form of product reviews on the Web. Then, each opinionated document will correspond to a review or opinion given by holder h_k in time t_k . Let S_{ik} be the set of all sentences in d_{ik} , with $S_{ik} = \{s_{ij1}, s_{ij2}, \dots, s_{ijm}\}$. Opinions on e_i in d_{ik} will be expressed on the *entity* itself and on a subset A_{ik} of its *aspects*. Similarly, each *aspect* of A_{ik} will appear on d_{ik} as a set of *aspect expressions* AE_{ijk} , a subset of AE_{ij} . The entity e_i will appear as a subset of different *entity expressions* $EE_{ik} \subseteq EE_i$. Thus, the set EX_{D_i} is defined as the set of all *aspect expressions* of all *aspects* and all *entity expressions* appearing in D_i . A sentence is related to one *aspect expression* or *entity expression* only if it appears in that sentence. Next, sentiment orientation needs to be determined for each pair (ex, s) only if any *aspect expression* or *entity expression* appears on it. After determining sentiment orientation, h_k and t_k of the corresponding document d_{ik} should simply be added in order to build each opinion tuple.

On the other hand, Liu's proposal indicates that it seems reasonable that frequently used nouns in product reviews are usually genuine and important *aspects expressions* because when people comment on different *aspects* of a product, the vocabulary that they use usually converges. Nevertheless, two main reasons explain the fact that many different expressions could indicate the same concept, particularly in the tourism domain:

- The economy principle in languages [26] indicates that they try to say a lot using few words. For example, the sentence "The hotel has good wifi." corresponds to a lexicalization, where the original expression, "The hotel has good internet access through wifi.", is shortened according to the economy principle.
- Each language presents systems that organize its concepts, also pursuing simplification. For that reason, many words in English (as in all other languages) simply are hyponyms of a determined hypernym. A hyponym is a word or phrase whose semantic field is included within that of another word, its hypernym. For instances, scarlet, vermilion, carmine, and crimson are all hyponyms of red (their hypernym), which is, in turn, a hyponym of color [27].

In practice, finding the *aspects* that are evaluated in a set of opinionated documents is a really complex task. In fact, detecting *aspect expressions* from a set of documents with opinions should be a completely different task than defining or finding the real *aspects* in them, because the amount of possible expressions appearing in a text is really huge. We have already shown that in the tourism product reviews, several expressions are in fact used. A different issue found in Liu's proposals is related to the concepts of sentence and *word distance*, that although widely used, are not clearly defined. Despite deeper linguistic analysis, here we will define a sentence as an ordered set of tokens, including words and punctuation. One token that appears in two different positions must be considered twice, as the positions where they appear are distinct. In other words, a sentence S will correspond to a set of unique tuples (*token, position*). Positions can only be in $\mathbb{N} \cup \{0\}$ and the difference between two adjacent components must be 1. As such, the concept of *word distance* between two elements of sentence S will correspond to the difference of the positions of the two tokens in S .

$$\text{Word Distance}(t_a, t_b) = |\text{position}(t_a) - \text{position}(t_b)| \text{ with } t_a, t_b \in S \quad (1)$$

As $\text{Word Distance}(t_a, t_b)$ is simply the absolute value of the difference between numbers in $\mathbb{N} \cup \{0\}$, $\text{Word Distance}(t_a, t_b)$ is a metric on the set S as it satisfies the conditions of non-negativity, identity of indiscernibles, symmetry and triangle inequality. Note that the minimal distance between 2 elements in S is 1, and it occurs between adjacent elements. The maximum distance corresponds to $|S| + 1$. Considering these definitions, we apply the technique developed by Hu and Liu in [21].

3.2. Determination of the Opinion Orientation

Taking Liu's work in [22] as inspiration, a set of rules to determine the sentence orientation was developed, always considering *opinion words* as a basis.

3.2.1. Word Orientation Rules

- **Word Rules:** Positive *opinion words* will intrinsically have a *score* of 1, denoting a normalized positive orientation, while negative ones will have associated a *score* of -1 . Every noun and adjective in each sentence that is not an *opinion word* will have an intrinsic *score* of 0 and will be called a *neutral word*.
- **Negation Rules:** A negation word or phrase usually reverses the opinion expressed in a sentence. Consequently, *opinion words* or *neutral words* that are affected by negations need to be specially treated. Three rules must be applied: Negation Negative \rightarrow Positive, Negation Positive \rightarrow Negative and Negation Neutral \rightarrow Negative. Negation words and phrases include: *no, not, never, n't, dont, cant, didnt, wouldnt, havent, shouldnt* (misspellings are here intentional). Also, some negation patterns are considered, including *stop + vb-ing, quit + vb-ing* and *cease + to + vb*.
- **Too Rules:** Sentences where the words *too, excessively* or *overly* appear, are also handled specially. When an *opinion word* or a *neutral word* appears near one of the mentioned terms, denoted *too words*, its orientation will always be Negative (*score* = -1).

3.2.2. Aspect Orientation Rules

Having mentioned rules that help in determining each word orientation in a sentence, it is now explained how all these orientations should be combined to determine the final orientation of a sentence on a particular *aspect*. This algorithm should only consider words marked as *opinion words* or *neutral words* as they are the only ones that will provide the orientation for each sentence.

- **Aspect Words Aggregation Rule:** Let s be a sentence that contains the set of *aspect expressions* $A = \{a_1, \dots, a_m\}$, each one of them appearing only one time in s . Also, let AW_i be the set of words that comprise aspect a_i , where $AW_i = \{aw_{i1}, aw_{i2}, \dots, aw_{in}\}$. Each aw_{ij} will be called an *aspect word* and it will correspond to an *aspect expression* a_i . If *scores* for each *opinion word* and *neutral word* in s are known, a *score* for each aw_{ij} in s is given by the following aggregation function:

$$\text{score}(aw_{ij}, s) = \sum_{ow_j \in s} \frac{\text{score}(ow_j)}{\text{Word Distance}(ow_j, aw_{ij})} \quad (2)$$

Where ow_j is an *opinion word* or *neutral word* in s , $Word\ Distance(ow_j, aw_{ij})$ is the word distance between the aspect word aw_{ij} and the opinion word ow_j in s . We take this function from Ding, Liu and Yu's work. However, their proposition lacked an explanation of how the function should be applied to *aspect expressions* that are composed of more than one word (which we call *compound*). We have seen that in tourism product reviews some *aspect expressions* are in fact *compound*. For instance, in the sentence "The hotel had a poor view of the beautiful lake." an *aspect expression* that should be extracted by Liu's algorithms is *lake view*. However, Liu's proposal does not explain how the orientation on this *aspect* should be obtained in the sentence. In order to consider these cases, we propose that the formula should not be used for each *aspect expression* but rather for each word in each expression. Orientations are aggregated according to the next rule.

- **Aspect Aggregation Rule:** For each *compound aspect expression* a_i in s , its orientation will be calculated considering the *scores* of all the words that compose it, $aw_{ij} \in AW_i$, according to the following equation.

$$score(a_i, s) = \sum_{aw_{ij} \in AW_i} score(aw_{ij}, s) \quad (3)$$

- **Position Aggregation Rule:** We have also seen that in tourism product reviews *aspect expressions* could appear more than once in a sentence. This case is not covered by Liu's proposals, but here we need a method to cover these cases. Supposing that a_i appears t times in s and knowing the score of each *aspect expression* appearance $a_i^k, k \in \{1, 2, \dots, t\}$, we propose that the final score of a_i , or $fscore(a_i, s)$, should be calculated by simply adding the values of the scores of all the a_i appearances in s , according to the following equation.

$$fscore(a_i, s) = \sum_{k=1}^t score(a_i^k, s) \quad (4)$$

Note that when a_i only appears one time in s , $fscore(a_i, s) = score(a_i, s)$. Then, for each *aspect expression*, if $fscore(a_i, s)$ is positive, the opinion is considered positive on a_i and if it is negative, the opinion is considered negative on a_i . If none of these cases occur, the sentence is considered neutral.

- **But Clauses Rules:** We use exactly the same rule that Liu proposes in [22]. This rule states that when a *but word* b appears in sentence s , s must be broken into two segments, the one before and the one after b . If the orientation of any *aspect word* aw_{ij} appearing in the sentence segment after b is zero, its orientation should then be determined using the segment before b , effectively negating it. We realized that a little ambiguity existed since in some of these cases aw_{ij} could appear outside the segment that is considered to determine the orientation of s . Here, we simply propose that aw_{ij} must be added at the final position of the corresponding segment in order to avoid the consistency issue.

3.3. Summarization

Liu's proposal seems fairly simple and effective for summarizing opinions. However, it lacks a robust way of measuring the importance of each evaluated *aspect*. In [23], *aspects* are ranked according to the frequency of their appearances in the reviews, but it is also declared that other types of rankings are also possible, like ranking *aspects* according to the number of reviews that express positive or negative opinions. Here, we attempt to measure the importance of each *aspect* simultaneously using the amount of positive and negative opinions of it. We also use that measure to rank *aspects*. The underlying assumption is that an *aspect* that has a lot of positive and negative opinions will be more important, since the high number of opinions of both orientations might indicate that customers are very interested in that *aspect*. In this way, the total number of times that an *aspect* appears is not only considered in measuring importance, but also the dispersion in the number of positive and negative opinions. Let P_i and N_i be the number of positive and negative opinions on aspect $a_i, i \in \{1, \dots, n\}$. Then, $PScore_i$ and $NScore_i$ will be the min-max normalized values of P_i and N_i , respectively. With this, we calculate the standard deviation of these *Scores* using:

$$STDS\ core_i = \sqrt{\frac{1}{2} \left(\left(PScore_i - \frac{PScore_i + NScore_i}{2} \right)^2 + \left(NScore_i - \frac{PScore_i + NScore_i}{2} \right)^2 \right)} \quad (5)$$

We define our new measure for each *aspect* a_i , called *Relative Importance*, as the min-max normalized value of its *STDScore* $_i$. We propose that aspect-based summaries should include bar charts and a table that shows the actual values of *PScore* $_i$, *NScore* $_i$ and *Relative Importance* for each *aspect expression*.

4. Experiments and analysis

The experiment we carried out consisted in evaluating how our extension performs when applied to tourism product reviews from Los Lagos, particularly, hotels and restaurants. Our work here mainly consisted in: (1) Generating annotated corpora or datasets to evaluate the performance of the algorithms for the selected products, using the site TripAdvisor as a source, (2) Measuring the performance of the algorithms and (3) designing and developing an application to extract opinions from these reviews and generate our proposed summarization charts. This application was implemented using Python, the Natural Language Toolkit (NLTK) for NLP tasks and the Django Framework. It also included modules that helped carrying out tasks (1) and (2).

4.1. Annotated Corpora

We built a web crawler and downloaded reviews originally written in English for hotels and restaurants in the *Lake District* from TripAdvisor. The downloaded reviews of each product were saved in a CSV file, which we used to randomly select, in each case, 100 reviews that were used to build the annotated corpora. The annotation process followed the spirit of what Liu proposes in [23] and [22]. Each review was tokenized into sentences using [28] and was manually annotated following a set of rules and a notation system that had previously been designed by us (for details see our corpora material¹.) However, sentences that seemed ambiguous or really difficult to tag were discussed with a second human annotator, an expert in linguistics. Once an agreement was achieved, the sentence was tagged according to that agreement. This marks an important difference between this study and other tagging procedures commonly carried out in literature, where different annotators tag the same corpus separately and only once the annotation procedure has finished are different results of the same corpus compared to define the final choice. This different approach was used here due to time constraints, since it seemed more efficient and was worth trying as a contribution to research in this field.

Table 1 gives a general description of the generated corpora. In both cases, almost 80% of the sentences contained opinions. This shows that opinionated sentences represent an important fraction of the total sentences, which somewhat validates the use of TripAdvisor as a source of opinions for hotels and restaurants. Nevertheless, as expected, non-opinionated sentences are also a considerable number, consequently introducing noise into the opinion-extraction process.

Feature	Hotels Corpus	Restaurants Corpus
Number of Reviews	100	100
Total Number of Sentences	789	470
Number of Opinion Sentences	609	368
Opinion Sentences/Sentences	77.19%	78.3%

Table 1. Corpora Details.

Table 2 gives details about the *aspect expressions* that were manually extracted. Following Liu's notation, we call those expressions explicit *aspect expressions* that appear as nouns or NPs in a sentence and implicit *aspect expressions* in all other cases. Results show that in both corpora explicit *aspect expressions* are the most common ones, representing around 70% of all the extracted expressions. When some *aspect expressions* appear in both an explicit and implicit manner, they were considered as explicit. On the other hand, extracted *aspect expressions* that are purely implicit are also an important number, being almost 20% in both cases. A simple review showed that most of these *aspects* were indicated by adjectives.

¹http://wi.dii.uchile.cl/publications/corpora_material.rar

Aspect Expressions Type	Hotels Corpus		Restaurants Corpus	
	Number	Percentage	Number	Percentage
Explicit	229	73.87%	161	67.93%
Explicit and Implicit	30	9.68%	26	10.97%
Implicit	51	16.45%	50	21.1%
Total	310	100%	237	100%

Table 2. Detail on *aspect expressions* found in corpora.

4.2. Algorithm performance

To measure the performance of the algorithms, three tasks were evaluated by comparing its results with the manually annotated corpora: (1) Explicit Aspect Extraction, to measure the effectiveness of explicit *aspect expression* extraction, (2) Subjectivity Classification, to evaluate the effectiveness of opinion sentence extraction and (3) Sentiment Classification, to measure the accuracy of orientation prediction of each pair (*ex, s*) (*aspect expression, sentence*) for the positive class. Here, we present the best general performance obtained by doing a sensitivity analysis regarding the most sensitive parameter - the minimum support rule to extract *aspect expressions* as defined in [21]. Precision, recall and f-measure were calculated for six different values of this parameter for each task. Then, the best model was chosen using f-measure. Table 3 shows the obtained values.

Corpus Index Name	Hotels Corpus		Restaurants Corpus		Average Performance		
	Precision	Recall	Precision	Recall	Precision	Recall	F-measure
Explicit Aspect Extraction	33%	29%	42%	37%	38%	33%	36%
Subjectivity Classification	79%	93%	81%	88%	80%	91%	85%
Sentiment Classification	89%	93%	91%	93%	90%	93%	92%

Table 3. Performance results.

These results show that performance on the *aspect extraction* task is fairly poor in the tourism domain. The algorithm is only capable of extracting almost 30% of the total explicit expressions for hotels and almost 40% for restaurants. Moreover, a high percentage of the extracted expressions do not correspond to real *aspect expressions* for both cases. On the other hand, sentiment classification shows fairly good results, but in this case most of the possible conclusions are difficult to prove because this task was only evaluated for those *aspect expressions* that were extracted. Since these expressions are somewhat the simplest ones, determining the sentiment orientation on them may be easier. Consequently, precision and recall could decrease when all *aspect expressions* are considered.

Results also support the properties of tourism product reviews presented in section 1. These *stories* in which reviewers mention objects that do not correspond to attributes or components of the product may explain the low precision obtained for the explicit *aspect extraction* task in both cases. For instance, in the case of hotels, users commonly refer to objects like time, day and city, which, although relevant for stories, tell nothing about the hotel. Also, nouns and NP sets that do not occur with relative high frequencies will probably need some special treatment in order to be extracted, keeping in mind that many expressions can be used to refer to the same *aspect*. In [21], Liu proposed a method to extract these infrequent *aspect expressions* by exploiting their relationships with frequent *opinion words*. Here, this method was not considered since in Liu's case, the extracted infrequent *aspect expressions* only represented an improvement of 15% for recall, at the cost of decreasing precision by almost 7%. However, given the poor results that have been obtained, it seems interesting to evaluate how this step would improve or worsen performance in this case. On the other hand, as Liu states in [21] and as seen in previous sections, the reason that probably explains precision being a little lower than recall in the task of subjectivity classification is the fact that there are many non-opinionated sentences in tourism product reviews. Since the algorithm labels some of these sentences as opinion sentences because they contain both product *aspect expressions* and some *opinion words*, precision decreases. Nevertheless, although these sentences may not show strong user opinions toward the product features, they may still be beneficial and useful [21].

4.3. Summary Visualization

The application we built permits users to see aspect-based summaries as proposed in section 3. Besides bar charts for each *entity* in the system, a table shows the actual values of the *Positive Score*, *Negative Score* and *Relative Importance* for each *aspect expression*; figure 1 shows an example. By clicking the name of each column, the table and the bar chart are sorted according to the clicked column (each click alternates between an ascending or descending sort.) By clicking one *aspect expression*, the user is redirected to a page showing specific information about it.

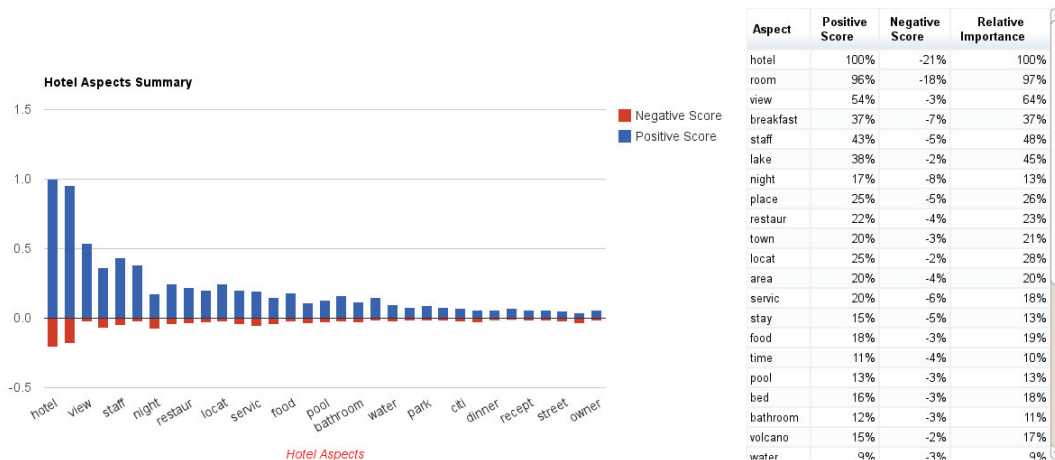


Figure 1. Proposed aspect-based summary for the entity *hotel* in the Lake District, using opinions extracted from hotel reviews on the Lake District (Chile) in TripAdvisor. In the chart, *aspect expression* bars are ordered according to *Relative Importance* in a descending manner.

5. Conclusion

In the first place, the proposed models to define and extract opinions from web documents present a simple, yet relatively effective manner of transforming the unstructured data about opinions available on the web. However, the algorithm for *aspect expressions* extraction, based on frequent nouns and NPs appearing in reviews, achieved a poor performance in the tourism domain. Results show that, in fact, multiple expressions are used to denote the same attribute or component of a tourism product in reviews. Therefore, not only the most frequent words need to be considered when extracting *aspect expressions* in order to achieve a better recall for this task. The fact that users tend to *tell stories* when writing reviews about tourism products led to poor precision in the task of extracting *aspect expressions* since in reviews a lot of objects that are not components or attributes of the product are mentioned. All the extracted expressions that are not components or attributes of a product need to be filtered. In this context, the use of ontologies, as in [29], [17] and [30], or other methods of studying relations between words, such as the one proposed in [12] or in [31]), could be very useful. Conversely, the application of NLP rules for determining semantic orientation proved to be very effective for extracted *aspect expressions*, achieving an average precision and recall of 90%. Nevertheless, since *aspect expressions* that were extracted only represent a small percentage of the ones that were manually detected, the method needs to be tested for all possible expressions on the topic of tourism in order to give a more conclusive analysis. This is proposed for future work.

On the other hand, one important downside of the proposed rules seems to be the fact that they are not domain sensitive. Specific sentences regarding context or domain dependent topics need to be specially treated. In the tourism domain, this could represent a major problem since a lot of opinions could imply a positive or negative sentiment depending on the product the opinion is given on. A method of dealing with these issues, although proposed in [22], was here left for future work. Using different state-of-the-art-methods to determine the sentiment orientation could also solve this problem. On the other hand, considering that in tourism product reviews a significant number of sentences do not contain opinions - which led to poor precision in the task of subjectivity classification - applying methods to filter these sentences seems crucial. Also, we realized that the annotation task could easily become very complex. Nevertheless, through the participation of a linguistics expert in the process,

it was possible to more accurately understand how opinions are given by users and how opinion linguistic corpora should be elaborated. Documenting any corpora with all the assumptions, rules, techniques or methodologies that were used when generating the input texts or annotating is a key factor to a better understanding for those who may use those corpora. This was a main downside found in Liu's case, considering that in the opinions domain any annotation process will always be a somewhat subjective task.

Acknowledgements

This work was supported partially by the FONDEF project D10I-1198, entitled WHALE: *Web Hypermedia Analysis Latent Environment* and the Millennium Institute on Complex Engineering Systems (ICM: P-05-004-F, CONICYT: FBO16).

References

- [1] D. Park, S. Kim, The effects of consumer knowledge on message processing of electronic word-of-mouth via online consumer reviews, *Electronic Commerce Research and Applications* 7 (4) (2009) 399–410.
- [2] H. Shin, D. Hanssens, K. Kim, B. Gajula, Impact of positive vs. negative e-sentiment on daily market value of high-tech products.
- [3] F. Zhu, X. Zhang, Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics, *Journal of Marketing* 74 (2) (2010) 133–148.
- [4] K. Lancaster, A new approach to consumer theory, *The journal of political economy* 74 (2) (1966) 132–157.
- [5] S. L. Smith, The tourism product, *Annals of tourism research* 21 (3) (1994) 582–595.
- [6] P. Kotler, et al., *Marketing management*, Stuttgart, Verlag Schäffer-Poeschel.
- [7] A. Gustafsson, M. Johnson, *Competing in a service economy*, John Wiley & Sons, 2006.
- [8] L. Zhuang, F. Jing, X.-Y. Zhu, L. Zhang, Movie review mining and summarization, in: *Conference on Information and Knowledge Management: Proceedings of the 15 th ACM international conference on Information and knowledge management*, Vol. 6, 2006, pp. 43–50.
- [9] S. Lichtenstein, P. Slovic, *The construction of preference*, Cambridge University Press, 2006.
- [10] K. Scherer, What are emotions? and how can they be measured?, *Social science information* 44 (4) (2005) 695–729.
- [11] Y. Lu, C. Zhai, N. Sundaresan, Rated aspect summarization of short comments, in: *Proceedings of the 18th international conference on World wide web*, ACM, 2009, pp. 131–140.
- [12] A. Popescu, O. Etzioni, Extracting product features and opinions from reviews, in: *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, 2005, pp. 339–346.
- [13] N. Archak, A. Ghose, P. Ipeirotis, Show me the money!: deriving the pricing power of product features by mining consumer reviews, in: *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, 2007, pp. 56–65.
- [14] R. Decker, M. Trusov, Estimating aggregate consumer preferences from online product reviews, *International Journal of Research in Marketing* 27 (4) (2010) 293–307.
- [15] L. Ku, Y. Liang, H. Chen, Opinion extraction, summarization and tracking in news and blog corpora, in: *Proceedings of AAAI-2006 Spring Symposium on Computational Approaches to Analyzing Weblogs*, no. 2001, 2006.
- [16] I. Titov, R. McDonald, A joint model of text and aspect ratings for sentiment summarization, *Urbana* 51 (2008) 61801.
- [17] L. Zhao, C. Li, Ontology based opinion mining for movie reviews, *Knowledge Science, Engineering and Management* (2009) 204–214.
- [18] H. Kim, K. Ganesan, P. Sondhi, C. Zhai, Comprehensive review of opinion summarization.
- [19] B. Liu, *Web data mining: exploring hyperlinks, contents, and usage data*, Springer Verlag, 2007.
- [20] B. Pang, L. Lee, Opinion mining and sentiment analysis, *Foundations and Trends in Information Retrieval* 2 (1-2) (2008) 1–135.
- [21] M. Hu, B. Liu, Mining opinion features in customer reviews, in: *Proceedings of the National Conference on Artificial Intelligence*, Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2004, pp. 755–760.
- [22] X. Ding, B. Liu, P. Yu, A holistic lexicon-based approach to opinion mining, in: *Proceedings of the international conference on Web search and web data mining*, ACM, 2008, pp. 231–240.
- [23] M. Hu, B. Liu, Mining and summarizing customer reviews, in: *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, 2004, pp. 168–177.
- [24] M. Hu, B. Liu, Opinion extraction and summarization on the web, in: *Proceedings Of The National Conference On Artificial Intelligence*, Vol. 21, Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2006, p. 1621.
- [25] B. Liu, M. Hu, J. Cheng, Opinion observer: analyzing and comparing opinions on the web, in: *Proceedings of the 14th international conference on World Wide Web*, ACM, 2005, pp. 342–351.
- [26] A. Vicentini, The economy principle in language, *Notes and Observations from Early Modern English Grammars*. *Mots, Palabras, Words* 3 (2003) 37–57.
- [27] V. Fromkin, R. Rodman, N. Hyams, *An introduction to language*, Wadsworth Publishing Company, 2010.
- [28] T. Kiss, J. Strunk, Unsupervised multilingual sentence boundary detection, *Computational Linguistics* 32 (4) (2006) 485–525.
- [29] A. Cadilhac, F. Benamara, N. Aussenac-Gilles, Ontolexical resources for feature based opinion mining: a case-study, in: *23rd International Conference on Computational Linguistics*, 2010, p. 77.
- [30] E. Vallés Balaguer, P. Rosso, A. Locoro, V. Mascardi, Análisis de opiniones con ontologías, *Polibits* (41) (2010) 29–36.
- [31] D. Bollegala, Y. Matsuo, M. Ishizuka, An integrated approach to measuring semantic similarity between words using information available on the web, in: *HLT-NAACL*, 2007, pp. 340–347.