SENTIMENT ANALYSIS IN TRIPADVISOR

A PREPRINT

Ana Valdivia*
Department of Computer Science
and Artificial Intelligence
Universidad de Granada
Granada, Spain 18014
avaldivia@ugr.es

M. Victora Luzón
Department of Software Engineering
Universidad de Granada
Granada, Spain 18014
luzon@ugr.es

Francisco Herrera
Department of Computer Science
and Artificial Intelligence
Universidad de Granada
Granada, Spain 18014
herrera@decsai.ugr.es

***K*eywords** Sentiment Analysis · TripAdvisor

1 Introduction

The number of Web 2.0 has recently experienced important growth. These websites emerged as an evolution of Web 1.0 or static websites. In Web 2.0 the user is not only a content consumer, but also is able to generate the content and collaborate with other users. In this way, the users take an active role and create a virtual community. There are different types of Web 2.0: blogs (Blogger or Wordpress), media content (Prezi, Youtube, Flickr), wikis (Wikipedia, WikiSpace), collaborative (Dropbox, GoogleDocs) and social networks (Twitter, Facebook, Google+). The burgeoning information explosion offered through the Web 2.0 has implied that end customers often check other user's opinions in forums, blogs and social networks before buying a product or contracting a service.

TripAdvisor emerged in 2004 as a Web 2.0 for the tourism domain. This user-generated content website offers a large amount of reviews from travelers' experiences regarding hotels, restaurants and tourist spots. TripAdvisor has since been ranked as the most popular site for trip planning. Nowadays, million of tourists arrange their holidays taking into account TripAdvisor reviews (see Figure1).

Sentiment Analysis (SA) is extremely useful when monitoring Web 2.0, allowing us to know public opinion of a large number of issues without the need for satisfaction inquiries [1, 2]. According to the Oxford dictionary, sentiment analysis is the process of computationally identifying and categorizing opinions expressed in a piece of text to determine whether the writer's attitude toward a particular topic, product, and so on is generally positive, negative, or neutral. The interest in sentiment analysis has increased significantly over the last few years due to the large amount of stored text in Web 2.0 applications and the importance of online customer opinions. As a result, more than 1 million research papers contain the term "sentiment analysis," and various start-ups have been created to analyze sentiments in social media companies.

Multiple studies on TripAdvisor exist, but there is no complete analysis from the sentiment analysis viewpoint. This article proposes TripAdvisor as a source of data for sentiment analysis tasks. We develop an analysis for studying the matching between users' sentiments and automatic sentiment-detection algorithms. Finally, we discuss some of the challenges regarding sentiment analysis and TripAdvisor, and conclude with some final remarks.

*Corresponding author

Figure 1: Granada (Paseo de los Tristes). TripAdvisor is the most popular site for planning a trip.

2 TripAdvisor and Sentiment Analysis

According to Wikipedia, TripAdvisor is an American travel website company providing reviews from travelers about their experiences in hotels, restaurants, and monuments. Stephen Kaufer and Langley Steinert, along with others, founded TripAdvisor in February 2000 as a site listing information from guidebooks, newspapers, and magazines. InterActiveCorp purchased the site in 2004, and one year later, spun off its business travel group, Expedia. After that, the website turned to user-generated content. It has since become the largest travel community, reaching 390 million unique visitors each month and listing 465 million reviews and opinions about more than 7 million accommodations, restaurants, and attractions in 49 markets worldwide. Figure 2 shows the Google search rate for TripAdvisor, illustrating its popularity around the world.

Because it has so much data, TripAdvisor has become extremely popular with both tourists and managers. Tourists can read the accumulated opinions of millions of everyday tourists. They can also check the popularity index, which is computed using an algorithm that accounts for user reviews and other published sources such as guidebooks and newspaper articles. This index runs from number 1 to the overall total number of restaurants, hotels, or other attractions within the city. Travelers can find the most interesting visitor attraction or most popular restaurant. Linked to this is the bubble rating (user rating), a 1–5 scale where one bubble represents a terrible experience and five bubbles an excellent experience. All reviewers are asked to use this scale to summarize their feedback. Together with this rating, users include their opinions, which can cover the performance of a restaurant, hotel, or tourist spot. Therefore, reading and analyzing reviews can help develop a business.

The World Travel & Tourism Council report shows that tourism generates 9.8 percent of the wider gross domestic product and supports 248 million jobs [2]. These numbers suggest that the tourism industry is the most important economic driver of many economies. Therefore, it's important to understand the main drivers of the tourist flow as well as tourists'opinions about a city's restaurants, hotels, and tourist attractions.

TripAdvisor has enough standing to be used as a text source [3] storing numerous reviews of tourist businesses around the world. Sentiment analysis extracts insights from this data. Sentiment classification, the best-known sentiment analysis task, aims to detect sentiments within a document, a sentence, or an aspect. This task can be divided into three steps: polarity detection (label the sentiment of the text as positive, negative, or neutral), aspect selection/ extraction (obtain the features for structuring the text), and classification (apply machine learning or lexicon approaches to classify the text). Sentiment analysis methods (SAMs), which are trained for sentiment polarity detection [4, 5, 6], can automatically detect sentiments from documents, sentences, or words. A large variety of SAMs address the different categories of texts (blogs, reviews, tweets, and so on). However, the analysis of feelings is not a perfect science,

[2] `www.wttc.org/-/media/files/reports/economic%20impact%20research/regions%202016/world2016.pdf.`

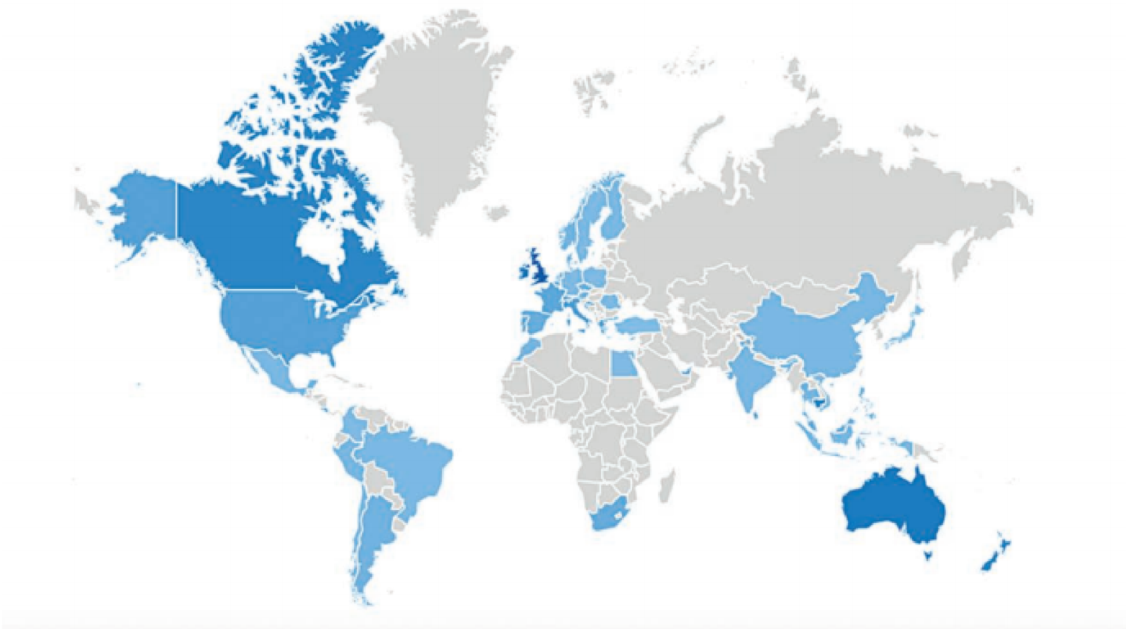

Figure 2: Map of the popularity of TripAdvisor searches in Google over the last five years.

especially when applied to the unstructured texts that predominate in social networks [7]. Human language is complex, so teaching a machine to detect different grammatical nuances, cultural variations, jargon, and misspellings in messages on the network is a difficult process, and it is even more difficult to automatically understand how the context can affect the message's tone. Because humans can apply a contextual understanding, we can intuitively interpret the intentionality of any writing. Computers, however, have difficulty understanding the context in which a phrase is expressed and detecting whether a person is being sarcastic or not.

A few sentiment analysis studies set TripAdvisor as a data source. For example, researchers have collected reviews about the TripAdvisor app in Google Play Store for extracting app features to help developers [8]. Others analyzed TripAdvisor's hotel reviews for classifying good and bad customer opinions [9]. Still others propose a system to summarize comments from travel social networks, such as TripAdvisor for analysis [10]. Similarly, other researchers developed a tool to analyze tourists' opinions of restaurants as well as hotels from a region in Chile [11].

3 TripAdvisor: A study for calibrating user's polarity

We scrape TripAdvisor webpages on three well-known monuments in Spain: Alhambra, Mezquita Córdoba, and Sagrada Familia. We consider user ratings of one and two bubbles as negative, three as neutral, and four and five as a positive sentiment. We then apply four SAMs (SentiStrength [12], Bing [13], Syuzhet [14], and CoreNLP [15]) and extract the overall polarity on each opinion.

Figure 3 shows the results. We observe that the distributions of polarities are different from user ratings. The user ratings bar plot shows that users tend to rate their visits to the three monuments positively, with more than 90 percent of ratings having four or five bubbles. SentiStrength and Syuzhet methods reach a similar distribution to the user rating. However, Bing and CoreNLP detect more negativity in the TripAdvisor opinions. In all cases, the number of neutral polarities is higher than the neutral ratings (three bubbles).

Next, we studied the distribution of bubble ratings over the negative SAM polarities. We thus analyzed the behavior of user feedback against the SAM evaluations. Figure 4 presents 12 bar plots (four SAMs for each of three monuments) containing the shares of all negative SAM polarity over the bubble evaluation (percent over the original user ratings). Analyzing this data, we observe that SentiStrength and Syuzhet detect at best 57.48 and 45.10 percent of negative reviews. However, they misclassify on average 20 percent of positive user reviews (three and four bubbles). On the other hand, Bing and CoreNLP methods detect as negative more negative user ratings, but misclassify over 30 percent of positive reviews. Bing and CoreNLP tend to highlight the negative opinions.

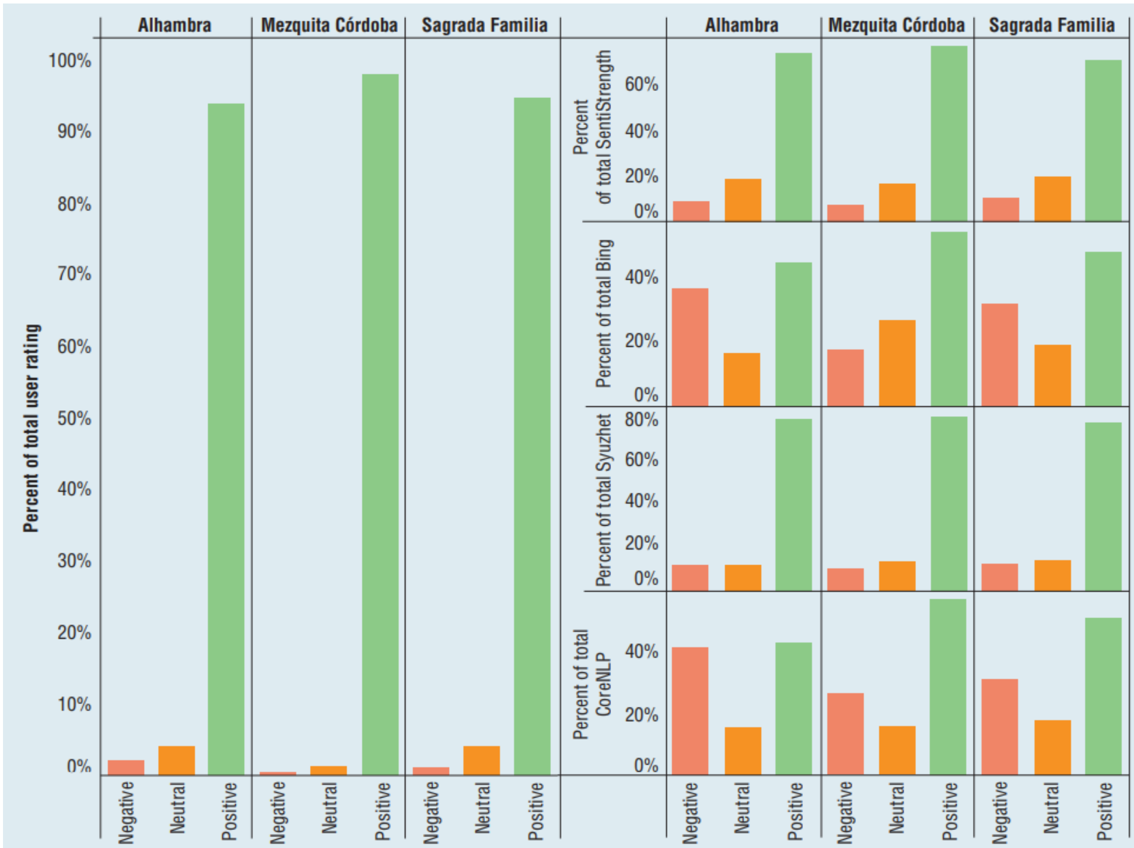

Figure 3: Distribution of sentiments between TripAdvisor users (bubble ratings) and four sentiment analysis methods (SAMs): SentiStrength, Bing, Syuzhet, and CoreNLP. Red indicates negative sentiments, orange neutral, and green positive.

In general, we observe that users tend to write negative sentences on positive user ratings, and vice versa. Therefore, we suggest not setting the user rate as a label sentiment for the whole review and analyzing the opinions in depth.

This study clearly shows the need to analyze opinions beyond user ratings. As a practical methodology, we propose following three steps: handle the negative opinions identified with SAMs via learning models, get a good clusterization according to consensus degrees among SAMs, and discover relationships among common aspects to characterize the cause behind the negative comments.

4 Challenges

Several challenges arise when TripAdvisor uses sentiment analysis, due to the specific content in TripAdvisor-based opinions. Although some related topics have been extensively studied in the literature, their adaptation to the context of TripAdvisor opinions requires revisiting them.

Aspect-based sentiment analysis (ABSA) is an important sentiment analysis task [16]. An aspect refers to an attribute of the entity, for example, hotel room cleanliness, the staff at a tourist spot, or the service at a restaurant. ABSA aims to identify the sentiment toward an aspect and extract fine grained information about specific TripAdvisor-based opinions (hotels, monuments, restaurants, and so on). Recent relevant studies are based on deep learning [17], which should be analyzed in the TripAdvisor context.

ABSA is helpful to business managers because it allows for the extraction of transparent customer opinions. Discovery knowledge techniques such as subgroup discovery [18] can be applied to discover relationships among common aspects and get aspect associations for both positive and negative opinions.

The detection of irony and sarcasm is a complex sentiment analysis task. The detection of ironic expressions in TripAdvisor reviews is an open problem that could help to extract more valuable information about the study's

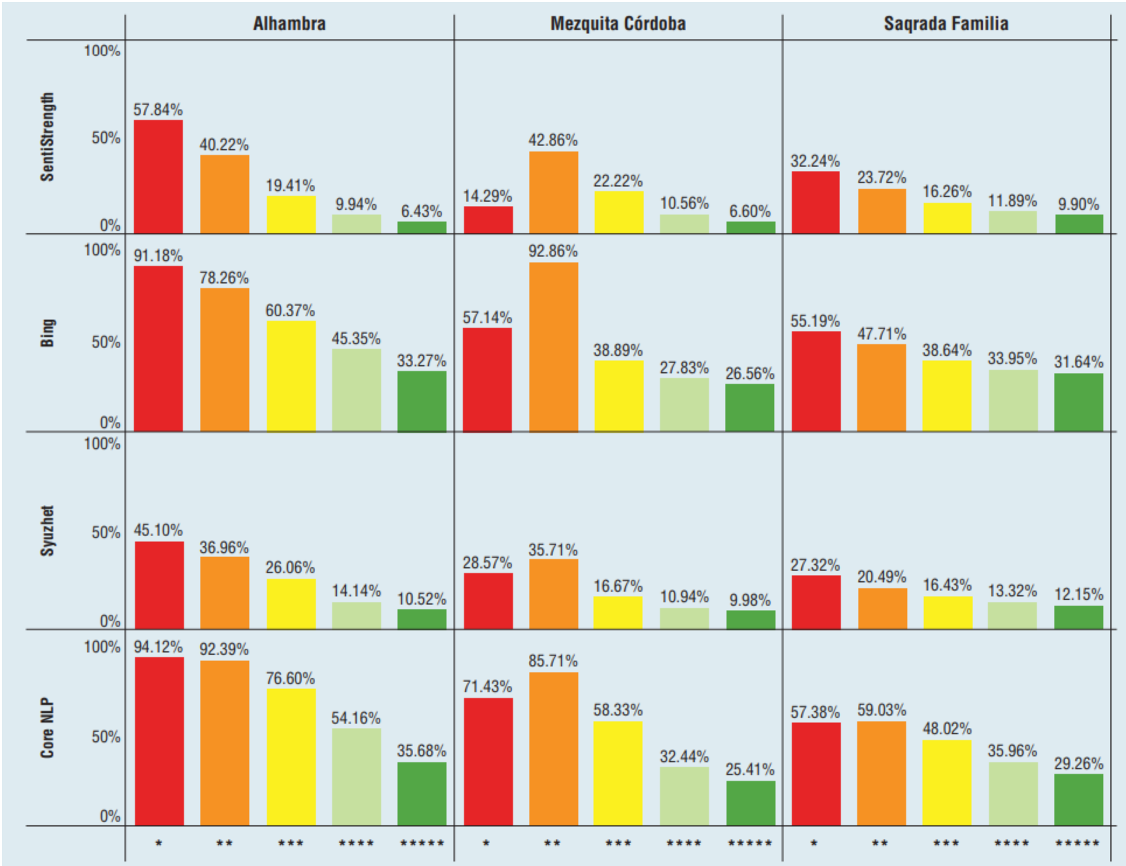

Figure 4: Distribution of SAMs' negative polarity by user rating in TripAdvisor.

subject [19]. Spam is another sentiment analysis-related concern. Some authors have developed studies to measure the credibility of TripAdvisor with satisfactory results [20].

A novel approach is the extraction of aspects/features from opinions to raise the issue as a bag of feature vectors, considering the problem as multi-instance learning [21]. This might provide a robust approach from the classification viewpoint.

Sentiment analysis is an incipient research field. It is difficult to determine how it will evolve in the future, although there is a general belief that this analysis needs to go beyond a simple classification of texts on a positive and negative one dimensional scale. Over the last few years, the list of sentiment analysis related challenges has grown (subjectivity classification, opinion summarization, opinion retrieval, and so on).

Through Web platforms such as TripAdvisor, tourists can openly describe their experiences and thus affect a business's viability. Therefore, the implementation of sentiment analysis techniques to mine sources of opinion is crucial to understanding the faults and assets of a tourist service. Given the large number of applications in the tourist domain, sentiment analysis has great potential to directly influence quality improvement in tourism.

Because of inconsistencies between user ratings and SAM evaluations, with users often writing negative sentences in positive opinions and vice versa, we need new approaches to fix the positive, negative, and neutrality via consensus among SAMs, as well as design models to discover relationships among common aspects to characterize the reasons behind negative comments.

Acknowledgments

This work has been supported by FEDER (Fondo Europeo de DEsarrollo Regional) and the Spanish National Research Project TIN2014-57251-P.

References

[1] Erik Cambria. Affective computing and sentiment analysis. *IEEE Intelligent Systems*, 31(2):102–107, 2016.

[2] Bing Liu. *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge University Press, 2015.

[3] Peter O'Connor. User-generated content and travel: A case study on tripadvisor. com. *Information and communication technologies in tourism 2008*, pages 47–58, 2008.

[4] Filipe N Ribeiro, Matheus Araújo, Pollyanna Gonçalves, Marcos André Gonçalves, and Fabrício Benevenuto. Sentibench-a benchmark comparison of state-of-the-practice sentiment analysis methods. *EPJ Data Science*, 5(1):1–29, 2016.

[5] Jesus Serrano-Guerrero, Jose A Olivas, Francisco P Romero, and Enrique Herrera-Viedma. Sentiment analysis: A review and comparative analysis of web services. *Information Sciences*, 311:18–38, 2015.

[6] Erik Cambria and Amir Hussain. *Sentic computing: a common-sense-based framework for concept-level sentiment analysis*, volume 1. Springer, 2015.

[7] Farhan Hassan Khan, Saba Bashir, and Usman Qamar. Tom: Twitter opinion mining framework using hybrid classification scheme. *Decision Support Systems*, 57:245–257, 2014.

[8] Emitza Guzman and Walid Maalej. How do users like this feature? a fine grained sentiment analysis of app reviews. In *Requirements Engineering Conference (RE), 2014 IEEE 22nd International*, pages 153–162. IEEE, 2014.

[9] Dietmar Gräbner, Markus Zanker, Gunther Fliedl, Matthias Fuchs, et al. *Classification of customer reviews based on sentiment analysis*. Citeseer, 2012.

[10] Srisupa Palakvangsa-Na-Ayudhya, Veerapat Sriarunrungreung, Pantipa Thongprasan, and Satit Porcharoen. Nebular: A sentiment classification system for the tourism business. In *Computer Science and Software Engineering (JCSSE), 2011 Eighth International Joint Conference on*, pages 293–298. IEEE, 2011.

[11] Edison Marrese-Taylor, Juan D Velásquez, and Felipe Bravo-Marquez. A novel deterministic approach for aspect-based opinion mining in tourism products reviews. *Expert Systems with Applications*, 41(17):7764–7775, 2014.

[12] Mike Thelwall. Heart and soul: Sentiment strength detection in the social web with sentistrength, 2013. *Cyberemotions: Collective emotions in cyberspace*, 2013.

[13] Minqing Hu and Bing Liu. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177. ACM, 2004.

[14] Matthew Jockers. Package 'syuzhet'. *URL: https://cran. r-project. org/web/packages/syuzhet*, 2017.

[15] Christopher Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, and David McClosky. The stanford corenlp natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, pages 55–60, 2014.

[16] Kim Schouten and Flavius Frasincar. Survey on aspect-level sentiment analysis. *IEEE Transactions on Knowledge & Data Engineering*, (1):1–1, 2016.

[17] Soujanya Poria, Erik Cambria, and Alexander Gelbukh. Aspect extraction for opinion mining with a deep convolutional neural network. *Knowledge-Based Systems*, 108:42–49, 2016.

[18] Martin Atzmueller. Subgroup discovery. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 5(1):35–49, 2015.

[19] Soujanya Poria, Erik Cambria, Devamanyu Hazarika, and Prateek Vij. A deeper look into sarcastic tweets using deep convolutional neural networks. *arXiv preprint arXiv:1610.08815*, 2016.

[20] Raffaele Filieri, Salma Alguezaui, and Fraser McLeay. Why do travelers trust tripadvisor? antecedents of trust towards consumer-generated media and its influence on recommendation adoption and word of mouth. *Tourism Management*, 51:174–185, 2015.

[21] Francisco Herrera, Sebastián Ventura, Rafael Bello, Chris Cornelis, Amelia Zafra, Dánel Sánchez-Tarragó, and Sarah Vluymans. *Multiple instance learning: foundations and algorithms*. Springer, 2016.