5th International Conference on Advanced Research Methods and Analytics (CARMA2023)
Universidad de Sevilla, Sevilla, 2023
DOI: http://dx.doi.org/10.4995/CARMA2023.2023.16429

Georeferencing sentiment scores to map and explore tourist points of interest

Luigi Celardo[1], Michelangelo Misuraca[2], Maria Spano[3]
[1]Department of Social Science, University of Naples Federico II, Italy, [2]Department of Business Administration and Law, University of Calabria, Italy, [3]Department of Economics and Statistics, University of Naples Federico II, Italy.

Abstract

Tourists are increasingly involved in co-creating attractions' symbolic images, sharing their experiences and opinions on websites like TripAdvisor and other similar rating and review platforms. In this paper, we propose a strategy for analyzing people' opinions about tourist points of interest, using an Ambient Geographic Information approach to georeference the polarity scores of reviews. Visualizing these scores on a map can be used to obtain helpful information for implementing strategic actions and policies of institutional and business actors involved in the tourist industry, as well as to help users plan their future experiences. A case study concerning the reviews of the restaurants in Naples (Italy) shows the effectiveness of the proposal.

Keywords: *social media, polarity score, georeferenced sentiment.*

1. Introduction

The diffusion of Web 2.0 changed in a few years each aspect of everyday life related to social representation and interaction. In a more flexible and disintermediated society, we are observing a growing consideration of how people – increasingly embedded in a digital ecosystem – communicate their feelings, opinions and experiences. This behavior is part of our reality, in accordance with the online and offline interconnectedness posited by Jurgenson (2019). This view implies a new communication paradigm in the definition of an experience of buying and consuming a product or service, including tourism-related ones. Today, social media guesting the narratives of people's experiences are used by the diverse actors involved in the tourist industry. Visitors, in particular, play a key role in co-creating tourist attractions' symbolic image, and the reputation generated by digital platforms is becoming progressively significant. Many individuals rely heavily on reviews and free-text comments on platforms such as TripAdvisor, Booking or Google when choosing a destination, reserving accommodations or a table in a restaurant, visiting cultural, scenic or amusement attractions. On the other hand, the use of people's opinion in tourist industry can contribute positively to the decision-making processes of institutions and businesses in fostering sustainability and understanding consumer behavior patterns.

In this scenario, the so-called *electronic word-of-mouth* becomes a primary source of information that deserve to be taken into account (Cheung & Thadani, 2012). It is necessary to monitor what visitors say about their experiences and feelings (Nambisan & Watt, 2011). Analyzing people's comments and reviews is essential since dissatisfied visitors are potentially dangerous, triggering a vicious circle of lousy reputation (Shirdastian, Laroche, & Richard, 2019), deterioration of symbolic value and significant financial losses (Luo, 2009). Nevertheless, these comments are composed as free text, and the information is encoded in a form that is difficult to process automatically. In a text mining framework, it is possible to pre-treat a textual body made of these comments and transform the unstructured data into structured data, performing statistical analyses to extract and manage the underlying knowledge base. Among the different tasks of text mining useful in the tourist industry, a significant role is played by the detection of the semantic orientation of texts, expressing the so-called *sentiment* in a numerical form. The *sentiment* resulting from the reviews written by tourists for a particular activity or attraction can be interpreted as a quantitative feature of a process involving their narratives about the experience.

Several alternative approaches have been proposed to calculate a score representing texts' negative/positive orientation and employ these scores in an opinion-mining strategy (Hemmatian & Sohrabi, 2019). Contributions proposed in the literature have been more oriented towards topic extraction (Zhao *et al.*, 2016) or classification (Kim & Lee, 2014), where instead the visualization of sentiment is still an open research topic (Kurcher *et al.*, 2018).

In a tourist domain, the geographical dimension has to be wisely considered to better analyze its socio-economic traits. In this work, we propose a strategy based on the computation of polarity scores for a set of reviews related to some points of interest and their spatial localization. Georeferenced data, typically represented by a set of geographic coordinates, allows visualizing and understanding spatial patterns and relationships that may not be apparent from other types of data. Latitude and longitude coordinates may be then used to georeference the sentiment and visualize the semantic orientation of each tourist point of interest on a map. Our strategy relies on the *Ambient Geographic Information* (AGI) approach (Stefanidis *et al.*, 2013), in which social media are used to understand the human landscape and its evolution over time. Starting from the AGI framework, we extended the concept to digital platforms like TripAdvisor. Other authors considered the importance of using sentiment data and geographical references. The semantic orientation of comments posted on Twitter have been geographically analyzed to study the density of unfavorable/favorable opinions across U.S. (Camacho *et al.*, 2021). The joint use of sentiment and geographical data has been also used to assess the socio-environmental impact of large-scale infrastructure projects (Li et *al.*, 2021). In a tourist domain, the *tourism sustainability index* (TSI), a synthetic indicator encompassing a dimension based on polarity scores, has been proposed to frame and georeference tourist satisfaction in accordance with the European Tourism Indicator System (De Marchi *et al*., 2022).

From a theoretical viewpoint, the sentiment scores and other metadata may be used to cluster the points of interest in a spatial perspective, better guiding the actions of operators and institutional actors and people interested in exploring an area to plan a visit or a journey. In the following, a case study based on the city of Naples (Italy) is presented to show the proposal's effectiveness. Naples is nowadays one of the most important tourist attractions in Italy. Its historic center has been on the UNESCO World Heritage List since 1995 but has only become a primary destination in the last few years. Here, we are particularly interested in evaluating restaurants because Neapolitan restaurants are considered part of the tourist experience (Vrontis *et al.*, 2021), as the city is famous for its cuisine and gastronomic heritage. Furthermore, the food & wine supply chain is an essential driver of the local economy, creating jobs and promoting the area (Della Corte *et al.*, 2015).

2. Materials and methods

The restaurant reviews used in this study have been scraped from the Italian TripAdvisor website, using `Naples` in the query and `restaurants` as the main category, and stored in a local repository together with some metadata: *name*, *address*, *latitude* and *longitude* (validated with the corresponding Google Maps `id place`), *rating*, *# of reviews*. At the current stage, we considered 774 activities (a share of 30% with respect to the total number of 2,634 restaurants in Naples). Moreover, we decided to set a limit of 1,000 reviews per restaurant – namely the most recent ones – obtaining a collection of 283,801 reviews.

To perform a lexicon-based sentiment analysis of this collection, we used an original customized lexicon of Italian terms. Most resources in the sentiment research area, like lexicons, labelled collections and NLP tools, are mainly available in English. The lack of linguistic resources is critical in the majority of studies, producing a so-called *lexical gap* (Chiavetta *et al.*, 2016). Thus, we built an Italian lexicon by merging the resources developed in the Sentix project (Basile & Nissim, 2013), the Opener project (Russo *et al.*, 2016), and other aptly screened studies (e.g., Bolasco & Della Ratta, 2004). The resulting lexicon contains 26,511 polarized terms with a value of +1 if positives and -1 if negatives.

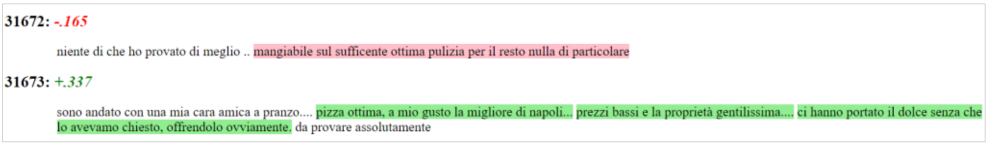
For our analysis, a light pre-treatment was applied to the reviews. Non-alphabetic characters and symbols – like numbers or emoticons – were removed to consider only content-bearing words. The polarity scores have been calculated with a sentence-level logic (Balbi *et al.*, 2018). Each review is segmented into its constituent sentences to consider the sentiment associated with each aspect of the described experience. Given a review r_i (i=1,…,n), its a_i sentences $\{s_{i1}, ..., s_{ik}, ..., s_{ia_i}\}$ are identified by considering as separators strong punctuation marks like full stops, question marks and exclamation marks. The k-th sentence s_{ik} is represented as a sequence of its p_k terms $\{w_{ik1}, ..., w_{ikj}, ..., w_{ikp_k}\}$. Each term w_{ikj} in the k-th sentence of the i-th review is compared with the terms in the lexicon, assigning -1 to negative terms and +1 to positive terms, respectively. Terms not listed in the lexicon are scored with a null value. The polarity of each term is then weighted considering negators (e.g., *mai*, *nessuno, nessuna*), amplifiers and de-amplifiers (e.g., *poco, molto, pochissimo*), adversative and contrasting terms (e.g., *ma*, *tuttavia*). This weighting scheme allows for emphasizing or dampening the negativity or positivity of each polarized term, leading to a more effective measure of semantic orientation (Vechtomova, 2017). The polarity score of each sentence $\mathrm{PS}_{s_{ik}}$ is obtained as the sum of weighted term scores $\mathrm{PS}_{w_{ikj}}$ on the sentence length:

$$\mathrm{PS}_{s_{ik}} = \frac{\sum_{j=1}^{p_k} \mathrm{PS}_{w_{ikj}}}{\sqrt{p_k}} \tag{1}$$

Since we are interested in obtaining a polarity score at a review level, we calculated an overall score PS_{r_i} for each text by a down-weighted zeros average of sentence polarities, giving a lower weight to sentences conveying a neutral sentiment:

$$\mathrm{PS}_{r_i} = \frac{\sum_{k=1}^{a_i} \mathrm{PS}_{s_{ik}}}{a_i^- + a_i^+ + \sqrt{\log(1+a_i^0)}} \tag{2}$$

where a_i^-, a_i^+ and a_i^0 are the numbers of sentences in r_i with a negative, positive, or neutral polarity, respectively. Figure 1 graphically depicts an example of the polarity score computation, reporting for each review the overall score together with negative aspects (in red) and positive aspects (in green) of the tourist experience.



31672: *-.165*

niente di che ho provato di meglio .. mangiabile sul sufficente ottima pulizia per il resto nulla di particolare

31673: *+.337*

sono andato con una mia cara amica a pranzo.... pizza ottima, a mio gusto la migliore di napoli... prezzi bassi e la proprietà gentilissima.... ci hanno portato il dolce senza che lo avevamo chiesto, offrendolo ovviamente. da provare assolutamente

Figure 1. Examples of reviews with polarity score and highlighted negative/positive aspects.

Since the scores PS_{r_i} assume values in a $]-\infty,+\infty[$ interval, we decided to rescale all the results in a $]0,1[$ interval to facilitate the interpretation. In the subsequent step of the strategy, sentiment scores have been used to characterize each restaurant. Specifically, we used latitude and longitude to plot the restaurants on a map and visualize the results of the sentiment analysis as hot spots. We categorized the sentiment scores into low ($\mathrm{PS}_{r_i} < 0.3$), medium ($0.3 \leq \mathrm{PS}_{r_i} \leq 0.6$), and high ($\mathrm{PS}_{r_i} > 0.6$), using a red-to-green color palette for the gradient. The different actors can use the resulting representation to explore the specific area under investigation, whereas researchers can include this information in more articulated analytic strategies. Georeferenced polarity scores can be used with other metadata to cluster points of interest identifying groups of activities that share similar characteristics or attributes. This can be particularly valuable both from an urban planning and tourist marketing side, where can be crucial to understand the characteristics of different neighborhoods or areas of a city and develop targeted strategies or interventions.

3. Some preliminary results

In Figure 2, we can see the polarity scores of Neapolitan restaurants georeferenced on the city map. The output obtained by applying the strategy can be interactively browsed by zooming on the different points. Here we used a static screenshot by way of illustration. Green areas represent the restaurants with a higher positive sentiment, whereas red areas represent the restaurants with a higher negative sentiment. The map shows a concentration of restaurants with a positive sentiment near the city's historical center and the waterfront near the so-called Riviera di Chiaia, rich in tourist attractions and gastronomic sites. The hot areas, associated with the red color, are mainly located in the peripheral or ex-industrial districts, such as Bagnoli (on the left side of the map) or San Giovanni a Teduccio (on the right side of the map). Although these districts of Naples have tourist potential (e.g., Bagnoli hosts the Science Museum, San Giovanni hosts the National Railway Museum), they have critical deficiencies in the sphere of transport, services and air quality.

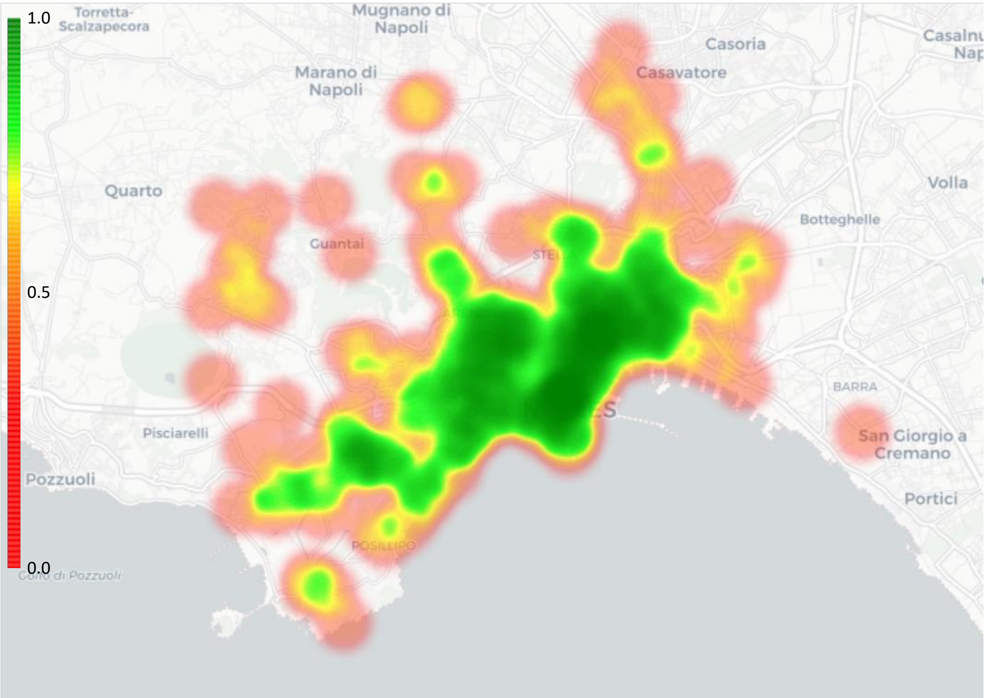


Figure 2. Visualization of restaurants' sentiment in the city of Naples.

According to a study carried out by La Rocca (2021) on the city of Naples, the poor accessibility of peripheral areas (in terms of public transport) and the degradation of urban spaces are, among others, the primary grounds for complaints, affecting negatively the intention of visitors to come back again to the city. Restaurants and shops generally suffer a lack of infrastructure and services (Buonanno *et al.*, 2009). Moreover, the absence of action to protect and enhance the historical and cultural heritage can reduce tourist attractiveness, as in the case of the San Giovanni district.

The complete results of the case study will be discussed more in detail elsewhere.

References

Balbi, S., Misuraca, M. & Scepi, G. (2018). Combining different evaluation systems on social media for measuring user satisfaction. *Information Processing & Management*, *54*(4), 674-685.

Basile, V., & Nissim, M. (2013). Sentiment analysis on Italian tweets. In A. Balahur, E. van der Goot & A. Montoyo (Eds.), *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis* (pp. 100-107). Association for Computational Linguistics.

Bolasco, S., & Della Ratta, F. (2004). Experiments on semantic categorisation of texts: analysis of positive and negative dimension. In G. Purnelle, C. Fairon & A. Dister (Eds.), *Le poid des mots. Actes des 7es Journées internationales d'Analyse statistique des Données Textuelles* (pp. 202-210). UCL Presses Universitaires de Louvain.

Buonanno, G., Morawska, L., Stabile, L., & Viola, A. (2009). PM concentrations in pizzerias: implications for occupational exposure. *Journal of hazardous materials*, *164*(2-3), 870-874.

Camacho, K., Portelli, R., Shortridge, A., & Takahashi, B. (2021) Sentiment mapping: point pattern analysis of sentiment classified Twitter data. *Cartography and Geographic Information Science*, *48*(3), 241-257.

Cheung, C.M.K., & Thadani, D.R. (2012). The Impact of Electronic Word-of-Mouth Communication: A Literature Analysis and Integrative Model. *Decision Support Systems*, *54*, 461-470.

Chiavetta, F., Lo Bosco, G., & Pilato, G. (2016). A Lexicon-based Approach for Sentiment Classification of Amazon Books Reviews in Italian Language. In T.A. Majchrzak, P. Traverso, V. Monfort, K.-H. Krempels (Eds.) *Proceedings of the 12th International Conference on Web Information Systems and Technologies* (pp. 159-170). Scitepress.

Della Corte, V., Sciarelli, M., Cascella, C., & Del Gaudio, G. (2015). Customer Satisfaction in Tourist Destination: The Case of Tourism Offer in the City of Naples. *Journal of Investment and Management*, *4*(1-1), 39-50.

De Marchi, D., Becarelli, R., & Di Sarli L. (2022). Tourism Sustainability Index: Measuring Tourism Sustainability Based on the ETIS Toolkit, by Exploring Tourist Satisfaction via Sentiment Analysis. *Sustainability*, *14*(13), 8049-8067.

Hemmatian, F., & Sohrabi, M.K. (2019). A survey on classification techniques for opinion mining and sentiment analysis. *Artificial Intelligence Review*, *52*, 1495-1545.

Jurgenson, N. (2019). *The Social Photo: On Photography and Social Media*. New York, NY: Verso Books.

Kim, K., & Lee, J. (2014). Sentiment visualization and classification via semi-supervised nonlinear dimensionality reduction. *Pattern Recognition*, *47*, 758-768.

Kurcher, K., Paradis, C., & Keren, A. (2018). The state of the art in sentiment visualisation. *Computer Graphics Forum*, *37*, 71–96.

La Rocca, R.A. (2021). Urban Accessibility and Tourist Activity: An Application to the Metropolitan City of Naples. In D. La Rosa, & R. Privitera (Eds.), *Innovation in Urban and Regional Planning - INPUT 2021* (pp. 583-591) Springer.

Li, Y., Zhang, Y., Tiffany, L. A., Chen, R., Meng, C., & Liu, J. (2021). Synthesizing social and environmental sensing to monitor the impact of large-scale infrastructure development. *Environmental Science & Policy*, *124*, 527-540.

Luo, X. (2009). Quantifying the Long-Term Impact of Negative Word of Mouth on Cash Flows and Stock Prices. *Marketing Science*, *28*(1), 148-165.

Nambisan, P., & Watt, J.H. (2011). Managing customer experiences in online product communities. *Journal of Business Research*, *64*, 889–895.

Russo, I., Frontini, F., & Quochi, V. (2016). *OpeNER Sentiment Lexicon Italian - LMF*, Institute for Computational Linguistics "A. Zampolli", National Research Council, Pisa. http://hdl.handle.net/20.500.11752/ILC-73.

Shirdastian, H., Laroche, M., & Richard, M.-O. (2019). Using big data analytics to study brand authenticity sentiments. The case of Starbucks on Twitter. *International Journal of Information Management*, *48*, 291-307.

Stefanidis, A., Crooks, A., & Radzikowski, J. (2013). Harvesting ambient geospatial information from social media feeds. *GeoJournal*, *78*(2), 319–338.

Vechtomova, O. (2017). Disambiguating context-dependent polarity of words: An information retrieval approach. *Information Processing & Management*, *53*, 1062-1079.

Vrontis, D., Basile, G., Tani, M. & Thrassou, A. (2021). Culinary attributes and technological utilization as drivers of place authenticity and branding: the case of Vascitour, Naples, *Journal of Place Management and Development*, *14*(1), 5-18

Zhao, Y., Qin, B., Liu, T., & Tang, D. (2016). Social sentiment sensor: A visualization system for topic detection and topic sentiment analysis on microblog. *Multimedia Tools and Applications*, *75*(15), 8843-8860.