MONITORING ONLINE PERCEPTION OF ENVIRONMENTAL ISSUES ON COASTS OF SICILY

Damiano De Marchi[1], Mirko Lalli[1], Alessandro Mancini[1]
[1] The Data Appeal Company - Travel Appeal, via Ippolito Pindemonte 63 - 50124, Florence (Italy), phone +39 345 156 3011, e-mail: damiano.demarchi@datappeal.io

Abstract – The analysis of big data on human experience (reviews, comments, ratings, etc.) can provide valuable insights to companies and institutions about market intelligence, since they have significant impact on consumers' purchase decisions. But there are other fields of applications. This pioneer study applied the artificial intelligence proprietary tools of The Data Appeal Company for a different aim: monitoring the online perception of environmental issues on 88 beaches of Sicily. Results proved that it is possible to monitor environmental situation even to sites where there are no other kind of monitoring, using as bases the free and available contents posted by humans online, processed and analyzed by artificial intelligence.

Introduction

The development of Social Media Companies and Web 2.0 in early 2000s has encouraged people to express their opinions about products/services. As e-commerce continued evolving, many online enterprises (Amazon, TripAdvisor, Booking.com) included means for registered users to be able to share opinions about their buying experiences, offering more sophisticated methods to enrich the review experience, e.g. including the rating or the profile and popularity of the reviewer. This process generates volumes of information, that are commonly referred to as Big Data: projections estimates in 2020 there will be around 40 zettabytes (40 trillion gigabytes of data), including social media and minor companies not solely dedicated to digital services [1].

Opinions are central to most human activities and hence, are one of the key drivers of human behaviors [2] and many studies suggest that online product reviews and related features have a significant impact on consumers' purchase decision and sales [3-5]. Reasons are multiple: on the one hand, consumers can obtain information before making their actual purchase decisions [6], on the other hand, companies attract consumers by providing an online platform that enables customers to exchange their consumption experiences [7].

Businesses and institutions gain valuable insights into the massive amounts of the information they have by applying tools and techniques of Big Data Analytics, i.e. the techniques utilized to examine and process Big Data so that hidden underlying patterns are revealed, relationships are identified, and other insights concerning the application context under investigation are exposed [8]. Opinions, sentiments, and emotions can be captured using the individual's writings, facial expressions, speech, and many other media [9] via Sentiment Analysis which is generally defined as the computerized process of recognizing, detecting, and determining the orientation of human opinion or emotion and its polarity [10]. In our daily life, there are many applications of Sentiment Analysis, mostly concentrated on market intelligence: measuring the degree of user satisfaction on products or services to

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)
Damiano De Marchi, Mirko Lalli, Alessandro Mancini, *Monitoring online perception of environmental issues on coasts of Sicily,* pp. 214-218, © 2020 Author(s), CC BY 4.0 International, DOI 10.36253/978-88-5518-147-1.21

improve their weaknesses or developing new products and services, forecasting of price changes according to news sentiments, etc. [11].

Background

The Data Appeal Company, formerly known as Travel Appeal, is an Italian scale-up founded in 2014 specialized in Data Science and Artificial Intelligence. With an initial main focus on the travel industry, it has built the world's largest and most efficient Travel & Location Intelligence Data Lake, mapping in deep detail all travel properties and Points of Interest (POI) and connect them with all data about human experience (text and visual contents, social conversations, reviews, prices, events, bookings) on over 80 online channels. Through its semantic engine, proprietary algorithms and Artificial Intelligence system it can read and process millions of data from different sources to find relevant information, collecting and analyzing online travel data in real time, transforming that data into updated and immediately applicable strategies.

Enlarging its data lake to any POI on a map, made the company evolve and also its solutions ecosystem (Dashboard, API and App) to be applied to any sector (banking, finance, real estate, retail, etc.) having a specific interest on location intelligence and reputation. This led to the idea to use its Artificial Intelligence towards developing sustainability to provide important information regarding all the dimensions involved - economics, environment, social dimensions - at any spatial level as perceived and experienced by humans (local community, visitors, tourists), with an application to the coasts of Sicily.

The Data Appeal Artificial Intelligence has been applied to the coasts of Sicily (largest island of Mediterranean See), analyzing over 15000 reviews (texts and scores) on Tripadvisor and Google from a sample of 88 public beaches in order to verify the level of cleanliness perceived by the users and the main factors that determine a positive or negative judgment in the general perception.

Proposed Methodology

The proposed methodology is a complicate process [12-13] that need several steps of analysis, that we briefly summarized in this paragraph.

The starting point of the analysis is to extract information from reviews: the aim is not only to provide a polarity score (sentiment) for each content, but to identify the main topics and the subjects (aspects) and judgments (opinions) connected with these topics. The topic to be analysed can also be very abstract: the strength of the algorithm used is a technique called Word2Vec allowing to represent words through a multidimensional numeric vector. This vector has an important characteristic: words used in similar contexts (e.g. coasts, beaches, …) have similar vectors, that is "neighbours" in a reference vector space. This closeness allows to enrich the analysis on a given topic with all those terms and / or sentences that cannot be predefined, but that emerge directly from the texts or better from the contexts. As an example, I might be interested in an analysis on the beaches and the algorithm suggests including in the analysis terms such as "beach", "coast", "bay", which widen the perimeter of

analysis. The ultimate goal is to identify all the reviews or phrases related to the topic of analysis.

The second step is to evaluate the polarity of these sentences in order to assert what people think about the topic extracted. In order to do this it has been used a Sentiment analysis model: it is a classic machine learning model in the NLP (Natural Language Processing) field that seeks non-linear dependencies between the various words to "understand" computationally the logics that represent satisfaction and, more generally, the polarity, of a generic text. It is a supervised model, more generally of a neural network that uses an embedding layer to numerically represent the words of a given dictionary, also in this case specific to language. One of the strengths of this algorithm is that it does not require opinions in the text, but manages to provide a multiclass score (positive, negative and neutral) of any text.

Details have to be analysed to:

- Identify key phrases and relevant topics (calculated on sentences that concern our topic of analysis and are therefore contextualized)
- Identify the sentiment of the sentences and consequently the sentiment of the theme analysed
- Identify the opinions connected with these sentences and relate them to the relevant themes of the first point
- Geolocate the POIs where the topic is "discussed" most, as well as the sentiment associated with these POIs
- Identify the temporal point of view the subject being analysed

Results need to be further analysed in order to understand the magnitude of the topics that talk about cleanliness in the review compared with other topics that people obviously discussed inside texts. This led to the creation of 2 classers: one that contains sentences talking about cleanliness (identified by the previous algorithm) and one containing all the other sentences: what emerged overall is that the arguments and opinions people discuss about are repeated.

The final steps include computing a n-gram model (n-gram is a contiguous sequence of n items): an n-gram model is a type of probabilistic language model for predicting the next item in such a sequence in the form of a $(n-1)$–order Markov model. In this way it is possible to know which are the most used terms by people inside reviews. Then the calculation of the magnitude of a specific n-gram, i.e. the percentage of document that contain that specific sequence of words.

Results

The top keywords on beaches have a strong relation to element of cleanliness as shown in Table 1.

It covers over 87 % of the reviews mentioning beach, sea, water, sand. This has a direct marketing effect for local businesses and tourism operators, but this led also that cleanliness is a distinctive element for the destination overall, with governance and management implications. In general, the sampled beaches were considered clean, but with a different level of cleanliness and a relative satisfaction. The research further analysed, when possible, the elements who generated the negative impact on cleanliness perception.

Table 1 - Keywords with more than 1000 mentions.

Keyword	Mentions	% related to element cleanliness
Spiaggia (beach)	6175	87 %
Mare (sea)	4235	87 %
Acqua (water)	1737	91 %
Sabbia (sand)	1134	88 %

In figure 1 it is possible to see results of this analysis: it appears that terms more associated with a class are a further distance from the diagonal line between the lower-left and upper-right corners: the presence of waste generated by humans is noticed by travelers. The judgments regarding the presence of garbage are linked to attitudes of carelessness on the part of beach guests that show "incivility" in the management of waste such as "plastic", "bottles", the main elements characterizing the negative impact on the cleanliness perception.

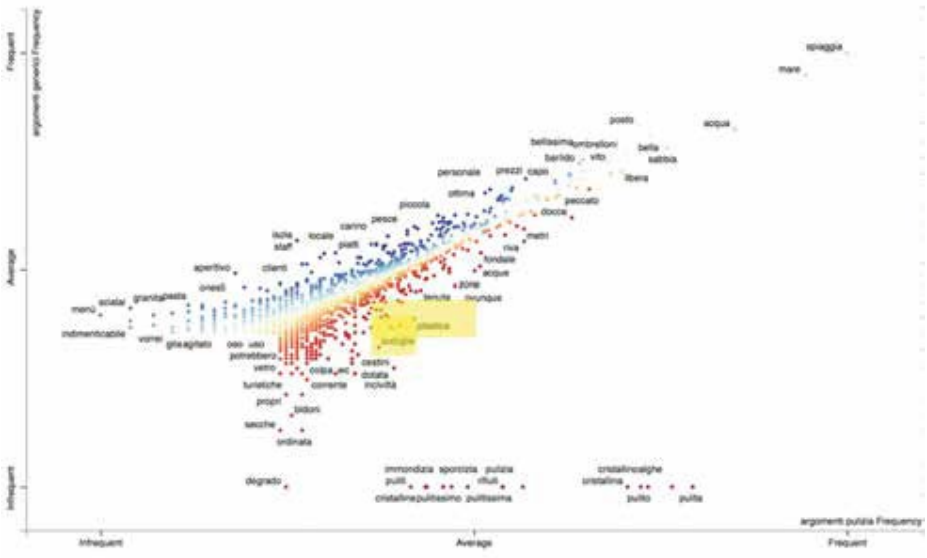

Figure 1 - Frequency of general keywords and cleanliness elements.

Conclusions

This study started with the idea of using reviews, ratings, posts, not for market intelligence proposes, but to foster sustainability. Results showed that this can lead to monitor environmental situation to sites where there wasn't any kind of monitoring, using as bases the free and available contents posted by humans online, processed and analyzed by Artificial

Intelligence. The Regional Government took the results as one of the preliminary studies to enact the new regional plastic-free law for beaches and coastal areas. This first application opens a brand-new world for Artificial Intelligence analysis of human experience big data: all information and tools provided by companies as The Data Appeal Company can become a fundamental source, combined to official statistics, to develop a really integrated and innovative information base towards monitoring sustainability and extending the current statistical frameworks beyond their economic focus, to incorporate environmental, and social dimensions and at relevant spatial levels: global, national and sub-national.

Acknowledgments

This study was supported by the Data Appeal Company.

References

[1] EMC and IDC (2014) - *The digital universe of opportunities*, Infobrief, April.

[2] Hu M., Liu B. (2004) - *Mining and summarizing customer reviews.* In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '04). Association for Computing Machinery, New York, NY, USA, pp. 168–177.

[3] Duan, W, Gu, B & Whinston, AB (2008) - *The dynamics of online word-of-mouth and product sales-An empirical investigation of the movie industry*, Journal of Retailing, vol. 84, no. 2, pp. 233-242.

[4] Forman C., Ghose A., Wiesenfeld, B. (2008) - *Examining the Relationship Between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets,* Information Systems Research 19, pp 291-313.

[5] Elwalda A., Lü K. Ali M. (2016) - *Perceived derived attributes of online customer reviews,* in Computers in Human Behavior 56 pp. 306-319.

[6] Zhu F., Zhang X. (2010) – *Impact of online consumer reviews on sales: the moderating role of product and consumer characteristics,* J. Mark. 74 pp. 133–148.

[7] Mudambi S.M., Schuff D. (2010) *What makes a helpful online review? A study of customer reviews on amazon.com,* MIS Q. 34 (2010) pp. 185–200.

[8] Iqbal R., Doctor F., More B., Mahmud S., Yousuf, U. (2016) - *Big Data analytics: Computational intelligence techniques and application areas.* International Journal of Information Management.

[9] Yadollahi A., Shahraki A., Zaiane O. (2017) - *Current State of Text Sentiment Analysis from Opinion to Emotion Mining.* ACM Computing Surveys. pp. 1-33.

[10] Kang D., Park, Y. (2014) - *Review based measurement of customer satisfaction in mobile service: Sentiment analysis and VIKOR approach.* Expert Systems with Applications, 41(4), pp. 1041–1050.

[11] Soleymani M.*et al.* (2017) - *A survey of multimodal sentiment analysis*, in: Image and Vision Computing, vol. 65, pp. 3-14.

[12] Neurocomputing vol. 403 (2020), Elsevier.

[13] Information Sciences vol. 532 (2020), Elsevier.