



Tourism destination management using sentiment analysis and geo-location information: a deep learning approach

Marina Paolanti¹ · Adriano Mancini¹ · Emanuele Frontoni¹ · Andrea Felicetti¹ · Luca Marinelli¹ · Ernesto Marcheggiani¹ · Roberto Pierdicca¹

Received: 26 January 2020 / Revised: 23 December 2020 / Accepted: 22 January 2021 /

Published online: 16 February 2021

© The Author(s) 2021

Abstract

Sentiment analysis on social media such as Twitter is a challenging task given the data characteristics such as the length, spelling errors, abbreviations, and special characters. Social media sentiment analysis is also a fundamental issue with many applications. With particular regard of the tourism sector, where the characterization of fluxes is a vital issue, the sources of geotagged information have already proven to be promising for tourism-related geographic research. The paper introduces an approach to estimate the sentiment related to Cilento's, a well known tourism venue in Southern Italy. A newly collected dataset of tweets related to tourism is at the base of our method. We aim at demonstrating and testing a deep learning social geodata framework to characterize spatial, temporal and demographic tourist flows across the vast of territory this rural touristic region and along its coasts. We have applied four specially trained Deep Neural Networks to identify and assess the sentiment, two word-level and two character-based, respectively. In contrast to many existing datasets, the actual sentiment carried by texts or hashtags is not automatically assessed in our approach. We manually annotated the whole set to get to a higher dataset quality in terms of accuracy, proving the effectiveness of our method. Moreover, the geographical coding labelling each information, allow for fitting the inferred sentiments with their geographical location, obtaining an even more nuanced content analysis of the semantic meaning.

Keywords Sentiment analysis · Geotagged social media · Deep learning · Tourism

✉ Marina Paolanti
m.paolanti@univpm.it

¹ Università Politecnica delle Marche, Ancona, Marche, Italy

1 Introduction

The proliferation of the Web 2.0 tools as well as the spread of social media in the last years, is bringing a significant increase in the use of user-generated contents (UGC). In this context, sentiment analysis is now making its way with substantial growth in several domains such as marketing and finance. It is a branch of affective computing research (Poria et al. 2017) and this area is now widely recognized as a new Natural Language Processing (NLP) line, which broadly analyses users' opinions, reviews or thoughts about topics, companies or experiences identifying its sentiment (Liu 2015). Sentiment analysis has the aim of evaluating whether a textual item expresses a positive or negative opinion, in general terms or about a given entity (Nakov et al. 2016). To get at the sentiment classification of textual items, a well-defined sentiment lexicon provides the list of sentiment words and phrases, each with an explicit sentiment score. Indeed, the sentiment lexicon significantly influences the performance of a lexicon-based Sentiment Analysis (Hogenboom et al. 2014). The sentiment lexicon is built from dictionary-based and corpus-based methods; these latter furtherly divided into statistical-based and semantics-based forms according to the specific techniques (Khan et al. 2016).

Research efforts have been devoted to developing algorithms and sentiment analysis methods, able to automatically detect the underlying sentiment of a text (Sun et al. 2017). Many companies are deploying these algorithms to make better decisions, understanding customers behaviour or thoughts about their company or any of their products (Valdivia et al. 2019).

While several studies have examined various users' reactions towards brands or companies (Paolanti et al. 2017; Ghiassi et al. 2013), the sentiment related to tourism at the destination level and its possible outcomes has not received much attention from scholars. Tourism has both positive and negative economic, social and environmental impacts on destinations. One of the main goals of sustainable tourism development is to minimize the negative social and ecological impacts of visitor flux, improving at the same time the experience of visitors and the quality of life of residents (Gursoy et al. 2019). Despite the potentials of sentiment analysis to this end, exploiting the massive tourism reviews and posts produced, the recourse to this is limited since the manual and comprehensive extraction of useful knowledge from the studies is costly and time-consuming (Kim et al. 2017; González-Rodríguez et al. 2016).

Currently, Twitter (<http://twitter.com/>) represents one of the most popular social networking platforms. It allows people to express their opinions, favourites, interests, and sentiments towards different topics and issues they face in their daily life. The messages, namely tweets, are real-time (Alharbi and de Doncker 2019). Twitter gives the possibility to access to the unprompted views of a broad set of users on particular products or events. Twitter has about 200 billion tweets per year (corresponding to 500 million tweets per day, 350,000 tweets per minute, and 6000 tweets per second) (Alharbi and de Doncker 2019). There is also a need to automate the identification and extraction of the spatial information about tourist flows from tweets. Accurately identifying localized flood extent is difficult,

especially in urban areas, and social media can help with gathering and disseminating information. These are some of the reasons for the choice of Twitter as a case study in this research.

Tourism tweets are informal and conversational. Thus, effective sentiment analysis techniques must be introduced to process such textual data. Furthermore, significant construction of a high-quality tourism-specific sentiment lexicon is of great value. Most of the existing methods of Twitter sentiment classification apply machine learning algorithms, such as Naïve Bayes, Maximum Entropy classifier (Max. Ent.) and Support Vector Machines (SVMs) to build a classifier from tweets with manually annotated sentiment polarity labels (Saleena 2018). In recent years, there has been a growing interest in using deep learning techniques, which can enhance classification accuracy, especially in cases when the amount of labelled data is thousands of examples.

Considering the motivations stated above, the research contributions of this manuscript are twofold. On one hand, we experimentally compare the performance of state of art deep learning methods to estimate the sentiment related to a well-known tourism destination in Southern Italy. In particular, the sentiment is classified by four specially trained Deep Neural Networks (DNNs). The networks used are the ones proposed by Kim (2014); Kim et al. (2016) and Zhang et al. (2015), and these are word- and character-based. Four DNNs architectures are chosen by testing different combinations of parameters and taking into consideration the best one. The variable parameters tested are the maximum characters length of tweets and the size of the dictionary (for word-based DNNs) or alphabet (for characters-based DNNs). On the other, by exploiting geographical information, we show that deep learning architectures allows to infer, with a data-driven approach, spatial, temporal and demographic features of tourist flows. Indeed, recent research trends demonstrates that the combination of sentiment analysis and geo-location information can contribute to more efficient planning of tourism destinations (Yan et al. 2020). The approach has been applied to a newly collected dataset of tourism-related tweets. The dataset as well as the implementation code, is publicly available at (<https://XXX.it/content/dataset-tweets-related-cilento>, *obscured for blinded purposes*). In contrast to many existing datasets, the true sentiment is not automatically judged by the accompanying texts or hashtags. Still, it has been manually estimated by human annotators, thus providing a more precise dataset. Our approach yielded promising results in terms of accuracy and demonstrated the effectiveness of the proposed method. Our dataset is specific for the tourism domain. However, the models chosen are generic and are the best current state of the art text classifiers. Furthermore, these networks have been tested with other datasets, in their reference papers: (1) Word-CNN (Kim 2014): tested on 6 different datasets, most of them for sentiment analysis; (2) Word-LSTM (Sutskever et al. 2014): tested on WMT'14 datasets; (3) Char-CNN (Kim et al. 2016): tested on English Penn Treebank dataset; (4) ConvNets (Zhang et al. 2015): tested on 8 different datasets, with classes ranging from 2 to 14.

The paper is organized as follows: Sect. 2 is an overview of the research status of textual sentiment analysis; Sect. 3 introduces our approach, also describing the dataset and algorithms; final sections present results (Sect. 5) and conclusions (Sect. 6) with future works.

2 Related works

Several methods are available in the literature that uses classifiers for Twitter sentiment analysis, with the main purpose of detecting polarity. In literature there are many reviews of approaches which show noteworthy differences from both methods and data sources standing point. Most existing studies to Twitter sentiment analysis can be divided into supervised methods (Saif et al. 2012; Kiritchenko et al. 2014; Da Silva et al. 2014; Hagen et al. 2015; Jianqiang and Xueliang 2015; Jianqiang 2016) and lexicon-based methods (Paltoglou and Thelwall 2012; Montejo-Ráez et al. 2014). Supervised methods are based on training classifiers (such as Naive Bayes, Support Vector Machine, Random Forest) using feature combinations such as Part-Of-Speech (POS) tags, word N-grams, and tweet context information features (such as hashtags, retweets, emoticon, capital words, etc). Lexicon-based methods determine the overall sentiment tendency of a given text by utilizing pre-established lexicons of words weighted with their sentiment orientations, such as SentiWordNet (Baccianella et al. 2010) and WordNet Affect (Miller 1995). They represent a pair of knowledge-based techniques, based on the resources of semantic knowledge to determine the polarity of feelings expressed by peers. The analysis of tweets looking for underlined feelings, the “sentiment” sent through the texts, is based on the identification of specific emotional words from a defined lexicon. The ease of application and accessibility gives the popularity of these methods. However, their performance depends on a comprehensive basis and a rich representation. Other methods exploit statistical approaches (Cambria and White 2014) but, despite their widespread use in the scientific environment, show performance strongly influenced by the availability of large training sets (Cambria and White 2014).

Recently deep learning methods (Tang et al. 2014), well established in machine learning (Bengio 2009; LeCun et al. 2015), are becoming most popular by replacing shallow feature representations as for example bag-of-words combined with support vector machines, used in the past as the primary methods for identifying and analyzing textual sentiment analysis. There are three main methods used to classify sentiment: Support Vector Machine (SVM), Naïve Bayes and Artificial Neural Network (ANN), and deep learning algorithms. The use of ANN comes with a more significant computational cost but better results than the other two approaches. SVM and Naïve Bayes methods have limited performance but with a reduced computational cost. In (Kim 2014), the author reports some experiments using a model based on Convolutional Neural Networks (CNN) to extract features and detect automatic sentiment analysis of Twitter messages. Experimental results on different datasets have demonstrated that the classification performance are improved by using CNNs. A new deep neural network model has been introduced by Santos et al. (2014) that implements Twitter sentiment analysis by using information from character-level to sentence-level and obtaining high accuracy of prediction. In Jianqiang et al. (2018), the authors use a convolution algorithm to train a deep neural network for improving the accuracy of Twitter sentiment analysis. The method uses statistical characteristics and semantic relationships between the words of each tweet.

The automatic sentiment evaluation has proved to play an important role even for tourism (Moreno-Ortiz et al. 2019); in this sector, the machine learning methods mainly used are based on SVM and Naïve Bayes approaches (Kirilenko et al. 2018; Alaei et al. 2019). Moreover, many of these works compare the performance among different machine learning methods. The work of Ye et al. (2009) compares the results obtained using three other machine learning methods: SVM, Naïve Bayes and k-Nearest Neighbors (k-NN) algorithm, considering reviews on seven famous destinations in Europe and in the United States, they obtained the best results with SVM classifier. In another work (Zhang et al. 2011), comparing only SVM and Naïve Bayes algorithms, the best results are obtained with a Naïve Bayes, obtaining a high accuracy in the sentiment classification of restaurant reviews.

There are other works that use lexicon-based approaches for tourist sentiment analysis. Serna et al. (2016) considered tweets concerning Summer and Easter holiday and used WordNet algorithm (Miller 1995) to extract emotions from tweets. SentiWordNet (Baccianella et al. 2010), an evolution of WordNet, has been successively used to classify sentiment in a tour reviews forum (Neidhardt et al. 2017).

However, few works focus on automatic sentiment analysis in tourism considering texts obtained from social networks as Twitter (Kirilenko et al. 2018). In the tourism domain Chua et al. (2014, 2016) have demonstrated how the analysis of geotagged social media data yields more detailed spatial, temporal and demographic information of tourist movements in comparison to the current understanding of tourist flow patterns in the region. The joint exploitation of sentiment and geo-information revealed to be a powerful instrument in case of critical events, like in Gonzalo et al. (2020) and for disaster management (Wu and Cui 2018); notwithstanding, a deeper understanding of its benefits for tourism destination planning is necessary, motivating the experiments showed in the following.

As stated in Adwan et al. (2020), the approaches adopted for Sentiment Analysis of Twitter data vary from lexicon-based and machine learning to graph-based approaches.

The lexicon based method uses sentiment dictionary with opinion words and match them with the data to determine polarity. Lexicon based approach can further be divided into two categories: Dictionary based approach (based on dictionary words i.e. WordNet or other entries) and Corpus based approach (using corpus data, can further be divided into Statistical and Semantic approaches). Machine learning based approach uses classification technique to classify text into classes, and methods like Naïve Bayes (NB), maximum entropy (ME), and support vector machines (SVM) have achieved great success in sentiment analysis. Recently, the prevalence of Deep Learning models, which are a subset of Machine Learning, are proven to be of high level task, especially for automatic analysis of Twitter data (Lim et al. 2020). Based on the pros and cons highlighted in Table 1, the approach adopted in this paper is based on Deep Learning for its ability to deal with a huge amount of data. Moreover, the DNNs can provide a lot of information by extracting the most relevant features. In this domain this aspect is fundamental, since we deal with a subjective task such as the sentiment.

Table 1 Comparative overview table with the key differences between the proposed frameworks for Twitter Sentiment analysis

	Learning-based	Lexicon-based	Hybrid-based	Graph-based
	Uses classification technique to classify text into classes	Leverages a list of words annotated by polarity or polarity score to determine the opinion score of a given text	Combines two or more approaches for Twitter sentiment analysis such as combining machine-learning and lexicon-based approaches	Represents tweets in graph structures, along with words, hashtags or users
Strengths	They take into consideration both the strength of an opinion as well as the the relevance of the feature the opinion is about for the general customer	It is more understandable and can be easily implemented	The main advantages are the lexicon/learning symbiosis, the detection and measurement of sentiment at the concept level and the lesser sensitivity to changes in topic domain	Probabilistic assumption that a sentiment and its rating are inter-dependent. Overcomes a well-known problem of other methods where a positive sentiment expressed with a word having a negative connotation is misinterpreted (e.g., low price)
Weaknesses	Needs a guideline and known aspects to work. Based on Wordnet database	Cannot deal with domain and context specific orientations	Reviews with a lot of noise (irrelevant words for the subject of the review) are often assigned a neutral score because the method fails to detect any sentiment	The correspondence between identified clusters and the actual aspects or ratings is not explicit (general drawback for unsupervised models)

Table 1 (continued)

	Learning-based	Lexicon-based	Hybrid-based	Graph-based
Automated Sentiment Analysis in Tourism	Support Vector Machine (SVM) and Naive Bayes are the most popular in tourism-related sentiment analysis (Claster et al. 2010)	Lexicon-based sentiment analysis algorithm designed with the main focus on real time Twitter content analysis. The algorithm consists of two key components, namely sentiment normalisation and evidence-based combination function, which have been used in order to estimate the intensity of the sentiment (Jurek et al. 2015)	System uses hybrid approach (lexicon and machine learning-based techniques) for analysing sentiment in the domain of travel and tourism (Phoo and Naw)	A DWWP system consisting of domain-specific words detection (DW) and word propagation (WP) is proposed. DW deals with the negligence of user-invented new words and converted sentiment words by means of AMI (Assembled Mutual Information). This WP approach incorporates manually calibrated sentiment scores, semantic and statistical similarity information, which improves the quality of sentiment lexicon in comparison with existing data-driven methods (Li et al. 2018)

3 Materials and methods

In this section, the textual sentiment analysis framework, as well as the dataset used for its evaluation, are introduced. Cilento is a well-known tourist venue located in Southern Italy, the vastness and the orthographic complexity of the area always are the strongest and the weaken points, at the same time, of this tourism venue. In this peculiar context, that since many years the local administrations struggled for agreeing on best fitting regional tourism strategy, to boost bold development and cohesion. Since late 2013 we have been engaged in a national project (TOOKMC: Transfer Of Organized Knowledge Marche-Cilento) funded by European and state agencies to foster the exchange of best practices in sustainable tourism between developed and under-developed regions in Italy. For the evaluation, trained DNNs have been used, and they have been comprehensively evaluated on a dataset of tweets collected for this work, related to Cilento. The framework shown at Fig. 1. The following subsections, provide further details about the data collection and the DNNs settings.

3.1 Dataset of tweets related to Cilento

The dataset is collected thanks to the Cilentotook project (<http://www.cilentodasoprire.it/>). The Tweets posted by tourists visiting Cilento have been saved, considering an extended period of data collection and a large geographical area. The data collection is not limited to the Cilento and the holiday season only, thanks to the fact that visitors keep on talking about their holidays even after they left holiday destinations, dropping useful information to our analysis. Tweets are collected by a custom Python script that relies on the tweepy library (tweepy Python library—<http://www.tweepy.org/> (last access 2021/01/30 21:45:16)). Tweepy enables an easy use of Twitter streaming API handling authentication, connection, session and message reading. Scripts run in a Docker container and apply a spatial filter based

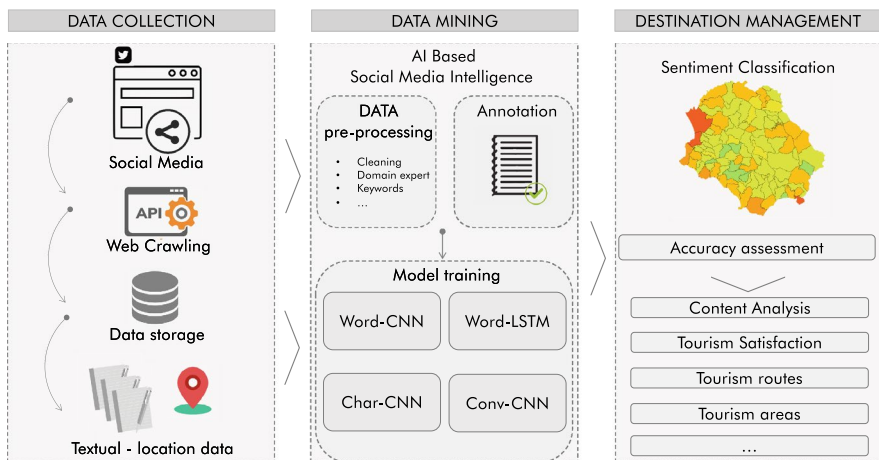


Fig. 1 Workflow of the proposed research framework

on the bounding box that retrieves only tweets in a well-known area exploiting the geographical feature of Twitter API (GeoObjects Twitter API—<https://developer.twitter.com/en/docs/tweets/data-dictionary/overview/geo-objects.html> (last access 2021/01/30 21:45:16)). Extracted tweets have been saved in a MongoDB database running on a Docker container. The data have been collected from May 2017 to May 2019, considering the length of tourism season in the area.

Twitter offers the users an option to “geotag” a tweet as it is published. The geotagging is based on the exact location, assigned to a Twitter Place. This latter can be thought of as a “bounding box” with latitude and longitude coordinates defining the extent of the location area. Such geographic metadata is known as “tweet location”. It gives the highest accuracy in terms of geographic location. The weakest point of using “tweet locations” is only 1–2% of tweets are usually geotagged. Tweets are organized in a database. Each tweet builds a record containing the text and the metadata. Additional information are as follow:

- `_id`: id unique of a tweet;
- `from_user`: username of the user who sent the tweet;
- `to_user_id`: id of the possible user to which the tweet has been sent;
- `loc.lat`: latitude from which the tweet has been sent;
- `loc.lon`: longitude from which the tweet has been sent;
- `created_at`: when the tweet has been sent;
- `text`: text of tweet.

The actual sentiment has been manually assessed by human annotators, to get the ground truth of collected tweets (Italian and English), providing a more accurate and less noisy dataset, if compared to automatically generated labels from hashtags. The dataset is composed of geotagged Twitter data related to Cliento’s area as follows:

- 3200 tweets with positive sentiment;
- 4720 tweets with neutral sentiment;
- 4842 tweets with negative sentiment.

Before splitting data in training/testing, we did a subsampling procedure, keeping all the classes balanced (1/3, 1/3, 1/3). After class balancing, they all have the same number of 3200 samples.

Two different reviewers (annotators) with a Master’s Degree in Computer Science evaluated the annotations, according to the following rules:

- Positive:
 - if it contains words (verbs, adjectives) that have in themselves positive meaning such as: “to love”, “to adore”, “pleasure”, “beautiful”;
 - if it contains words with a negative meaning that have negation in them;
 - if it contains at least two “!” strengthens the positive idea;
 - if it contains typical symbols of liking such as: “happy faces”, “thumbs up”, “hearts”.

- Neutral: if there is no precedence towards Positive or Negative sentiment.
- Negative:
 - if it contains words (verbs, adjectives) that have negative meanings such as: “hate”, “ugly”, “disgusting”;
 - if it contains words with a positive meaning that have a negation meaning;
 - if it contains at least twice a point “!” strengthens the negative idea;
 - if it contains typical symbols such as: “ad facets”, “thumbs down”;

Conflicts or disagreements between the annotators happened only for few revised tweets. A joint agreement was always found after a further mutual confrontation.

3.2 DNNs for textual sentiment analysis

To compare the sentences classification performance along with different parameter settings, four DNNs architectures have been implemented. Two DNNs are based on word-level and are: *Word-CNN* (a Convolutional Neural Network (CNN) Kim 2014) and *Word-LSTM* (a Long Short-Term Memory (LSTM) recurrent neural networks Sutskever et al. 2014). The others are DNNs are character-based: *Char-CNN* (a character-level Convolutional Neural Network (CharCNN) Kim et al. 2016) and *ConvNets* (a character-level Convolutional Networks (ConvNets) Zhang et al. 2015).

For the *Word-CNN* and the *Word-LSTM*, a dictionary has been created using all terms (not-replied) of the entire dataset of text, with the data structure: “index”, “word”, “frequency”. The index is the positioning of the word calculated on the base of its frequency of use compared to the entire dataset. Consequently, in top positions there will be the most frequent words.

To achieve the best performance of the proposed architectures three models of CNN have been tested (Kim 2014):

- Static: the model uses pre-trained vectors. All the word vectors according to Kim (2014) will not be modified, i.e. are kept static.
- Non-static: as the previous model, but the pre-trained vectors are refined for each task.
- Random: all word vectors are randomly initialized according to the reference paper (Kim 2014) and then modified during training.

In particular, the pre-trained vectors derived from the publicly available word2vec vectors that were trained on 100 billion words from Google News. The vectors have dimensionality of 300 and were trained using the continuous bag-of-words architecture (Mikolov et al. 2013b). Words that are not present in the set of pre-trained words are randomly initialized.

The preliminary phase of processing is summarized as follows:

- A vocabulary of words is generated;

- A tweet consists of a set of words. Each word is transformed into a word vector, i.e. a vector of fixed size, formed only by real numbers. So, each tweet is converted in an array of real numbers (array of word vectors), everyone identified by an ID code (ID of each word).
- The arrays of numbers are the inputs of a CNN, that through an embedding layer extracts by each ID a set of features (embedding_dim);
- These features are the inputs of two parallel convolutional layers with different kernel dimension, in order to extract different features.
- Finally, the features of parallel arms are recombined and sent to a fully connected final layer for the following classification step.

In the second approach, for the character-based DNNs, in the preliminary phase of processing, a dictionary or alphabet has been created using all terms not-replied of the entire dataset of text, with the data structure: “index”, “word” or “character”, “frequency”. Not-replied means that the dictionary will only be composed of unique words, i.e. without repeated words, also referring to Kim (2014). The index is the positioning of the word or characters calculated on the base of its frequency of use compared to the entire dataset. Consequently, the most recurrent (higher frequency) words or characters get top positions. Each phrase is mapped into a real vector domain, a technique when working with text called word embedding (Mikolov et al. 2013a, b). This is a technique where words are encoded as real-valued vectors in a high dimensional space, where the similarity between words in terms of meaning translates to closeness in the vector space. Keras (<http://keras.io>) provides a useful way to convert positive integer representations of words into a word embedding by an embedding layer.

The networks have been trained from scratch on our dataset. There are no other publicly available datasets of tweet related to tourism domain that contain tweets both in Italian and English. For this reason, we have not preferred to adopt fine-tuning by pre-training the networks on different datasets. For each model, our dataset has been split in 80% training and 20% validation. The partitions were automatically generated, keeping the tweet sentiment balanced in both. The hyperparameters (learning rate, batch size, epochs, optimization algorithm, loss function) used for the training are different in each model and the default ones in the repositories have been used. No optimization of the hyperparameters has been done. The fine-tuning of the hyperparameters has been done in both types of network. For the Dictionary-Based, different sizes of the dictionary have been tested, varying the number of words that populate it, obviously using the most used words in the dataset. For all the other hyperparameters of the networks, the values described in the reference papers have been used. For example, each word is mapped onto a 32 length real valued vector is the same configuration adopted by the Character-based network. The word vector size has been set to 32, since the dimension has been evaluated experimentally after several tests. The total number of words that we are interested in modeling is also limited to the most frequent words, and zero out the rest. Finally, the sequence length (number of words) in each phrase varies, so we have constrained each phrase to be N-words, truncating longer phrase and pad the shorter phrase with zero values. This choice depends

Table 2 Performance comparison of the DNNs architectures

DNNs	Parameters		Metrics			
	Sentences length	Dict/Alph length	Accuracy	Precision	Recall	F1-score
Word-CNN	128	1000	0.776	0.776	0.776	0.776
	128	3000	0.778	0.778	0.778	0.778
	128	5000	0.767	0.767	0.767	0.767
	128	10000	0.776	0.776	0.776	0.776
	200	1000	0.779	0.779	0.779	0.779
	200	3000	0.767	0.767	0.767	0.767
	200	5000	0.766	0.766	0.766	0.766
	200	10000	0.780	0.780	0.780	0.780
	280	1000	0.774	0.774	0.774	0.774
	280	3000	0.776	0.776	0.776	0.776
	280	5000	0.760	0.760	0.760	0.760
	280	10000	0.784	0.784	0.784	0.784
Word-LSTM	128	1000	0.750	0.750	0.750	0.750
	128	2000	0.751	0.751	0.751	0.751
	128	3000	0.751	0.751	0.751	0.751
	128	5000	0.752	0.752	0.752	0.752
	128	10000	0.731	0.731	0.731	0.731
	200	1000	0.758	0.758	0.758	0.758
	200	2000	0.751	0.751	0.751	0.751
	200	3000	0.751	0.751	0.751	0.751
	200	5000	0.752	0.752	0.752	0.752
	200	10000	0.724	0.724	0.724	0.724
	280	1000	0.755	0.755	0.755	0.755
	280	2000	0.757	0.757	0.757	0.757
	280	3000	0.755	0.755	0.755	0.755
	280	5000	0.758	0.758	0.758	0.758
280	10000	0.724	0.724	0.724	0.724	
Char-CNN	128	70	0.843	0.843	0.843	0.843
	128	100	0.847	0.847	0.847	0.847
	128	200	0.846	0.846	0.846	0.846
	128	345	0.846	0.846	0.846	0.846
	200	70	0.840	0.840	0.840	0.840
	200	100	0.846	0.846	0.846	0.846
	200	200	0.848	0.848	0.848	0.848
	200	345	0.845	0.845	0.845	0.845
	280	70	0.843	0.843	0.843	0.843
	280	100	0.842	0.842	0.842	0.842
	280	200	0.847	0.847	0.847	0.847
	280	345	0.845	0.845	0.845	0.845

Table 2 (continued)

DNNs	Parameters		Metrics			
	Sentences length	Dict/Alph length	Accuracy	Precision	Recall	F1-score
ConvNets	128	70	0.761	0.761	0.761	0.761
	128	100	0.769	0.769	0.769	0.769
	128	200	0.786	0.786	0.786	0.786
	128	345	0.776	0.776	0.776	0.776
	200	70	0.777	0.777	0.777	0.777
	200	100	0.782	0.782	0.782	0.782
	200	200	0.783	0.783	0.783	0.783
	200	345	0.779	0.779	0.779	0.779
	280	70	0.793	0.793	0.793	0.793
	280	100	0.788	0.788	0.788	0.788
	280	200	0.798	0.798	0.798	0.798
	280	345	0.799	0.799	0.799	0.799

on an exploratory evaluation. The results of each experiment are reported in Tables 1 and 2. The aim is to verify that the performance depends on the length. It was observed that some tweets are composed of several sentences in which sentiment is deduced in the first, the other periods contain opinions to support.

Moreover, we tried different combinations of M and N to do different tests, so different quantities for the reference alphabet have been tested. Concerning the neural networks based on character level, more experiments have been executed by varying the number of characters of the alphabet and the maximum number of characters for each tweet. Unique numbers codify the characters of the alphabet based on their frequency of using. Then when a limit on the number of characters that can be used is imposed, the less frequently used characters are deleted. This last was an experimental evaluation. Our aim is to evaluate how much the classification performance depended on the size of the alphabet (for character based model) and the dictionary (for word based model). The results have demonstrated that some of these emoji appear only once in the dataset and therefore are insignificant.

In Kim (2014) and Zhang et al. (2015) the one-hot vector encoding was used, while another type of coding has been used in our work. Given an array of characters, i.e. the tweet, this array will first be recreated in reverse order and then encoded into an array of integers. The number zero is used to distinguish both the spaces and the characters deleted in a limited alphabet. The non-space character is the same in Zhang et al. (2015). Moreover, all the emoji present in tweets are added to the alphabet. Each emoji is a character, but used as a word in the first

two networks. The word embedding output is the input of the second part of each DNN architectures.

3.3 Evaluation metrics

The sentiment analysis is evaluated based on DNNs performance relative to human coding. Human annotators were asked to assess the sentiment expressed in our datasets using the same guideline, so that each textual record would have two independent human classifications. We have chosen annotators skilled of sentiment analysis, uniformly trained and provided with a straight set of rules to classify the textual features as: negative, neutral, or positive. The raters did not know the other rater's classifications. The DNNs performance has been evaluated against human raters based on the following measures:

- Three measures typically used in deep learning-based classification: accuracy, precision, and recall (LeCun et al. 2015).
- The Cohen coefficient (κ) is a statistical indicator that measures the agreement between two categorical decisions. κ varies between -1 and 1 . $\kappa < 0$ indicates absence of concordance while $\kappa = 1$ indicates maximum concordance (Cohen 1960).
- The Kendall coefficient (τ) is a statistical indicator used to measure the ordinal association between two measured quantities. τ varies between -1 and 1 . If $\tau = 1$, the agreement between the two measures is perfect. $\tau = -1$ indicates perfect disagreement. $\tau = 0$ indicates the perfect independence of the two measures (Kirilenko et al. 2018).
- The ratio of opposite classifications (O) means that one rater giving positive and the other giving negative classification of the same item to the total number of items.

The human raters' performance has been evaluated with the same measures applied to their individual classification vectors.

4 Results

This section reports the results of the experiments conducted on our dataset. The experiments relied only on tweets that received both annotators' agreement on the overall visual and textual sentiment. By removing tweets with ambiguous sentiment, we increase the quality of the dataset and ensure the validity of the experiments.

Moreover, the DNNs have been tested using 3 versions: static, non-static and random. However, only the results of the best approach (non-static) were put in place, using pretrained "word2vec" vectors. It carries out a transfer learning procedure: it starts with pretrained vectors and refines them with training using our dataset. Instead, in the other two cases: the static is very dependent by the old dataset; while in the random, for the training the vectors are randomly initialized.

The four DNNs architectures have been chosen by testing different combinations of parameters and taking into consideration, and the best performers are: Word-CNN, Word-LSTM, Char-CNN and ConvNets. The variable parameters tested are the maximum characters length of tweets and the size of the dictionary (for word-based DNNs) or alphabet (for characters-based DNNs). The full length of a tweet on Twitter is 280 characters. Most of the time, tweets with this length contain more periods, and it is usually from the first that we already recognize the sentiment of the whole tweet. The experiment performed demonstrates how much the grading performance strictly depends on the length. It has been observed that some tweets are composed of several periods in which sentiment is deduced in the first, the other periods contain opinions to support. Therefore, we conducted more experiments by reducing the maximum length of tweets.

Depending on whether the architecture was words or characters based, we have performed further experiments testing the dictionary size or the alphabet variation, bearing in mind that more extensive dictionaries with words showing lower frequency can be counterproductive to the sentiment classification. For each testing phase, we evaluated the performance in terms of accuracy, precision, recall and F1-score.

Table 2 reports the results of the experiments performed. In particular, for the *Word-CNN*, it is possible to notice that considering the same length of the sentence in characters, a larger dictionary, decreases the accuracy. This trend can be justified by the fact that words repeated less frequently in tweets have lower discriminative power. As a consequence, the additional vectors of features, associated with these words, extracted from the word embedding layer and added to the input of the subsequent layers, cause a degradation of the performance. On the other hand, limited dictionaries does not allow for collecting enough discriminative words. This is due to the dilution effect of the “stopwords”, such as articles, prepositions, conjunctions, with lower discriminative effectiveness amongst the most frequent words. The truncation to the first N tweets determines a slight increase in the final accuracy. This trend may be since to identify the sentiment of a tweet the informative content of a limited part of the text is enough. On the other hand, excessive truncation causes the loss of too much information content, leading to the wrong classification of the sentiment.

The results obtained for the *Word-LSTM* show a larger dictionary increases the accuracy of sentences with the same length, conrarily to what happens with CNN. From this point of view, it seems that a Recurrent Neural Network can identify less discriminative vectors of features because they are associated with words that are either too frequent (stopword) or not very frequent. In general, the results of the Recurrent Neural Network are lower than those of CNN.

For the *Char-CNN* and *ConvNets*, the results demonstrate that with the same length of the sentence in characters, bigger alphabets increase accuracy, contrary to what happens with dictionary-based CNNs. In this case, the truncation of the alphabet by eliminating the less used characters reduces the accuracy. In this case, the truncation to the first N characters of a tweet does not seem to contribute to a significant changing in accuracy, unless the truncation is excessive to eliminate too much information content that leads to the wrong classification of the sentiment. Then,

the better result is obtained for a character-based network than a dictionary-based network. This aspect is because tweets are a source of typos and intentional alteration of words to emphasize content. This is why character-by-character analysis can be more robust since, in the convolutional layers, a kernel could filter out these artefacts. Unlike a dictionary-based approach, these artefacts are considered different words.

Since sentiment estimation is a subjective task where different persons may assign different sentiments to tweets, inter-annotator-agreement is a common approach to determine the reliability of a dataset and the difficulty of the classification task. We have used more robust agreement indices such as the Cohen coefficient (κ) and the Kendall coefficient (τ) (Kirilenko et al. 2018). We calculated κ , τ and O to measure the agreement between the manual annotation (ground truth) and the results of the DNNs classification. The raters had no knowledge of the other rater's classifications. The results are summarized in Table 3. Moreover, these results concerning the different inter-annotator agreement measurements show that on average the 4 approaches have obtained good results: in fact, they always reach values above 0.6, both in terms of Cohen coefficient and Kendall coefficient. The best approach was Char-CNN, with Kendall coefficient higher than 0.76, Cohen coefficient reaching 0.8, and finally with ratio of opposite classifications not exceeding 0.023.

5 Discussion and implications

We exploit the geographical information to improve the understanding of the analysis results. In such a way, we have inferred the semantic meaning of a tweet with the geographic location, to get to at the sentiment-based assessment. Visualization of data is essential since it allows the users to have more accessible and understandable complex data. To display information clearly and efficiently, tools used for data visualization are: statistical graphics, plots, information graphics and other. However, numerical data use dots, line, bars, cartogram to visualize a piece of quantitative information. Geo-localized and classified tweets can be represented on a map using the Qgis software (<https://www.qgis.org/it/site/>). A geodatabase containing tweets has explicitly been created. In particular, after being classified, tweets have been exported as a table with the same fields of the original mongoDB table, by adding sentiment field that contains the sentiment for each tweet: positive, negative and neutral. Moreover, it is possible to load a geographic map and overlap its with points that identify tweets that were posted in that place. Longitude and latitude indicate the coordinate of each tweet. Different colours are used to differ the sentiment. Figure 2a shows tweets posted in the area of Campania region. The five macro-regions of Cilento are shown in purple, while the remaining part of the Campania region is shown in green. Different colours are used to highlight different sentiment of tweets: Green for Positive, Yellow for Neutral, Red for Negative.

The geodatabase also allows performing queries such as the average sentiment of tweets in a confined and irregular area like a region. Then the result can be represented by a region's map by a colour ramp from green to red when the tweets carry positive or negative sentiments respectively. Figure 2b shows the map that indicates

Table 3 Comparison of agreement between two human and DNNs of the tweet in the dataset of tweets related to Cilento

DNNs	Parameters		Inter-Annotator agreement		
	Sentences length	Dict/Alph length	κ	τ	O
Word-CNN	128	1000	0.664	0.724	0.030
	128	3000	0.666	0.719	0.034
	128	5000	0.651	0.699	0.038
	128	10000	0.664	0.697	0.042
	200	1000	0.669	0.731	0.029
	200	3000	0.650	0.706	0.034
	200	5000	0.649	0.707	0.033
	200	10000	0.671	0.699	0.044
	280	1000	0.660	0.720	0.031
	280	3000	0.664	0.717	0.035
	280	5000	0.640	0.690	0.040
	280	10000	0.675	0.706	0.041
	Word-LSTM	128	1000	0.625	0.676
128		2000	0.626	0.677	0.043
128		3000	0.627	0.676	0.044
128		5000	0.627	0.674	0.043
128		10000	0.597	0.646	0.050
200		1000	0.637	0.688	0.040
200		2000	0.626	0.679	0.042
200		3000	0.627	0.686	0.039
200		5000	0.628	0.680	0.041
200		10000	0.586	0.629	0.058
280		1000	0.633	0.688	0.040
280		2000	0.635	0.696	0.034
280		3000	0.633	0.687	0.039
280		5000	0.636	0.699	0.035
280		10000	0.586	0.646	0.049
Char-CNN	128	70	0.765	0.811	0.018
	128	100	0.770	0.808	0.021
	128	200	0.769	0.808	0.021
	128	345	0.769	0.809	0.020
	200	70	0.760	0.802	0.020
	200	100	0.769	0.805	0.022
	200	200	0.771	0.805	0.023
	200	345	0.767	0.806	0.021
	280	70	0.765	0.808	0.019
	280	100	0.763	0.798	0.023
	280	200	0.771	0.808	0.021
280	345	0.767	0.808	0.020	

Table 3 (continued)

DNNs	Parameters		Inter-Annotator agreement		
	Sentences length	Dict/Alph length	κ	τ	O
ConvNets	128	70	0.641	0.683	0.045
	128	100	0.653	0.693	0.044
	128	200	0.679	0.724	0.037
	128	345	0.664	0.699	0.041
	200	70	0.665	0.706	0.039
	200	100	0.674	0.713	0.037
	200	200	0.675	0.716	0.037
	200	345	0.669	0.709	0.040
	280	70	0.689	0.735	0.032
	280	100	0.682	0.721	0.036
	280	200	0.697	0.731	0.036
	280	345	0.698	0.732	0.035

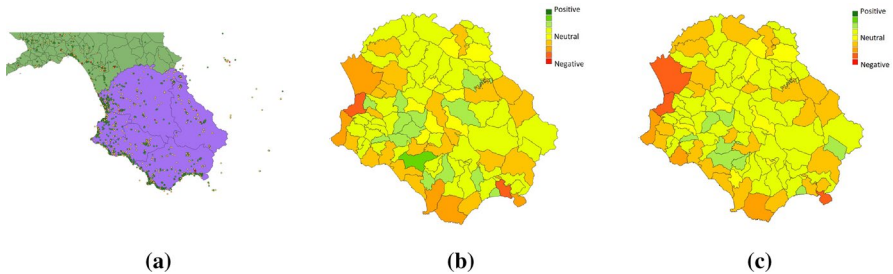


Fig. 2 Data visualization of the sentiment analysis of a geotagged tweet. **a** The representation of tweets posted in Campania region, **b** the map with the average sentiment of the tweets sent by each municipality of Cilento, and **c** the map with the average sentiment of the tweets concerning a place within each municipality of Cilento

the average sentiment of the tweets sent by each municipality of Cilento. The choice of a nonlinear colour scale is to emphasize the negative tweets. If a linear scale had been used the negative sentiment would not have been perceived due to the lesser number of negative tweets compared with positive and neutral ones.

Figure 2b shows the map that indicates the average sentiment of the tweets sent by each municipality of Cilento. For example, if a municipality is red, it means that all tweets sent from that region had a strictly negative average sentiment. Figure 2c shows a map depicting the average sentiment of the tweets concerning a place within

each municipality of Cilento. For example, if a municipality is red, it means that all the tweets in the dataset talking about that region have a strictly negative average sentiment. The colour scale has been chosen nonlinear to emphasize the weight of negative tweets.

More in deep, the developed graph shows the existing relations among different macro-regions. Green (positive), yellow (neutral) and red (negative) circles represent tweets talking about the specific macro-region; the dimension is proportional to the number of tweets. The arches represent instead the “direction” of the tweets, where the thickness is proportional to the number of tweets. The graphical representation shows directly and immediately the results of the analysis. It emerged that, for all macro-regions it is common to talk about itself, mainly with positive sentiment. Conversely, tweets oriented towards other macro-region are less frequent and, when it occurs, with a neutral sentiment. However, there are cases of neighbouring macro-regions, where one speaks a lot about the other with negative sentiment. Thus, the content analysis of tweets is not only functional for obtaining the sentiment, instead it is aimed at understanding which type of tourist experience or destination the user refers to at the time of the publication on Twitter. The results discussed in this study have some practical implications. We believe that the mapping of tourist sentiment can allow both destination managers and policymakers to obtain valuable information about the online reputation of the destination, conceived as a brand. According to Ghafari et al. (2017) the destination brand reputation significantly influences the brand image and brand loyalty. Moreover, they found that the destination brand reputation is one of the dimensions of the overall destination brand equity (Keller et al. 2008). This is also consistent with what Mazurek (2019) reviewed, who highlighted the importance of properly managing the brand reputation of a tourist destination to improve its attractiveness. Based on our results, except in some critical locations where the presence of negative tweets is more significant, the presence of mainly neutral or positive tweets could help improve the online destination brand reputation. Further investigations have been devoted to compare the tweets about inland area and the coast to deduce essential insights and statistics of the more frequent keywords used in tweets. This analysis is based on different time slots of 2 h during the day. From these evaluations, it is possible to infer that during the summer period, the time slots that present a positive sentiment concern the emotions caused by the landscapes at sunset and sunrise. The most frequent words are sunset and sunrise. Negative sentiment, on the other hand, is generated by weather conditions and environmental instability. The most frequent words were “piove”, “abbandonati”, “rifiuti”. During the winter period, the most frequent words are linked to the Christmas holidays, such as “Christmas”, “cenadinatale”, “auguridinatale”. The negative sentiment of that period is linked to scams, especially in holiday homes. Figures 3 and 4 depict some statistics and compare the sentiment related to the tweets about the inland and the coast area.

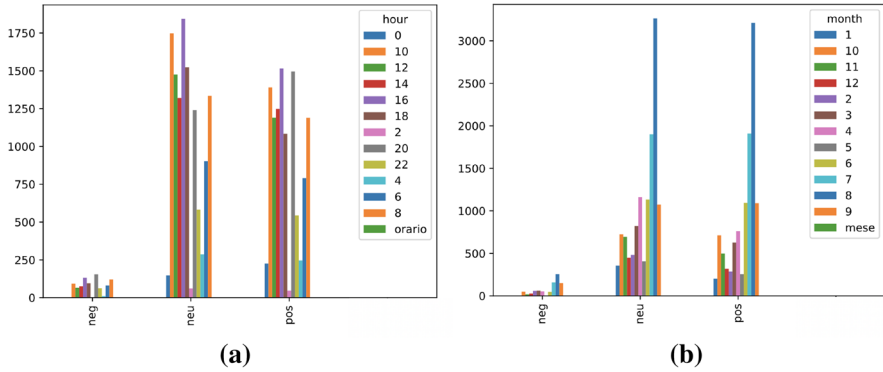


Fig. 3 Sentiment comparison of the tweets related to the inland area. Figure 3a depicts the sentiment based on hour analysis

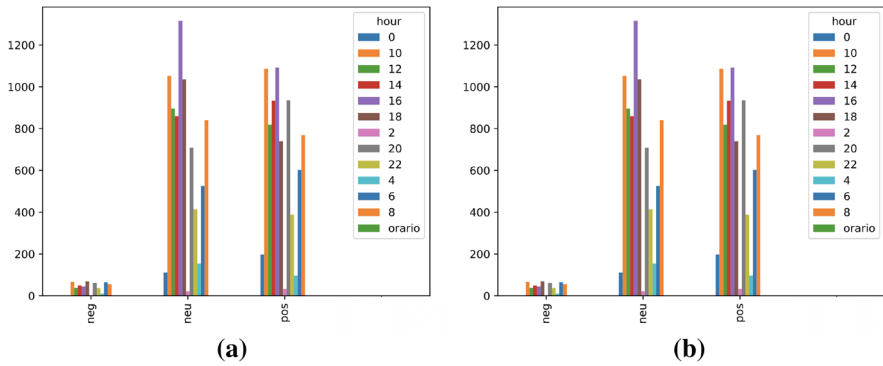


Fig. 4 Sentiment comparison of the tweets related to the coast area. Figure 3a depicts based on month analysis

6 Conclusion and future works

Sentiment analysis of social media content represents a challenging but rewarding task enabling companies to gain deeper insights into tourist flows. A deep learning approach has been introduced that classifies the sentiment of tweets related to Cilento a regional tourist attraction in Southern Italy. Four trained DNNs identify the sentiment of a tweet. The experiments on a newly collected dataset yield high accuracy and demonstrate the effectiveness and suitability of our approach. This last is useful to collect and investigate the spatial, temporal and demographic features of tourist flows, enables relatively sophisticated descriptions of tourist movement, as well as the demographic profiles of tourist groups. Thus, through the use of these insights, tourism operators, as well as regional policymakers, could adopt a wider vision of the tourists perceptions. In parallel, as the online user-generated content plays an important role in planning trips and making decisions about destinations

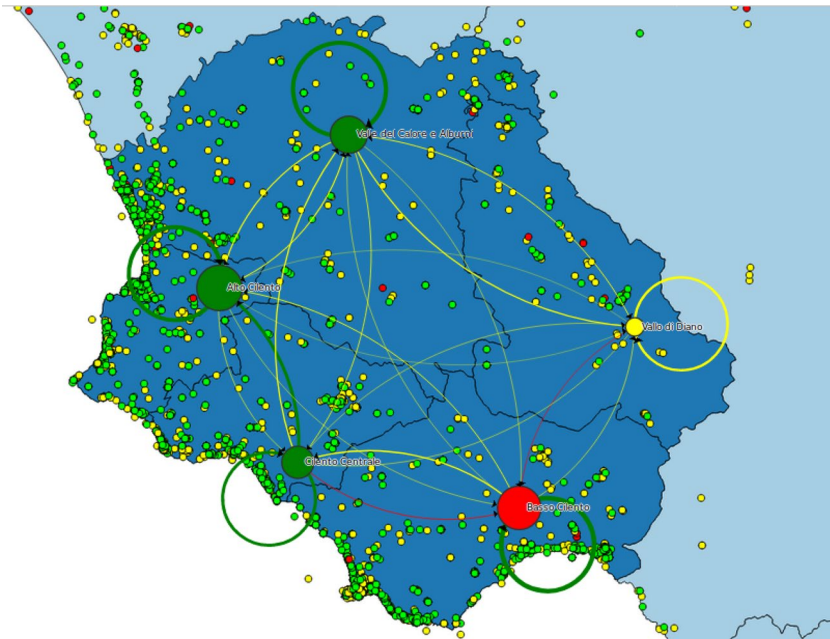


Fig. 5 Graphical representation of the existing relations among different macro-regions

(Litvin et al. 2008; Yoo and Gretzel 2011), they contribute to shape the brand reputation of a given tourism destination. Moreover, it is important to underline that focusing only on the data provided by sentiment analysis could be reductive or even misleading compared to the real scenario. A more accurate analysis is, therefore, necessary which should also include the content of each tweets of the sample. With this in mind, the continuous monitoring of sentiment, combined with content analysis, would allow tourism players to make better marketing decision (Valdivia et al. 2019). As future development, we plan to deepen this analysis with the aim of understanding, through a further content analysis, the reasons behind a given sentiment. We believe that this approach could contribute to improving the traditional customer satisfaction or dissatisfaction models applied to the tourism industry, which can provide only a partial view of the whole tourism experience (Alegre and Garau 2010) (Fig. 5).

Further investigation will be devoted to improving our approach by employing a larger dataset and extracting additional informative features. Moreover, we will extend the evaluation by comparing the proposed DNNs with other existing systems for textual sentiment analysis. In our work it was interesting to test the approaches using a predominantly bilingual dataset to test the generalizability of the methods. Another step will be the test of the same approaches with datasets containing only one language.

Funding Open Access funding provided by Università Politecnica delle Marche.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adwan O, Al-Tawil M, Huneiti A, Shahin R, Zayed AA, Al-Dibsi R (2020) Twitter sentiment analysis approaches: a survey. *Int J Emerg Technol Learn* 15(15):79–93
- Alaei AR, Becken S, Stantic B (2019) Sentiment analysis in tourism: capitalizing on big data. *J Travel Res* 58(2):175–191
- Alegre J, Garau J (2010) Tourist satisfaction and dissatisfaction. *Ann Tour Res* 37(1):52–73
- Baccianella S, Esuli A, Sebastiani F (2010) Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. *Lrec* 10:2200–2204
- Bengio Y et al (2009) Learning deep architectures for AI. *Found Trends Mach Learn* 2(1):1–127
- Cambria E, White B (2014) Jumping NLP curves: a review of natural language processing research. *IEEE Comput Intell Mag* 9(2):48–57
- Chua A, Marcheggiani E, Servillo L, Moere AV (2014) Flowsampler: visual analysis of urban flows in geolocated social media data. In: International conference on social informatics, pp 5–17. Springer
- Chua A, Servillo L, Marcheggiani E, Moere AV (2016) Mapping cileto: using geotagged social media data to characterize tourist flows in southern italy. *Tour Manag* 57:295–310
- Cluster WB, Cooper M, Sallis P (2010) Thailand–tourism and conflict: Modeling sentiment from twitter tweets using naïve bayes and unsupervised artificial neural nets. In: 2010 second international conference on computational intelligence, Modelling and Simulation, pp 89–94. IEEE
- Cohen J (1960) A coefficient of agreement for nominal scales. *Educ Psychol Meas* 20(1):37–46
- Da Silva NF, Hruschka ER, Hruschka ER Jr (2014) Tweet sentiment analysis with classifier ensembles. *Decis Support Syst* 66:170–179
- Dos Santos C, Gatti M (2014) Deep convolutional neural networks for sentiment analysis of short texts. In: Proceedings of COLING 2014, the 25th international conference on computational linguistics: technical papers, pp 69–78
- Ghafari M, Ranjbarian B, Fathi S (2017) Developing a brand equity model for tourism destination. *Int J Bus Innov Res* 12(4):484–507
- Gonzalo AR, Pablo AH, Aldo M (2020) Sentiment analysis of twitter data during critical events through Bayesian networks classifiers. *Future Gener Comput Syst* 106:92–104
- Hagen M, Potthast M, Büchner M, Stein B (2015) Twitter sentiment detection via ensemble classification using averaged confidence scores. In: European conference on information retrieval, pp 741–754. Springer
- Jianqiang Z (2016) Combing semantic and prior polarity features for boosting twitter sentiment analysis using ensemble learning. In: 2016 IEEE first international conference on data science in cyberspace (DSC), pp 709–714. IEEE
- Jianqiang Z, Xiaolin G, Xuejun Z (2018) Deep convolution neural networks for twitter sentiment analysis. *IEEE Access* 6:23253–23260
- Jianqiang Z, Xueliang C (2015) Combining semantic and prior polarity for boosting twitter sentiment analysis. In: 2015 IEEE international conference on Smart City/SocialCom/SustainCom (Smart-City), pp 832–837. IEEE
- Jurek A, Mulvenna MD, Bi Y (2015) Improved lexicon-based sentiment analysis for social media analytics. *Secur Inform* 4(1):1–13

- Keller KL, Parameswaran M, Jacob I (2008) Strategic brand management: building, measuring and managing
- Kim Y (2014) Convolutional neural networks for sentence classification. In: [arXiv:1408.5882](https://arxiv.org/abs/1408.5882) (arXiv preprint)
- Kim Y, Jernite Y, Sontag D, Rush AM (2016) Character-aware neural language models. In: Thirtieth AAAI conference on artificial intelligence
- Kirilenko AP, Stepchenkova SO, Kim H, Li X (2018) Automated sentiment analysis in tourism: comparison of approaches. *J Travel Res* 57(8):1012–1025
- Kiritchenko S, Zhu X, Mohammad SM (2014) Sentiment analysis of short informal texts. *J Artif Intell Res* 50:723–762
- LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521(7553):436
- Li W, Guo K, Shi Y, Zhu L, Zheng Y (2018) DWWP: domain-specific new words detection and word propagation system for sentiment analysis in the tourism domain. *Knowl-Based Syst* 146:203–214
- Lim WL, Ho CC, Ting CY (2020) Sentiment analysis by fusing text and location features of geo-tagged tweets. *IEEE Access* 8:181014–181027
- Litvin SW, Goldsmith RE, Pan B (2008) Electronic word-of-mouth in hospitality and tourism management. *Tour Manag* 29(3):458–468
- Mazurek M (2019) Brand reputation and its influence on consumers behavior. *Contemp Issues Behav Financ* 20:45–52
- Mikolov T, Chen K, Corrado G, Dean J (2013a) Efficient estimation of word representations in vector space. [arXiv:1301.3781](https://arxiv.org/abs/1301.3781) (arXiv preprint)
- Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J (2013b) Distributed representations of words and phrases and their compositionality. *Adv Neural Inf Process Syst* 20:3111–3119
- Miller GA (1995) Wordnet: a lexical database for English. *Commun ACM* 38(11):39–41
- Montejo-Ráez A, Martínez-Cámara E, Martín-Valdivia MT, Ureña-López LA (2014) A knowledge-based approach for polarity classification in Twitter. *J Assoc Inf Sci Technol* 65(2):414–425
- Moreno-Ortiz A, Salles-Bernal S, Orrequia-Barea A (2019) Design and validation of annotation schemas for aspect-based sentiment analysis in the tourism sector. *Inf Technol Tour* 21(4):535–557
- Neidhardt J, Rümmele N, Werthner H (2017) Predicting happiness: user interactions and sentiment analysis in an online travel forum. *Inf Technol Tour* 17(1):101–119
- Paltoglou G, Thelwall M (2012) Twitter, myspace, digg: unsupervised sentiment analysis in social media. *ACM Trans Intell Syst Technol* 3(4):66
- Paolanti M, Kaiser C, Schallner R, Frontoni E, Zingaretti P (2017) Visual and textual sentiment analysis of brand-related social media pictures using deep convolutional neural networks. In: International conference on image analysis and processing, pp 402–413. Springer
- Poria S, Cambria E, Bajpai R, Hussain A (2017) A review of affective computing: From unimodal analysis to multimodal fusion. *Inf Fusion* 37:98–125
- Saif H, He Y, Alani H (2012) Semantic sentiment analysis of twitter. International semantic web conference. Springer, Berlin, pp 508–524
- Saleena N et al (2018) An ensemble classification system for twitter sentiment analysis. *Proced Comput Sci* 132:937–946
- Serna A, Gerrikagoitia JK, Bernabé U (2016) Discovery and classification of the underlying emotions in the user generated content (UGC). In: Information and communication technologies in tourism 2016. Springer, pp 225–237
- Sun S, Luo C, Chen J (2017) A review of natural language processing techniques for opinion mining systems. *Inf Fusion* 36:10–25
- Sutskever I, Vinyals O, Le, QV (2014) Sequence to sequence learning with neural networks. In: Advances in neural information processing systems, pp 3104–3112
- Tang D, Wei F, Qin B, Liu T, Zhou M (2014) Coooollll: a deep learning system for twitter sentiment classification. In: Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014), pp 208–212
- Valdivia A, Hrabova E, Chaturvedi I, Luzón MV, Troiano L, Cambria E, Herrera F (2019) Inconsistencies on tripadvisor reviews: a unified index between users and sentiment analysis methods. *Neurocomputing*
- Wu D, Cui Y (2018) Disaster early warning and damage assessment analysis using social media data and geo-location information. *Decis Support Syst* 111:48–59

- Yan Y, Chen J, Wang Z (2020) Mining public sentiments and perspectives from geotagged social media data for appraising the post-earthquake recovery of tourism destinations. *Appl Geography*. <https://doi.org/10.1016/j.apgeog.2020.102306>
- Ye Q, Zhang Z, Law R (2009) Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. *Expert Syst Appl* 36(3):6527–6535
- Yoo KH, Gretzel U (2011) Influence of personality on travel-related consumer-generated media creation. *Comput Hum Behav* 27(2):609–621
- Zhang X, Zhao J, LeCun Y (2015) Character-level convolutional networks for text classification. In: *Advances in neural information processing systems*, pp 649–657
- Zhang Z, Ye Q, Zhang Z, Li Y (2011) Sentiment classification of internet restaurant reviews written in Cantonese. *Expert Syst Appl* 38(6):7674–7682

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.