Tourists' Perceptions of Climate: Application of Machine Learning to Climate and Weather Data from Chinese Social Media

Y. G. TAO,^a F. ZHANG,^a W. J. LIU,^a AND C. Y. SHI^b

^a School of History Culture and Tourism, Jiangsu Normal University, Xuzhou, Jiangsu Province, China

^b School of Geography, Geomatics and Planning, Jiangsu Normal University, Xuzhou, Jiangsu Province, China

(Manuscript received 1 March 2021, in final form 24 June 2021)

ABSTRACT: Understanding tourists' perceptions of climate is essential to improving tourist satisfaction and destination marketing. This paper constructs a sentiment analysis framework for tourists' perceptions of climate using not only continuous climate data but also short-term weather data. Based on Chinese social media platform Sina Weibo, we found that Chinese tourists' perceptions of climate change were at an initial stage of development. The accuracies of word segmentation between sentiment and nonsentiment words using ROST content mining (CM), BosonNLP, and GooSeeker were all high, and the three gradually decreased. The positively expressed sentences accounted for 79.80% of the entire text using ROST emotion analysis (EA), and the sentiment score was 0.784 at the intermediate level using artificial neural networks. The results indicate that the perceived emotional map is generally consistent with the actual climate and that cognitive evaluation theory is suitable to study text on climate perception.

KEYWORDS: Atmosphere; Weather modification; Machine learning

1. Introduction

Climate and weather are important considerations for tourists and have an important influence on the global tourism sector (Scott et al. 2008; Dwyer et al. 2009; Gössling et al. 2016; Mushawemhuka et al. 2018). As early as 1991, a British survey found that 73% of respondents considered "good weather" to be the main reason for overseas tourism (Mintel International Group 1991). Changes in weather patterns as a result of climate change will have widespread consequences for tourism demand (Hübner and Gössling 2012). Climate and weather affect tourists' satisfaction and intentions to revisit (Scott et al. 2008; Kim et al. 2017; Jeuring 2017; Caldeira and Kastenholz 2018). Curnock et al. (2019) also analyzed the relationship among tourists' emotions, threat perception, and values related to their attitude toward climate change. According to TripAdvisor reviews, a tourism destination's climate and weather influence tourists' selection and enjoyment of that destination (Fitchett and Hoogendoorn 2018). Whereas a growing number of studies address the potential impacts of climate change on tourism, little is known about how the tourism sector accesses, uses, and analyzes the available weather and climate information (Nalau et al. 2017). Throughout the existing research on tourists' climate and weather perception, the data sources primarily come from field surveys, and online reviews are less frequently used.

In recent years, user-generated content (UGC) provided by tourists on social media has become an important source of information that affects tourism development (Capatina et al. 2018). Relative to traditional survey data, big data sources such

as UGC have gained popularity in tourism research due to their large scale, diversity, high speed, authenticity, initiative, and low cost (IBM 2016; Fuchs et al. 2014; Alaei et al. 2017). Summarizing public opinions based on social media data and conducting sentiment analysis on these data provide a new perspective for research on tourism satisfaction, such as tourists' experiences of destination environments (Kirilenko et al. 2017; Becken et al. 2017). Moreover, social media websites can be used as a data source for mining public opinion on a variety of subjects, including climate change (Dahal et al. 2019). Furthermore, human behavior in the domain of social media is affected by weather (Zimmerman and Kruschwitz 2017). However, few studies have focused on tourists' perceptions of climate or weather (Fitchett and Hoogendoorn 2018). Sentiment analysis of online reviews may provide new data and methods for researching tourists' perceptions of climate and weather.

China has diverse climates and weather that range from frigid to tropical (Zhang 1991). Most areas are currently negatively affected by climate change. In 2018, the number of domestic tourists in China was 5.539 billion, and the number of outbound tourists was 150 million, ranking first in the world (Finance Department of the Ministry of Culture and Tourism 2019). In recent years, the number of monthly active users on China's social media platform Sina Weibo has surpassed that of Twitter, making Sina Weibo the largest independent social media company in the world (Sina Technology 2017). Moreover, tourists enjoy sharing their travel experiences on social media (Alaei et al. 2017). Hence, China has abundant social media data with which to study tourists' perceptions of climate.

Denotes content that is immediately available upon publication as open access.

Corresponding author: Y. G. Tao, taoyuguo123@163.com

Publisher's Note: This article was revised on 6 October 2021 to replace Fig. 3 with a revised version that should have been included when originally published.

DOI: 10.1175/WCAS-D-21-0039.1

© 2021 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

This study mainly uses the machine-learning method of the semisupervised artificial neural network (ANN) to analyze tourists' sentiment about climate perception. The analysis is based on climate- and weather-related comments on Sina Weibo concerning 229 representative 5A-class scenic spots in mainland China. This study attempts to address the following questions. By relying on cognitive evaluation theory and an emotion analysis method, we construct a framework that uses climate (continuous and stable) and weather (short-term and erratic) data to analyze tourists' perceptions of climate. Given the various values attributed to environmental issues due to tourists' diverse cultural backgrounds, discrepancies in perceptions of climate may exist (Scott et al. 2008; Wilkins et al. 2017); therefore, based on microscopic data of the scenic area, what characteristics and patterns do the maps of tourists' climate perception at the macro- to mesolevels have? Do these perceptions reflect the climate of the entire region, including the destinations?

2. Literature review

a. Tourists' perceptions of climate and weather

The climate is continuous and stable, whereas the weather is short term and erratic (U.S. Environmental Protection Agency 2013). During approximately the past 30 years, tourists' perceptions of climate and weather have been a focus of academic attention. Climate and weather are important factors in shaping destination image and the choice of holiday destinations (Lohmann and Kaim 1999; Gössling and Hall 2006; Gössling et al. 2016). Tourists believe that the impact of weather on their experience may vary in different scenarios. For example, in Scandinavia and in the Arctic, most tourists consider aspects of weather to have relatively small behavioral impacts (Denstadli et al. 2011; Denstadli and Jacobsen 2014), whereas tourists in Lisbon believe that meteorological conditions have a considerable impact on tourists' activities.

Tourists have a certain preference for climate and weather conditions, and this preference may vary with sociodemographic, place of origin, and other variables. Tourists typically prefer clear skies and dislike low-visibility weather (Forland et al. 2013). However, studies conducted in the Alps show that nature-oriented tourists do not care about the amount of sunshine (Pröbstl-Haider et al. 2015). Weather preferences and tolerances deviate between high and low Arctic destinations, illuminating subjective ideas of what constitutes acceptable weather (Jacobsen et al. 2011). The similarities and differences among Canadian, New Zealand, and Swedish tourists' perceptions of climate conditions were investigated (Scott et al. 2008). For tourists in Melbourne, Australia, and Warsaw, Poland, factors such as the interviewees' origin affected tourists' preference for high temperature conditions (Lindner-Cendrowska and Blazejczyk 2016; Lam et al. 2018). International tourists were more resilient to the weather than domestic tourists visiting the U.S. Great Lakes (Rutty and Scott 2016). The young segment of the market was more flexible in adapting to episodes of extremely high temperatures (Gomez-Martin et al. 2014). Fitchett and Hoogendoorn (2018) explored the influence of these factors on tourists' sensitivity to the climate of a destination

using a commentary on climatic factors in TripAdvisor reviews for 19 locations in South Africa.

The climate and weather quality of the destination perceived by tourists is the key factor affecting the overall travel satisfaction and loyalty of many tourists. During summer in Lisbon, Portugal, the maximum air temperature is found to have a significant negative effect on overall satisfaction (Caldeira and Kastenholz 2018). Perceived weather quality directly and indirectly affects tourist satisfaction and revisit intention and is correlated with the perceived quality of physical attributes and services (Kim et al. 2017). For domestic camping tourists in the Netherlands, weather salience is positively related to attitudes toward tourism and holiday satisfaction (Jeuring 2017). Extreme weather events, as the key risk to visitors, are mentioned as the most likely effects of climate change for island destinations in the Caribbean (Hübner and Gössling 2012; De Urioste-Stone et al. 2016). At the zoos in Phoenix, Arizona, and Atlanta, Georgia, visitors seem averse to thermal extremes (Perkins and Debbage 2016). Climate and weather information published in the media affects the travel plans of many tourists (Rutty and Scott 2010).

In a comprehensive framework for studying tourists' perceptions of climate change developed by Gössling et al. (2012), in addition to perceived preferences, extreme climate events, and so on, public perception polarity is also an important issue surrounding tourists' perceptions of climate change impacts. Studies on weather perceptions and demand responses to extreme events also need to consider the emotional states of tourists, and rain and extreme weather events can involve considerable negative emotions (Gössling et al. 2016). However, this research is relatively weak. Automated sentiment analysis of big data in tourism can extract public opinion (Kirilenko et al. 2017); however, in terms of tourists' perception of climate and weather, whether such data can support such an analysis is inconclusive. Climate and weather are closely related, and some studies have combined the two as objects of visitor perception (Rutty and Scott 2010; Hübner and Gössling 2012; De Urioste-Stone et al. 2016). One of the differences between the two is the time scale, and climate represents average weather conditions (Gössling et al. 2012). However, the existing data on tourists' perceptions of climate and weather are basically from their own field survey data. If we can use network weather data to study tourists' perceptions of climate, we cannot only increase the amount of data but also realize the use of short-term data to study continuous problems.

b. Sentiment analysis

In recent years, sentiment analysis has been a major concern of the academic community (Schmunk et al. 2014). As a component of text mining research, sentiment analysis is mainly performed using technology such as natural language processing (NLP) and artificial intelligence (AI) (Witte and Mülle 2006). The purpose of sentiment analysis is to extract the sentiment tendency of a particular object according to unstructured text (García et al. 2012). The method is formulated as a classification problem with binary polarities (positive and negative) or ternary polarities (positive, neutral, and negative). Sentiment analysis methods can be divided into lexicon matching methods and machine-learning methods (Feldman 2013; Chiu et al. 2015). The lexicon matching method calculates

scores based on the sentiment expressed by the words that are restricted by language rules and are extracted based on a predefined dictionary of positive and negative words. The most widely used sentiment dictionaries include WordNet-Affect, SentiWordNet, and SenticNet (Cambria 2016). In the case of the machine-learning method, a training corpus of emotional annotation texts is supplied to the machine-learning algorithms. This method not only assesses the sentiment polarity of emotional keywords but also considers the frequency of the co-occurrence and polarity of other arbitrary keywords (Cambria 2016). Machine-learning approaches rely on statistical models and are useful for analyzing large rather than small amounts of text. In addition, a hybrid approach combining the lexicon matching and machine-learning approaches can comprehensively analyze two sentiment polarities to obtain an overall sentiment polarity (Alaei et al. 2017; Kirilenko et al. 2017).

c. Tourist sentiment analysis

Although research using sentiment analysis can be traced back to the 1970s (Broß 2013), this method began receiving increasing attention from tourism researchers only recently. Most of the research focuses on the method itself, with less emphasis on the use of methods to solve problems (Kirilenko et al. 2017). Researchers represented by Marrese-Taylor analyzed the sentiment polarity of reviews of hotels and tourist destinations with lexicon matching methods based on Chinese and English dictionaries such as WordNet, SentiWordNet, and HowNet (Tan and Wu 2011; Marrese-Taylor et al. 2013; Misopoulos et al. 2014; Liu et al. 2019). According to the application of geometry, probability, and AI, machine-learning methods can be classified into many categories, including support vector machines (SVMs), naïve Bayes (NB), ANNs, decision-tree-based models, ensemble models, boosting models, and others. SVM, NB, and ANN are relatively widely used in tourism research. SVM aims to find the best linear separation between positive and negative sentiment data. NB estimates the probabilities of sentiments based on document attributes. ANN and derivative deep learning algorithms process data through a “neuron” self-organizing network to simulate the biological brain to recognize emotions. Some studies have compared different machine-learning approaches (Claster et al. 2010, 2011, 2013). Other studies utilized supervised SVM and NB to explore the emotional experiences of tourists to European and American destinations based on reviews posted on tourism websites (Ye et al. 2009). A representative example of early studies that used hybrid approaches to tourist sentiment analysis is the study by Waldhör and Rind (2008). Kirilenko et al. (2017) conducted a more comprehensive comparative study using word frequency matching, supervised SVM and NB, and the unique machine-learning method of unsupervised deep learning and showed that great challenges exist, regardless of the approach used.

Most data used in tourist sentiment analysis are obtained from tourism websites such as TripAdvisor and Expedia (Gräbner et al. 2012; Broß 2013; Marrese-Taylor et al. 2013; Schmunk et al. 2014; Misopoulos et al. 2014; Xiang et al. 2015). In comparison with the few studies that extract data from the social media platform Twitter (Claster et al. 2010, 2011, 2013; Shimada et al. 2011), research based on another important social

media site, Sina Weibo, is rare (Li et al. 2018). Furthermore, the primary language of much social media data is English, whereas few data are available in other languages (Alaei et al. 2017; Kirilenko et al. 2017). Some studies have tried to use data from Chinese websites. For example, certain studies have used the lexicon matching method (Tan and Wu 2011; Liu et al. 2019), whereas other studies have used supervised SVM and NB to analyze data from the China National Knowledge Infrastructure (Zheng and Ye 2009; Zhang et al. 2011). Nevertheless, few studies have reported using a supervised or semisupervised ANN.

For the content of tourist sentiment analysis, several studies have discussed part-of-speech tagging. Sentiment analysis addresses the problem of polarity classification, whether binary or ternary. Binary classification initially assumes customer reviews to be subjective, thereby determining the sentiment polarity as “positive” or “negative.” That is, sentiment analysis processes only subjective reviews. If a review is objective, the sentiment analysis terminates (Alaei et al. 2017). However, customer reviews can be subjective or objective. The former type is based on opinions, personal feelings, beliefs, and judgments about entities or events. The latter type is based on facts, evidence, and measurable observations (Feldman 2013), which exclude words that are distinctly designated as positive or negative in the dictionary (Alaei et al. 2017). Many studies confirm that objective comments consisting of objective vocabularies and descriptions such as family are also parts of sentiment analysis. Adjectives, nouns, verbs, and adverbs should all be considered in sentiment analysis, which reflects the quality of customers’ emotional experiences. Therefore, a binary classification must be extended to a ternary classification that involves a third “objective” category (Kahn et al. 2007; Pang and Lee 2008). The vocabularies thus constitute different text sizes at the word, phrase, sentence, and paragraph levels (Hu et al. 2017). In ternary classification, the classifier implicitly performs a classification to distinguish objective and subjective sentences, providing labels analogous to “positive,” “negative,” or “neutral.” Therefore, sentiment analysis considers both subjective and objective comments, forming a unity of subjectivity and objectivity (Kahn et al. 2007; Alaei et al. 2017; Kirilenko et al. 2017). Although it is challenging to find implicit emotions in an objective comment such as “It took an hour to receive our luggage!,” ternary classification can be an important goal to pursue in future sentiment analysis (Schmunk et al. 2014; Alaei et al. 2017).

In addition to part-of-speech tagging, most studies using sentiment analysis have focused on restaurants or catering (Zhang et al. 2011; Gräbner et al. 2012; Broß 2013; Marrese-Taylor et al. 2013; Xiang et al. 2015), and a few studies have examined destinations (Ye et al. 2009; Shimada et al. 2011; Claster et al. 2010, 2011, 2013; Liu et al. 2019). Apart from general analyses of restaurants, catering services, and destinations, some studies have attempted monographic studies of hotel catering services, room facilities, and locations (Chiu et al. 2015). Other studies have further explored themes such as destination environments and festivals (Becken et al. 2017; Saura et al. 2018a,b; Kirilenko et al. 2017). Based on data from Twitter, Becken et al. (2017) used the lexicon matching method to perform sentiment analysis on reviews of the Great Barrier

FIG. 1. Sentiment analysis framework of tourists' climate perception using weather and climate reviews.

Reef in Australia. Saura et al. (2018a) conducted a sentiment analysis of reviews of Spanish hotels using unsupervised SVM. Some studies have demonstrated that using valuable climate change data posted on Twitter for sentiment analysis to understand public attitudes is an area worthy of further research (Kirilenko et al. 2015). However, in-depth studies that use social media data to analyze tourists' perceptions of climate as a natural environmental element are scarce.

3. Research methods and data source

a. Research methods

Emotion is an experience tendency influencing tendency perception (Arnold 1960; Lazarus and Folkman 1984). Therefore, this paper establishes a tourist sentiment analysis framework for climate perception in which the evaluation includes two processes: emotional subject evaluation and text reevaluation. Behind the evaluation is a classification system composed of positive and negative and subjective and objective words or reviews. As compared with the few climate reviews of tourist destinations, there are many weather reviews. Considering that climate represents the average weather status (Gössling et al. 2012), the sentiment analysis of reviews on the average weather over a long period is arguably an analysis of climate. Continuous and stable results can be obtained through the sentiment analysis of a large number of different weather reviews, and the purpose of using weather reviews to study climate perception is achieved (Fig. 1). Therefore, through sentiment analysis of spot weather and climate reviews, we can provide new and abundant data and methods for climate perception research and achieve the purpose of integrating micro- and macrorsearch.

To conduct content analysis, we relied on the widely used dictionary-based ROST content mining (CM) Chinese and English word segmentation technology and two applied machine-learning algorithm-based Chinese word segmentation techniques, GooSeeker and BosonNLP, to analyze word frequency in the comments. We also used the semantic network graphs generated by ROST CM to examine and compare the content and structure of comments about different destinations.

ROST emotion analysis (EA) uses a sentiment dictionary-based method to analyze the proportion of sentences with different polarity in the text. Using the results, we employed the machine-learning method of a semisupervised ANN by editing the Python 3.0 program to call the Boson platform (<https://bosonnlp.com>) based on the Chinese corpus. In other words, we used a hybrid research method based on the application of ANN supplemented by ROST EA. The accuracy of the ANN method is higher. The machine-learning method is based on a 10-million-level training corpus, and the larger the dataset in the training corpus, the higher the accuracy of the sentiment analysis method (Schmunk et al. 2014; Ye et al. 2009). This method not only can diagnose general words such as nouns and verbs but also can analyze proper nouns such as slang and network words. The test showed that the emotional score of the comment "Beijing smog is impossible to avoid" is 0.250, but the emotional score of the comment "Beijing can make people avoid smog" is 0.804. The test result for the comment "A Tour of Shenyang Botanical Garden: Sculptures and Fossils" showed that the sentiment score was 0.938. The above test shows that the model can effectively identify sentences with a high degree of similarity and can also effectively measure the implicit sentiment of sentences. On the sentiment score of 0 to 1, 10 scales are divided isometrically into the following: extremely poor, very poor, poor, somewhat poor, slightly below average, slightly above average, somewhat good, good, very good and excellent.

b. Data source

Sina Weibo was launched in August 2009. In 2013, the number of registered users was 536 million (Sina Technology 2013). Analogous to social media sites such as Facebook and Twitter, it has a large text corpus and is an important data source for sentiment analysis (Confente 2014). This study used the web crawling software "BaZhuaYu" (Octopus) to collect data from Sina Weibo. The collection period was set from 1 August 2009 to 31 August 2018. Given a lack of comments on climate (including climate change), we set the keywords as "weather" or "climate" + "name of the scenic area." We included the following steps to clean the data. First, we deleted

duplicate data. Second, for the year selection, because the numbers of reviews in 2009, 2010, and 2018 were either small or did not include all months and thus affected the calculation of annual or quarterly sentiment scores, we disregarded these years. Third, we removed noisy data. Sina Weibo includes accounts of government agencies and enterprises in addition to those of individuals. Because the dissemination of online public opinion is from individuals, we eliminated information from weather forecasts and advertisements from the government and private enterprises. Fourth, concerning the selection of scenic areas, we eliminated places with fewer than 10 reviews among the 258 scenic spots. The denoising process was mainly completed manually, reflecting the semisupervised nature of this machine-learning method. Although the process was time and labor consuming, it improved the accuracy and provided a great amount of data, generating 58 263 comments for approximately 229 scenic areas from 1 January 2011 to 31 December 2018, totaling 4.44 million words. On average, each scenic area had 254 reviews with a mean length of 76 words.

The number of comments in 2011 increased to 6006, and Weibo was burgeoning. In 2013, the number of reviews reached a record high of 10 950. From 2011 to 2013, Weibo flourished as the number of comments increased by 82%. The plummeting of the number of reviews in 2014 indicates that the distribution of comments had returned to rationality. Nonetheless, the rapid growth in reviews in recent years shows that tourists have paid close attention to the weather and climate. In terms of months, more comments were made from March to October. April and October had the highest volumes of reviews, which were 40.20% and 33.46% higher than the monthly average, respectively. For both provinces and scenic areas, approximately 80% of the desirable places are in the south, and a few are in the north.

4. Results

a. Content analysis

Most existing research on tourist sentiment analysis has considered only subjective sentiment vocabularies such as “happy” and “regret” (Alaei et al. 2017; Kirilenko et al. 2017). Some research has proposed that objective sentiment vocabularies such as “family” and “war” should also be included in sentiment analysis. The former type of vocabulary includes mainly adjectives, whereas most of the latter type are nouns (Kahn et al. 2007). Whether subjective or objective, the vocabularies can be further categorized into positive and negative types. Therefore, the vocabularies used in reviews can be subdivided into five classes: positive subjective, positive objective, negative subjective, negative objective, and nonsentiment. For instance, forest, sunrise, landscapes, and diversity are positive sentiment expressions that describe the quality of the natural environment (Saura et al. 2018b). Low temperatures, high temperatures, precipitation, humidity, and cloud cover are examples of negative sentiment concerning weather conditions (Baylis et al. 2018). However, relevant research

has neither comprehensively calculated the word frequencies of these vocabularies nor conducted a systematic content analysis.

As compared with the two Chinese word frequency analysis software programs (the most widely used ROST CM and the increasingly recognized BosonNLP), GooSeeker can analyze a larger dataset. Therefore, we used GooSeeker to count the high-frequency words concerning Chinese tourists' perceptions of climate for domestic attractions, totaling 58 263 comments. The results show that the frequency fitting curve of the top 100 words is consistent with a power function distribution ($R^2 = 0.93$). The nonsentiment word “weather” ranked first, followed by the positive subjective sentiment words “good,” “happy,” “like,” “suitable,” “comfortable,” and “beautiful,” which ranked 6th, 46th, 57th, 67th, 70th, and 91st, respectively. The negative subjective sentiment words “bad” and “pity” ranked 53rd and 77th, respectively. In terms of both frequencies and rankings, positive subjective sentiment vocabularies predominated over negative subjective sentiment vocabularies, indicating that positive images prevailed among tourists' perceptions of climate. Moreover, the proportions of sentences that included the negative sentiment words “high temperature,” “extreme,” and “warming” that directly reflect climate change appeared in only 0.87%, 0.18%, and 0.10%, respectively. China's climate is characterized by severe weather (Zhang 2016). The low frequencies of negative sentiment words suggest that Chinese tourists paid little attention to climate risk.

We calculated the proportions of the different vocabulary types and analyzed the accuracies of GooSeeker, ROST CM, and BosonNLP. This study examined two provincial-level administrative units located in the Yangtze River delta as the two cases for analysis, with Shanghai as the urban type and Anhui as the rural type (Anhui and Shanghai ranked highest and lowest among the sentiment scores of provincial-level administrative units). Weather-related data (shaded) are much higher than climate-related data (underlined), and the former can significantly increase the scale of the latter (Table 1). Table 1 shows that the proportions of nonsentiment, positive subjective sentiment, positive objective sentiment, negative subjective sentiment, and negative objective sentiment words concerning climate in Shanghai were 88.62%, 5.38%, 4.52%, 0.56%, and 0.92%, respectively. For Anhui, these proportions were 68.38%, 0.83%, 21.91%, 0.00%, and 8.88%, respectively. The similarities between the two groups of proportions are as follows. Most words were nonsentiment. Positive sentiment words appeared considerably more frequently than did negative sentiment words, which is consistent with the finding that tourism reviews are dominated by positive polarity words (Kirilenko et al. 2015; Liu et al. 2019; Kirilenko and Stepchenkova 2017; Yan et al. 2018). Because the proportion of objective sentiment words was high, close attention was required. The differences between the two groups of comments are that there were significantly more subjective sentiment words about Shanghai than about Anhui and that there was a much higher number of objective sentiment words about Anhui than about Shanghai. However, it is nonetheless premature to judge the

TABLE 1. High-frequency words in comments indicating tourists' perceptions of climate and deviations in (left) Shanghai and (right) Anhui (top 50). Here, OB represents the actual number of observations, RO indicates the deviation of ROST CM, GO indicates the deviation of GooSeeker, and BO indicates the deviation of BosonNLP. Two asterisks are used to indicate positive subjective sentiment words, one asterisk indicates positive objective sentiment words, two daggers are used to indicate negative subjective sentiment words, one dagger indicates negative objective sentiment words, and unlabeled words are nonsentimental. Boldface type indicates data that are generally closely related to weather, and italic font indicates data that are generally closely related to climate.

High-frequency word	OB	RO	GO	BO	High-frequency word	OB	RO	GO	BO
Shanghai	985	19	77	23	Huangshan Mountain	3689	178	193	181
Weather	919	44	27	42	Weather	2849	188	23	192
Pearl	704	1	13	1	Scenic areas*	558	281	296	290
Oriental	700	0	9	0	Daytime*	494	1	1	1
The Bund	200	0	31	0	Sunrise*	474	8	30	8
Zoo	197	3	12	2	Cloudy[†]	471	104	471	104
Wild**	179	1	8	0	High temperature[†]	427	7	5	7
Science museum	170	0	49	0	Temperature	363	1	3	2
Animal	113	1	27	4	Low temperature[†]	362	6	6	6
Center	61	0	18	1	Sea of clouds*	355	0	20	0
Unfold	61	0	0	0	Tourism*	347	23	25	23
Night view*	60	1	10	1	Anhui	316	50	40	50
See no	58	3	7	22	Night	309	0	0	0
Evening	52	0	1	0	Forecast	309	1	1	1
Rain[†]	51	0	12	0	Scenic site*	292	3	5	3
Mood	49	3	0	2	Sunset*	264	2	31	2
Place	49	0	0	0	Tianzhu Mountain	234	9	234	10
Happy**	48	2	1	3	Jiuhua Mountain	225	1	57	2
Tomorrow	48	1	1	1	Tourist*	222	2	1	2
First time	45	0	13	16	Tomorrow	221	5	1	7
Time	43	0	1	0	Southwestern	214	0	0	0
Afternoon	43	2	3	3	Light rain*	209	6	8	8
Good weather*	41	5	13	5	Northeastern	178	0	0	0
Then	40	1	0	1	East wind	165	0	0	0
Subway	34	9	2	9	Good weather*	162	76	162	74
Sunny	32	0	0	0	Wuhu	157	4	3	4
Children*	32	2	2	1	Breeze	152	0	0	0
Pity ^{††}	31	0	0	0	Temperature	149	11	11	11
Town God's Temple	31	1	0	1	Heaven*	149	0	1	0
Huangpu River	30	2	0	5	<i>Climate</i>	148	9	13	10
Second	30	8	0	8	Time	147	4	3	4
Forecast	29	0	0	0	Today	144	1	1	2
Photograph	28	0	0	0	Sunny**	139	0	134	0
Mansion	28	2	6	9	Southeastern	132	0	0	0
Sightseeing*	27	2	2	2	Tomorrow	127	2	0	1
So good**	27	2	2	3	Beautiful scenery*	122	2	23	2
Global	26	0	2	0	Night	121	0	0	0
Variety**	26	0	1	0	Rain[†]	116	13	113	13
Sun*	25	1	1	1	Huangshan City	113	1	3	1
Sunshine*	25	6	6	6	Probability	112	0	1	0
Finance	25	0	2	0	Northwestern	111	0	0	0
Reason	25	0	6	0	Scenery*	109	10	13	11
Tourism*	22	5	6	6	North wind	107	0	0	0
Plaza	21	0	1	0	On the mountain	106	26	20	26
Start off	21	0	0	0	Moderate rain[†]	103	11	13	11
Go out	21	3	0	3	Sunshine*	101	0	1	0
One-day tour*	20	0	0	0	Wind direction	97	0	0	0
Happiness**	20	0	0	0	Precipitation	96	1	1	1
Come back	20	0	0	0	Wind speed	91	0	0	0
This time	20	0	0	0	West wind	91	0	0	0
Total	5572	130	372	181	Total	16 658	1047	1967	1070

sentiment levels and the sentiment discrepancies of the two provinces based on the above information.

Table 1 displays the segmentation accuracies of GooSeeker, ROST CM, and BosonNLP, which were 97.67%, 93.32%, and 96.75%, respectively, for Shanghai and 93.71%, 88.19%, and 93.58%, respectively, for Anhui. Both groups thus exhibited a high segmentation accuracy. ROST CM yielded the highest segmentation accuracy, BosonNLP was the second highest, and GooSeeker was the lowest. The main reason that GooSeeker exhibited a low accuracy is that it cannot identify proper nouns such as “Tianzhu Mountain.” Statistics on multiple scenic areas further confirmed the characteristics of high segmentation accuracy and the accuracy rankings of ROST CM, BosonNLP, and GooSeeker. Additionally, because BosonNLP introduces its sentiment dictionary into ROST CM for word frequency analysis, the two methods process the same text size, which is much smaller than that of GooSeeker. The above analysis helps explain why many studies have used ROST CM for Chinese word frequency analysis, including the study conducted by Cong et al. (2014). This analysis can also inform software selection in future research.

b. Sentiment analysis

Among the employed methods, GooSeeker does not have the sentiment analysis function, ROST EA uses the sentiment dictionary-based method to identify the polarity of sentences, and the machine-learning method ANN on the Boson platform can evaluate sentiment scores for each review or for the entire set of reviews labeled with time and location data. This paper uses a hybrid research method based on an ANN supplemented by ROST EA.

1) ROST SENTIMENT ANALYSIS

The results showed that positive, neutral, and negative sentiment sentences accounted for 79.80%, 7.88%, and 12.32% of the comments, respectively. The proportion of positive sentiment sentences was very high, providing support for the previous conclusion that positive expressions dominated the comments. However, we also found that the percentage of negative sentiment sentences was not low. Therefore, careful attention must be given to the data because negative sentiment is more likely than positive sentiment to affect the images of destinations. Neutral comments were the least frequent, indicating that tourists' perceptions of climate tend to reflect clear-cut attitudes based on distinct likes and dislikes. Due to the variety of research objects, it is meaningless to compare these ratios with other results. However, summarizing the proportion of each polarity would help identify the characteristics of the sentences. As with the results of this study, comments about the environments of Spanish hotels, Swiss hotels, and the Great Barrier Reef in Australia expressed more positive than negative sentiment (Saura et al. 2018a,b; Becken et al. 2017). Therefore, we inferred that tourists prefer to share beautiful environments on social media, including food, accommodation, transportation, touring, shopping, entertainment, and other experiences. In addition, considering that under exceptional circumstances, tourists may express more negative than positive sentiment about destinations (Ye et al. 2009),

not all comments on destinations are dominated by positive sentiment.

2) ANN SENTIMENT ANALYSIS

The score of the accuracy was 0.87. Because of differences in the research objects, methods, and data types, the accuracies were widely divergent (Alaei et al. 2017). Early research found that by adjusting the unsupervised NB method to analyze 116 tweets about tourism cities, the accuracy could reach 0.78 to 0.92 (Shimada et al. 2011). Recent studies have shown that when employing supervised SVM, supervised NB, and unsupervised deep learning methods to analyze 762 475 fragments of tweets, the accuracies can be 0.50–0.51, 0.54–0.55, and 0.39, respectively (Kirilenko et al. 2017). Using 0.7 as a threshold for the accuracy of sentiment analysis (Donkor 2014), the accuracy of 0.87 is high in this study, with the number of reviews being 58 263.

(i) National sentiment score

The overall sentiment score of Chinese tourists' perceptions of climate for 5A-class scenic spots in mainland China was 0.784. This score falls in the intermediate level of “good,” which is average. The meaning of 0.784 is the same as the meaning of the ratio of the positive sentiment obtained with ROST EA in the previous section, which is 79.8%. The two datasets mutually support each other's correctness and validity. The results show that the comments were dominated by positive images. The results sustain the conclusion of more positive than negative sentiment vocabularies and are consistent with the relevant research (Kirilenko et al. 2015; Kirilenko and Stephenkova 2017; Liu et al. 2019).

From the perspective of time evolution, the sentiment scores from 2011 to 2018 were 0.804, 0.781, 0.793, 0.806, 0.788, 0.771, 0.766, and 0.777, respectively. All of them were positive, but they demonstrated a significant downward trend in satisfaction rates. The highest value appeared in 2014, and the lowest value was in 2017 (Fig. 2). For the period from 2014 to 2017, when Weibo's comments returned to a rational distribution, the R^2 of the linear regression model reached 0.95, indicating a distinct regularity in this downward trend. In terms of quarterly change, the sentiment scores from the first to the fourth quarter were 0.795, 0.782, 0.777, and 0.790. The season with the highest score differed by only 2.43% from that with the lowest score. In general, few differences among seasons were observed.

Good consistency exists between the rhythm of sentiment values and the actual climate. For example, from 2014 to 2017, the annual sentiment scores showed a trend of “up–down–down–down.” However, during this same period, the China Climate Bulletin showed that the extreme weather in 2014 was less extreme than that in 2013. The mean national temperature in 2015 was the highest since 1961. The number of extremely high temperature events in 2016 was significantly higher than that in 2015. The average national temperature in 2017 was 0.84°C higher than that in typical years.

(ii) Regional sentiment scores

We drew a sentiment map of China's tourism climate based on the sentiment scores by province. Figure 3 illustrates that

FIG. 2. Sentiment scores and reviews of tourists' perceptions of climate in China from 2011 to 2018.

the eastern boundary of the provinces, with the lowest sentiment scores crossing Northeast and Southwest China, varied along the two sides of the Hu Line (Heihe–Tengchong Line). The sentiment scores were higher east of the line than west of the line, demonstrating the general characteristics of high scores in the east and low scores in the west. Apart from the sentiment scores, there was a significant disparity in the number of reviews on either side of the line. The number of reviews in Xinjiang, Inner Mongolia, Qinghai, Tibet, Gansu, Ningxia, Sichuan, and other provinces west of the line accounted for only approximately 10% of all reviews. Although the percentage is different from the 5% population ratio west of the Hu Line (Hu 1983), it reflects the imbalance between East and Western China.

In addition to the characteristics of high scores in the east and low scores in the west, the sentiment map also exhibits the features of spatial continuity and discontinuity. The continuity of this map is less than that of the normal climate map. Together with Shanghai, the five provinces of Qinghai, Gansu, Tibet, Inner Mongolia, and Sichuan, which had lower sentiment scores, constitute a plateau area. On the one hand, the third lowest-ranking provinces of Guizhou, Hunan, Jiangxi, and Fujian form a narrow southern area. On the other hand, the third highest-ranking provinces of Shanxi, Henan, and Hubei constitute the Central Plains area. The provinces of Jilin and Liaoning constitute the northeastern area, and the high-ranking provinces of Zhejiang and Jiangsu constitute the eastern area.

FIG. 3. Sentiment map of climate among tourists in China.

Comparing the divided areas with the scheme of climatic regionalization in China proposed by the Chinese Academy of Sciences (Zhang 1991), we found that the plateau area corresponds to the Qinghai–Tibet region that exhibits a plateau climate and the northwestern area that exhibits a temperate continental climate. The southern part corresponds to Central and East China, which have a subtropical monsoon climate. The Central Plains and the northeastern districts cover Central, North, and Northeast China, which possess a temperate monsoon climate. The eastern boundary of the northeast–southwest contiguous zone, which is composed of the lowest and the penultimate low sentiment score, basically coincides with the Hu Line and approximately corresponds to the eastern boundary of the plateau alpine and temperate continental climate zones. In China, the monsoon climate is predominant in the east, while the continental climate is predominant in the northwest, and many climate types exist (Zhang 2016). These two characteristics are reflected in Fig. 3. The distribution characteristics of theoretical values at the national level and the presence of multiple contiguous areas at the regional level are consistent with the actual climate conditions and provide the first evidence for the credibility of the method that uses sentiment analysis of social media data to understand tourists' perceptions of climate.

However, there remain some “enclave” provinces, such as Shanghai, Chongqing, and Guangxi, which demonstrate a significant difference in sentiment scores from the surrounding areas, thus creating spatial discontinuity. This discontinuity is caused by tourism factors such as the characteristics of tourism activities, differences in natural or cultural attributes of attractions, and the local microclimate of scenic areas. We use the lowest-level region of Shanghai and the highest-level region of Chongqing as examples to explain the impact of these factors. Most comments about Shanghai concerned a tour to the Oriental Pearl TV Tower. This attraction has high-visibility requirements, whereas low-visibility conditions have a significant negative impact on visitors' perceptions of the weather (Denstadli and Jacobsen 2014). Therefore, the characteristics of tourism activities are responsible for the low sentiment score. In contrast, the 5A-class scenic spots in Chongqing are mostly natural landscapes and are in the Sichuan Basin, which has a humid subtropical climate. The attributes of landscapes and the microclimate environment resulted in high sentiment scores. Based on spatial continuity and discontinuity and the seasonality of tourism activities, tourists' perceptions of climate in multiple scenic areas within the region can reflect the perceptions of climate for the entire region except for scenic spots.

According to the actual climate divisions and provincial administrative division integrity, China can be divided into eight districts: Northeast China, North China, East China, Central China, South China, Southwest China, Northwest China, and the Qinghai–Tibet District (Zhang 1991; Xie and Wu 2008). The sentiment scores of the eight districts are listed from high to low with values of 0.802, 0.794, 0.793, 0.790, 0.769, 0.764, 0.762, and 0.743. North China had the highest score, which is consistent with the tourism reputation of 5A-class scenic spots in China (Yunyou original). The scores for Central, Northeast, and East China did not differ greatly. Scores for South, Northwest, and Southwest China were in the same

category. Southwest China ranked second to last, which may be attributed to its rainy weather, as tourists generally dislike rainy days (Denstadli and Jacobsen 2014; Dubois et al. 2016). Qinghai–Tibet is the world's “third pole,” and its sentiment score was significantly lower than that of the other areas.

Table 2 illustrates that the sentiment scores of Northeast, North, East, Central, and Northwest China did not vary greatly among seasons, indicating that tourists were not sensitive to changes in climate in different seasons. A few large areas show discernable periods that are either suitable or inappropriate for traveling. Specifically, the southwestern region and the southern part of China are most pleasant to visit in the first quarter and fourth quarter, respectively. The reason may be that warm weather in the cold season is more satisfying than in the warm season (Gössling et al. 2012). It was not advisable to travel to the Qinghai–Tibet region during the first quarter, as it had an emotional score of 0.707 in this quarter. Relevant research has shown that tourists are more intolerant of low temperatures than of high temperatures (Denstadli and Jacobsen 2014; Dubois et al. 2016).

(iii) Scenic sentiment score

The top 10 scenic areas in the emotional score are evenly distributed in North and South China. Most of them are natural scenic areas, and a few are cultural landscapes. The lowest 10 sites are clustered around Northwest or North China with approximately the same number of natural and cultural scenic areas. Notably, the two lowest-ranking attractions are the natural desert landscapes on the Northwest Plateau. Similar to the Oriental Pearl TV Tower in East China, the low sentiment score for the Three Gorges Dam in Central China is probably due to the requirement of the high visibility for tourism activities. Low ratings for other scenic areas may be related to rainy weather or low temperatures. Baylis et al. (2018) found that low temperatures, high temperatures, precipitation, humidity, and cloud cover had significant negative effects on tourists' perceptions of weather or climate. In addition, we found that the average sentiment scores of the natural and cultural scenic areas were 0.786 and 0.780, respectively. The former was only slightly higher than the latter, indicating that tourists did not form distinct perceptions of climate for different scenic areas.

5. Discussion

In terms of the application of our methods, we actively explored the ANN method to analyze data from Chinese social media and conducted sentiment analysis to understand tourists' perceptions of climate and weather. This research nevertheless represents only an initial attempt in this direction, and further theoretical consolidation and examination of scientific effectiveness are required, which are difficult, specifically with regard to the nuances between different types of negative emotions, such as irritation and annoyance (Gössling et al. 2016). Similarly, determining the criteria for distinguishing between objective and nonemotional words is challenging. In the future, the achievements of this method can be compared with the results of artificial intelligence methods that rely on the Baidu AI and Tencent Wenzhi platforms.

TABLE 2. Sentiment scores of tourists' perceptions of climate for regions in China.

	Northeast	North	East	Central	South	Southwest	Northwest	Qinghai-Tibet
First quarter	0.806	0.808	0.802	0.804	0.767	0.787	0.776	0.707
Second quarter	0.793	0.805	0.789	0.793	0.755	0.745	0.767	0.734
Third quarter	0.787	0.792	0.783	0.784	0.762	0.754	0.764	0.759
Fourth quarter	0.794	0.805	0.788	0.795	0.802	0.776	0.758	0.741

In terms of data collection, “name of scenic area” + “weather” or “climate” were used as the keywords for finding data. Although the data obtained all reflected the climate of tourist destinations, not all comments were generated by tourists. A small portion of the comments were from non-tourists such as residents. As this portion does not strictly comply with the requirements for analyzing tourists' perspectives in particular, some data were screened manually, which resulted in more accurate but less efficient data collection. In the future, noisy data can be processed synchronously in combination with machine-learning methods (Zhang et al. 2019). Additionally, the data in this paper contain only text, not other forms of data such as pictures and videos. Integrating these multisource data in future studies will improve the accuracy of scientific research (Deng and Li 2018).

For the analysis of the results, because the data corresponded to a brief period, the perceived characteristics, particularly the evolutionary trends, were not comprehensive. Subsequent research should obtain longer-term data or add more different spatial objects to facilitate the formation of more comprehensive and scientific conclusions. This paper focused on describing perceptions rather than providing an in-depth analysis of the reasons behind them. For example, there is no highlighted increase in sentiment scores on Weibo during the period of 2012–14, which contradicts the overall downward trend. As noted by Gössling et al. (2012), how negative experiences dominate holiday experiences and affect holiday satisfaction is unclear. Therefore, it is necessary to strengthen the theoretical and mechanistic analyses. In short, although this study is pioneering in using sentiment analysis methods and weather data to analyze tourists' perception of the climate, direct comparison is not easy to implement due to scarce related research. Perceived characteristics are very consistent with the facts, which provides a great reference value for follow-up research.

6. Conclusions

This study employed the sentiment analysis approach to establish an analytical framework to explore a new path for studying tourists' perceptions of climate using climate and weather data. This study provides insight for the use of new weather data to study tourists' climate perception, supports the paradigm shift of research in this field from small data such as questionnaires to online big data, and realizes the unification of micro- and macroresearch using microdata. This approach also contributes to the research on tourists' climate perception shift from traditionally relying primarily on field surveys or statistically small data to relying on big data. The application of

semisupervised ANNs to Chinese social media data enriches the method of tourist sentiment analysis. The idea of using weather data to build climate perception models is instructive for other regions. The results can provide a reference for climate information and crisis management of climate change at destinations.

We propose a four-quadrant emotion word classification system composed of positive and negative and subjective and objective sentiments. The results show that tourists' perceptions of climate are dominated by positive images. Considering that objective comments composed of nonemotional words may also contain emotions, these comments should also be included among the categories of sentiment analysis. Chinese tourists pay inadequate attention to climate risk. Hence, further bolstering awareness and actions concerning the climate crisis is necessary. We utilized ROST CM, BosonNLP, and GooSeeker to systematically analyze high-frequency words. Among the three software programs, ROST CM yielded the most accurate results, while GooSeeker produced the least accurate results. However, GooSeeker can process the largest volumes of texts, which should be considered in related research.

The theoretical results of tourists' climate perception are consistent with the actual situation and, to a certain extent, prove that cognitive evaluation theory is suitable for the study of the influence of texts on climate perception. Tourists' climate perceptions reflect not only the scenic spots but also the entire region outside of scenic spots at macro-mesolevels. ANN generates a perception sentiment score of 0.784, which corresponds to the intermediate level of “good.” The sentiment scores of Chinese tourists on the climate perception of the destination reveal a downward trend over time. The sentiment scores of large areas are characteristically high in the east and low in the west, and spatial continuity and discontinuity coexist. The climate-awareness perception map shows explicit climate traces accompanied by a distinct feature of tourism. Some summer or winter resorts also have higher evaluation values in the web texts during the corresponding seasons. This finding shows that the sentiment evaluation value of the web text is consistent with the actual experience. The results enrich the connotation of cognitive evaluation theory and provide a reference for tourists' decision-making and destination marketing.

Acknowledgments. This research was funded by the National Natural Science Foundation of China, Grants 41571131, 42071168, and 41701147. This research is also supported by a project funded by the Priority Academic Program Development

of Jiangsu Higher Education Institutions (PAPD). Thanks are extended to Z. H. He, Y. Lu, C. Han, and L. Feng for collecting data.

Data availability statement. Readers can contact the corresponding author (taoyuguol23@163.com) to obtain data.

REFERENCES

- Alaei, A. R., S. Becken, and B. Stantic, 2017: Sentiment analysis in tourism: Capitalizing on big data. *J. Travel Res.*, **58**, 175–191, <https://doi.org/10.1177/0047287517747753>.
- Arnold, M. B., 1960: *Emotion and Personality: Psychological Aspects*. Vol. 1, *Emotion and Personality*, Columbia University Press, 296 pp.
- Baylis, P., N. Obradovich, Y. Kryvasheyeu, H. H. Chen, L. Coviello, E. Moro, M. Cebrian, and J. H. Fowler, 2018: Weather impacts expressed sentiment. *PLOS ONE*, **13**, e0195750, <https://doi.org/10.1371/journal.pone.0195750>.
- Becken, S., B. Stantic, J. Chen, A. R. Alaei, and R. M. Connolly, 2017: Monitoring the environment and human sentiment on the Great Barrier Reef: Assessing the potential of collective sensing. *J. Environ. Manage.*, **203**, 87–97, <https://doi.org/10.1016/j.jenvman.2017.07.007>.
- Broß, J., 2013: Aspect-oriented sentiment analysis of customer reviews using distant supervision techniques. Ph.D. dissertation, University of Berlin, 394 pp.
- Caldeira, A. M., and E. Kastenholz, 2018: It's so hot: Predicting climate change effects on urban tourists' time-space experience. *J. Sustainable Tourism*, **26**, 1516–1542, <https://doi.org/10.1080/09669582.2018.1478840>.
- Cambria, E., 2016: Affective computing and sentiment analysis. *IEEE Intell. Syst.*, **31**, 102–107, <https://doi.org/10.1109/MIS.2016.31>.
- Capatina, A., A. Micu, A. E. Micu, R. Bouzaabia, and O. Bouzaabia, 2018: Country-based comparison of accommodation brands in social media: An fsQCA approach. *J. Bus. Res.*, **89**, 235–242, <https://doi.org/10.1016/j.jbusres.2017.11.017>.
- Chiu, C., N. H. Chiu, R. J. Sung, and P. Y. Hsieh, 2015: Opinion mining of hotel customer-generated contents in Chinese weblogs. *Curr. Issues Tourism*, **18**, 477–495, <https://doi.org/10.1080/13683500.2013.841656>.
- Claster, W. B., H. Dinh, and M. Cooper, 2010: Naïve Bayes and unsupervised artificial neural nets for Cancun tourism social media data analysis. *Proc. Second World Congress on Nature and Biologically Inspired Computing*, Kitakyushu, Japan, IEEE, 158–163, <https://doi.org/10.1109/NABIC.2010.5716370>.
- , M. Cooper, and P. Sallis, 2011: Thailand—Tourism and conflict: modeling sentiment from twitter tweets using naïve Bayes and unsupervised artificial neural nets. *Second Int. Conf. on Computational Intelligence*, Bali, Indonesia, IEEE, 89–94, <https://doi.org/10.1109/CIMSiM.2010.98>.
- , P. Pardo, M. Cooper, and K. Tajeddini, 2013: Tourism, travel and tweets: Algorithmic text analysis methodologies in tourism. *Middle East J. Manage.*, **1**, 81–99, <https://doi.org/10.1504/MEJM.2013.054071>.
- Confente, I., 2014: Twenty-five years of word-of-mouth studies: A critical review of tourism research. *Int. J. Tourism Res.*, **17**, 613–624, <https://doi.org/10.1002/jtr.2029>.
- Cong, L., B. Wu, A. M. Morrison, H. Shu, and M. Wang, 2014: Analysis of wildlife tourism experiences with endangered species: An exploratory study of encounters with giant pandas in Chengdu, China. *Tourism Manage.*, **40**, 300–310, <https://doi.org/10.1016/j.tourman.2013.07.005>.
- Curnock, M., and Coauthors, 2019: Shifts in tourists' sentiments and climate risk perceptions following mass coral bleaching of the Great Barrier Reef. *Nat. Climate Change*, **9**, 535–541, <https://doi.org/10.1038/s41558-019-0504-y>.
- Dahal, B., S. A. P. Kumar, and Z. L. Li, 2019: Topic modeling and sentiment analysis of global climate change tweets. *Soc. Network Anal. Min.*, **9**, 24, <https://doi.org/10.1007/s13278-019-0568-8>.
- Deng, N., and X. Li, 2018: Feeling a destination through the “right” photos: A machine learning model for DMOs' photo selection. *Tourism Manage.*, **65**, 267–278, <https://doi.org/10.1016/j.tourman.2017.09.010>.
- Denstadli, J. M., and J. K. S. Jacobsen, 2014: More clouds on the horizon? Polar tourists' weather tolerances in the context of climate change. *Scand. J. Hosp. Tourism*, **14**, 80–99, <https://doi.org/10.1080/15022250.2014.886096>.
- , —, and M. Lohmann, 2011: Tourist perceptions of summer weather in Scandinavia. *Ann. Tourism Res.*, **38**, 920–940, <https://doi.org/10.1016/j.annals.2011.01.005>.
- De Urioste-Stone, S. M., L. Le, M. D. Scaccia, and E. Wilkins, 2016: Nature-based tourism and climate change risk: Visitors' perceptions in mount desert Island, Maine. *J. Outdoor Recreat. Tourism*, **13**, 57–65, <https://doi.org/10.1016/j.jort.2016.01.003>.
- Donkor, B., 2014: Sentiment analysis: Why it's never 100% accurate. <http://brnr.me/sentiment-analysis-never-accurate>.
- Dubois, G., J. P. Ceron, S. Gössling, and C. M. Hall, 2016: Weather preferences of French tourists: Lessons for climate change impact assessment. *Climatic Change*, **136**, 339–351, <https://doi.org/10.1007/s10584-016-1620-6>.
- Dwyer, L., D. Edwards, N. Mistilis, C. Roman, and N. Scott, 2009: Destination and enterprise management for a tourism future. *Tourism Manage.*, **30**, 63–74, <https://doi.org/10.1016/j.tourman.2008.04.002>.
- Feldman, R., 2013: Techniques and applications for sentiment analysis. *Commun. ACM*, **56**, 82–89, <https://doi.org/10.1145/2436256.2436274>.
- Finance Department of the Ministry of Culture and Tourism, 2019: Basic situation of the tourism market in 2018. Government of China, http://www.gov.cn/shuju/2019-02/13/content_5365248.htm.
- Fitchett, J., and G. Hoogendoorn, 2018: An analysis of factors affecting tourists' accounts of weather in South Africa. *Int. J. Biometeor.*, **62**, 2161–2172, <https://doi.org/10.1007/s00484-018-1617-0>.
- Forland, E. J., J. K. S. Jacobsen, J. M. Denstadli, M. Lohmann, I. Hanssen-Bauer, H. O. Hygen, and H. Tømmervik, 2013: Cool weather tourism under global warming: Comparing Arctic summer tourists' weather preferences with regional climate statistics and projections. *Tourism Manage.*, **36**, 567–579, <https://doi.org/10.1016/j.tourman.2012.09.002>.
- Fuchs, M., W. Höpken, and M. Lexhagen, 2014: Big data analytics for knowledge generation in tourism destinations—A case from Sweden. *J. Destin. Market. Manage.*, **3**, 198–209, <https://doi.org/10.1016/j.jdmm.2014.08.002>.
- García, A., S. Gaines, and M. T. Linaza, 2012: A lexicon based sentiment analysis retrieval system for tourism domain. *e-Rev. Tourism Res.*, **10**, 35–38.
- Gomez-Martin, M. B., X. A. Armesto-Lopez, and E. Martinez-Ibarra, 2014: Spanish tourists (young segment market) are flexible in adapting to episodes of extremely high temperatures. *Int. J. Biometeor.*, **58**, 781–797, <https://doi.org/10.1007/s00484-013-0659-6>.

- Gössling, S., and M. Hall, 2006: Uncertainties in predicting tourist flows under scenarios of climate change. *Climatic Change*, **79**, 163–173, <https://doi.org/10.1007/s10584-006-9081-y>.
- , D. Scott, C. M. Hall, J. P. Ceron, and G. Dubois, 2012: Consumer behaviour and demand response of tourists to climate change. *Ann. Tourism Res.*, **39**, 36–58, <https://doi.org/10.1016/j.annals.2011.11.002>.
- , B. Abegg, and R. Steiger, 2016: “It was raining all the time!”: Ex post tourist weather perceptions. *Atmosphere*, **7**, 10, <https://doi.org/10.3390/atmos7010010>.
- Gräbner, D., M. Zanker, G. Fliedl, and M. Fuchs, 2012: Classification of customer reviews based on sentiment analysis. *Information and Communication Technologies in Tourism*, M. Fuchs, F. Ricci, and L. Cantoni, Eds., Springer, 460–470.
- Hu, Y. H., 1983: *The Demographic Distribution in China*. East China Normal University Press, 52 pp.
- , Y. L. Chen, and H. L. Chou, 2017: Opinion mining from online hotel reviews—A text summarization approach. *Inf. Process. Manage.*, **53**, 436–449, <https://doi.org/10.1016/j.ipm.2016.12.002>.
- Hübner, A., and S. Gössling, 2012: Tourist perceptions of extreme weather events in Martinique. *J. Destin. Market. Manage.*, **1**, 47–55, <https://doi.org/10.1016/j.jdmm.2012.09.003>.
- IBM, 2016: The four V's of big data. <https://datasciencedegree.wiscnsin.edu/wp-content/uploads/2015/05/4-Vs-of-big-data.jpg>.
- Jacobsen, J. K. S., J. M. Denstadli, M. Lohmann, and E. J. Forland, 2011: Tourist weather preferences in Europe's Arctic. *Climate Res.*, **50**, 31–42, <https://doi.org/10.3354/cr01033>.
- Jeuring, J. H. G., 2017: Weather perceptions, holiday satisfaction and perceived attractiveness of domestic vacationing in the Netherlands. *Tourism Manage.*, **61**, 70–81, <https://doi.org/10.1016/j.tourman.2017.01.018>.
- Kahn, J. H., R. M. Tobin, A. E. Massey, and J. A. Anderson, 2007: Measuring emotional expression with the linguistic inquiry and word count. *Amer. J. Psychol.*, **120**, 263–286, <https://doi.org/10.2307/20445398>.
- Kim, S., J. H. Park, D. K. Lee, Y. H. Son, H. Yoon, S. Kim, and H. J. Yun, 2017: The impacts of weather on tourist satisfaction and revisit intention: A study of South Korean domestic tourism. *Asia Pac. J. Tourism Res.*, **22**, 895–908, <https://doi.org/10.1080/10941665.2017.1357640>.
- Kirilenko, A. P., and S. Stepchenkova, 2017: Sochi 2014 Olympics on Twitter: Perspectives of hosts and guests. *Tourism Manage.*, **63**, 54–65, <https://doi.org/10.1016/j.tourman.2017.06.007>.
- , T. Molodtsova, and S. O. Stepchenkova, 2015: People as sensors: Mass media and local temperature influence climate change discussion on Twitter. *Global Environ. Change*, **30**, 92–100, <https://doi.org/10.1016/j.gloenvcha.2014.11.003>.
- , S. O. Stepchenkova, H. Kim, and X. Li, 2017: Automated sentiment analysis in tourism: Comparison of approaches. *J. Travel Res.*, **57**, 1012–1025, <https://doi.org/10.1177/0047287517729757>.
- Lam, C. K. C., M. Loughnan, and N. Tapper, 2018: Visitors' perception of thermal comfort during extreme heat events at the Royal Botanic Garden Melbourne. *Int. J. Biometeor.*, **62**, 97–112, <https://doi.org/10.1007/s00484-015-1125-4>.
- Lazarus, R. S., and S. Folkman, 1984: *Stress, Appraisal, and Coping*. Springer, 445 pp.
- Li, J., L. Xu, L. Tang, S. Y. Wang, and L. Li, 2018: Big data in tourism research: A literature review. *Tourism Manage.*, **68**, 301–323, <https://doi.org/10.1016/j.tourman.2018.03.009>.
- Lindner-Cendrowska, K., and K. Blazejczyk, 2016: Impact of selected personal factors on seasonal variability of recreationist weather perceptions and preferences in Warsaw (Poland). *Int. J. Biometeor.*, **62**, 113–125, <https://doi.org/10.1007/s00484-016-1220-1>.
- Liu, Y., K. X. Huang, J. G. Bao, and K. Q. Chen, 2019: Listen to the voices from home: An analysis of Chinese tourists' sentiments regarding Australian destinations. *Tourism Manage.*, **71**, 337–347, <https://doi.org/10.1016/j.tourman.2018.10.004>.
- Lohmann, M., and E. Kaim, 1999: Weather and holiday preferences image, attitude and experience. *Tourism Rev.*, **54**, 54–64, <https://doi.org/10.1108/eb058303>.
- Marrese-Taylor, E., J. D. Velásquez, F. Bravo-Marquez, and Y. Matsuo, 2013: Identifying customer preferences about tourism products using an aspect-based opinion mining approach. *Procedia Comput. Sci.*, **22**, 182–191, <https://doi.org/10.1016/j.procs.2013.09.094>.
- Mintel International Group, 1991: *Special Report—Holidays*. Leisure Intelligence Mintel International Group.
- Misopoulos, F., M. Mitic, A. Kapoulas, and C. Karapiperis, 2014: Uncovering customer service experiences with Twitter: The case of airline industry. *Manage. Decision*, **52**, 705–723, <https://doi.org/10.1108/MD-03-2012-0235>.
- Mushawemhuka, W., J. M. Rogerson, and J. Saarinen, 2018: Nature-based tourism operators' perceptions and adaptation to climate change in Hwange National Park, Zimbabwe. *Bull. Geogr. Socio-Econom. Ser.*, **42**, 115–127, <https://doi.org/10.2478/bog-2018-0034>.
- Nalau, J., S. Becken, S. Noakes, and B. Mackey, 2017: Mapping tourism stakeholders' weather and climate information-seeking behavior in Fiji. *Wea. Climate Soc.*, **9**, 377–391, <https://doi.org/10.1175/WCAS-D-16-0078.1>.
- Pang, B., and L. Lee, 2008: Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.*, **2** (1–2), 1–135, <https://doi.org/10.1561/15000000011>.
- Perkins, D., and K. Debbage, 2016: Weather and tourism: Thermal comfort and zoological park visitor attendance. *Atmosphere*, **7**, 44, <https://doi.org/10.3390/atmos7030044>.
- Pröbstl-Haider, U., W. Haider, V. Wirth, and B. Beardmore, 2015: Will climate change increase the attractiveness of summer destinations in the European Alps? A survey of German tourists. *J. Outdoor Recreat. Tourism*, **11**, 44–57, <https://doi.org/10.1016/j.jort.2015.07.003>.
- Rutty, M., and D. Scott, 2010: Will the Mediterranean become “too hot” for tourism? A reassessment. *Tour. Hosp. Plann. Dev.*, **7**, 267–281, <https://doi.org/10.1080/1479053X.2010.502386>.
- , and —, 2016: Comparison of climate preferences for domestic and international beach holidays: A case study of Canadian travelers. *Atmosphere*, **7**, 30, <https://doi.org/10.3390/atmos7020030>.
- Saura, J. R., P. Palossanchez, and M. A. R. Martin, 2018a: Attitudes expressed in online comments about environmental factors in the tourism sector: An exploratory study. *Int. J. Environ. Res. Public Health*, **15**, 553, <https://doi.org/10.3390/ijerph15030553>.
- , A. Reyes-Menendez, and C. Alvarez-Alonso, 2018b: Do online comments affect environmental management? Identifying factors related to environmental management and sustainability of hotels. *Sustainability*, **10**, 3016, <https://doi.org/10.3390/su10093016>.
- Schmunk, S., W. Höpken, M. Fuchs, and M. Lexhagen, 2014: Sentiment analysis: Extracting decision-relevant knowledge from UGC. *Information and Communication Technologies in*

- Tourism 2014*, Z. Xiang and I. Tussyadiah, Eds., Springer International Publishing, 253–265.
- Scott, D., S. Gössling, and C. R. D. Freitas, 2008: Preferred climates for tourism: Case studies from Canada, New Zealand and Sweden. *Climate Res.*, **45**, 61–73, <https://doi.org/10.3354/cr00774>.
- Shimada, K., S. Inoue, H. Maeda, and T. Endo, 2011: Analyzing tourism information on Twitter for a local city. *ACIS Int. Symp. on Software Network Engineering*, Seoul, South Korea, IEEE, 61–66.
- Sina Technology, 2013: Weibo's financial report for Q1 2013. <https://tech.sina.com.cn/i/2013-05-17/04598348853.shtml>.
- , 2017: Weibo's financial report for Q1 2017. <https://tech.sina.com.cn/i/2017-05-16/doc-ifyfeius8004650.shtml>.
- Tan, S., and Q. Wu, 2011: A random walk algorithm for automatic construction of domain-oriented sentiment lexicon. *Expert Syst. Appl.*, **38**, 12 094–12 100, <https://doi.org/10.1016/j.eswa.2011.02.105>.
- U.S. Environmental Protection Agency, 2013: *Glossary of Climate Change Terms*. <http://www.epa.gov/climatechange/glossary.html>.
- Waldhör, K., and A. Rind, 2008: E-blog analysis—Mining virtual communities using statistical and linguistic methods for quality control in tourism. *Information and Communication Technologies in Tourism 2008*, P. O'Connor, W. Höpken, and U. Gretzel, Eds., Springer, 453–462.
- Wilkins, E., S. D. Uriostestone, A. Weiskittel, and T. Gabe, 2017: Weather sensitivity and climate change perceptions of tourists: A segmentation analysis. *Tourism Geogr.*, **20**, 273–289, <https://doi.org/10.1080/14616688.2017.1399437>.
- Witte, R., and J. Mülle, 2006: Text Mining: Wissensgewinnung aus natürlichsprachigen Dokumenten (Text mining: Gaining knowledge from natural language documents). Universität Karlsruhe Fakultät für Informatik Institut für Programmstrukturen und Datenorganisation Doc., 200 pp.
- Xiang, Z., Z. Schwartz, J. H. Gerdes, and M. Uysal, 2015: What can big data and text analytics tell us about hotel guest experience and satisfaction? *Int. J. Hosp. Manage.*, **44**, 120–130, <https://doi.org/10.1016/j.ijhm.2014.10.013>.
- Xie, Z., and B. Wu, 2008: Tourism spatial structure of resource-based attractions in China. *Dili Kexue*, **28**, 748–753.
- Yan, Q., S. Zhou, and S. Wu, 2018: The influences of tourists' emotions on the selection of electronic word of mouth platforms. *Tourism Manage.*, **66**, 348–363, <https://doi.org/10.1016/j.tourman.2017.12.015>.
- Ye, Q., Z. Zhang, and R. Law, 2009: Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. *Expert Syst. Appl.*, **36**, 6527–6535, <https://doi.org/10.1016/j.eswa.2008.07.035>.
- Zhang, C., Z. Huang, F. Cao, and X. Y. Chen, 2019: Recognise me from outside to inside: Learning the influence chain of urban destination personalities. *Tourism Manage.*, **70**, 390–403, <https://doi.org/10.1016/j.tourman.2018.09.005>.
- Zhang, J. C., 1991: *Climate of China*. China Meteorological Press, 457 pp.
- Zhang, M. D., 2016: *China National Geographic Encyclopedia*. Beijing United Publishing Company, 50 pp.
- Zhang, Z., Q. Ye, Z. Zhang, and Y. J. Li, 2011: Sentiment classification of Internet restaurant reviews written in Cantonese. *Expert Syst. Appl.*, **38**, 7674–7682, <https://doi.org/10.1016/j.eswa.2010.12.147>.
- Zheng, W., and Q. Ye, 2009: Sentiment classification of Chinese traveler reviews by support vector machine algorithm. *Int. Conf. on Intelligent Information Technology Application*, Nanchang, China, IEEE, 335–338.
- Zimmerman, S., and U. Kruschwitz, 2017: Speaking of the weather: Detection of meteorological influences on sentiment within social media. *Computer Science and Electronic Engineering Conf.*, Colchester, United Kingdom, IEEE, 1–7.