A. Ballatore and B. Adams (2015). Extracting Place Emotions from Travel Blogs.
Proceedings of the Association of Geographic Information Laboratories for Europe (AGILE),
Lisbon, June 2015, pp. 1–5.

Extracting Place Emotions from Travel Blogs

Andrea Ballatore
Center for Spatial Studies
University of California, Santa Barbara
Santa Barbara, CA
aballatore@spatial.ucsb.edu

Benjamin Adams
Centre for eResearch
University of Auckland, New Zealand
b.adams@auckland.ac.nz

Abstract

Place is a central category in the human experience. Across cultures, individuals describe experiences, express opinions, narrate stories set in and about places. The web provides a large, dynamic corpus of documents describing places from a myriad of viewpoints. Emotions and their expression play an important role in these representations of places, making some places appear joyful and beautiful, and others scary, sad, or even disgusting. In this paper we propose to tap the corpus of place descriptions from the emotional viewpoint, aiming at the development of a framework to model, extract, and analyze emotions relative to places. As first steps in this direction, we focus on place classes, i.e. the types of places that are discussed, such as *city*, *forest*, and *road*. To identify such classes, we design the Place Vocabulary, a linked semantic resource that contains nouns in English that are used to identify natural and built places. Subsequently, we propose a natural language processing technique to extract a multi-dimensional model of place emotion, based on the vocabulary in WordNet-Affect. The technique is applied to a corpus of about 100,000 travel blog posts from *travelblog.org*, enabling the exploration of the emotional structure of place classes.

Keywords: sentiment analysis; emotion analysis; place; place vocabulary

1 Introduction

Places are constructed and experienced along multiple dimensions, ranging from physical to social structures. Across cultures, individuals describe experiences, narrate stories that are set in places, and express direct opinions about places. Individuals develop complex attachments to their homeland and can develop sentiments of revulsion for places associated with traumas. As Bell [4] pointed out,“we experience in places the sentiments of sociality, sentiments of liking and disliking, trust and fear, renewal and loss, connection and disconnection, belongingness and foreignness, justice and injustice” (p. 832). This affective space constitutes a crucial element in the social construction and experience of place [18].

On the web, an enormous, dynamic, heterogeneous corpus of opinions about places keeps expanding across newspapers, blogs, and social media. The measurement of public sentiment about a given topic in such outlets has already tangible applications in politics, marketing, and business analytics [10]. However, existing techniques aim at quantifying sentiment on a negative-positive spectrum, ignoring more nuanced aspects of emotions [5]. While some research has targeted this corpus from a geographic perspective [14, 1], no attempt has been made to explore places with respect to their emotional structure.

In this paper we propose to fill this gap by tapping this corpus to model, extract, and analyze the space of emotions, at the intersection of natural language processing, emotion analysis, and geography. To move towards this goal, we harness ideas and resources from the area of sentiment detection and opinion mining, which rely on natural language processing to aggregate and summarize writers' opinions and emotions from raw text. We start the exploration of the emotional dimension of place by focusing on *place classes* (e.g., city and park), as opposed to place instances, such as Los Angeles and Griffith Park. To identify such classes, we designed and curated the *Place Vocabulary*, a reusable semantic resource that contains nouns in English that are used to refer to natural and built places. The Place Vocabulary is the result of a selective merge of the GeoNames ontology, WordNet, and the DBpedia ontology, and is freely available online.

To explore the complex relationship between emotions and places, we propose a natural language processing technique to extract a multi-dimensional space of place emotion. This approach relies on the emotion vocabulary defined in the WordNet-Affect [16], which provides terms associated with six basic emotions (anger, disgust, fear, joy, sadness, and surprise). The approach constructs a multi-dimensional vector space based on term co-occurrences within a context window around terms that indicate place classes. As a case study, we apply the proposed technique to a corpus of approximately 100,000 travel blog posts, eliciting emotions for a set of place classes.

In the reminder of this paper, we review relevant literature (Section 2). Section 3 outlines the Place Vocabulary, while in Section 4 we describe our technique to extract emotion vectors from a corpus of text documents, illustrating its application on a case study. Finally, Section 5 draws conclusions and indicates directions for future research.

2 Related work

The study of human emotions is a long-standing issue in philosophy, cognitive science and psychology [17]. In the 1970s, Plutchik [13] developed an influential psycho-evolutionary theory of emotions. Subsequently, Ekman [7] developed the idea that humans experience six cross-cultural, universal emotions, i.e., love, joy, surprise, anger, sadness, and fear, which can be combined into many secondary and tertiary emotions. Each of these emotions can be experienced at a different intensity. From a computational viewpoint, such a multidimensional theory of emotion can be translated into vector space, in which an emotional state is a vector with each value referring to the intensity of one particular emotion [6].

The issue of automatic detection of expressions of emotions in natural language is the core effort in the area of sentiment analysis and opinion mining. This strand of natural language processing aims at the classification of a textual document into a subjective category (e.g., good/bad) or along a single continuous dimension from negative to positive [12]. Domain-specific sentiment analysis techniques have been tailored to online reviews of hotels and travel destinations, adopting syntactic parsing [5], as well as supervised machine learning approaches [20]. Opinions can be explicit or implicit in a text, the latter being especially difficult to extract from simple methods based on syntactic matching. Recently, alternative models have been developed, such as Sentilo, a frame-based model for sentiment analysis [15].

Surprisingly, this large body of work, surveyed by Liu [10], ignores entirely the prominent role that location and place play in human experience and cognition. Emotion analysis can be considered as a specific type of sentiment analysis that adopts multi-dimensional models, for example using Ekman's six basic emotions as orthogonal dimensions [7]. Such emergent techniques of emotion analysis have been applied to identify emotional patterns in books and emails, including suicide notes and love letters [11]. Likewise, digital humanities researchers apply emotion analysis to works of literature [9].

In the context of geographic information science (GIScience), sentiment and emotion analysis have received limited attention. Current efforts in this area tackle the study of the concept of place by identifying recurring patterns in online datasets. Notably, Purves et al. [14] proposed a framework to extract lexical descriptions of places from user-generated content, identifying elements, qualities, and activities for different place classes. The study of place can greatly benefit from the quantification of emotional patterns in large textual corpora, improving the modeling and prediction of a variety of geographic phenomena, as is already happening in finance and marketing [10].

3 Place Vocabulary

To study the emotional structure of places, we developed a vocabulary of place nouns in English, freely available online.[1] This vocabulary focuses on *place classes*, as opposed to gazetteers and other resources that include *place instances*. This semantic resource was constructed by tapping existing data sources, including DBpedia, GeoNames, and WordNet. These sources were selected as part of general-purpose geographic commons, structured around lightweight ontologies and vocabularies [2].

DBpedia is a structured dataset extracted from Wikipedia, whose ontology contains a hierarchy of 113 place categories.[2] The open gazetteer GeoNames classifies places based on a hierarchy of 660 types, ranging from very general nouns to very specific geographic jargon.[3] Finally, WordNet [8] is a well-established lexical database whose hierarchy of geographic concepts provides 1,329 place nouns. WordNet includes information about terms' polysemy and several semantic relations.

These three sources were merged following the linked data approach described in [3], resulting in a vocabulary of 1,875 place nouns. Each term is defined based on its provenance. For example, the noun 'hospital' has a unique ID (*pv775*) and is linked to the corresponding concepts in the data sources:

```
id: "pv775"
label: "hospital"@en
dbp: http://dbpedia.org/ontology/Hospital
gn: http://www.geonames.org/ontology#S.HSP
wn: http://wordnet-rdf.princeton.edu/wn31/hospital-n
```

This way, this vocabulary provides ready access to the semantic content of the data sources, such as subclasses, superclasses, multi-lingual definitions. For ease of use, the vocabulary is currently available in JSON format. The Place Vocabulary was used in the travel blog case study to identify references to places at the class level in raw text.

4 Emotion extraction for place classes

To extract the emotional structure of place classes, we started by devising a computational technique. Following the general approach by Mohammad [11], we aim at extracting a representation of target terms in an emotion space, allowing exploration and comparison. In computational terms, given a place class t (e.g., *hospital*) and an emotion vocabulary E, we define a function $f(t,E,C)$ that extracts an emotion vector $\vec{e}_t$ from a corpus C of text documents. The remainder of this section defines the technique and illustrates its application on a case study of a corpus of travel blogs.

Travel blog corpus. For this case study, we selected a corpus of crowdsourced travel blog entries freely available on the web. In total, we collected approximately 100,000 travel blog entries from *travelblog.org*, a popular travel blogging website. In each post, an author describes their travel experiences occurred in one or more places. The main locations associated with each entry are entered by the author from a hierarchical vocabulary of place names organized by continent, country, and administrative regions. The entries cover the entire planet and span a time frame from 2004 to 2014.

Such travel posts provide one affective perspective (among many) on places. It works well as a use case for emotion analysis because travelogues are designed to convey the experiential perception of the individual traveler, more so than encyclopedia entries or other sources. Sentiment analysis on a large set of travelogues can potentially lend insight into research theorizing about the role of the tourist in shaping the perception of types of places like religious buildings, city centers, and natural areas [19].

The corpus was subject to a pre-processing step. The raw text was POS-tagged with the Stanford Log-linear Part-Of-Speech

[1] http://github.com/andrea-ballatore/PlaceVocabulary

[2] http://dbpedia.org/ontology/Place

[3] http://www.geonames.org/ontology

Tagger, and lemmatized with Stanford Core NLP.[4] The resulting corpus includes only posts in English and longer than 500 words, for a total of 101M tokens.

Identification of place classes. The Place Vocabulary (see Section 3) was used to identify references to place classes. The matching was conducted on lemmatized nouns, excluding proper nouns (i.e. NNP and NNPS). The matching identified about 332,000 place class occurrences, with *city*, *town*, *road*, *beach*, *hotel*, *street*, and *hostel* having the highest number of occurrences (roughly from 100,000 to 50,000). As expected, the distribution of such occurrences follows a power law, with few terms receiving most of the references, followed by a tail of rare terms. Indeed, polysemy affects this kind of lexical matching (e.g., *bank* as a financial institution or as a river bank). As word sense disambiguation is an unsolved problem and outside the scope of this study, we manually removed from the analysis relatively rare terms (references $<$ 20,000), as well as highly polysemic and vague noun classes , including *place*, *lot*, *side*, and *water*. As a result of this process, as set of 26 place nouns was included in the study.

Context windows. Define a context window W_t for term t as the set of the terms co-occurring within a distance. The selection of terms was restricted to common nouns, verbs, adjectives, and adverbs. For example, when $W_t = 2$ for *hostel*, the context window contains the four underlined terms:

> ...The {following day, the *hostel* where we stayed had} relatively clean beds, but ...

To obtain a rich representation of the context in which place classes are used, we extracted all the context windows of size 5, therefore collecting the 10 co-occurring terms for each reference of a place class t. For instance, *beach* obtained a context window containing 635,093 terms, of which 29,383 unique, across the entire corpus.

Emotion vocabulary. As an emotion vocabulary E for our approach, we adopted WordNet-Affect [16], a widely used resource that contains linguistic tokens associated with six basic emotions (anger, disgust, fear, joy, sadness, and surprise). For example, adjectives *fearful*, *scary*, and *hysterical* are associated with fear, while nouns *repugnance*, *revulsion*, and *horror* express disgust. These tokens are expressed with their POS tags. For each place term t and for each emotion e, we calculated an *emotion score* S as the ratio between the number of emotion terms in the context window W_t and the total number of terms. The emotion score can be decomposed into specific emotions:

$$S = \frac{|W_t \cap E|}{W_t} = S_{anger} + S_{fear} + S_{disgust} + S_{joy} + S_{sadness} + S_{surprise}$$

To reduce the noise caused by polysemy, we created a stop-word list to exclude emotion tokens that typically occur in a different, non-emotion laden sense.[5]

[4] http://nlp.stanford.edu/software/corenlp.shtml

[5] The list includes: *rag*, *score*, *get*, *close*, *occupy*, *entrance*, *move*, *gravel*, *scene*, *fit*, *catch*, and *chill*.

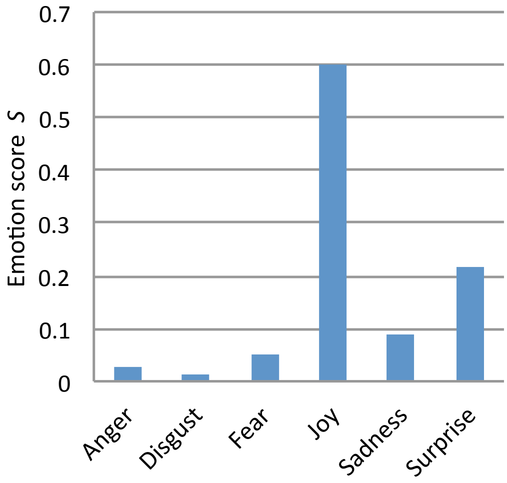

Figure 1: Overall emotion scores for 26 place classes in the travel blog corpus for context window equal to 5.

Emotion vectors. The emotion scores were computed for each of the 26 place classes, analyzing 10.3M words found in the context windows. In total, 158,231 emotion terms from WordNet-Affect were identified. As shown in Figure 1, 60% of terms were associated with joy, followed by surprise (22%), sadness (9%), and the remainder 9% divided between anger, disgust, and fear. This indicates that the travel blog corpus contains overwhelmingly experiences of place filled with joy and surprise, as is largely expected in the context of tourism, where travelers tend to visit locations known for their positive features. Each place class was represented as an emotion vector $\vec{e}_t$, capturing the emotion scores for the six dimensions.

Discussion. The place classes that obtained highest scores, both globally and for each dimension, are reported in Table 1. The most intense overall responses were observed for *restaurant*, *city*, and *museum*, which suggests a strong centrality of these place classes in the tourist experience. Looking more in detail at the emotion vector for *restaurant* in Table 2, it is possible to observe the range of semantic fields that characterize the emotional configuration of the experiences occurring in restaurants. These terms suggest an overwhelming dominance of joy and surprise, related to the experience of food, sociality, and comfort. By contrast, a much smaller role is played by negative emotions, ranging from anger (e.g., frustration and irritation with service) to outright disgust (e.g., sickness and disgust for bad food). Through these emotion vectors, the proposed method enables the quantification and operationalization of emotional responses in text corpora.

5 Conclusion and future work

In this paper we have outlined a computational approach to extract and model emotions for places, focusing in particular on place classes. Our method used crowdsourced linked data to generate a place vocabulary, which was used in combination with

Emotion score S		*anger*		*fear*		*disgust*		*joy*		*sadness*		*surprise*	
restaurant	.024	road	.013	road	.033	sea	.016	restaurant	.328	road	.049	city	.104
city	.021	street	.012	street	.017	hotel	.005	city	.299	town	.035	beach	.082
museum	.02	city	.011	mountain	.016	hostel	.005	beach	.233	hotel	.035	mountain	.078
hill	.019	hostel	.011	city	.016	road	.004	town	.233	hostel	.031	island	.078
beach	.017	hotel	.011	hotel	.016	school	.004	museum	.186	city	.031	building	.069
sea	.016	airport	.01	hostel	.016	city	.004	hostel	.186	beach	.029	restaurant	.068
pool	.016	beach	.009	airport	.015	beach	.004	street	.183	street	.027	lake	.061
mountain	.016	town	.008	building	.015	restaurant	.003	hotel	.182	tree	.027	hill	.057
street	.016	station	.007	town	.014	market	.003	road	.163	station	.027	temple	.056
hostel	.015	house	.007	beach	.013	station	.003	hill	.153	airport	.023	market	.056

Table 1: Place classes with highest emotion scores *S* (top 10 for each emotion)

Place class	*Emotion*	*S*	*Top emotion terms*
restaurant	anger	.007	bother, mad, angry, amok, harass, annoying, evil, madness, frustrating, frustrated
	fear	.007	awful, horrible, terrible, scary, horrid, shadow, dismay, ugly, afraid, anxious
	disgust	.003	sick, obscene, wicked, disgusting, disgust, offensive, yucky, horror, revolting, sickening
	joy	.328	great, good, enjoy, happy, like, love, friendly, taste, fascinating, comfortable
	sadness	.015	bad, dark, poor, sadly, sorry, sad, weight, harass, regret, miserable
	surprise	.068	amazing, fantastic, wonderful, surprisingly, surprise, wonder, beat, amaze, amazingly

Table 2: Emotion vector for *restaurant*, including emotion scores *S* and lemmatized top emotion terms.

the WordNet-Affect emotion vocabulary to identify text snippets about places with emotion content. Based on these matches we have generated emotion vectors for 26 place classes along six emotions: anger, fear, disgust, joy, sadness, and surprise. As a case study, we used about 100,000 travel posts, though the method can be directly applied to other kinds of corpora.

The work presented here represents a promising starting point for further research into the emotion analysis of place. However, there are several limitations and opportunities for improvement. Because simple lexical matching is used for emotion terms in the WordNet-Affect vocabulary, the model is very sensitive to polysemy and ignores compound words and expressions. Moreover, the generation of emotion scores does not consider the intensity conveyed by a term (e.g., *frustrated* carries the same weight as *furious*). The travel blog corpus is evidently biased towards tourist experiences, and the results should be compared with different corpora, containing for example news stories.

In future work, our approach will be applied to the study of spatial and temporal variation of place emotions. Spatial properties of expressions (e.g., inside, outside, above, etc.) will be considered to refine the attribution of an emotion to place classes. More importantly, we will extend the model to identify emotions beyond single tokens, looking at compound nouns and n-grams. By applying the approach to place instances, we will be able to compute an emotion-based semantic similarity of places, and to identify typical cases and outliers based on their emotional structures. The validation of the model will be conducted on observable social outcomes, such as crime and health, assessing the power and the limitations of emotion analysis applied to the experiential heterogeneity of place.

References

[1] B. Adams, G. McKenzie, and M. Gahegan. Frankenplace: Interactive thematic mapping for ad hoc exploratory search. In *Proceedings of the 24th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 2015.

[2] A. Ballatore, D. Wilson, and M. Bertolotto. A Survey of Volunteered Open Geo-Knowledge Bases in the Semantic Web. In G. Pasi, G. Bordogna, and L. Jain, editors, *Quality Issues in the Management of Web Information*, volume 50 of *Intelligent Systems Reference Library*, pages 93–120. 2013.

[3] A. Ballatore, M. Bertolotto, and D. Wilson. Linking geographic vocabularies through WordNet. *Annals of GIS*, 20 (2):73–84, 2014.

[4] M. M. Bell. The Ghosts of Place. *Theory and society*, 26 (6):813–836, 1997.

[5] M. Cataldi, A. Ballatore, I. Tiddi, and M.-A. Aufaure. Good location, terrible food: Detecting feature sentiment in user-generated reviews. *Social Network Analysis and Mining*, 3(4):1149–1163, 2013.

[6] T. Danisman and A. Alpkocak. Feeler: Emotion classification of text using vector space model. In *AISB 2008 Convention Communication, Interaction and Social Intelligence*, volume 1, pages 53–59, 2008.

[7] P. Ekman. An argument for basic emotions. *Cognition & emotion*, 6(3-4):169–200, 1992.

[8] C. Fellbaum. WordNet. In R. Poli, M. Healy, and A. Kameas, editors, *Theory and Applications of Ontology: Computer Applications*, pages 231–243. Springer, Berlin, 2010.

[9] R. Heuser and M. Algee-Hewitt. Mapping the Emotions of London in Fiction, 1700-1900: A Crowdsourcing Experiment. In P. Murrieta-Flores, D. Cooper, and C. Donaldson, editors, *Literary Mapping in the Digital Age*. Ashgate, Farnham, UK, 2015. [Forthcoming].

[10] B. Liu. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1):1–167, 2012.

[11] S. M. Mohammad. From once upon a time to happily ever after: Tracking emotions in mail and books. *Decision Support Systems*, 53:730–741, 2012.

[12] B. Pang and L. Lee. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2 (1-2):1–135, 2008.

[13] R. Plutchik. A general psychoevolutionary theory of emotion. In R. Plutchik and H. Kellerman, editors, *Emotion: Theory, Research, and Experience: Vol. 1. Theories of Emotion*, pages 3–33. Academic, New York, 1980.

[14] R. Purves, A. Edwardes, and J. Wood. Describing place through user generated content. *First Monday*, 16(9), 2011. [Online].

[15] D. R. Recupero, V. Presutti, S. Consoli, A. Gangemi, and A. G. Nuzzolese. Sentilo: Frame-based sentiment analysis. *Cognitive Computation*, pages 1–15, 2014.

[16] C. Strapparava and A. Valitutti. WordNet Affect: an Affective Extension of WordNet. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (Lisbon)*, volume 4, pages 1083–1086, 2004.

[17] K. Strongman. *The Psychology of Emotion: From Everyday Life to Theory (5th ed.)*. Wiley & Sons, Chichester, UK, 2003.

[18] Y.-F. Tuan. *Space and place: The perspective of experience*. University of Minnesota Press, Minneapolis, MN, 1977.

[19] J. Urry and J. Larsen. *The Tourist Gaze 3.0*. SAGE Publications, London, 2011.

[20] Q. Ye, Z. Zhang, and R. Law. Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. *Expert Systems with Applications*, 36 (3):6527–6535, 2009.