Is This My Place? A Twitter Dataset To Recognize The Sense Of Place

Giovanni Siragusa[1], Livio Robaldo[2], and Luigi Di Caro[1]

[1] Computer Science Department, Universitá degli Studi di Torino, Italy
{siragusa, dicaro}@di.unito.it
[2] Legal Innovation Lab Wales, Swansea University
livio.robaldo@swansea.ac.uk

Abstract. The concept of *Place* in Geography and related fields has a complex definition involving different facets, ranging from its aspect and setting (e.g., buildings, shops, and the surrounding environment), to its functional semantics and to the sentiments it may evoke. Over the last years, researchers focused on identifying explicit and hidden semantics behind places and their reciprocal similarities. In this article, we used the Cresswell's definition of *place*, which takes into account the sentiments people feel about it, i.e., the so-called *Sense Of Place* (SOP). As our contribution, we first constructed a novel Twitter-based dataset by harvesting tweets referring to 4 different cities having English as their mother-tongue language (New York, San Francisco, Wellington and Sydney). Then, the collected tweets have been labelled by three annotators in terms of expressed SOP. Finally, we used the constructed dataset to train classical Machine Learning (ML) models, evaluating them on a novel (test) set of annotated tweets regarding the city of London. Results demonstrate the validity of the SOP-based data construction in terms of both human agreements and ML accuracy levels.

Keywords: Place · Twitter · Sentiment Analysis

1 Introduction

Artificial Intelligence (AI) is experiencing growth in activity, both in academia and in industry. Current AI-based technologies mostly use Machine Learning (ML) to process documents and replicate legal decision-making.

Although ML provides valid solutions, its overall usefulness is limited in that it tends to disregard specific semantic aspects and legal reasoning.

ML is based on *statistical reasoning*: new cases are classified by similarity with the cases included in the training set. As a result, performance are intrinsically limited. Furthermore, and most important of all, as it is well-known ML tends to behave like a "black box" unable to explain its decisions and it can therefore lead to biases and other discriminatory outcomes: ML trained on biased datasets tend to replicate the same biases on new inputs.

In order to overcome the limits of ML, lot of recent research has been devoted to investigate approaches in symbolic AI. The idea is to plug into the

ML-based system human-understandable symbols, i.e., concepts and other logical constructs, that enable forms of *logical reasoning*. Besides leading to an improvement in the performance, symbolic AI is the path to explainability, needed to contrast biases in decision-making [22].

Nevertheless, as it is well-known symbolic AI is requires lot of efforts, as (human) domain experts must be involved in the creation of symbolic representations such as ontologies or logical rules. Concept are highly *context-sensitive*: their meaning can vary depending on the context in which they are used so that even domain experts can struggle to find an agreement among their definitions. This is particularly true for "general-purpose" abstract concepts such as the concept of "Place", the one on which this paper focuses on (see 1.1 below).

Therefore, the transition from ML-based systems to sustainable approaches in AI fully based on computational logic is still a very long journey. To underpin the transition, lot of contemporary approaches propose to use concepts *in combination with* the statistical inferences of ML [5]. The research presented in this paper is one of these approaches: it presents a ML-based analysis of Twitter comments centered on the concept of "Place".

1.1 The concept of "Place"

Place is a concept frequently mentioned in every-day life. We use it in sentences as *"This is my favourite place"*, *"I finally found my place in the world"*, or *"Lay a place at the table for Mr. T"*. In the commonsense language, the term *place* is used to refer to a city (e.g., New York), a public space (e.g., Central Park), a shop or even to the seat we usually take at the table. These examples suggest that the concept of *place* is broad and ranges from a punctual space to a wide area.

In the Geography domain, the concept of *place* assumes a more structured representation. In particular, according to Cresswell [12,11] this concept involves three different aspects:

- **Location:** the physical absolute point in the space, identified by a set of coordinates;
- **Locale:** the visible features and settings of a place, such as streets, shops, parks and so on;
- **Sense Of Place:** the set of emotions and feelings that a place inspires in people. These sentiments can be subjective when they are based on someone's personal biography, or they can be shared when a group of people feels the same sentiment towards a place.

According to the intuition of Cresswell, a complete definition of *place* asks for a systematic analysis of all the three aspects listed above. This analysis starts with the identification of a place to focus on, and, secondly, the collection of the features and settings that constitute the place of interest and that could have an impact on how people feel it, such as shops, house styles and streets. Such analysis could be conducted both at micro-level, analyzing a single street or house, or at macro-level, analyzing an entire district or city. Finally, the *Sense Of Place* (SOP) can be "extracted" by analyzing a large set of observations by the people living and experiencing the chosen place.

Nowadays, social networks are part of our everyday life, and people are used to disclose their opinions, activities and emotions through them. Therefore, scientific community started to consider people in social networks as social sensors for different fields such as politics, economics and sociology [18]. Thus, it is possible to harvest and analyze their posts to discover high-level aspects. Moreover, the possibility to associate a geographical reference to a post (the so called geotags) or to infer the location starting from significant hashtags allowed scientists to develop map-based data analysis which can also be used to identify disaster-affected areas or regions with high crime rates, as respectively in the works of Cerutti et al. [9] and Ristea et al. [17].

Since users' posts contain both textual and location (the geotag) information, they suit well to define the SOP of a chosen place. Thus, following the idea of Siragusa and Leone [19], in this article we present a dataset of user posts, adopting Twitter as our source of information. Such dataset could be used to train a classifier in order to discern tweets that express a general sentiment from those expressing a sentiment towards a place (i.e., the SOP). Our idea is that an automatic classifier of SOP messages will allow to swiftly obtain large quantity of data, enabling geographers and data-analysts to understand the relationship between people and places and exploring the process that shapes those relationships. For constructing the dataset, we chose four different cities (San Francisco, Wellington, Sydney, and New York) that have English as their mother-tongue language and we collected tweets referring to them. We then asked to three annotators to judge whether the tweets were expressing some SOP with respect to the selected cities. Finally, we trained 4 different state-of-the-art Machine Learning classifiers: Support Vector Machine (SVM), Naive Bayes (NB), Decision Trees (DT) and AdaBoost (AB) to recognize the SOP-tweets, showing good performance especially with SVM.

The paper is structured as follows: Section 2 describes how the tweets were collected and labelled to create the dataset; Section 3 describes how we used the dataset to train the classifiers. Section 4 contains related works concerning the extraction, labelling and analysis of *place*-based data. Section 5 concludes the article with some final remarks.

2 Tweets Extraction

For creating our dataset, we collected from Twitter[3] those tweets referring to the following 4 cities: New York, San Francisco, Wellington and Sydney. We used the Tweepy[4] library for the extraction, and we collected only the tweets containing an hashtag that represented one of the cities of interest. For each city, the hashtags we looked for were:

[3] https://twitter.com
[4] https://www.tweepy.org

New York: #NewYork, #NY, #newyork;
San Francisco: #SanFrancisco, #SF, #sanfrancisco;
Wellington: #Wellington, #wellington;
Sydney: #Sydney, #NSW, #sydney.

We collected a total of more than 2 millions of tweets, even if with a varied distribution over the 4 cities. Then, we applied to each tweet the sentiment classifier of TextBlob[5] (because the SOP is strictly entwined with the sentiment), obtaining its sentiment expressed in the range [-1.0, 1.0], with 0 indicating a neutral tweet. Using the sentiment scale, we filtered out all the tweets that did not express a (positive or negative) sentiment, i.e only those with a zero score. We decided to keep the tweets that are close to zero (e.g., 0.2) since they may express a SOP. Table 1 reports the number of remaining tweets for each city.

City	# Total Tweets	# Tweets with sentiment
San Francisco	10,631	1043
New York	1,274,647	422
Sydney	24,761	1223
Wellington	1,003,007	518
Total	2,313,046	3206

Table 1. The table reports the number of collected tweets of each city and the number of tweets expressing some sentiment (positive or negative).

We then asked to three annotators to make a judgement, under the form of Y (corresponding to *Yes*) and N (corresponding to *No*), to find the tweets expressing a SOP. In particular, the annotators had to follow the following guidelines:

- a tweet is tagged with the label Y if it expresses a feeling, an emotion or a sentiment (either positive or negative) towards a city or an area inside it (e.g., a neighborhood or a park);
- a tweet is tagged with the label N if it does not express any feeling, emotion or sentiment (either positive or negative) towards a city or an area inside it;
- a tweet that seems to have been written to promote trips to a city or to promote tourist activities should not be tagged with Y since it does not correspond to a sincere emotion, having instead only advertising purposes;
- a tweet expressing a political exposure towards the people who govern the city does not express a feeling towards the city itself, so it should be labelled with N.

We computed the inter-annotators agreement using Fleiss' kappa coefficient [13], and the agreement of each annotators pair using Cohen's kappa coefficient [10]. Figure 1 reports Fleiss's kappa on each city, where San Francisco resulted to be the city with the highest agreement (close to 0.7) between the three annotators

[5] https://textblob.readthedocs.io/en/dev/

according to the SOP expressed by the corresponding tweets. Figures 2, 3 and 4 report Cohen's kappa coefficient for each couple of annotators. From these, we can notice that Annotator 1 and Annotator 3 had high agreement, in contrast with Annotator 2.

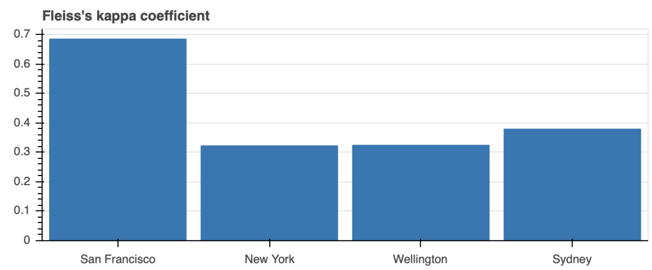

Fig. 1. The images shows the Fleiss' kappa coefficient for all the four cities.

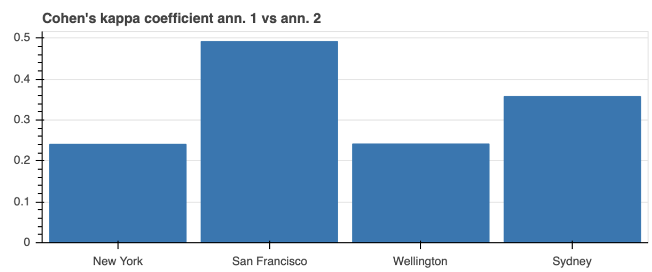

Fig. 2. The image shows the Cohen' Kappa for annotator 1 and annotator 2.

We assigned the label Y or N with each tweet using a majority vote scheme (i.e., if two annotators expressed Y on the same tweet, we labelled it with Y to indicate a SOP, and viceversa). Table 2 shows some tweets with their annotation. The dataset can be found at the following link: https://bit.ly/3s5KF58.

Table 3 reports the number of tweets of each city that have a sentiment, and how many of them actually express a SOP after the application of the majority vote scheme.

Finally, we analyzed the most frequent words in both SOP and non-SOP tweets. More in detail, we applied the following pre-processing pipeline: first, we removed URLs, names (words starting with @) and ReTweets (RT) tags, keeping only those hashtags linking to salient information; then, we lowercased all words and stemmed them, using the NLTK library[6]. We then generated a wordcloud

[6] https://www.nltk.org

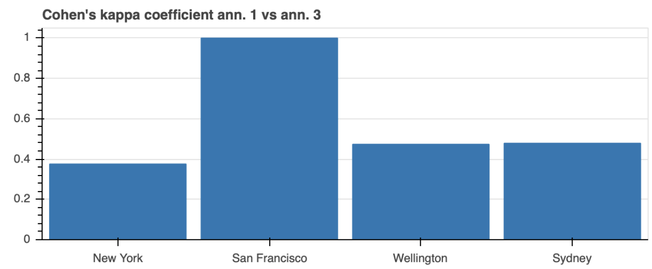

Fig. 3. The image shows the Cohen' Kappa for annotator 1 and annotator 3.

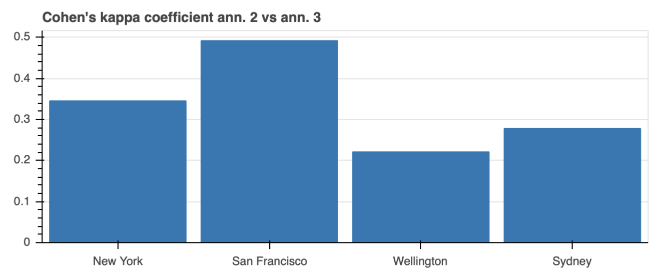

Fig. 4. The image shows the Cohen' Kappa for annotator 2 and annotator 3.

using the library WordCloud[7]. Figures 5 and 6 show the most frequent words for both kinds of tweets. Comparing the two figures, we can see that both contain words conveying sentiments as *love*, *amazing*, *happy*, *beautiful*, and so forth. However, SOP-tweets also contain words such as *love people*, *people living*, *place*, *spot* and *live*, highlighting how those tweets are more focused on the city, its districts and citizens; non-SOP tweets, instead, show more general sentiments.

3 Automatic Detection of SOP

In this article, one contribution regards the idea of automatically distinguishing tweets expressing a general sentiment from those having a Sense of Place (SOP).

We used the dataset described in Section 2 to perform two types of evaluation. To check if it is possible to train a classifier to recognize other tweets that express a SOP, we performed a 5-fold cross evaluation training the classifier on 4 folds and testing on the remaining one. We processed each tweet in the following way: first, we removed names (words having @ in the tweets), URLs, and RT (ReTweetted) tags, leaving the hashtags since they sometimes express a sentiment towards a place; finally, we lowercased all words. The resulting tweets are then passed to a pre-trained Sentence-Bert [16] model in order to obtain their vector representation, which is used to train these classifiers:

[7] https://amueller.github.io/word_cloud/

Tweet	Label
Great weekend with Hannah Nguyen-Chang in #SF. Think we over-done it this Saturday as she's experiencing lots of pa...	SOP
We are on sale in #SanFrancisco Access Hollywood Live (NBC) says "You HAVE to check them out! That was tremendous!"	non-SOP
From https://t.co/MDjtLDWTeX - One of the greatest things about #Wellington is that it's full of art. As Jeff Tweedy of @people said, "I think art is a consolation regardless of its content. It has the power to move and make you feel like you're not."	SOP
Hype is getting hyped this weekend! We are so grateful to have fantastic customers in #Newyork, #Sanfrancisco and...	non-SOP

Table 2. The table reports some tweets with their label.

City	#Tweets	#SOP-Tweets
San Francisco	1043	171
New York	422	7
Sydney	1223	73
Wellington	518	42
Total	3206	293

Table 3. The table reports, for each city, the number of tweets expressing a sentiment and the number of tweets that contain a Sense Of Place.

- a Support Vector Machine (SVM) with radial basis function kernel;
- a Gaussian Naive Bayes (NB);
- a Random Forest (RF) classifier with a minimum of five examples for each leaf. It has also the advantage of explaining why a tweet is classified as SOP or non-SOP since each node of the tree is composed of simple human-readable rules;
- and an AdaBoost (AB) classifier.

For each classifier, we performed a grid search to find the optimal parameters, using the Scikit-learn[8] implementation of the classifiers.

We compared them with two baseline: i) a random classifier which randomly assigns the label SOP and non-SOP to tweets; and ii) a majority class classifier which assigns the majority class (non-SOP label) to all tweets. Table 4 reports the Accuracy, Precision, Recall and F-measure of each classifier. We also considered the dataset unbalance in computing the metrics. From the table, we can notice that both SVM and RF are mostly able to recognize SOP-tweets.

We also tested the classic Term Frequency - Inverse Document Frequency (TF-IDF) vectorization to train the classifiers, adding two further steps to the cleaning pipeline: stopwords removal and word stemming. The classifiers trained with this latter vectorization performed worst on all metrics, loosing about 2

[8] https://scikit-learn.org/stable/

Fig. 5. The image shows a wordcloud of the tweets that contain a Sense Of Place.

Classifier	Accuracy	Precision	Recall	F-measure
Random	50.0	50.0	50.0	50.0
Majority Class	0.17	0.001	0.17	0.003
SVM	**83.45**	**83.80**	**81.39**	**83.45**
Gaussian NB	79.35	79.75	79.36	79.28
AB	75.08	75.17	75.09	75.06
RF	80.71	80.97	80.72	80.67

Table 4. The table reports the cross-fold evaluation of the classifiers.

percentage points for SVM and RF, and about 4 percentage points for Gaussian NB and AB.

The problem of the previous evaluation is that we do not know how well a classifier trained on the dataset will perform on new tweets. For this reason, we collected an additional set of tweets about the city of London using the following hashtags: *#London*, *#LND* and *#london*. Following the same steps explained in Section 2, we extracted only those having a sentiment, obtaining a dataset composed of 1388 tweets. We then asked to the three annotators to judge them following the guidelines provided for the previous labelling task. After applying the majority voting scheme, we obtained that only 73 tweets expressed a SOP. The Fleiss' kappa coefficient [13] for the three annotators was a bit lower, i.e. 0.34, with a Cohen's kappa of 0.4 for Annotators 1 and 3, 0.15 for Annotators 1 and 2, and 0.23 for Annotators 2 and 3.

Finally, we trained the classifiers on the 4-cities dataset and then tested them on the new London tweets. Table 5 reports the results, where the Majority Class

Fig. 6. The image shows a wordcloud of non-SOP tweets.

classifier shows the highest accuracy due to the high unbalance of the dataset towards the non-SOP label.

We can notice that the Gaussian NB classifier, despite ranked third in the previous evaluation, performs well in the London dataset, meaning that it is able to transfer its knowledge on SOP tweets on other datasets; AB classifier, instead, has poor performance on both datasets, demonstrating to be not suitable for this task. Finally, the SVM classifier performs well against the other ones for both datasets; however, it has a drop in performance for SOP-tweets only, obtaining 21.78 for the Precision, 84.72 for the Recall and 34.66 for the F-measure. Using the TF-IDF vectorization, we noticed a drop in the F-measure of 2 percentage points for all classifiers compared to the ones of Table 5. These results demonstrate the complexity of the SOP classification task, being more fine-grained with respect to the standard sentiment detection.

Classifier	Accuracy	Precision	Recall	F-measure
Random	50.0.	50.0.	50.0	50.0
Majority Class	**94.8**	47.4	50.0	48.7
SVM	83.41	60.39	**84.03**	62.58
Gaussian NB	86.17	**61.22**	81.54	**64.33**
AB	75.19	56.92	78.38	55.33
RF	84.57	60.08	80.04	62.41

Table 5. The table reports the results on London dataset.

We then analyzed the errors made by the SVM classifier (false positives and false negatives). We found that the false positives are misclassified due to the cleaning phase. In detail, since we clean the tweet removing names and RT tags, the resulting text could resemble the one of a SOP-tweet, especially in the case of TF-IDF vectors where the stopwords are removed. Table 6 reports some examples.

Tweet	Clean Text	TF-IDF's Bag-Of-Words
Loved every minute of being on the #panel - what an amazing bunch!! @UELAlumni always delivers the best!! #student #alumni #university #London	`loved every minute of being on the #panel - what an amazing bunch ! ! always delivers the best ! ! #student #alumni #university #london`	`love everi minut #panel amaz bunch alway deliv best #student #alumni #univers #london`
#LONDON: Thanks London its been emotional! See you all again next year #WeRunThisCity	`#london : thanks london its been emotional ! see you all again next year #werunthiscity`	`#london thank london emot see next year #werunthisc`
#london #londonpride highly appreciated the hospitality! See you soon #england	`#london #londonpride highly appreciated the hospitality ! see you soon #england`	`#london #londonprid highli appreci hospit see soon #england`

Table 6. The table reports some tweets of the London dataset classified as false positive, and their pre-processed text passed to Bert.

Concerning the false negatives, this is the most interesting case. We found two types of error (reported in Table 7):

Annotation error: there are tweets that have been classified as SOP in the dataset, but they do not contain a sentiment towards a place; thus the classifier correctly recognized them. These are borderline cases, where the sentiment expressed by the user regards some activity they performed (or attended) at the city; for instance, the tweet "*It's going to be a great weekend @LondonTattooCon look forward to seeing the best #tattooartist*" has been labelled as SOP by the annotators, but it expresses a positive sentiment regarding the intention to attend the tattoo convention. Since these tweets are complex cases, our objective as future work is to discuss with annotators and linguistic experts in order to improve the guidelines and the quality of the dataset;

Classification error: similarly to the false positive cases, some tweets resemble the negatives ones because they do not have any word expressing a clear sentiment towards a place.

Tweet	Error type
It could be a magical place to go if you fancy a cheeky Butterbeer or some Pumpkin Juice. #HarryPotter	classification
Lovely morning to start the 2nd #DevOpsDays in #London as we talk all things #DevOps again.	annotation
Spent a great #tabletopRPG evening NMDLondon tonight Too many AWESOME GMs to fit in one video #NoMoreDamsels #Dung...	annotation

Table 7. The table reports some tweets of the London dataset classified as false negative.

4 Related Works

The specification of a geographical reference in a shared post is nowadays an habit for the users of the most famous social networks as *Twitter*[9], *Facebook*[10] and *Instagram*[11]. The geographical references can appear in a post either in the form of a geotag or an hashtag that expresses the name of the location (a city, a restaurant, a public park, etc.). Such information have been used by scientists in order to develop geographical and temporal based studies in the field of Geography and Data Science. In addition to these well-known platforms, other new location-based services emerged in the last years. Among them, we mention *Trendsmap*[12] which shows on a map the latest trends emerging from Twitter, *Ushahidi*[13] which collects and visualises information about crisis witnesses providing the users the possibility to respond and *FixMyStreet*[14] which allows the UK citizens to signal streets problems (pot holes, unsafe walls, not working lampposts) to the local authorities. *FirstLife*[15] [3] is a more interactivity-oriented service which focuses its attention on the user intended as citizen, giving them the possibility to interact with a map on which they can share events, news and even aggregate people. Moreover, the data are associated with a temporal dimension which allows users to filter and order the information according to time.

In the context of geo-oriented social networks, it is easy to define the concept of place expressed by Cresswell [11,12]. Both *location* and *locale* are coded in the geographical map showed to the user, while the *sense of place* (SOP) is expressed by the user via posts. Furthermore, the concept of place has evolved together with technology, becoming the conceptual fusion of space and experience [15]. The advent of Covid-19 has fasted the projection of the concept of place from the physical world to the virtual one, since our freedoms of movement were taken away [6]; For instance, virtual meeting allowed to gaze the interior of our boss' (or employee) house, disclosing aspects of one another's home that reveal

[9] www.twitter.com
[10] www.facebook.com
[11] www.instagram.com
[12] www.trendsmap.it
[13] www.ushahidi.com
[14] www.fixmystreet.com
[15] www.firstlife.org

the sense of place [14]. For these reasons, we decided to use Twitter as source to extract SOP messages, where the hashtag of a city (e.g., #tokyo) defines the space boundaries of the place, and the post content its sense. Our work is similar to Siragusa and Leone [20], where the authors used Latent Dirichlet Allocation [7] (LDA) to extract those tweets that contain the sense of place. Instead of applying the LDA model, we decided to create a SOP dataset by hiring annotators; such dataset could be used to train a classifier to recognize SOP tweets or to conduct data-analysis.

Other works explored both the use of Twitter and the concept of place. Besbris et al. [4] studied the spatial stigma, i.e. how the neighbours of a place perceive it in a negative way due to crime, disorders, poverty and even racial isolation; and the residents of such place may embody the negative characteristics. Zhu et al. [23] tried to quantify the semantic of places according to their interaction with streets. They found that the interaction is beneficial to identify the semantic. Sakaki et al. [18] used the intuition of considering the users as social sensors in order to implement event detection. Cataldi et al. [8] extracted in real time the most emerging topics expressed by the community based on the interests of a specific user in a particular temporal frame, while Allisio et al. [2] exploited the temporal and spacial information associated with the tweets in order to produce a daily estimation of the degree of happiness of the main Italian cities. Adams et al. [1] developed a system based on Latent Dirichlet Allocation [7] (LDA) to extract place characteristics from a travel blog. They also correlated those characteristics to find similar places. Steiger et al. [21] applied LDA on collected tweets to extract real-world characteristics. They also tracked the topics frequency along a week, finding that people tend to talk about *home* and *work*.

5 Conclusion

In this article, we presented a dataset of tweets that express a Sense Of Place (SOP). We also showed that is possible to train a classifier on such dataset to label new tweets.

We started from the idea presented in Siragusa and Leone [19], where the authors used Twitter as a resource to collect users' posts and then find those ones that express a SOP. Differently from them, we employed three annotators to label all the tweets. Then, we created a dataset taking those tweets labelled as expressing or not a SOP related to a city of reference. Finally, we trained four Machine Learning classifiers, showing both it is possible to automatically recognize the SOP and that it is a difficult task. However, as explained in the introduction, most of the tested state-of-art Machine Learning classifiers behave as "black boxes", being unable to explain why a tweet has been classified with a SOP (or non-SOP) label. The only exception is the Random Forest, which produces simple rules (e.g., length of the sentence greater than a threshold) connected via conjunction or disjunction logical operators. Although these rules can be read by human experts, they tend to be too specific or situational, making difficult to generalize them.

As future work, we would like to extend the size of the dataset in order to train Neural Networks models. We think that these models could suit well for this kind of task since they are capable to deeply analyzing the meaning of each word of the sentence and its interaction with other ones. A preliminary result of the positive impact of these models has been showed by the Bert vectorization of the tweets. Another research path is the adoption of symbolic AI in conjunction with Machine Learning models to improve both their accuracy and explainability. In this way, the rules defined by the models could be used by data-analysts and geographers to explain why a tweet has been labelled with a certain tag. Furthermore, such rules will allow to understand the errors performed by the classifier, either improving it or the annotation process by refining the guidelines.

References

1. Adams, B., McKenzie, G.: Inferring thematic places from spatially referenced natural language descriptions. In: Crowdsourcing geographic knowledge, pp. 201–221. Springer (2013)
2. Allisio, L., Mussa, V., Bosco, C., Patti, V., Ruffo, G.F.: Felicittà: Visualizing and estimating happiness in italian cities from geotagged tweets. In: CEUR WORKSHOP PROCEEDINGS. vol. 1096, pp. 95–106. CEUR Workshop Porceedings (2013)
3. Antonini, A., Boella, G., Buccoliero, S., Calafiore, A., Di Caro, L., Giorgino, V., Ruggeri, A., Salaroglio, C., Sanasi, L., Schifanella, C.: First life, from the global village to local communities. In: 1st IASC Thematic Conference on Urban Commons (2015)
4. Besbris, M., Faber, J.W., Rich, P., Sharkey, P.: The geography of stigma: Experimental methods to identify the penalty of place. In: Audit studies: Behind the scenes with theory, method, and nuance, pp. 159–177. Springer (2018)
5. Bibal, A., Lognoul, M., de Streel, A., Frénay, B.: Legal requirements on explainability in machine learning. Artificial Intelligence and Law **29**(2) (2021)
6. Bissell, D.: A changing sense of place: Geography and covid-19. Geographical Research **59**(2), 150–159 (2021)
7. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. Journal of machine Learning research **3**(Jan), 993–1022 (2003)
8. Cataldi, M., Caro, L.D., Schifanella, C.: Personalized emerging topic detection based on a term aging model. ACM Transactions on Intelligent Systems and Technology (TIST) **5**(1), 7 (2013)
9. Cerutti, V., Fuchs, G., Andrienko, G., Andrienko, N., Ostermann, F.: Identification of disaster-affected areas using exploratory visual analysis of georeferenced tweets: application to a flood event. Association of Geographic Information Laboratories in Europe: Helsinki, Finland p. 5 (2016)
10. Cohen, J.: A coefficient of agreement for nominal scales. Educational and psychological measurement **20**(1), 37–46 (1960)
11. Cresswell, T.: Place. In: International Encyclopedia of Human Geography. vol. 8, pp. 169–177. Elsevier (2009)
12. Cresswell, T.: Place–part i. The Wiley-Blackwell companion to human geography pp. 235–244 (2011)

13. Fleiss, J.L.: Measuring nominal scale agreement among many raters. Psychological bulletin **76**(5), 378 (1971)
14. Hay, I.: Zoom and place: Video conferencing and virtual geography. South Australian Geographical Journal **116**(1), 7–11 (2020)
15. Merschdorf, H., Blaschke, T.: Revisiting the role of place in geographic information science. ISPRS International Journal of Geo-Information **7**(9), 364 (2018)
16. Reimers, N., Gurevych, I.: Sentence-bert: Sentence embeddings using siamese bert-networks. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics (11 2019), `http://arxiv.org/abs/1908.10084`
17. Ristea, A., Kurland, J., Resch, B., Leitner, M., Langford, C.: Estimating the spatial distribution of crime events around a football stadium from georeferenced tweets. ISPRS International Journal of Geo-Information **7**(2), 43 (2018)
18. Sakaki, T., Okazaki, M., Matsuo, Y.: Earthquake shakes twitter users: real-time event detection by social sensors. In: Proceedings of the 19th international conference on World wide web. pp. 851–860. ACM (2010)
19. Siragusa, G., Leone, V.: Such a wonderful place: extracting sense of place from twitter. In: European Semantic Web Conference. pp. 397–405. Springer (2018)
20. Siragusa, G., Leone, V.: Such a wonderful place: extracting sense of place from twitter. In: European Semantic Web Conference. pp. 397–405. Springer (2018)
21. Steiger, E., Westerholt, R., Resch, B., Zipf, A.: Twitter as an indicator for whereabouts of people? correlating twitter with uk census data. Computers, Environment and Urban Systems **54**, 255–265 (2015)
22. Waltl, B., Vogl, R.: Explainable artificial intelligence - the new frontier in legal informatics. Jusletter IT. **22** (2018)
23. Zhu, R., McKenzie, G., Janowicz, K.: Are streets indicative of place types? (2019)