ItaGLAM: A corpus of Cultural Communication on Twitter during the Pandemic

Gennaro Nolano, Carola Carlino, Maria Pia di Buono, Johanna Monti
UniOr NLP Research Group
"L'Orientale" University of Naples
Italy
{gnolano, ccarlino, mpdibuono, jmonti}@unior.it

Abstract

This paper describes the compilation and annotation of **ItaGLAM**, a corpus of tweets written by Italian Galleries, Libreries, Archives and Museums (GLAMs) during the lockdown period in Italy due to the COVID-19 pandemic. ItaGLAM has been annotated with a set of labels which may be useful to identify different types of communication. Furthermore, the collected data have been used to train a set of classifiers.

The results are analyzed to evaluate the information flow between GLAM and users and to analyze cultural communication on the Web.

1 Introduction

Over the last years, Social Networks have become one of the most popular platforms for sharing experiences and opinions through the use of simple strings of text (Zhao and Rosson, 2009). Indeed, this way of communicating has become an essential interaction tool, not only among private users, but also among companies to engage with their audience and to promote their brands (Alturas and Oliveira, 2016).

The use of social networks has also been adopted by museums, that, over time, have changed their way of communicating with their audience[2]. In particular, in regards to the GLAM sector, a new trend has been observed in recent years: the use of the web as a way to create and foster an online community (Langa, 2014; Allen-Greil and MacArthur, 2010).

While during the first decade of this century museum professionals considered the exhibition of collections on social networks (Laws, 2015) as 'excessive', nowadays the use of these platforms has become the norm. As Amanatidis et al. (2020) pointed out in their study about the use of social networks (and in particular Instagram) by museums in the Greek culture scene: 'social media has become a key factor in the way that cultural organizations communicate with their public in supporting the marketing of performing art organizations'.

Such centrality makes the Social network a potentially effective means that allows GLAMs to reach a wide and heterogeneous audience and to adapt to it. Therefore, we believe that the analysis of the cultural communication implies an analysis of how cultural corporations interact with the audience through social networks.

After considering the most used social networks (namely Facebook, Instagram, Twitter) in the cultural sector, we have decided to focus our research on the use of Twitter, which has already been proven to be a solid basis to analyze institutional communication, as Preoţiuc-Pietro et al. (2015) have highlighted.

Therefore, the main aim of our research is to investigate how Italian GLAMS have extraordinarily (Giraud, 2020) interacted with their audience during the lockdown in Italy due to the COVID-19 pandemic (NEMO - Network of European Museum Organisations, 2020), i.e. in the period from the 8th of March to the 5th of May 2020 (as per *DPCM March 11 2020*).

Over this time many cultural initiatives have been launched with the aim of strengthening the dialogue with the audience and make sure that, despite the impossibility of any kind of physical access, the connection between GLAMs and their

[2]https://www.osservatori.net/it/ricerche/comunicati\-stampa/laumento\-del\-livello\-di\-interesse\-per\-le\-attivita\-online\-dei\-musei\-incentivato\-dal\-covid\-19\-e\-gli\-investimenti\-per\-migliorare\-i\-servizi\-offerti

visitors would not be interrupted[3].

It has been observed that during the aforementioned period, while GLAMs institutions have drastically increased their use of Facebook, Instagram and Twitter, the latter one was the only Social Network for which an increase in the interaction audience-institution has been registered (Politecnico di Milano, 2020).

The study of communicative intents by GLAMs through Social Networks in the Italian language is still novel and, as such, best practices and tools to use still need to be tested and honed. In particular, there is still the need for an annotated corpus and a classifier that can be used on large amounts of data.

Despite the time frame taken into account is relatively short (covering 58 days in total), we think that investigating how Italian GLAMs used the web when it was the only form of communication at their disposal, represents a good training ground to test our practices and to train and evaluate different kinds of classifiers useful also in future works.

The paper is organized as follows: Section 2 describes the related works in the analysis of communication on Twitter by cultural institutions. Section 3 introduces the methodology used in this analysis: namely, it describes the creation of the corpus, the creation and use of the annotation set, and the training and evaluation of different classifiers. Finally, in Section 4 we explain the results of the research.

2 Related Work

The large amount of data available on Twitter makes this platform ideal for several studies. As such, during the years tweets have been used in several research projects regarding disaster response (Zahra et al., 2019), content classification (Dann, 2010; Stvilia and Gibradze, 2014) and, in particular, sentiment analysis (O'Connor et al., 2010; Gamallo and Garcia, 2014; Talbot et al., 2015). Despite these efforts, only a few studies have focused on the classification of communicative intents of organizations and institutions on Twitter, like Lovejoy and Saxton (2012) and Foucault and Courtin (2016), who focused on French tweets written during the MuseumWeek event.

Similar kinds of study can be found in researches dealing with Italian tweets, with several contributions dealing with sentiment analysis (Basile and

Nissim, 2013; Cimino et al., 2014) and automatic misogyny identification (Anzovino et al., 2018). To the best of our knowledge, no work has been done so far on communicative intent classification for Italian tweets.

3 Methodology

The task of tweet classification has turned out to be rather challenging for various reasons, many inherent to the platform itself. First and foremost, tweets are very short texts (with the maximum length of 280 characters), and with an average token count of 16.80 in our corpus.

Secondly, it is not unusual to find tweets composed only of hashtags, or URLs. While URLs by themselves are rarely if ever useful in a classification task, hashtags could represent a source of information only if they are used according to their original communicative intent or to the initiative to which they are related.

In the following subsections we describe how:

- the corpus was created;

- the annotation set has been chosen and then applied;

- the classifiers have been trained and tested.

3.1 Dataset

Because of the COVID-19 outbreak, the Italian Government (as many others around the world) imposed a lockdown policy, which lasted from the 8th March to the 5th May 2020 (58 days in total as per *DPCM March 11 2020*).

During this period of time, museums and art galleries adopted several strategies to continue engaging with their audience in order to maintain the communication alive, and to grant access to digital cultural heritage media. As already mentioned in Section 1, they increased the scope of their communication on the main social platforms, i.e. Facebook, Twitter and Instagram.

In this context, the focus of our analysis is the use of Twitter. The communication on Twitter is characterised by the use of certain `hashtags`, which have been used by GLAMs to propose several types of initiatives to their audience. Initially, the set of hashtags we used was made up of 33 hashtags promoted and used by Italian GLAMs and Italy's Ministry of Cultural Heritage and Activities (Italian: *Ministero per i Bene e le Attività Culturali e per il Turismo - MiBACT*), and

---

[3]https://icom.museum/en/news/how-to-reach-and-engage-your-public-remotely/

selected on the basis of their popularity according to the Twitter *trend topics* (TT)[4].

Among these hashtags, `#museitaliani` (and its graphic variation `#museiitaliani`) is the only one already existing before the pandemic, and subsequently adapted by museums for the initiatives proposed during the pandemic; while others, such as `#artyouready` and `#emptymuseum` have been created *ad hoc* during the lockdown period to describe specific initiatives.

By using these hashtags as a queue in the public Twitter API[5] we have created a corpus with a total of 23,716 tweets.

To better focus on the tweets and their intents concerning cultural communication, we have decided to filter out of the corpus any hashtag with less than 1,000 occurrences. We have thus obtained a queue of six hashtags (`#artyouready`, `#emptymuseum`, `#museitaliani`, `#museichiusimuseiaperti`, `#laculturanonsiferma`, `#laculturaincasa`) and a corpus of 15,988 tweets.

This corpus has been filtered once again so that only unique tweets (i.e. no retweets) written in Italian have been kept. By using a list of GLAMs manually extracted from the corpus, we have then extracted out of the remaining 8,038 tweets those written by a GLAM institution, thus ending up with our final corpus of 3,429 tweets published by 213 Italian cultural institutions. Table 1 shows the occurrences of the hashtags in the final corpus.

Hashtag	# Occ.
`#artyouready`	367
`#emptymuseum`	373
`#museitaliani`	906
`#museichiusimuseiaperti`	1560
`#laculturanonsiferma`	668
`#laculturaincasa`	283
Total	**4,157**

Table 1: Number of occurrences for each hashtag.

3.2 Annotation Process

In order to define the intents of GLAMs towards the users, the corpus has been annotated with four communication categories first presented by Courtin et al. (2015), and then used by Foucault and Courtin (2016), and Juanals and Minel (2018)

to annotate the information flow on a social network during a cultural event.

The annotation has been done at tweet level, using a set of labels composed as follows:

- **Sharing Experience - SE**: tweets that share an experience, an opinion or one's feeling
  Example: *Eccoci qui oggi a ricordare e a raccontare come i musei chiusi non siano chiusi e i musei vuoti non siano vuoti. Forza!*
  (Here we are today, reminding and telling how closed museums are not actually closed and empty museums are not actually empty. Come on!);

- **Promoting Participation - PP**: tweets that require some kind of activity from the users, either in real life or on-line
  Example: *Art you ready? Domani partecipa anche tu al contest di @ museitaliani condividendo con noi le tue foto dei musei privi di persone. Cerca fra i ricordi, seleziona la foto, e condividi con # artyouready # MuseumFromHome # iorestoacasa. Ti aspettiamo!*
  (Art you ready? Take part in tomorrow's @ museitaliani contest by sharing with us your photos of empty museums. Search through your memories, choose the photo and share it with # artyouready # MuseumFromHome # iorestoacasa. We are waiting for you!);

- **Interacting with the Community - ItC**: tweets through which Institutions create and foster their communities by directly interacting with the users
  Example: *Siete stati davvero tanti ad accogliere l'invito a partecipare al flashmob # artyouready e tutti avete postato foto meravigliose! Ecco i tre scatti selezionati tra i più belli*
  (So many of you accepted to take part in the # artyouready flashmob, and you all posted great photos! Here are the three shots selected among the most beautiful ones);

- **Promoting-Informing - PI**: tweets that promote or inform other users about activities, exhibitions, or about any sort of information on the museum.
  Example: *Il castello di Fénis si trova in Valle d'Aosta circondato da una doppia cinta di mura merlate è caratterizzato da torri quadrate e cilindriche con feritoie e caditoie.*
  (Fénis Castle is located in Aosta Valley, with its double crenellated surrounding walls, it is

---

[4]This process is described in details in Carlino et al. (2020)
[5]https://developer.twitter.com/en

characterized by square and cylindrical towers with loopholes and storm drains).

A fifth category **N/A** has been included in order to classify tweets that do not fit in any of the aforementioned categories, like the ones composed of only hashtags.

Following this set of categories and our guidelines, the tweets have been annotated using the open source platform INCEpTION[6], and a first round of annotation has been carried on 400 tweets, double annotated by a domain expert and a non-expert in order to calculate the Inter-Annotator Agreement (IAA).

The use of a non-expert was necessary so that the annotation would not have been influenced by any external knowledge (for example the original meaning behind the various hashtags).

The resulting Fleiss' Kappa has revealed to be moderately good at 0.629, which is considered sufficient for the task at hand. As it can be seen from the confusion matrix in Figure 1, the agreement is very strong on **PI** and **ItC**, moderately strong on **SE**, and very weak on **PP**.

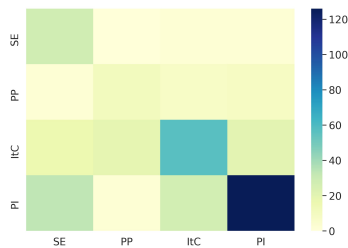Furthermore, 89 tweets have been deemed unus-

Figure 1: Confusion matrix for the agreement on every label.

able as they have been tagged with the label **N/A**, therefore, they have been removed from the corpus.

Table 2 presents the number of occurrences for each label for the remaining 3,340 tweets. These results show an issue regarding the label **PP**, that is severely underrepresented in the corpus. The effects of this underrepresentation on our classifiers will be explained in detail in Section 4, and the analysis of possible solutions will be the focus of future work.

3.3 Intent classification

In order to train the classifiers, the corpus has been preprocessed so that all tweets are lowercase, and all punctuation marks, URLs, numbers and

---

[6]https://inception-project.github.io/

Label	# Occ.	% corpus
SE	843	25.24%
PP	165	4.94%
ItC	1110	33.23%
PI	1222	36.59%

Table 2: Number of occurrences and percentage in the corpus for each label.

stopwords[7] removed. The cleaning process has been done via the NLTK package for Python[8], which has also been used for tokenization.

The experiments have been conducted on six classifiers: five more traditional classifiers trained on a TF-IDF vectorized text (created using the machine learning library for Python Scikit-learn[9]), and a Feed Forward Neural Network[10] created with Keras[11] and trained on a 100-dimensions GlOve[12] embedded text.

The set of classifiers is thus the following: a Naive Bayes (**NB**, also used as baseline); a Support Vector Classifier (**SVC**); a the K-Nearest Neighbors classifier (**KNN**); a Decision Tree (**DT**); a Multilayer Perceptron (**MLP**) and a Neural Network classifier (**NN**).

The dataset was split using the *train_test_split* tool found in the sklearn library for Python, which splits the data into random train and test subsets given a test set size. With test size set at 0.3, the training set is composed of 2,338 tweets, and the testing set is composed of the remaining 1,002 tweets.

In order to evaluate the classification task, the values of precision, recall and F1 have been all weighted by the number of samples of each label. The final results are shown in Table 3.

Classifier	P	R	F1
NB	0.69	0.66	0.64
SVC	**0.70**	**0.68**	**0.67**
KNN	0.70	0.39	0.35
DT	0.56	0.55	0.55
MLP	0.66	0.66	0.66
NN	0.64	0.63	0.63

Table 3: System results.

---

[7]The list of stopwords used is the default one for Italian found in the NLTK package. Furthermore, the term 'Twitter' has been added to this list after the first experiments.

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

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

[10]Parameters: 4 layers, dropout=0.7, Adam Optimizer

[11]https://keras.io/

[12]https://nlp.stanford.edu/projects/glove/

4 Evaluation and Result Analysis

The results show that the methodology adopted in this work can be useful in better understanding how cultural institutions communicate on the Web. The tools used in this specific task are adequate in annotating and automatically classifying the way cultural institutions communicate on the Twitter platform.

That being said, the results shown in Section 3 demonstrate that our experiments can still be improved.

Firstly, the increase in the size of the dataset would surely enhance the performances of the classifiers. In particular, this should be done focusing on the label PP, that, as it can be observed in Table 4, is the less frequent among the four.

Furthermore, while the precision for the label **PP** is usually higher than the average (note how it reaches 1.00 in our baseline), its recall is very low, even for our **SVM classifier**, which shows the best results overall. The intuition here is that, while it is usually easy for the classifiers to understand which tweet has the **PP** label, they are also very "picky", and cannot really learn all the features needed in order to classify this label against the others.

Classifier		SE	PP	ItC	PI
NB	P	0.66	**1.00**	0.58	0.61
	R	0.37	0.02	0.74	**0.76**
SVC	P	0.66	**0.88**	0.72	0.60
	R	0.48	0.52	0.65	**0.81**
KNC	P	0.36	0.69	**0.74**	0.68
	R	**0.79**	0.32	0.43	0.39
DT	P	0.47	0.43	**0.51**	0.50
	R	0.51	0.50	**0.53**	0.48
MLP	P	0.59	**0.71**	0.67	0.68
	R	0.62	0.54	**0.68**	0.67
NN	P	0.61	**0.77**	0.60	0.71
	R	0.64	0.47	**0.71**	0.61

Table 4: Precision (P) and Recall (R) for each label.

Other possible solutions to this issue can be the use of techniques such as resampling and cost-based methods.

Secondly, by focusing on the textual features of the tweets, we can further investigate where improvements can be made.

In particular, looking at the top 5 tf-idf scores for each label (Table 5), we notice that the selected hashtags may occur in all types of tweets with a low difference among their scores. Such a low deviation does not contribute enough

Token	SE	PP	ItC	PI
#museichiusimuseiaperti	**1.55**	2.47	2.12	1.99
#iorestoacasa	1.66	1.92	**1.51**	1.71
#museitaliani	2.43	2.35	**2.05**	2.33
#laculturanonsiferma	2.88	2.47	2.83	**2.35**
#emptymuseum	3.02	**1.97**	3.29	3.4
#artyouready	2.9	**1.71**	3.04	3.23
#laculturaincasa	4.1	-	3.47	**2.96**
flashmob	-	**2.1**	-	-
mibact	3.34	**2.4**	2.7	3.09
oggi	3.03	-	-	**2.78**
youtube	-	-	3.16	-
cultura	4.18	-	**3.12**	4.18

Table 5: Top 5 word by their tf-idf score on each label.

to the classification process, as shown by `#museichiusimuseiaperti` values which are seemingly strong enough as a feature to differentiate **PP** against the others, but does not do a good job differentiating the other labels against each other.

Those data could give us some insight on how museums communicate through the Twitter platform. Indeed, usually, GLAMs tend to use the same hashtags regardless of their communicative intents (even when the hashtag used was initially linked to certain initatives), which was already expected with some general hashtags, like `#iorestoacasa`.

The effects of possible removal or reweighting of these hashtags needs to be further explored.

5 Conclusion and Future Work

In this work, we have described our project for classifying communicative intents in tweets written by Italian GLAMs during the COVID-19 lockdown. Through the experiments and the following analysis we have shown how this task can be challenging.

As future work we will focus on: increasing the size of the corpus, integrating statistical techniques to help dealing with imbalanced labels, and finally improving the selection and reweighting of the features (in particular concerning the hashtags). Another topic which needs further investigation concerns the use of different kinds of textual embeddings, which might improve the result.

Once honed, the metholody and the tools we have used in this research could become an important asset in better understanding and analyzing cultural communication on the Web.

Acknowledgments

Authorship Attribution is as follows: Gennaro Nolano is author of Section 3.3 and 4, Carola Carlino is author of Section 3, 3.1 and 5, Maria Pia di Buono is author of Section 2 and 3.2, and Johanna Monti is author of Section 1.

References

Dana Allen-Greil and Matthew MacArthur. 2010. Small towns and big cities: How museums foster community on-line. *Museums and the Web 2010: Proceedings. Toronto: Archives & Museum Informatics. Published March*, 31(2010):84–96.

Bráulio Alturas and Liliana Oliveira. 2016. Consumers using social media: impact on companies' reputation. In *Proceedings of the Academy of Marketing Conference 2016: Radical Marketing*. Academy of Marketing.

Dimitrios Amanatidis, Ifigeneia Mylona, Spyridon Mamalis, and Irene Eirini Kamenidou. 2020. Social media for cultural communication: A critical investigation of museums' instagram practices. *Journal of Tourism, Heritage & Services Marketing*, 6(2):38–44.

Maria Anzovino, Elisabetta Fersini, and Paolo Rosso. 2018. Automatic identification and classification of misogynistic language on twitter. In *Natural Language Processing and Information Systems*, pages 57–64.

Valerio Basile and Malvina Nissim. 2013. Sentiment analysis on italian tweets. In *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*.

Carola Carlino, Gennaro Nolano, Maria Pia di Buono, and Johanna Monti. 2020. Laculturanonsiferma–report su uso e la diffusione degli hashtag delle istituzioni culturali italiane durante il periodo di lockdown. *arXiv preprint arXiv:2005.10527*.

Andrea Cimino, Stefano Cresci, Felice Dell'Orletta, and Maurizio Tesconi. 2014. Linguistically–motivated and lexicon features for sentiment analysis of italian tweets. In *The 4th Conference for Evaluation of NLP and Speech Tools for Italian (EVALITA)*.

Antoine Courtin, Brigitte Juanals, Jean-Luc Minel, and Mathilde de saint leger. 2015. A tool-based methodology to analyze social network interactions in cultural fields: The use case "museumweek". In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*.

Stephen Dann. 2010. Twitter content classifcation. In *First Monday*, volume 15.

Nicolas Foucault and Antoine Courtin. 2016. Automatic classification of tweets for analyzing communication behavior of museums. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 3006–3013, Portorož, Slovenia, May. European Language Resources Association (ELRA).

Pablo Gamallo and Marcos Garcia. 2014. Citius: A naive-Bayes strategy for sentiment analysis on English tweets. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 171–175, Dublin, Ireland. Association for Computational Linguistics.

Claudia Giraud. 2020. Il Ministero della Cultura presenta un report sulla reputazione online dei musei durante il Covid, 07. Available online at `https://www.artribune.com/professioni-e-professionisti/2020/07/il-ministero-della-cultura-presenta-un-report-sulla-reputazione-online-dei-musei-durante-il-covid/`, last accessed 10/09/2020.

Brigitte Juanals and Jean-Luc Minel. 2018. Analysing cultural events on twitter. In Irene Garrigós and Manuel Wimmer, editors, *Current Trends in Web Engineering*, pages 213–222, Cham. Springer International Publishing.

Lesley A Langa. 2014. Does Twitter help museums engage with visitors? In *iConference 2014 Proceedings*.

Ana Sánchez Laws. 2015. *Museum websites and social media: issues of participation, sustainability, trust and diversity*, volume 8. Berghahn Books.

K. Lovejoy and Gregory D. Saxton. 2012. Information, community, and action: How nonprofit organizations use social media. In *Economic & Social Impacts of Innovation eJournal*.

NEMO - Network of European Museum Organisations. 2020. Survey on the impact of the COVID-19 situation on museums in Europe: Final Report, 12. Available online at `https://www.ne-mo.org/fileadmin/Dateien/public/NEMO_documents/NEMO_COVID19_Report_12.05.2020.pdf`, last accessed 10/09/2020.

Brendan O'Connor, Ramnath Balasubramanyan, Bryan R. Routledge, and Noah A. Smith. 2010. From tweets to polls: Linking text sentiment to

public opinion time series. In *ICWSM*, volume 11, pages 1–2.

Politecnico di Milano. 2020. La reputazione online di musei, parchi archeologici, istituti e luoghi della cultura italiani - Report relativo al mese di maggio 2020., 5. Available online at `http://musei.beniculturali.it/notizie/notifiche/report-maggio-2020-sulla-reputazione-online-dei-musei`.

Daniel Preoţiuc-Pietro, Vasileios Lampos, and Nikolaos Aletras. 2015. An analysis of the user occupational class through Twitter content. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1754–1764, Beijing, China. Association for Computational Linguistics.

Besiki Stvilia and Leila Gibradze. 2014. What do academic libraries tweet about, and what makes a library tweet useful? In *Library & Information Science Research*.

Ruth Talbot, Chloe Acheampong, and Richard H Wicentowski. 2015. Swash: A naive bayes classifier for tweet sentiment identification. In *SemEval@NAACL-HLT*.

Kiran Zahra, Muhammad Imran, and Frank Ostermann. 2019. Automatic identification of eyewitness messages on twitter during disasters. In *Information Processing and Management*, volume 57.

Dejin Zhao and Mary Beth Rosson. 2009. How and why people twitter: the role that micro-blogging plays in informal communication at work. In *Proceedings of the ACM 2009 international conference on Supporting group work*, pages 243–252.