Impact Estimation of Emergency Events Using Social Media Streams

(Article begins on next page)

Impact Estimation of Emergency Events Using Social Media Streams

Giacomo Blanco* , Edoardo Arnaudo*† , Dario Salza* , and Claudio Rossi*

* LINKS Foundation
Turin, Italy
{*name*}.{*surname*}@*linksfoundation.com*

† Politecnico di Torino
Turin, Italy
{*name*}.{*surname*}@*polito.it*

***Abstract*—In recent years, Social Media platforms have attracted millions of users, becoming a primary communication channel. They offer the possibility to massively ingest and instantly share big volumes of user-generated content before, during, and after emergency events. Being able to accurately quantify the impact of such hazardous events could greatly help all organizations involved in the emergency management cycle to adequately plan the required recovery operations. In this work, we propose a novel Natural Language Processing approach built on rule-based algorithms able to estimate, from tweets posted during natural hazards, the impact of emergency events in terms of affected population and infrastructures. We implement our approach in an operational environment and present its validation on a publicly released dataset of more than 1.4K manually annotated tweets, showing an overall weighted F1 score of 0.77.**

***Index Terms*—Social media, impact estimation, emergency events, natural language processing, rule based algorithm**

I. Introduction

Natural disasters have become increasingly frequent in recent years, together with their impact on infrastructures and population [1]. During emergencies, information is the key component leading towards situational awareness and significantly contributing to a more efficient and effective organization of disaster management operations. In the aftermath of a disaster, official sources often produce reliable reports upon time-consuming on-site inspections, providing manually validated data, albeit with a significant delay.

However, in recent years, the surge of mobile devices and social platforms has indirectly enabled a large flow of information with capillary distribution around the globe.

Monitoring social media usage during disasters has been proven to be effective in estimating some important properties of the event obtaining important results in extraction of situational awareness information [2] and understanding natural hazard effects on society such as business recovery time after disaster [3]. Provided that an accurate geocoding mechanism of the tweets is provided, it is also possible to precisely locate the area affected by the event in a timely manner: by analysing location of tweet messages posted during an earthquake it is possible to estimate the impact area [4] and even its epicenter with high accuracy [5]. Similar outcomes can be obtained for other hazard types, such as hurricanes, where for instance the activity of Twitter users is often positively correlated with their path [6]. The volume of disaster-related tweets produced in an area is shown to be strongly correlated also with the level of damage caused by the event itself [7]–[10]. However, most of these solutions show two main drawbacks: first, they tend to focus on large-scale information, extracting estimates and averages from a global aggregation, losing smaller, yet still relevant impact data. Second, they are often specifically designed to deal with a particular event type, making them less effective in different domains.

Given these issues, in this work we propose a system able to analyze a collection of tweets related to the same event providing both large and small-scale estimates of damages provoked by hazard, intended as the number of affected entities, following a purpose specific taxonomy (see Table I).

The system has not been specifically designed for a certain kind of disaster, but instead to be general enough to handle hazards of any type. We evaluate the performances of the proposed system on a set of manually annotated tweets posted during several natural hazards, providing both overall performances and per-disaster evaluations. Our contributions can thus be summarized as follows:

- we propose a generic and modular impact estimation system which processes disaster-related tweets to provide both large- and small-scale damage estimates of hazardous events.
- we thoroughly evaluate our approach on a collection of

TABLE I
Impact Estimation taxonomy

Infrastructure	***Population***
Road	Dead
Bridge	Injured
Residential	Missing
Power Network	Evacuated
Water Network	Rescued
Cultural Heritage	Hospitalized
Burned Area (Km^2)	Other

This work was developed in the context of the Horizon 2020 projects SAFERS (grant agreement n.869353), SHELTER (grant agreement n.821282) and APPRAISE (grant agreement n.101021981).

manually labeled tweets, made available to the public[1], highlighting strengths and shortcomings of our approach.

II. Related works

A. Event Detection

By design, our impact estimation system does not require the preliminary event detection and tweet grouping to be done in any specific way. Several solutions to this problem have been devised over the years in literature, allowing for different alternatives suitable to this context. A first viable approach is proposed by [11], where tweets are collocated within a spatial grid, tracking disaster related keywords frequencies over time. A statistical-based probability is then computed for a potential disaster event. Slightly different solutions have been proposed in the context of information retrieval: following an announcement published by external and official sources, these systems monitor social media streams looking for event-related tweets. For instance, the system proposed in [12] is triggered by the detection of extreme weather events and starts monitoring a set of crisis-related keywords in the designated area. The occurrence of the event is then confirmed after a volume-based statistical test. Using a similar approach, [13] proposes a framework that tracks and retrieves useful information about a given incident using social media messages, after an emergency service has broadcast the occurrence of an incident. In our deployment we leverage the fully automated system described in [14], performing online clustering of the incoming tweets using an ensemble of textual and geographical features being able to cover with low latency a global area with local resolution.

B. Impact estimation

Several works in literature have demonstrated the benefits of using social media streams for rapid damage assessment, focusing on different aspects of impacts. Considerable research efforts have been made in damage assessment for earthquake events: a Latent Dirichlet Allocation (*LDA*) approach [15] has been employed to classify damage levels reported in tweets posted during a California earthquake, using the predictions in a damage assessment model proposed specifically for earthquakes by *FEMA*[2] to obtain a damage map. Alternatively, a system based on machine learning has been exploited to classify the damage level, on a scale from *0* to *3*, reported in damage-related tweets, testing it in the context of an earthquake [16]. A similar study [17] proposes the use of a long short-term memory (*LSTM*) model to automatically process documents describing damages sustained by buildings after an earthquake to classify the reported damage level following the *ATC-20*[3] taxonomy. The study in [18] represents the first system aimed at estimating the damage of an event with more detail: the system leverages a set of machine learning-based text classifiers, able to distinguish informative content in social media, categorize the information type and extract from these tweets potentially useful data according to the information type. A similar work has been proposed in [19], where a set of algorithms have been employed to extract from an input text pairs of type (*attribute-value*) which respectively contains the kind of affected category and its associated estimate mentioned in the text.

III. Event detection pipeline

While the proposed impact estimation system remains agnostic with respect to the preceding natural language processing pipeline, in this work we describe its use and details in a pipeline for real-time analysis from social media streams, whose general schema is reported in Fig.1. Excluding the impact estimation, the architecture is composed of three subsequent steps: data ingestion (i), information extraction (ii), and event detection (iii). The first is an open endpoint to the Twitter filtering API, which allows to receive a continuous stream of tweets containing a given set of keywords and language. The second step enriches textual information through text classification, such as information content, and Named Entity Recognition (NER), for instance the geographic location they refer to. The last module allows instead to aggregate the individual information provided by the stream of documents according to topic similarity and spatio-temporal proximity. The result is a set of clusters, representing the occurrence of disaster events of any kind. The output of this module, the collection of tweets for each detected event, is then fed to the impact estimation system which will extract impact information from every tweet to then aggregate it in second stage, producing the overall estimate associated to the event.

A. Tweet processing

In document processing step, we employ a serial pipeline following [20], composed of two serial steps: information extraction and localization. Tweets are provided by real-time data ingestion module that continuously gathers emergency-related posts, queried from *Twitter* through a predefined list of keywords for each type of supported hazard (earthquake, flood, storm and wildfire) and language (Italian, English and Spanish). Given the amount of information shared on social media, the first step is dedicated to filtering out useless and unwanted content, eliminating tweets containing a series of blacklisted words, or with a word count below a predefined threshold which has been set to 10. Each tweet is further processed to remove emoticons, URLs and other unwanted characters, then tokenized and encoded using MUSE [21], a set of pretrained FastText multilingual embeddings. The vector representations of each tweet are then fed into a series of deep learning models. First, a lightweight Convolutional Neural Network (CNN) provides a multi-label classification over a fixed taxonomy of information types, further discarding documents classified as irrelevant. Second, a Long-Short Term (LSTM) Memory network [22] extracts a series of named entities following the Ontonotes classification [23], focusing on location names. The use of an aligned latent space allows

[1] data and code available at https://github.com/gblanco10/impact-estimation
[2] https://www.fema.gov/
[3] https://www.atcouncil.org/

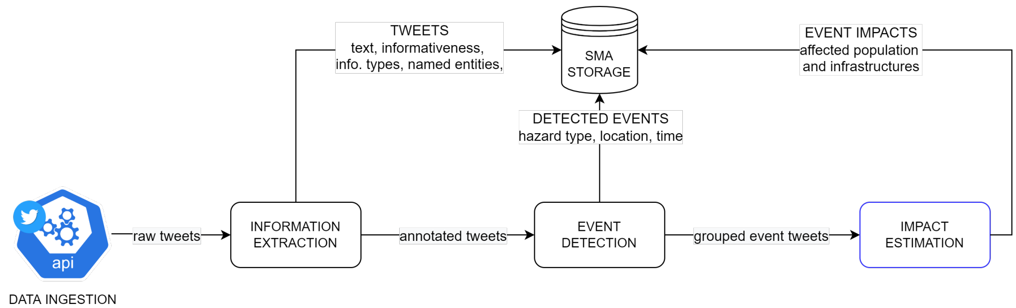


Fig. 1. Real-time pipeline for social media analysis with the employment of our modular impact estimation system.

for a more efficient training focusing on available data, without sacrificing the performance even on low-resource languages (e.g., Italian, Spanish). Last, each geographical entity detected in the text is geocoded using *OpenStreetMaps* services [24], in order to provide an estimated location for each document. The final output is thus represented by the textual content of the tweet itself, a unique identifier of the author, useful to further disambiguate content, the generated annotations and a set of geographical bounding boxes and timestamps associated with the single document.

B. Event detection

To collect tweet-level information fragments into a unique comprehensive event overview, we adopt an online clustering pipeline, following the work in [14]. The module is based on two main concepts: *clusters* and *events*. The former simply represents a collection of similar elements and is exploited as internal data structure, the latter can be seen as a *flagged* cluster that satisfies a set of required criteria, and constitutes the output of the system, that is multiple sets of tweets each associated to a hazard type, a start and end time, and a geographical area.
The detection process, parallel and independent for each language and hazard type combination, follows in the first part a greedy incremental clustering approach: tweets, received from the previous module are individually preprocessed, further extracting and refining features, and attached to one or multiple clusters according to a similarity metric based on geographical proximity and correspondence of keyword and hashtags. Cyclically, this iteration is suspended and several cluster management operations are performed, starting from defragmentation, that finds and merges independent active clusters that may describe the same event, having over time become similar or neighboring. Subsequently, non-event clusters are evaluated for flagging, considering primarily the number of unique authors in the clusters and some content information metrics. After a time period, tweets of non-event clusters expire, and are removed from the cluster, that is purged if consequently empty. Event clusters are similarly time-limited to improve content coherence, but leveraging an idling and chaining mechanism the events are unbounded in duration.

IV. Impact Estimation algorithm

The impact estimation system proposed in this work aims to precisely estimate the impact of a natural disaster of any kind in a set of categories regarding both population and infrastructures (reported in Table I), starting from a set of tweets all related to the same disaster. For each event it estimates whether a given category has been affected or not, and in case provide the number of impacted entities in the category. It works in a two stages fashion: in the first one the information about disaster impact is extracted from each tweet. During this information extraction stage, it exploits a multilingual rule-based algorithm (provided that the necessary rules have been written for the target language, currently supporting English, Italian and Spanish) which detects the mention of impacted categories in input text. The system then associates each detected category with a numeric estimate present in text and validates this association thanks to a language model. In second stage the system aggregates the fine-grained information extracted from all event tweets to provide the final estimate associated with the whole event. Fig.3 reports all operations executed to obtain the event impact estimate starting from a collection of tweets.

A. Impact extraction

The overall schema of this stage is presented in Fig.2. The process is applied independently to every single tweet belonging to event currently under analysis. The first step, right after the ingestion of real-time information, is an additional cleaning and formatting phase, while the subsequent ones extract from the cleaned but unstructured textual input a structured output containing the information about affected entities. To do this, a rule-based algorithm is employed, which provides position in text of each mention of possibly impacted category. As the actual estimate associated to a given category is seldom reported in a standard format, but rather is often written in many different forms, a candidate estimate generation step produces a set of candidate numerical expressions possibly related to each detected category. After that, possibly ambiguous assignations are resolved, outputting the final estimation for current tweet.

Concerning the first operation, input text has already been preprocessed by the information extraction step of pipeline

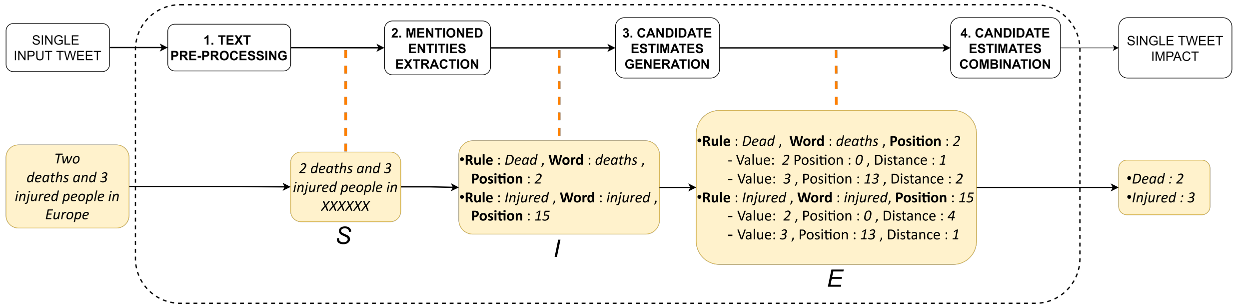


Fig. 2. Impact extraction building blocks. The yellow blocks represent the intermediate result after each block for an example input text.

in Fig.1, but additional operations specific to this phase are required for impacts extraction: any number written in characters is converted in digits and detected named entities are masked in text. In addition, a dependency parser [25] is exploited to create a word dependency graph and detect sentence boundaries (respectively G and S in Fig.3).

After preprocessing, the system identifies the categories of affected entities mentioned in text exploiting the rule-based algorithm (step 2 in Fig.2). This is applied to each input sentence and returns the positions of the keywords that triggered all the matching rules, allowing the analysis of a specific text portion to find the associated impact measure (see I in Fig.3). In order to provide an estimate for each detected impacted category, the system retrieves, once per category, the sentence in which the entity was found together with any numeric expression in the sentence itself (step 3 in Fig.2). Among these numbers, if any, we expect to find a numerical quantity describing an impact estimate for the observed category. Moreover, the system also performs a check on each digit found to assess whether it is a good candidate estimate or not: we are interested in raw measurements, meaning that a good estimate must not be part of either a date or time string, and it must not refer to scientific measurements or currencies. The system collects all valid candidate estimates found in the sentence, keeping for each of them the impact category they are possibly referred to, the value, the position in input text and the distance from related mentioned entity (see E in Fig.3). If no correct candidate estimate has been found in the sentence, a placeholder value of *0* is used. This describes the situation in which the system has detected an impact category, but it has not been able to find its associated estimate. Once the detected candidate estimates have been collected, the system produces a final estimate for each mentioned category sequentially, following their appearance order in text (step 4 in Fig.2). Among all potential estimates associated with each category, the first one satisfying a set of conditions is selected as result for the given category. The conditions are as follows:

- There is a path in dependency graph G connecting mentioned category and candidate estimate.
- The position of a candidate estimate precedes any other mentioned category, after currently observed one.
- The position of a candidate estimate follows any other mentioned category, preceding current one (intuitively, it must reside between other categories, if present).
- A candidate estimate has not been already associated to any other impact category.

Fig. 3. Impact Estimation algorithm on single event data

Input: current event tweets collection T
Output: current event impact estimate R
Stage 1: Impact Extraction phase
for all $t \in T$ **do**
    $G, S \leftarrow$ PREPROCESS(t)
    for all $s \in S$ **do**
        $I \leftarrow$ APPLY_RULES(s)
        $E \leftarrow$ FIND_CANDIDATE_ESTIMATES(I, s)
    end for
    for all $e \in E$ **do**
        if VERIFY_CANDIDATE(e, G) **then**
            $V($CATEGORY$(e)) \leftarrow$ TIME(t), VALUE(e)
            break
        end if
    end for
end for

Stage 2: Impact Aggregation phase
for all $c \in C$ **do**
    if LENGTH$(V(c))= 0$ **then**
        $R(c) \leftarrow 0$
    else
        $R(c) \leftarrow$ MOST_FREQ(MOST_REC$(V(c))$)
    end if
end for
return R

The placeholder value of *0* is also set for any category without a candidate estimate satisfying these conditions. The reasoning for these conditions stands in how reporting quantities in text is normally structured: in fact, it is often the case that the number of affected entities is close to related entity (before or after) or at most in the same sentence, but it is not common to find this numerical quantities once another affected entity has been mentioned in text, even less so to have the same number linked to two different entities. At the end of this stage, the system has assigned an estimate for each detected impact

TABLE II

IMPACT ESTIMATION - CLASSIFICATION RESULTS PER EVENT TYPE AND MICRO-AVERAGE OVERALL PER CLASS.

Category	*Event type*	*A*	*P*	*R*	*F1*	*Support*
Road	Fire	0.99	1.00	0.66	0.80	6
	Flood	0.99	0.50	1.00	0.67	2
	Overall	**0.99**	**0.75**	**0.75**	**0.75**	**8**
Bridge	Earthquake	1.00	1.00	1.00	1.00	2
	Fire	1.00	1.00	1.00	1.00	3
	Overall	**1.00**	**1.00**	**1.00**	**1.00**	**5**
Residential	Earthquake	0.97	0.25	1.00	0.40	2
	Storm	0.98	0.96	1.00	0.98	149
	Flood	0.98	0.81	0.98	0.89	48
	Fire	0.97	0.50	0.83	0.62	6
	Overall	**0.98**	**0.88**	**0.99**	**0.93**	**205**
Power Network	Flood	1.00	1.00	1.00	1.00	5
	Storm	1.00	1.00	1.00	1.00	3
	Overall	**1.00**	**1.00**	**1.00**	**1.00**	**8**
Water Network	Flood	**1.00**	**1.00**	**1.00**	**1.00**	**1**
Cultural Heritage	Flood	**1.00**	**1.00**	**1.00**	**1.00**	**4**
Burned Area	Fire	0.99	1.00	0.84	0.91	19
	Flood	0.99	0.00	0.00	0.00	1
	Overall	**0.99**	**1.00**	**0.80**	**0.89**	**20**

Category	*Event type*	*A*	*P*	*R*	*F1*	*Support*
Dead	Earthquake	1.00	1.00	1.00	1.00	175
	Storm	0.99	1.00	0.99	0.99	272
	Flood	0.99	0.99	0.99	0.99	442
	Overall	**0.99**	**0.99**	**0.99**	**0.99**	**889**
Injured	Earthquake	1.00	1.00	1.00	1.00	142
	Storm	1.00	1.00	1.00	1.00	276
	Flood	1.00	1.00	1.00	1.00	7
	Overall	**1.00**	**1.00**	**1.00**	**1.00**	**425**
Missing	Flood	0.99	0.99	1.00	0.99	354
	Storm	0.99	0.75	1.00	0.86	3
	Overall	**0.99**	**0.99**	**1.00**	**1.00**	**357**
Evacuated	Earthquake	1.00	1.00	1.00	1.00	3
	Storm	1.00	1.00	1.00	1.00	1
	Fire	0.96	0.10	1.00	0.18	1
	Overall	**0.99**	**0.35**	**1.00**	**0.52**	**5**
Rescued	Flood	**0.99**	**0.75**	**1.00**	**0.86**	**3**
Hospitalized	Storm	**0.99**	**1.00**	**0.50**	**0.66**	**2**
Other	Flood	0.99	0.36	0.80	0.50	5
	Storm	1.00	1.00	1.00	1.00	2
	Overall	**0.99**	**0.46**	**0.85**	**0.60**	**7**

category in the given tweet and updated the general status associated to the current event (V in Fig.3). A schematic view of the information extraction process together with the results produced by each phase, decorated by an example input text, is reported in Fig.2-3.

B. Impact aggregation

The information extracted in first stage (V in Fig.3) has, for each category reported in Table I, all impact measures extracted from event related tweets sorted according to tweet creation time. Impact aggregation phase combines then the information extracted from each tweet to produce an overall estimate associated to the whole event, using the following approach for each impact category: if there are no measures, it is established that the event has not provoked an impact in the given category. Otherwise, the impact measure in the observed category is given by the most frequent value among the most recent *100* ones. If two measures appear with same frequency, it is selected the largest one.

The description of the complete process of impact estimation for a single disaster event is reported in *Stage 2* of Fig.3. The information about crisis events shared on social media platforms can easily be noisy and rapidly change over time: limiting the choice to the most recent values helps to ignore the initial estimates, which are usually inaccurate, while choosing the most frequent value allows to overlook eventual noisy estimations.

V. PERFORMANCE EVALUATION

The proposed system has been evaluated on a test set of *1416* English tweets (retrieved and processed with the pipeline reported in Fig.1) posted during six different crisis events happened in *2021*: flood in Nepal, flood in Germany, flood in North Carolina, storm in Czech Republic, earthquake in China and fire in California. Among a wide set of detected events, these were carefully chosen because considered representative on a global scale considering a yearly scenario. Additionally, only crises big enough to contain impact information have been selected. Most of them are related to flood disasters, given their high frequency, while earthquakes have been selected since is the hazard that accounts for the highest number of people potentially exposed [1]. Other disaster types have been included for sake of completeness, to evaluate the system against different crises. Details about test set construction are provided in Table III.

The main objective of the evaluation is to measure system ability in detecting affected entities mentioned in text, focusing on damaged infrastructures and affected people, and estimating their number. To this purpose, the selected tweets were manually annotated using the following criteria: each tweet is associated with a possibly empty set of impact categories, following the taxonomy reported in Table I. A category is associated with a tweet whenever the former is mentioned in the text. Each present category is also associated with a numerical impact, expressed as the number of entities affected in that category. In case the tweet does not clearly express the number of affected entities, the corresponding impact is set to *0*.

TABLE III

TEST DATASET PROPERTIES. START AND END COLUMNS REPRESENT THE TIME OF RESPECTIVELY THE FIRST AND LAST TWEET ASSOCIATED TO THE EVENT RATHER THAN THE START AND END TIME OF THE EVENT ITSELF.

Event	**Start**	**End**	**# Tweets**
Earthquake in China	2021-05-21 15:53	2021-05-22 15:51	183
Flood in Nepal	2021-06-16 06:32	2021-06-20 02:49	182
Storm in CZE	2021-06-24 17:22	2021-06-27 09:00	373
Flood in Germany	2021-07-15 07:50	2021-07-17 05:41	226
Flood in North Carolina	2021-08-18 20:39	2021-08-21 02:57	194
Fire in California	2021-08-23 20:33	2021-08-28 15:02	258

Even though the system works at the event level, providing an estimate after aggregation of information extracted from each tweet, performance validation has been done at tweet level only, given that it represents the most critical phase of the process. Working at event level implies instead the necessity of ground truth information of impacted categories at this level of granularity: despite the presence of disaster databases such as *EM-DAT* [26], they usually do not report impacts

TABLE IV
IMPACT ESTIMATION - REGRESSION METRICS PER EVENT TYPE AND MICRO-AVERAGE OVERALL

Category	*Event type*	*MAE*	*MSE*	*Support*
Road	Fire	25.00	1201.00	6
	Flood	0.00	0.00	2
	Overall	**18.75**	**900.75**	**8**
Bridge	Earthquake	0.50	0.50	2
	Fire	10.00	150.00	3
	Overall	**6.20**	**90.20**	**5**
Residential	Earthquake	0.00	0.00	2
	Storm	6.19	803.25	149
	Flood	2.43	211.44	48
	Fire	16.67	1666.67	6
	Overall	**5.56**	**682.11**	**205**
Power Network	Flood	0.00	0.00	5
	Storm	0.00	0.00	3
	Overall	**0.00**	**0.00**	**8**
Water Network	Flood	0.00	0.00	**1**
Cultural Heritage	Flood	1.00	1.00	**4**
Burned Area	Fire	1168.51	1.29e7	19
	Flood	8000.00	6.40e7	1
	Overall	**1510.08**	**1.55E7**	**20**

Category	*Event type*	*MAE*	*MSE*	*Support*
Dead	Earthquake	18.22	5147.95	175
	Storm	3.12	416.74	272
	Flood	1.02	43.16	442
	Overall	**5.05**	**10282.42**	**889**
Injured	Earthquake	0.89	11.07	142
	Storm	8.83	1060.18	276
	Flood	0.29	0.57	7
	Overall	**6.04**	**692.20**	**425**
Missing	Flood	3.30	74.52	354
	Storm	0.00	0.00	3
	Overall	**3.27**	**73.90**	**357**
Evacuated	Earthquake	6944.00	1.46e8	3
	Storm	140.00	19600.00	1
	Fire	24500.00	6.00e8	1
	Overall	**9124.40**	**2.08e8**	**5**
Rescued	Flood	**0.00**	**0.00**	**3**
Hospitalized	Storm	**1.50**	**4.50**	**2**
Other	Flood	1800.20	1.62e7	5
	Storm	19500.00	7.60e8	2
	Overall	**6857.28**	**2.29e8**	**7**

with a detailed taxonomy. Moreover, these databases leverage different sources before reporting the event estimate, while our system relies only on information posted on social media, therefore comparing our predictions with values reported in database would inevitably evaluate the accuracy of information posted on Twitter rather than the impact estimation itself. Results of impact category classification are displayed in Table II, highlighting, for each class, the values of *Accuracy* (*A*), *Precision* (*P*), *Recall* (*R*) and *F1 score* metrics measured on the test set both separated by hazard type and overall, together with the number of tweets mentioning the category (reported in *Support* column). Results highlight how population-related categories are more represented in test set with respect to infrastructure ones. The system appears to correctly detect impact categories in the majority of cases, and at the same time very rarely produces false positive predictions. Recall and precision values highlight in fact relatively good and balanced performances, regardless of the category and hazard type, reaching an overall weighted *F1* score of *0.77* on the entire test set. After the system has detected the presence of an impact category, it is also asked to provide a numerical estimate of entities affected, which can be considered a regression problem. *Mean Absolute Error* (*MAE*) and *Mean Squared Error* (*MSE*) have been collected to characterize the performances in this additional task. The metrics reported in Table IV have been evaluated considering only estimates provided when the given category was actually mentioned in text. Results show an acceptable *MAE* in each category, except for *Evacuated* and *Other* labels, where the few tweets mentioning these categories were poorly estimated. Performances per hazard type show again consistency for those categories sufficiently represented in test set, while underrepresented categories show varying performances. Given the low amount of validation data available, caused by the peculiarity of these categories, it is not possible to draw strong conclusions about the performances in such categories without further manual labeling.

VI. CONCLUSION AND FUTURE WORKS

This paper proposes a system capable of providing precise impact estimations of ongoing crisis events: obtained performance demonstrates the feasibility of its use in a real-world scenario, as a tool to support the planning of ongoing rescue operations. However, despite its promising results, we believe that the current system can still improve in both tasks, *i.e.*, the detection of affected entities and the estimation of numerical quantities. In fact, while the rule-based approach can be tailored to any use case, general definitions are typically prone to false positive predictions, while careful rule setting to cover all possible cases may become a never-ending task. Given recent advances in literature, a machine learning-based algorithm for automatic detection of specific entities may obtain even better results than the rule-based approach, provided that a robust ground-truth is available. Moreover, impact estimates in each category are tightly coupled with word dependencies extraction: erroneous dependencies may result in wrong or misplaced predictions, creating a chain reaction of misleading associations. In this scenario, the design and implementation of a more resource intensive, albeit more precise, deep learning-based dependency extractor could also be beneficial for this task.

REFERENCES

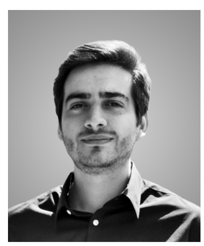
[1] M. Pesaresi, D. Ehrlich, T. Kemper, A. Siragusa, A. Florczyk, S. Freire, and C. Corban, "Atlas of the human planet 2017: Global exposure to natural hazards," European Commision, Luxembourg (Luxembourg), Policy assessment KJ-NA-28556-EN-N (online),KJ-NA-28556-EN-C (print),KJ-NA-28556-EN-E (ePub), 2017.

[2] J. Yin, S. Karimi, A. Lampert, M. A. Cameron, B. Robinson, and R. Power, "Using social media to enhance emergency situation awareness: Extended abstract," in *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015*, Q. Yang and M. J. Wooldridge, Eds. AAAI Press, 2015, pp. 4234–4239. [Online]. Available: http://ijcai.org/Abstract/15/602

[3] R. Eyre, F. Luca, and F. Simini, "Social media usage reveals recovery of small businesses after natural hazard events," *Nature Communications*, vol. 11, 04 2020.

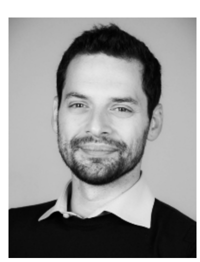
[4] Y. Wang, S. Ruan, T. Wang, and M. Qiao, "Rapid estimation of an earthquake impact area using a spatial logistic growth model based on social media data," *Int. J. Digit. Earth*, vol. 12, no. 11, pp. 1265–1284, 2019. [Online]. Available: https://doi.org/10.1080/17538947.2018.1497100
[5] Y. Liang, J. Caverlee, and J. Mander, "Text vs. images: on the viability of social media to assess earthquake damage," in *22nd International World Wide Web Conference, WWW '13, Rio de Janeiro, Brazil, May 13-17, 2013, Companion Volume*, L. Carr, A. H. F. Laender, B. F. Lóscio, I. King, M. Fontoura, D. Vrandecic, L. Aroyo, J. P. M. de Oliveira, F. Lima, and E. Wilde, Eds. International World Wide Web Conferences Steering Committee / ACM, 2013, pp. 1003–1006. [Online]. Available: https://doi.org/10.1145/2487788.2488102
[6] Y. Kryvasheyeu, H. Chen, N. Obradovich, E. Moro, P. V. Hentenryck, J. Fowler, and M. Cebrian, "Rapid assessment of disaster damage using social media activity," *Science Advances*, vol. 2, no. 3, p. e1500779, 2016. [Online]. Available: https://www.science.org/doi/abs/10.1126/sciadv.1500779
[7] F. Yuan and R. Liu, "Feasibility study of using crowdsourcing to identify critical affected areas for rapid damage assessment: Hurricane matthew case study," *International Journal of Disaster Risk Reduction*, vol. 28, pp. 758–767, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S221242091830150X
[8] Y. Faxi and L. Rui, "Mining social media data for rapid damage assessment during hurricane matthew: Feasibility study," *Journal of Computing in Civil Engineering*, vol. 34, 01 2020.
[9] D. Wu and Y. Cui, "Disaster early warning and damage assessment analysis using social media data and geo-location information," *Decis. Support Syst.*, vol. 111, pp. 48–59, 2018. [Online]. Available: https://doi.org/10.1016/j.dss.2018.04.005
[10] M. Dou, Y. Wang, Y. Gu, S. Dong, M. Qiao, and Y. Deng, "Disaster damage assessment based on fine-grained topics in social media," *Comput. Geosci.*, vol. 156, p. 104893, 2021. [Online]. Available: https://doi.org/10.1016/j.cageo.2021.104893
[11] A. Dittrich and C. Lucas, "Is this twitter event a disaster?" in *17th AGILE Conference on Geographic Information Science*, 2014.
[12] C. Rossi, F. S. Acerbo, K. Ylinen, I. Juga, P. Nurmi, A. Bosca, F. Tarasconi, M. Cristoforetti, and A. Alikadic, "Early detection and information extraction for weather-induced floods using social media streams," *International journal of disaster risk reduction*, vol. 30, pp. 145–157, 2018.
[13] F. Abel, C. Hauff, G. Houben, R. Stronkman, and K. Tao, "Semantics + filtering + search = twitcident. exploring information in social web streams," in *23rd ACM Conference on Hypertext and Social Media, HT '12, Milwaukee, WI, USA, June 25-28, 2012*, E. V. Munson and M. Strohmaier, Eds. ACM, 2012, pp. 285–294. [Online]. Available: https://doi.org/10.1145/2309996.2310043
[14] D. Salza, E. Arnaudo, G. Blanco, and C. Rossi, "A 'glocal' approach for real-time emergency event detection in twitter," in *ISCRAM 2022 Conference Proceedings – 19th International Conference on Information Systems for Crisis Response and Management*, 2022.
[15] B. Resch, F. Usländer, and C. Havas, "Combining machine-learning topic models and spatiotemporal analysis of social media data for disaster footprint and damage assessment," *Cartography and Geographic Information Science*, vol. 45, pp. 1–15, 08 2017.
[16] L. Li, M. Bensi, Q. Cui, G. B. Baecher, and Y. Huang, "Social media crowdsourcing for rapid damage assessment following a sudden-onset natural hazard event," *Int. J. Inf. Manag.*, vol. 60, p. 102378, 2021. [Online]. Available: https://doi.org/10.1016/j.ijinfomgt.2021.102378
[17] S. Mangalathu and H. V. Burton, "Deep learning-based classification of earthquake-impacted buildings using textual damage descriptions," *International Journal of Disaster Risk Reduction*, vol. 36, p. 101111, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S221242091830983X
[18] M. Imran, S. Elbassuoni, C. Castillo, F. Diaz, and P. Meier, "Extracting information nuggets from disaster- related messages in social media," in *10th Proceedings of the International Conference on Information Systems for Crisis Response and Management, Baden-Baden, Germany, May 12-15, 2013*, T. Comes, F. Fiedrich, S. Fortier, J. Geldermann, and T. Müller, Eds. ISCRAM Association, 2013. [Online]. Available: http://idl.iscram.org/files/imran/2013/613_Imran_etal2013.pdf
[19] S. Panem, M. Gupta, and V. Varma, "Structured information extraction from natural disaster events on twitter," in *Proceedings of the 5th International Workshop on Web-scale Knowledge Representation Retrieval & Reasoning, Web-KR@CIKM 2014, Shanghai, China, November 3, 2014*, Y. Zeng, S. Kotoulas, and Z. Huang, Eds. ACM, 2014, pp. 1–8. [Online]. Available: https://doi.org/10.1145/2663792.2663794
[20] S. Piscitelli, E. Arnaudo, and C. Rossi, "Multilingual text classification from twitter during emergencies," in *IEEE International Conference on Consumer Electronics, ICCE 2021, Las Vegas, NV, USA, January 10-12, 2021*. IEEE, 2021, pp. 1–6. [Online]. Available: https://doi.org/10.1109/ICCE50685.2021.9427581
[21] G. Lample, A. Conneau, M. Ranzato, L. Denoyer, and H. Jégou, "Word translation without parallel data," in *International Conference on Learning Representations*, 2018.
[22] G. Lample, M. Ballesteros, S. Subramanian, K. Kawakami, and C. Dyer, "Neural architectures for named entity recognition," *CoRR*, vol. abs/1603.01360, 2016. [Online]. Available: http://arxiv.org/abs/1603.01360
[23] R. Weischedel, S. Pradhan, L. Ramshaw, J. Kaufman, M. Franchini, M. El-Bachouti, N. Xue, M. Palmer, J. D. Hwang, C. Bonial *et al.*, "Ontonotes release 5.0," 2012.
[24] OpenStreetMap contributors, "OpenStreetMap Nominatim," https://nominatim.openstreetmap.org, 2021.
[25] M. Honnibal and M. Johnson, "An improved non-monotonic transition system for dependency parsing," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Lisbon, Portugal: Association for Computational Linguistics, Sep. 2015, pp. 1373–1378. [Online]. Available: https://aclanthology.org/D15-1162
[26] EM-DAT, "Em-dat: The international disaster database," http://www.emdat.be/, 2009.

Giacomo Blanco received the B.Sc. and M.Sc. degrees in Computer Engineering all from *Politecnico di Torino*, (IT), in 2018 and 2020 respectively. His research interests are data science and machine learning, with focus on computer vision and natural language processing. He is currently working at LINKS Foundation as applied researcher on the development of intelligent systems able to increase situational awareness and support decision process during disasters management.



Edoardo Arnaudo graduated summa cum laude in 2019 from the *Università degli Studi di Torino*, obtaining a M.Sc. degree in Computer Science with specialisation in Machine Learning and Artificial Intelligence. In the same year, he started working at LINKS Foundation, in Turin (IT), as applied researcher in the *AI, Data & Space* domain in several multidisciplinary projects, with a focus on aerial and remote sensing applications. Since 2020, he is pursuing a Ph.D. at the *Politecnico di Torino*.

Dario Salza obtained a M.S.E. degree with specialization in Data Science and Engineering from *Télécom Paris* in 2020, and graduated summa cum laude in the same year from *Politecnico di Torino* with a M.Sc. degree in Computer Engineering. In his current role as applied researcher at LINKS Foundation he has worked in the domains of computer vision, data science and time-series analysis, operating research and business intelligence, with applications such as in the context of intelligent emergency management.



Claudio Rossi obtained a M.S. degree in Electrical and Computer Engineering from the *University of Illinois at Chicago (UIC)* in 2005, also graduating summa cum laude in Electronics from *Politecnico di Torino*. He worked as a project manager at Fiat Group Automobiles (FGA) and he obtained his PhD from *Politecnico di Torino* in 2014 while working as research intern at Telefonica I+D. He is now Program Manager at LINKS Foundation, in Turin, where he leads the research program *AI for Industry and Security*, coordinating several R&D projects.