Monitoring disaster impact: detecting micro-events and eyewitness reports in mainstream and social media

Hristo Tanev
European Commission
Joint Research Centre
hristo.tanev@ec.europa.eu

Vanni Zavarella
European Commission
Joint Research Centre
vanni.zavarella@ec.europa.eu

Josef Steinberger
University of West Bohemia
jstein@kiv.zcu.cz

ABSTRACT

This paper approaches the problem of monitoring the impact of the disasters by mining web sources for the events, caused by these disasters. We refer to these disaster effects as "micro-events". Micro-events typically following a large disaster include casualties, damage on infrastructures, vehicles, services and resource supply, as well as relief operations. We present natural language grammar learning algorithms which form the basis for building micro-event detection systems from data, with no or minor human intervention, and we show how they can be applied to mainstream news and social media for monitoring disaster impact. We also experimented with applying statistical classifiers to distill, from social media situational updates on disasters, eyewitness reports from directly affected people. Finally, we describe a Twitter mining robot, which integrates some of these monitoring techniques and is intended to serve as a multilingual content hub for enhancing situational awareness.

Keywords

Natural language processing, machine learning, crisis computing, disaster effects, social media

INTRODUCTION

Natural disasters have caused a total of $1.5 trillion in damage worldwide between 2003 and 2013, according to a study by the United Nations Food and Agriculture Organization (FAO), which finds they caused more than 1.1 million deaths and affected the lives of more than two billion people[1]. Effects of the disasters range from tragic loss of human lives and injured people to interrupted services, damage of infrastructures, as well as shortage of basic resources.

It is nowadays well recognized (Vieweg et al. 2010) that spreading real-time information about the damages caused by the disasters, as well as about the ongoing response measures on the field, can significantly contribute to the relief process and mitigate the overall impact, by enhancing Situational Awareness of stakeholders (victims, responders, donors).

When a heavy disaster hits a region, usually people rely on online mainstream media, radio and TV to look for updates about the situation in their surroundings. Social media have a very important role during such events, since people share real-time information about ongoing developments and at the same time seek for situational

[1] http://www.fao.org/3/a-i5128e.pdf.

updates from response organizations or other users. Unfortunately, tracking the descriptions of damages in online media is a challenging task, because of the overflow of information available, as well as the overlapping descriptions of the same facts, and even more in social media where language variation, re-posted, irrelevant or unreliable content are pervasive.

This paper addresses the problem of detecting natural disaster-related *micro-events* in online mainstream and social media sources. As micro-events caused by natural disasters, we consider casualties (dead, injured, trapped people), destruction or damage on buildings and other infrastructures, disruption of transport and other services, shortage or provision of resource supplies such as electricity, water and fuel, as well as recovery work and operations for rescuing of people. The term *micro* refers to the local consequences of such events – a damaged building or interruptions of service usually affect a small fraction of the local community. We propose solutions for automatic learning of grammar models for micro-event detection and apply these models in a real world application detecting micro-events in tweets published during natural disasters.

First, we present a novel algorithm for learning of event detection grammars. We then describe how we use it to induce in an unsupervised manner a grammar for detecting disaster-related micro-events from online media sources, starting from an unannotated news corpus. To the best of our knowledge, there are no other machine learning approaches which target the detection of such micro-events in news. We also describe the creation of a news corpus annotated for micro-events, which we use for evaluation.

Afterwards, we describe a similar algorithm that, by using weak supervision, is able to learn, from social media data, event patterns categorized into a predefined set of disaster-related event types. The resulting grammars can be used to "make sense" of the large mass of messages exchanged by social media users during disasters, namely for discarding the large fraction of off-topic content, and further classify the disaster related messages. The method is multilingual and has shown high accuracy in extracting micro-events from a stream of Twitter status messages.

Social media have been described as being used as an information "multichannel" during disasters: the general public seeks for official updates from authorities as well as personal updates from their communities, while response agencies disseminate critical information to the public and try to gather crowd-sourced situational updates from the field (Imran et al. 2015). We focus here in the latter type of social media messages which are often referred as "eyewitness content" and we define it as "information originating from eyewitnesses to the event or to response/recovery operations, or from their family, friends, neighbours", from (Olteanu et al. 2015).

The frequency of eyewitness reports on social media during disasters largely varies across events and event types. (Olteanu et al. 2015) asked crowd-sourced workers to annotate for an attribute "Information Source" around 25000 tweets from a set of 26 events, including natural hazards and man-made disasters. They found an average of 9% status messages reporting eyewitness accounts. A similar analysis we performed on a small sample (200 tweets) of a corpus collected for the August 2016 Central Italy earthquake[2] confirmed this finding: 8.7% of the tweets categorized as "Infrastructure Damage" were eyewitness reports, while 92% in the same category were considered Informative messages. Therefore, we believe that being able to distil eyewitness situational updates out of the vast majority of news media and institutional content echoed through social media and to re-direct it to target recipients may help accelerating the information flow and the overall response process. We tackle this task by deploying statistical Machine Learning classifiers, trained on data from previous disaster events.

Finally, we describe a simple Twitter channel that integrates most of the techniques described above: it dynamically tracks target disaster events by filtering, tagging and re-posting Twitter status messages, serving as a social media content hub for situation monitoring and impact assessment.

RELATED WORK

The grammar learning method is based on a previous algorithm presented in (Tanev, 2014). We have reported similar work on Social media in (Tanev and Zavarella, 2013). Disaster-related terminology was the focus of different previous studies, such as (Thywissen, 2006) and (Gunn, 2003).

Automatic detection of natural disasters and other security-related events was discussed in previous

[2] https://en.wikipedia.org/wiki/August_2016_Central_Italy_earthquake

publications, such as (Piskorski et al., 2011). Event scenario template filling was introduced as a task at the Message Understanding Conference (MUC) (Grishman et Sundheim, 1996) for a description of this task. Exploitation of social media for enhancing the information extracted from online news was discussed earlier in (Tanev et al., 2012). Most of the work on event detection from social media focused on English language only, a noticeable exception being (Zielinski and Bügel, 2012).

Algorithms for grammar induction were presented in different publications, look at (Palka and Zachara, 2015) and (Athanasopoulou et al., 2014) among the others. A survey of grammar induction approaches is presented in (D'Ulizia et al., 2011).

The usefulness of social media during disasters was discussed in (Lindsay 2011; Fugate 2011), (Howell et al., 2011) and others. Many publications discuss the automatic analysis of Twitter messages, published during disasters. Most of the work focused on message retrieval, classification, extraction of information, aggregation: see among the others (Imran et al., 2013, 2015) and (Middleton et al., 2014).

Within this field, a number of attempts have been performed to capture first-hand situational reports from disaster affected areas. Among computational models, (Morstatter et al., 2014) have approximated our task with the one of detecting tweets originating from within the region of the crisis, by analysing the linguistic features that differentiate them from the ones from outside the region. They could automatically collect annotated training data by relying on tweet geocoding metadata, however their definition of eyewitness is incompatible with ours. (Verma et al., 2011) explored the non-equivalent task of automatically identifying tweets for situational awareness (SA). They used a combination of shallow language features and the output of a number of classifiers, for Subjectivity, Personal/Impersonal, etc.

(Fang et al., 2016) present an experiment where they trained a SVM classifier on a large dataset extremely skewed towards negative instances (118K vs. 400); their feature set is partially overlapping with ours, but they additionally integrated 300 dimension Word Embeddings for each word in the tweets, computed from a 198M tweet corpus, which dramatically increased the performance. (Pekar et al., 2016) contains an empirical feature analysis of a classifier most similar to ours and ran on the same dataset; we built upon their empirical findings in our approach.

UNSUPERVISED LEARNING OF EVENT DETECTION PATTERNS

Method

We have developed an unsupervised algorithm for learning of grammar rules which is based on the algorithm presented in (Tanev, 2014). Our algorithm has the following basic steps:

1. Take a text collection *C* whose main topic is from a given domain (in this case natural disasters)

2. Find a set of terms *T* with the highest TF.IDF in the disaster-specific text collection *C*. TF.IDF returns the terms which are frequent in this domain-specific collection, but otherwise are rare in other texts. These terms are characteristic terms for the disaster domain. As an example, let's consider the *T* being the set *earthquake, floods, damaged, destroyed facilities, schools, universities, church, destroyed, damaged, flooded, collapsed, toppled,* etc.

3. Following the grammar learning algorithm described in (Tanev and Zavarella, 2013) we generate grammars from the terms in *T* and the news article corpus *C*.

a. Represent the terms from *T* as distributional feature vectors, whose dimensions are the n-grams appearing immediately on the left and on the right of the terms.

b. Cluster the terms, using the cosine similarity between their feature vectors. In this way words which appear in similar contexts will finish in the same cluster. According to the Harris' distributional hypothesis, those words that occur in the same contexts tend to have similar meanings. Following our toy example, the terms in *T* will form two clusters *T1*={*earthquake, floods*}, *T2*={*damaged, destroyed, flooded, collapsed, toppled*} and *T3={church, universities, school, facilities,…}*.

c. Find the pairs of clusters whose terms co-occur in the domain specific corpus *C* next to each other or are divided by prepositions and other stop words. Let's assume that in the corpus *C* we have the phrase *earthquake damaged*. Since *earthquake* belongs to the first cluster *T1* and *damaged* to *T2*, then a generalization pattern

[T1 T2] will be derived among the others.

d. Select only the rules which are composed of a cluster containing verbs and a cluster which does not contain verbs. The goal of this step is to leave only event detection rules. This constraint is novel with respect to the original algorithm.

e. Create a finite-state grammar rule from each selected word cluster pair. In order to provide higher recall, the grammar rules allow for a couple of additional alphanumerical tokens between the terms from the cluster pairs.

The logic behind this algorithm is the following: By capturing just the pairs of terms which have high TF.IDF for the specific collection, we are getting phrases which are specific for this text collection. Moreover, we limit ourselves to the phrases which contain one verb and one not verb word. In this way, our algorithm generates patterns which recognize domain specific event descriptions. Our empirically-derived hypothesis is that most of the specific events will be effects of the disasters and recovery operations, as well as deployment of emergency-response teams.

Experiments and Evaluation

We applied our grammar learning algorithm on a news corpus from the disaster domain. In particular, we used the set of not annotated news articles about floods which happened in different parts of the world in the beginning of 2016, another set of new articles about the tropical storms Erika (August 2016), Joaquin and Patricia (in 2015), as well as the quake which have hit the coasts of Sumatra in March 2016. The overall size of this corpus was 381 articles. In order to evaluate our algorithm, we used a small test corpus of 21 articles about the 2016 earthquake in Ecuador. This test corpus contains 198 annotated micro-events. The grammar which we have created has extracted 321 event-description phrases. In our evaluation we considered correct each event phrase which overlaps by one or more words with a phrase from the annotated corpus.

Precision and Recall were found to be 42% and 66% respectively. Considering that our algorithm is unsupervised, these numbers are encouraging, but clearly the results can be improved further. In particular, with some minor efforts we can boost the precision significantly by performing a manual inspection on the learned results and leaving only the relevant clusters and cluster pairs.

Our algorithm has detected phrases which express killings and injuries of people, damage and collapse of buildings, deployment of emergency crew, as well as recovery and rescue operations.

Table 1. A news article about the earthquake in Ecuador and the extracted phrases

News article	Rescuers start to arrive after Ecuador quake 1666 "QUITO, Ecuador The Latest on the earthquake in Ecuador. (all times local): 7:45 a.m. Authorities in Ecuador are mobilizing resources and help is getting to the ground after a long night of fear and uncertainty caused by a magnitude-7.8 earthquake that killed at least 77 people. Vice President Jorge Glas is overseeing efforts until President makes an emergency return from a visit to Rome. Glas arrived Sunday morning in Manta along the coast along with dozens of rescuers. The city's airport is badly damaged, but is receiving relief flights. National airline TAME has already organized two humanitarian airlifts with members of the Red Cross and police reinforcements. More than a dozen roads have been closed due to damage from the earthquake, making it harder for rescuers to reach where they are needed most. The Transportation Ministry says that the hardest hit was Manabi province, near the epicentre. Eight major roads there were either closed or partially collapsed from landslides or strong movements of the earth.
Extracted phrases (correct ones are marked with *)	**rescuers start, *start to arrive after Ecuador quake, uncertainty caused, caused by a magnitude-7.8 earthquake, *earthquake that killed, *killed at least 77 people, *President makes an emergency return, emergency return from a visit, National airline TAME has already organized, *dozen roads have been closed, *closed due to damage from the earthquake needed most, Ministry says that the hardest hit, *roads there were either closed or partially collapsed, *closed or partially collapsed from landslides*

Here are some other examples of relevant phrases which our algorithm managed to detect in the test corpus: *dozen roads have been closed, building above the shop collapsed, collapsed structures are a hotel, family were among those killed, people had been reported as missing, difficult for emergency workers to reach the towns* *hardest hit*, *continued to pick at rubble looking for survivors and victims*, *caused numerous victims and great damage*, *closed due to damage from the earthquake*, *roads have been closed*, *people have died*, *closed or partially collapsed from the landslides*. Most of the erroneously extracted event phrases were about the fact that the quake occurred, announcements and incomplete phrases about disaster effects. In our final evaluation we did not consider such phrases as correct, although they are not completely wrong either. Such phrases can be used to pinpoint to disaster effects which will be mentioned afterwards.

TOWARDS CREATING A CORPUS OF DISASTER-RELATED MICRO EVENTS

We created a news corpus annotated for micro-events which resulting from natural disasters. First, we have collected from the Web news about major earthquakes, wild fires, landslides, and floods. Then, we annotated the mentions of each micro-event belonging to one of the following event types: damage on people, damage and abandoning of infrastructures, vehicles, and services, such as transport, electricity, water and gas and more. The corpus is available at http://nlp.kiv.zcu.cz/projects/dime.

We did not enumerate the types of micro events, since creation of a micro-event taxonomy is beyond the scope of this paper, however we envisage to do that in our future work. Instead, we created a gold annotation for each type of natural disaster - flood, earthquake, forest fire and landslide and we used this gold standard as a reference, when we annotated the other articles. The final version of the corpus is envisaged to encompass around 200 articles and the version we have produced currently has around 70 articles.

SUPERVISED LEARNING OF EVENT DETECTION GRAMMARS

Although the unsupervised grammar learning algorithm, described in the previous subsection generated a substantial number of interesting patterns, the main problem is that these patterns are not labeled with the event types they detect. That is, the system which uses such patterns cannot distinguish between deaths, building damages, road closures, airport closures, and other disaster effects, although it can pinpoint to the text which describes those events. In order to perform more topically focused extraction of event patterns we carried out experiments with another semi-automatic learning algorithm.

We applied the algorithm described in (Tanev and Zavarella, 2013) in order to learn patterns consisting of a combination of words about a resource or entity and event description phrases, like in the following sample rule for Building Damage events:

```
event_rule :> ( leftPattern & [CLASS:"Damage"]
       participatingEntity & [CLASS:"Building", SURFACE:#surf] )
-> event & [CLASS:"Damage", PARTICIPANT:#surf ]
& PossibleSlotFor(Damage,Building)
```

where `PossibleSlotFor(X,Y)` is a Boolean operator checking whether the `participatingEntity` class `Y` is one the possible classes which may fill the slots of a pattern of class `X`. The method is provided with a conceptual specification of the target domain (a specification of domain entities and the event relations they participate in), a few initial words for those classes and an unannotated tweet corpus. It continues by expanding entity classes, learning entity modifiers and event patterns, which can then be encoded in simple finite-state rules.
In this way we have learned lexical patterns in English, Italian and Spanish for the following disaster-related events:

- Building damage
- Service interruption (transport services, hospitals, schools, and others)
- Emergency crew deployment
- Shortage and distribution of food, water, shelters, medical resources or fuel

Subsequently, we used the learned patterns to filter a stream of Twitter status messages, retrieved from the Twitter public Streaming API. We have conducted preliminary accuracy estimation on a set of the 20 retweeted messages in English on 25 November 2016 and we have found that 80% of the detected events are correct. This evaluation is based on a quite small test set and it lacks generality, since most of the events belonged to the classes “Building damage” and “Emergency crew deployment”, nevertheless it shows that the algorithm is potentially useful for gathering information about ongoing developments after disasters hit. We deploy such patterns in the Twitter application we present below.

DETECTING EYEWITNESS TWITTER STATUS MESSAGES

We wanted to evaluate how accurately one could automatically detect the relatively tiny fraction of eyewitness social media content communicated during disasters and separate it from the large amount of messages that talk largely about the same event, by deploying statistical Machine Learning text classifiers, trained on Twitter data taken from previous disaster events.

Features

In order to deal with the multilingual content communicated through social media one has to rely on shallow language features, as many largely spoken languages in the world are under-resourced with respect to language processing tools, such as POS taggers, parsers, etc. For this reason, compared to other previous works (e.g. Pekar et al., 2016) we experimented with a minimal set of language features that could be mostly produced using language-independent methods, and we test our approach on English and Italian.

Lexical:

uni-grams: we extract a ranked list of the *n* most relevant unigrams features (*n* is around 3100 and 1900 for English and Italian, respectively) from the corpus of all tweets for a target language (regardless of the label), by using a simple TF/IDF scoring, removing features containing Twitter metadata reserved characters. We apply them with no prior score-based weighting. The method applies a stopword list as the only language-specific resource and is thus easily portable across languages.

n-grams: analogously we extract a ranked list of the most relevant n-gram phrases, with no prior restriction on the length of the phrase.

number of unigrams: as some evidence was found by (Power et al. 2015) that first-hand emergency reports tend to be comparatively shorter, we apply a numeric feature for the token-based length of the tweet.

Stylistic:

Personal: similarly to (Pekar et al., 2016) we create a Boolean feature encoding whether the tweet text contains terms from a small lexicon of deictic expressions, including first-person pronouns/adjectives and adverbials (e.g. I, our, here, etc.). The assumption here is that an eyewitness account will more likely be written from a first-person perspective. We hand coded these small lexicons for both English and Italian.

All-caps: as word capitalization is used in social media jargons as an interjection marker to attract the reader's attention, we assume a first-hand account from a disaster affected person will more likely be all-caps. We create a Boolean feature indicating if the tweet contains only capital letter words.

Twitter Metadata:

hashtags: we add to the vector space a feature for each hashtag found in the dataset (e.g. *#Yolanda*), and then assign it 1 if it occurs in an instance tweet and 0 otherwise.

containsHashTags: this encodes the fact that an instance tweet contains at all hashtags or not.

user mentions: analogously, we create a Boolean feature for each user mention (e.g. *@denvernews*) found in the dataset.

containsMention: this encodes the fact that a instance tweet contains user mentions at all.

urls: adding URLs in a tweet is a form of linking previously published content and is thus potentially a distinguishing feature of non-first-hand situation reports. However, URLs may link multimedia content originating from the tweet's author himself, as different for example from linking news media or government sites. In order to account for URLs qualitatively, we add features for all URLs collected from the dataset.

containsURLs: encodes the fact that a tweet contains URLs at all or not.

RetweetCount: a numerical feature for the number of times a message has been retweeted. We assume here that novel situational updates from the affected people should gain comparatively more attention and being more re-posted.

fromMobileDevice: we check the Twitter message client application in order to determine whether it was posted from a mobile device or not.

IsReply: we assume that eyewitness reports are not likely to be posted as replies to previous messages.

Semantic:

Category Definitions: We use Boolean features each encoding the matching of one of 13 rule-based category definitions relevant to the disaster domain. Namely, we use the disaster-related event patterns learned by the algorithm described in the previous Section, and deploy them in a more lenient mode by simply combining predicate and entity lexicons by Boolean operators, with proximity restrictions. Example category features are "Infrastructure Damage", "Service Interruption", "Shelter Needed", etc.

The rationale here is that first-hand reports during a crisis, as distinct for example from mere sympathy expressions, will likely be classified in one or more of the micro-event topics the affected people are mostly concerned about. Moreover, as the micro-events and the corresponding lexical resources are rather general across disaster instances and disaster types, this feature set should boost classifier accuracy generalization across disasters. However, the list of modelled micro-event topics is not exhaustive, therefore we do not expect this to largely boost Recall.

Experimental setup and Evaluation

We worked with data from the labeled part of the CrisisLexT26 corpus (Olteanu et al., 2015). This corpus comprises over 24k multilingual tweets produced during 26 diverse mass emergency events (including floods, wildfires, bombings, etc.) from 2012-2013 collected through Twitter Sample API and annotated for 3 semantic dimensions: Informativeness, Information Type, and Information Source. This latter attribute, our target class in this experiment, in CrisisLexT26 ranged over the nominal values Eyewitness, Government, NGO, Business, Traditional/Internet Media, Outsiders; as we want a binary classification task, we merged the latter labels into one, so we are left with only "Eyewitness" and "Non-Eyewitness" in the dataset. Because we used the data to extract lexical features, we had to further filter by language, which was done by using an in-house language detection module, based on continuously updated language specific frequency tables (Steinberger et al., 2013). Moreover, as we need to query Twitter APIs for retrieving meta-data features and some of the messages or accounts from the dataset are not accessible or existing anymore, we ended up with 2 datasets of 10342 tweets (950 positive instances) and 465 tweets (52 positives) for English and Italian, respectively. After some preliminary tests we measured that the performance of learning algorithms was rather poor for such skewed datasets, therefore we balanced them (50/50%) by randomly under-sampling the negative instances before training the classifiers, using Weka Spread Subsample algorithm implementation (Hall et al., 2009).

We evaluated to what extent classifiers trained on our feature set were able to generalize to unseen test sets, both intra- and across event types. To this purpose, we compared the performance of a standard 10-fold cross validation, where the feature distribution in the test and training data is likely to be similar (we label this evaluation mode CrossEv), with a "hold-out" scenario, where we train the system on data from a subset of the events and test it on a separate subset of events, either of the same event type (SplitIntr) or a different one (SplitAcross). The two latter experiments emulate better a real-world scenario, where eyewitness messages from a new unseen disaster need to be detected by a classifier that has been trained on previous events.

Table 2 reports Precision, Recall and F1-measure for 3 different classifiers we experimented with: Naïve Bayes, Support Vector Machine and Random Forest, as implemented in Weka (Hall et al., 2009). Experiments were performed both in English and Italian language, for the 3 scenarios illustrated (CrossEv, SplitIntr, SplitAcross).

There is an evident performance drop for the SplitIntr and SplitAcross scenarios, particularly for the Precision, which confirms the well-known issue with domain adaptation for tweet classification tasks (see Li et al., 2015). Figures for Italian are somehow difficult to interpret, as performance unexpectedly grows from SplitIntr through SplitAcross across all three classifiers, which may be partially due to limited dataset size.

Table 2: Performance of 3 eyewitness report classifiers on the same feature set, for 10-fold cross-evaluation, intra-type hold-out and cross-type hold-out evaluation.

		NaïveBayes			SVM			RandomForest		
		P	R	F1	P	R	F1	P	R	F1
EN	CrossEv	0.700	0.843	0.765	0.798	0.748	0.772	0.808	0.774	**0.790**
	SplitIntr	0.148	0.413	0.218	0.215	0.370	**0.272**	0.162	0.348	0.221
	SplitAcross	0.116	0.683	**0.198**	0.116	0.341	0.173	0.117	0.293	0.167
IT	CrossEv	0.639	0.750	**0.690**	0.567	0.654	0.607	0.573	0.827	0.677
	SplitIntr	0.098	0.889	**0.177**	0.093	1.000	0.171	0.088	1.000	0.162
	SplitAcross	0.239	0.981	0.385	0.240	1.000	**0.387**	0.188	1.000	0.317

Overall, results significantly outperform the previous works on the same dataset. For example for English we nearly double F1 for the best classifier from (Pekar et al., 2016) in CrossEv scenario, and raise the best F1 from around 7 up to 17 and up to 27 for the SplitIntr and SplitAcross scenarios, respectively, while we still make use of comparatively fewer and less deep features. However, direct comparison is not possible as they were not balancing the dataset. As we explained, the lexical features can be easily extracted from unlabeled data: this suggests that tackling the domain adaptation issue may be achievable in our approach by simply integrating unlabeled instances belonging to the new target event into the dataset.

Moreover, results for Italian and English are roughly comparable, which provides evidence that our choice of using only shallow linguistic features makes the method quite scalable across languages.

Overall, all features showed low to medium correlation with target labels (≤.3 and ≤.4 for English and Italian, respectively). This is probably due to the fact that the whole corpus is heterogeneous and distribution over disaster types and disaster instances is unbalanced, which is confirmed by some top ranking but highly specific hashtags (e.g. #Modena, #yycflood). Nonetheless, some of the *Personal* features (*my*, *io*, *mia*) consistently ranked among the highest across the two languages, together with some topic related unigrams (*crash*, *flood*, *dead*, *terremoto*, *vittime*). Interestingly, some perception and cognitive state verb unigrams (*sentito*, *svegli*) were highly correlated with the positive class in Italian, which suggests that a new dedicated semantic feature class may be added. Finally, the *containsURLs* feature turned out to be unexpectedly correlated with positive class for both languages.

BUILDING A TWITTER APPLICATION FOR MONITORING DISASTER IMPACT

We deployed the methods illustrated so far in a simple on-line application based on the Twitter platform. Disastrobot (https://twitter.com/DisastroBot) is a Twitter channel that dynamically tracks target disaster events worldwide by filtering, tagging and re-posting Twitter status messages, therefore serving as a social media content hub for situation monitoring and impact assessment. The system is multilingual and currently comprises resources for English, Italian and Spanish.

The channel maintains a list of 5 target disaster events (currently only Quakes, Floods and Hurricanes), which is dynamically updated by processing an RSS feed from the GDACS portal (http://www.gdacs.org/) and applying heuristic functions of the event Severity, Affected Population and duration of the monitoring. For each target event, a pre-filtering is automatically created by submitting to the Twitter public Streaming API a query comprising:

1. A number of geographical bounding boxes heuristically computed from the disaster geo-coordinates and severity;
2. A set of domain-specific terms/hashtags capturing the type of the event (e.g. *#quake, inundación*)

The tweets continuously collected are then processed by the event detection grammars described earlier. Each tweet matching one of the 13 event types is then retweeted on the channel timeline, labeled with a corresponding hashtag (e.g. *#DSTRServiceInterruption*, *#DSTRFoodShortage*, etc.). By using hashtags we add an event meta-information layer to the textual content and allow users to visualize situational updates for selected event types (by just clicking on the type hashtag). Additionally, topic-classified tweets are processed by a trained classifier for eyewitness report detection, and a corresponding *#DSTREye* label is possibly added.

Fig.1 below shows sample tweets published on Disastrobot on 25th of November 2016.

Figure 1. Snapshots from the Disastrobot Twitter channel timeline.

CONCLUSIONS

We presented and evaluated a novel unsupervised algorithm for event detection and we have applied it on news articles, in order to detect descriptions of the effects of disasters. The algorithm accuracy can be improved further, but even without manual supervision, it succeeds in detecting a significant amount of micro-events. Integrating some supervision through bootstrapping learning iterations can be also used to continuously enhance the grammar model.

We also presented a Twitter application which is based on a weakly-supervised algorithm for learning of event-detection rules. This application detects disaster-related micro-events, reported on Twitter in three different languages. Compared to previous approaches to tweet topic classification our method does not require extended annotated data: however, we are planning to use the relatively accurate output of our learned rules to generate event-annotated training data for statistical classifiers, in view of extending the current system recall. In the same direction, we intend to experiment with applying our unsupervised algorithm to Social Media data and with assigning topic labels to the non-categorized output patterns: this could be done by computing similarity with the patterns extracted by the second algorithm.

For both event and eyewitness report detection, we are currently exploring methods for adapting the feature space of statistical classifiers to the contexts of target disasters, by dynamically integrating unannotated training data from the incoming event streams. This should mitigate the domain adaption performance drop and is still a missing component in view of making the system usable in operational scenarios.

Finally, we aim to generalize the methods in order to process content across different social media platforms.

ACKNOWLEDGMENTS

The work was partly supported by Project MediaGist, EU's FP7 People Programme (Marie Curie Action), no. 63078.

REFERENCES

Athanasopoulou, G., Klasinas, I. , Georgiladakis, S., Iosif, E. and Potamianos, A. (2014) Using lexical, syntactic and semantic features for non-terminal grammar rule induction in Spoken Dialogue Systems, IEEE Spoken Language Technology Workshop (SLT), South Lake Tahoe, NV, 2014, pp. 596-601

Fang, R., Liu, X., Li, Q., Nourbakhsh, A. and Shah, S. (2016) Witness Identification in Twitter, Conference on Empirical Methods in Natural Language Processing, 65.

Fugate, C. (2011) "Understanding the power of social media as a communication tool in the aftermath of disasters." Senate Committee on Homeland Security and Governmental Affairs, Subcommittee on Disaster Recovery and Intergovernmental Affairs.

Grishman, R. and Sundheim. B. (1996) Message Understanding Conference-6: A Brief History, COLING. Vol. 96. 1996.

Gunn, S. W. A. (2003) The language of disasters: a brief terminology of disaster management and humanitarian action Basics of International humanitarian missions (2003): 37-40.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The WEKA data mining software: an update. ACM SIGKDD explorations newsletter, 11(1), 10-18.

Homeland Security (2014) Using social media for enhanced situation awareness and decision support. Virtual Social Media Working Group and DHS First Responders Group.

Howell, G. V., and Taylor, M. (2011) When a crisis happens, who turns to Facebook and why, Asia Pac Publ Relat J 12 (2011): 1-8.

Imran, M., Elbassuoni, S. M., Castillo, C., Diaz, F. , and Meier, P. (2013). Extracting information nuggets from disaster-related messages in social media. Proc. of ISCRAM, Baden-Baden, Germany.

Imran, M., Mitra, P. and Srivastava, J. (2016). Cross-language domain adaptation for classifying crisis-related short messages. arXiv preprint arXiv:1602.05388.

Imran, M., Castillo, C., Diaz, F. and Vieweg, S. (2015). Processing social media messages in mass emergency: A survey. ACM Computing Surveys (CSUR), 47(4), 67.

Li, Hongmin, Nicolais Guevara, Nic Herndon, Doina Caragea, Kishore Neppalli, Cornelia Caragea, Anna Squicciarini and Andrea H. Tapia. (2015). Twitter Mining for Disaster Response: A Domain Adaptation Approach. ISCRAM 2015.

Lindsay, B. R. (2011). Social media and disasters: Current uses, future options, and policy considerations.

Middleton, S. E., Middleton, L., and Modafferi, S. (2014). Real-time crisis mapping of natural disasters using social media. IEEE Intelligent Systems, 29(2), 9-17.

Murray, V. and Ebi, K. L. (2012). IPCC special report on managing the risks of extreme events and disasters to advance climate change adaptation (SREX).

Olteanu, A., Vieweg, S. and Castillo, C. (2015). What to expect when the unexpected happens: Social media communications across crises. In Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (pp. 994-1009). ACM.

Palka, D. and Zachara, M. (2015). Automatic Grammar Induction for Grammar Based Genetic Programming. In

International Conference on Artificial Intelligence and Soft Computing (pp. 350-360). Springer International Publishing.

Pekar, V., Binner, J., Najafi, H. and Hale, C. (2016). Selecting Classification Features for Detection of Mass Emergency Events on Social Media. In Proceedings of the International Conference on Security and Management (SAM) (p. 192). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp).

Piskorski, J., Tanev, H., Atkinson, M., Van Der Goot, E. and Zavarella, V. (2011). Online news event extraction for global crisis surveillance. In Transactions on computational collective intelligence V (pp. 182-212). Springer Berlin Heidelberg.

Power, R., Robinson, B., Colton, J. and Cameron, M. (2015). A case study for monitoring fires with Twitter In Proceedings of the 12th International Conference on Information Systems for Crisis Response and management (ISCRAM).

Steinberger, R., Pouliquen, B. and Van der Goot, E. (2013). An introduction to the Europe Media Monitor family of applications. arXiv preprint arXiv:1309.5290.

Tanev, H., Ehrmann, M., Piskorski, J. and Zavarella, V. (2012). Enhancing Event Descriptions through Twitter Mining. In ICWSM.

Tanev, H. and Zavarella, V. (2014). Multilingual Lexicalisation and Population of Event Ontologies: A Case Study for Social Media. In Towards the Multilingual Semantic Web (pp. 259-274). Springer Berlin Heidelberg.

Tanev, H. (2014). Learning Textologies: Networks of Linked Word Clusters. In Text Mining (pp. 25-40). Springer International Publishing.

Thywissen, K. (2006). Core terminology of disaster reduction. Measuring vulnerability to natural hazards: Towards disaster resilient societies, United Nations University Press, Hong Kong.

Verma, S., Vieweg, S., Corvey, W. J., Palen, L., Martin, J. H., Palmer, M. and Anderson, K. M. (2011). Natural Language Processing to the Rescue? Extracting" Situational Awareness" Tweets During Mass Emergency. In

ICWSM.

Vieweg, S., Hughes, A. L., Starbird, K. and Palen, L. (2010). Microblogging during two natural hazards events: what twitter may contribute to situational awareness. In Proceedings of the SIGCHI conference on human factors in computing systems (pp. 1079-1088) ACM.

Zielinski, A., Bügel, U., Middleton, L., Middleton, S. E., Tokarchuk, L., Watson, K. and Chaves, F. (2012). Multilingual analysis of twitter news in support of mass emergency events. In EGU General Assembly Conference Abstracts (Vol. 14, p. 8085).