Jointly Learning from Social Media and Environmental Data for Typhoon Intensity Prediction

Hamada M. Zahera  
DICE, Paderborn University  
FCI, Menoufia University  
hamada.zahera@uni-paderborn.de

Mohamed Ahmed Sherif  
DICE, Paderborn University  
AKSW, Leipzig University  
mohamed.sherif@uni-paderborn.de

Axel-Cyrille Ngonga Ngomo  
DICE, Paderborn University  
AKSW, Leipzig university  
axel.ngonga@uni-paderborn.de

ABSTRACT
Existing technologies employ different machine learning approaches to predict disasters from historical environmental data. However, for short-term disasters (e.g., earthquakes), historical data alone has a limited prediction capability. In this work, we consider social media as a supplementary source of knowledge in addition to historical environmental data. Further, we build a joint model that learns from disaster-related tweets and environmental data to improve prediction. We propose the combination of semantically-enriched word embedding to represent entities in tweets with their semantics representations computed with the traditional word2vec. Our experiments show that our proposed approach outperforms the accuracy of state-of-the-art models in disaster prediction.

CCS CONCEPTS
• Information systems → Web mining; • Computing methodologies → Machine learning approaches;

KEYWORDS
Social Media Analysis, Semantic Embedding, Joint Model

1 INTRODUCTION
Disaster prediction and early warnings are crucial when mitigating the impact of disasters. However, there are many factors that still limit the accuracy of the current prediction algorithms such as the lack of complete data on natural hazards, monitoring instruments and the highly dynamic nature of natural hazards [12]. Interestingly, social media plays an increasingly significant role in disaster management and communication. Several studies leveraged shared information on social media to reduce the impact of disasters and deliver faster responses [7, 13]. Emergency managers use social media to engage with the public quickly and widely. Such correlations are valuable for supporting decision makers in emergency response processes. Previous works used data mining techniques to extract such correlation for support decision making in emergency responses. For example, Amrita et al. [2] applied a wavelet analysis to track the progression of disasters from social media data.

In this work, we propose a joint model to classify the intensity of typhoons (also called a typhoon’s category or class) by learning from environmental data and social media. We were inspired by previous works (see, e.g., [10, 16]) which suggest that the joint learning of multiple models can significantly outperform standalone models. Our proposed approach consists of two cascaded jointly-trained models. The first model (dubbed Feature Extractor) analyzes typhoon-related tweets and computes statistical features (i.e., tweets volume and sentiments variances). To capture tweet sentiments, we employ a semantics-enriched word embedding in which entities are recognized and represented as semantics vectors.

The second model (dubbed Typhoon Classifier) takes an input of the combined features extracted by the first model and the environmental data. Both models are trained jointly through a shared loss function and their learning parameters are optimized using the same gradient descent. We employ architectures based on Deep Neural Networks (DNN) and Recurrent Neural Networks (RNN) as our baseline approaches.

We summarize our contributions in this paper as follows: (1) we propose a joint model to classify typhoon intensities from social media and environmental data. (2) we study the impact of incorporating semantics embedding from knowledge graphs to enrich tweets representation. Our experiments show that feeding our model with semantics representation of the entities included in the tweets improved the system overall performance. (3) we provide a new disaster dataset of tweets and environmental dataset (dubbed TED), which comprises environmental data of 70 typhoons and their corresponding tweets. The dataset as well as our implementation are available at the project website.

2 RELATED WORKS
Social media analysis for disaster management. Several studies leveraged the role of social media in disasters management [1, 13]. For example, Yury et al. [6] proposed a social media-based framework to estimate damages initially from social media. Their system analyzed users activities on Twitter before, during and after the hurricane Sandy. Their results showed a strong correlation between activities on social media and hurricane path. Similarly, in this paper, we studied social media behaviors during different

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

K-CAP’19, 19-21 November, Marina del Rey, USA  
© 2019 Association for Computing Machinery.  
ACM ISBN 978-x-xxxx-xxxx-x/YY/MM. . . $15.00  
https://doi.org/10.1145/nnnnnnn.nnnnnnn  
https://github.com/dice-group/joint-model
typhoons, our analysis showed an implicit correlation based on tweets volume and sentiments. Social media has also been shown to be a rapid event detector from the crowd. For example, Takeshi et al. [13] proposed a probabilistic approach to predict the center and path of natural hazards based on geo-based tweets. They analyzed the behavior of social media users during different cases of earthquakes and used such correlation as detectors for earthquakes. Although previous works have explored the role of social media in crisis events and enhancing situational awareness, few research studies showed how tweets sentiments have been used as discriminative features to improve event prediction.

**Semantics-enriched data mining.** Different research works employ semantics nuances to boost the performance in machine learning (ML) tasks. For instance, the authors of [3] extended the traditional bag-of-words models with a bag-of-concepts model extracted from semantics knowledge graphs (e.g. DBpedia). However, the authors represented the presence of concepts as vector of indices within a concept space. In contrast, in our approach, we leverage semantics embedding from knowledge graph which projects concepts (i.e., entities) and their relationship for conceptualized data representation.

**Joint-Learning models.** Recently, joint learning models have achieved superior performances in complex tasks. For instance, Jishu et al. [11] proposed a jointly-training RNN-based models to extract keyphrases from disaster-related tweets. Through intersection of two RNN models via joint learning, the experimental results clearly demonstrated significant performances in comparison with existing baseline approaches. In a related task of image classification, Wanli et al. [9] proposed a unified deep model that jointly learn from different components in image classification task. Inspired by these works, we propose our joint training model to learn features from social media and environmental data in comparison to individual or ensemble models employed in event detection.

## 3 THE APPROACH

We propose two architectures of jointly-trained models (i.e. Feature Extractor+Typhoon Classifier). Figure 1 shows the architecture of our first model (LSTM+DNN). In Figure 2, we present the RNN-based architecture of Typhoon Classifier in our second model (LSTM+RNN). Note that, the architecture of Feature Extractor remains the same as in Figure 1.

### 3.1 Problem Formulation

Let $D = \{ (x_1, y_1), \ldots, (x_n, y_n) \}$ be a typhoon environmental dataset, where $x_i \in X$ represents a typhoon instance with $m$ environmental features (e.g., time-tamp, wind-speed, sea level pressure and gust) and $y_i \in Y$ corresponds to typhoon’s category (e.g., either a tropical-depression, tropical-storm, typhoon or super-typhoon). For each $x_i$, we collect all related tweets posted within its time-stamp, we dub such tweets as $T$. Then, we analyse these tweets to extract statistical features (i.e., tweets-based features) such as tweets volume and variances of tweets’ sentiments. Finally, we combine these features with typhoon’s environmental data in one input vector.

**Definition 1. Task Description.** Our goal is to design a classification model able to learn features from $D$ and $T$ to predict the typhoon category ($Y$). We build our approach as jointly-training of two cascaded models $F_1$ and $F_2$ dubbed Feature Extractor and Typhoon Classifier. To ensure the joint training of both models, we combine their loss functions ($L_{F_1}, L_{F_2}$) in one shared loss function as follows:

$$L_{joint} = \lambda L_{F_1} + L_{F_2} + \lambda L_{F_2}$$

(1)

Where $\lambda$ is a control parameter used to balance the individual loss functions ($L_{F_1}, L_{F_2}$). To compute the training losses, we use the cross-entropy function as a loss function as follows:

$$L = \frac{1}{n} \sum_{i=1}^{n} [y_i \log(H(x_i)) + (1 - y_i) \log(1 - H(x_i))]$$

(2)

Where $y_i$ and $H(x_i)$ donate target and predicted typhoon categories respectively for typhoon instance $x_i$.

**Definition 2. Joint-Learning.** Let $\phi_1$ and $\phi_2$ be the learning parameters of Feature Extractor and Typhoon Classifier. Both $\phi_1, \phi_2$ are optimized concurrently as follows: Assume two consecutive batches of data $B_i$ and $B_{i+1}$, the learning parameters in $B_i$ are optimized using the same gradient descent by backpropagating the gradients to both models as illustrated in Figure 1. In $B_{i+1}$, the computation of statistical features by Feature Extractor model are hence further adapted not only from the losses in its outputs, but also from the losses in the final output by Typhoon Classifier model. Therefore, Feature Extractor feeds adaptive features from social media to Typhoon Classifier, which improves the final accuracy.

### 3.2 Semantics-enriched Word Embedding

Our analysis$^2$ of tweets during different typhoons suggests that we can use tweets volume and sentiments as additional features in our model. To perform sentiment analysis, we used the continuous skip-gram approach [8] to train our word embedding model on typhoon tweets and the standford sentiment140 dataset$^3$. While generic word embeddings trained on generic large-scale datasets could have been used here, they often do not capture domain-specific knowledge and semantic nuances. In contrast, domain-adapted word embeddings are effective in the field of the context in which they are trained as they capture domain-specific knowledge [14].

---

$^2$For more details about our tweets analysis see https://git.io/tweets-analysis

$^3$http://cs.stanford.edu/people/alecmgo/trainingandtestdata.zip
Now, given tweet \( t = (w_1, w_2, e_1, \ldots, e_j, w_k) \), with \( w \) words and \( e \) entities. We map each word \( w_i \) to its embedding vector in \( \mathbb{R}^{1 \times d} \) with \( d \) dimension. Unlike classical word embeddings, we represent entity \( e \) with its corresponding vectors from the knowledge graph of ConceptNet [15], where entities and their relationships are projected into the same embedding space.

For each tweet, we build an embedding matrix \( M \in \mathbb{R}^{s \times |d|} \), where \( s \) is number of words per tweet. Each row \( i \) of \( M \) represents the word2vec embedding of the \( w_i \) at the corresponding position \( i \) in a tweet. Our word2vec model has a dimension \( d \) of 200, vocabulary size of 47,137 words and 1,152 recognized entities. Due to variable lengths of tweets, we fixed \( s \) to the average number of words per tweet to maintain a regular embedding matrix. For this reason, we truncated longer tweets and padded shorter tweets with zeros.

### 3.3 Feature Extractor

Feature Extractor is the first part of our joint model that aims to extract statistical features from tweets. We employ an LSTM-based model [5] to handle words sequences in tweets. In particular, we use an embedding layer to map words to their embedding vectors. Then, we employ one LSTM layer with 64 units and use softmax function in output layer. We use the following equations to compute the LSTM output \( (h) \):

\[
\begin{align*}
 i_t &= \sigma(x_t^T \cdot x_t + h_{t-1}^T \cdot h_{t-1} + b_i) \\
 f_t &= \sigma(x_t^T \cdot x_t + h_{t-1}^T \cdot h_{t-1} + b_f) \\
 o_t &= \sigma(x_t^T \cdot x_t + h_{t-1}^T \cdot h_{t-1} + b_o) \\
 g_t &= \text{tanh}(x_t^T \cdot x_t + h_{t-1}^T \cdot h_{t-1} + b_g) \\
 m_t &= f_t \otimes m_{t-1} + i_t \otimes g_t \\
 h_t &= o_t \otimes \text{tanh}(m_t)
\end{align*}
\]

Where \( i, m, f, o \) and \( \otimes \) are the input, memory, forget and output gates at time \( t \). \( \theta \) represents the weights and biases vectors. \( \sigma, \otimes \) are the sigmoid function and produce-wise multiplication. The final outputs of LSTM are probabilities of tweet sentiments (positive and negative) computed by softmax function in Equation 4. We extract and combine statistical features from LSTM outputs with typhoon data as \( D \in \mathbb{R}^{n \times (m+c+\ldots+e)} \), \( c, \ldots, e \) represents the statistical features: tweets count, variance of negative sentiments and variance of positive sentiments.

\[
P(y = j | z) = \frac{e^{z_j}}{\sum_{k=1}^{k} e^{z_k}}
\]

where \( P(y = j | z) \) is the probability of typhoon category \( j \), given input vector \( z = \theta^T x \) and \( k \) is number of typhoon categories.

### 3.4 Typhoon Classifier

Typhoon Classifier takes input features from Feature Extractor to predict the typhoon intensity as a final output. We explore two different deep architectures (i.e., the DNN and RNN) models as a typhoon classifier as follows:

**The DNN Model.** We employ two dense layers as the basis of our DNN model. Each layer has 16 linearly-rectified units (with ReLU activation function). In addition, we use dropout rates (0.60 after first layer and 0.75 after second layer) to avoid overfitting and increase the model architecture as the Typhoon Classifier.

To classify typhoon intensities, we use softmax function in output layer to compute probabilities for all typhoon categories as in Equation 4. At the end, the category with the highest probability is returned as the model output.

**The RNN Model.** We explore an RNN-based model as a Typhoon Classifier to model the sequences in typhoon data. Specifically, we use one LSTM layer with 64 units and a dropout rate of 0.75 as shown in Figure 2. In the DNN model, we compute the outputs of the RNN model using softmax function and return the class with highest probability as final output.

### 4 EXPERIMENTS

We conducted several sets of experiments to benchmark the performance of our approach and the baselines. We aim to answer the following questions:

- Q1. To which extent, social media can improve the performance of the state-of-the-art disaster prediction approaches?
- Q2. What is the impact of semantic embedding in tweets representation on the performance of our proposed approach?

We compared our approach with different baselines including traditional ML and deep neural models. We selected the SVM classifier as our traditional baseline classifier as it outperforms the other traditional classifiers [3]. For our experiment, we implemented an RBF-based kernel SVM classifier which is trained with typhoon environmental data.

We also used two benchmarks from the deep neural based models (i.e., DNN and RNN), where both achieved good performances in disaster-related research [4]. The architecture of DNN and RNN we implemented are explained in Section 3.4. We specify our proposed approaches used in the experiments as follows:

- **LSTM+DNN (word embedding):** Our first approach of two deep models (LSTM and DNN) are jointly trained with combined features from environmental data and word embeddings.
- **LSTM-DNN (semantic embedding):** This model is the same as LSTM+DNN (word embedding), but we consider semantic-enriched word embedding for data representation.
- **LSTM-RNN (word embedding):** Our second approach of two deep models (LSTM and RNN) trained with combined features from environmental data and word embeddings.
- **LSTM-RNN (semantic embedding):** This model is the same as LSTM+RNN (word embedding), with considering semantic-enriched word embedding for data representation.

---

For more details about our experiments see https://git.io/jointmodel-experiments
which helps to fit the features by joint training and improve the accuracy, precision, recall and F1-score with the semantics-based embedding over those based only on word embedding. In particular, the accuracy of our joint-models are improved by up to 3% in both LSTM+DNN and LSTM+RNN models compared to word embedding based models.

5 CONCLUSION
We propose a joint model that learns from social media and environmental data to improve disasters prediction. Unlike previous works, we extract adaptive features based on joint training of two deep models (i.e. LSTM+DNN or LSTM+RNN). Further, we study the impact of applying semantically-enriched data representation on the performance of our system. We employed semantics embedding from the external knowledge graph of ConceptNet. Our evaluation showed significantly improved accuracies in our approaches when compared to state-of-the-art baselines.

ACKNOWLEDGMENTS
This work has been supported by the project LIMBO (Grant no. 19F2029D), OPAL (no. 19F2028A) and SOLIDE (no. 13N14456).

REFERENCES