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

ACM Reference Format:

Hamada M. Zahera, Mohamed Ahmed Sherif, and Axel-Cyrille Ngonga Ngomo. 2019. Jointly Learning from Social Media and Environmental Data for Typhoon Intensity Prediction. In *Proceedings of ACM Conference (K-CAP'19).* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

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[1]

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

[1] https://github.com/dice-group/joint-model

K-CAP'19, 19-21 November, Marina del Rey, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

typhoons, our analysis showed a 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-word 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, *Jishnu 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 $D = \{\langle x_1, y_1 \rangle \dots \langle x_n, y_n \rangle\}$ 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

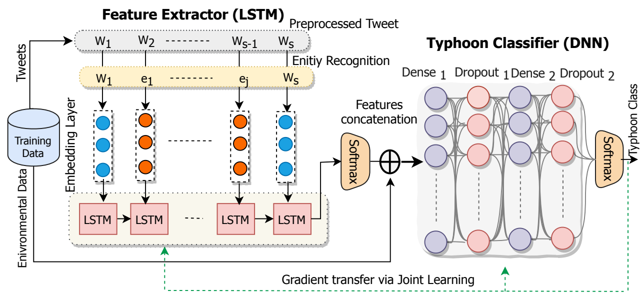

Figure 1: Our joint model (LSTM+DNN). Entities vectors (in orange) are obtained ConceptNet and Words vectors (in blue) from our word embedding.

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:

$$L_{joint} = \lambda_{F_1} \cdot L_{F_1} + \lambda_{F_2} \cdot L_{F_2} \tag{1}$$

Where λ 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:

$$L = \frac{1}{n} \sum_{i=1}^{n} [y_i \log(H(x_i))) + (1 - y_i) \log(1 - H(x_i))] \tag{2}$$

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

Definition 2. Joint-Learning. Let ϕ_1 and ϕ_2 be the learning parameters of *Feature Extractor* and *Typhoon Classifier*. Both ϕ_1 , ϕ_2 are optimized concurrently as follows: Assume two consecutive batches of data B_t and B_{t+1}, the learning parameters in B_t are optimized using the same gradient descent by backpropagating the gradients to both models as illustrated in Figure 1. In B_{t+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 $t = (w_1, w_2, e_1, \ldots, e_j, w_s)$, 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 a 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 ($\boldsymbol{h}$):

$$\begin{aligned}
\boldsymbol{i_t} &= \sigma(\theta_{xi}^T.x_t + \theta_{hi}^T.h_{t-1} + b_i) \\
\boldsymbol{f_t} &= \sigma(\theta_{xf}^T.x_t + \theta_{hf}^T.h_{t-1} + b_f) \\
\boldsymbol{o_t} &= \sigma(\theta_{xo}^T.x_t + \theta_{ho}^T.h_{t-1} + b_o) \\
\boldsymbol{g_t} &= \tanh(\theta_{xg}^T.x_t + \theta_{hg}^T.h_{t-1} + b_g) \\
\boldsymbol{m_t} &= f_t \otimes m_{t-1} + i_t \otimes g_t \\
\boldsymbol{h_t} &= o_t \otimes \tanh(m_t)
\end{aligned} \tag{3}$$

Where i, m, f and o are the input, memory, forget and output gates at time t. θ and b represents the weights and biases vectors. σ, $\otimes$ 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 $D \in \mathbb{R}^{n\times[m+c,v_-,v_+]}$. c, v_-, v_+ represents the statistical features: tweets count, variance of negative sentiments and variance of positive sentiments.

$$P(y = j|z) = \frac{e^{z_j}}{\sum_{i=1}^{k} e^{z_k}} \tag{4}$$

$P(y = j|z)$ 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 robustness. 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.

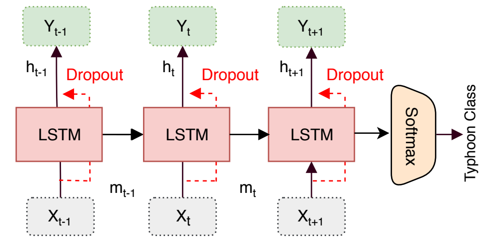

Figure 2: RNN architecture as the *Typhoon Classifier* .

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. Like 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[4] to benchmark the performance of our approach and the baselines. We aim to answer the following questions:

Q_1. To which extent, social media can improve the performance of the state-of-the-art disaster prediction approaches?

Q_2. 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 outperform 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 represenation.

[4]For more details about our experiments see https://git.io/jointmodel-experiments

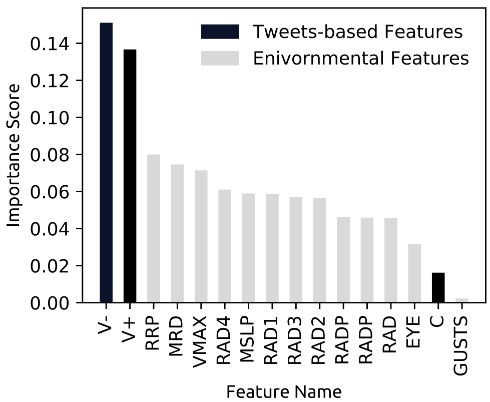

Figure 3: Importance of environmental and tweets-based features computed by RandomForest algorithm.

Table 1: Evaluation on test dataset using *Accuracy* (A), *Precision* (P), *Recall* (R) and *F1-Score* (F).

Model	Description	A	P	R	F
SVM	Baseline	0.57	0.34	0.57	0.43
DNN	Baseline	0.75	0.80	0.75	0.78
RNN	Baseline	0.80	0.82	0.80	0.81
LSTM+DNN	Word emb.	0.87	0.89	0.87	0.88
LSTM+DNN	Semantic emb.	0.91	0.92	0.92	0.91
LSTM+RNN	Word emb.	0.86	0.87	0.86	0.85
LSTM+RNN	Semantic emb.	0.89	0.90	0.89	0.89

4.1 Discussion and Result Analysis

To answer Q_1**,** we evaluated the baseline models with features extracted from the environmental data. Further, we used the same features, in addition to the features extracted from relevant tweets to train our proposed models. Table 1 shows how challenging is the task of typhoons intensity prediction, where the best baseline model (i.e., the RNN model) produces an accuracy of 0.80. In contrast, our proposed models clearly demonstrate a significant performance where tweets-based features were incorporated with environmental data. In particular, the best accuracy is achieved by our proposed model LSTM+DNN, where it outperforms the respective baseline model (DNN) by 11.7% on average (LSTM+DNN average accuracy of 0.87 compared to 0.75 for the DNN). The other proposed model of LSTM+RNN also outperforms the RNN model by 5.8% on average (0.86 average accuracy for LSTM+RNN compared to 0.80 for RNN).

As shown in Figure 3, tweet-based features were found to be more important than environmental features. As discussed in Section 3.1, tweet-based features are adaptive to prediction losses in our models which helps to fit the features by joint training and improve the accuracy of final prediction.

To answer Q_2**,** we investigated the impact of incorporating semantics embedding on the performance of our proposed approach. Our experiments showed improved performance in terms of 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. 19F2029I), OPAL (no. 19F2028A) and SOLIDE (no. 13N14456).

REFERENCES

[1] Adam Acar and Yuya Muraki. 2011. Twitter for crisis communication: lessons learned from Japan's tsunami disaster. *International Journal of Web Based Communities* 7, 3 (2011), 392–402.
[2] Amrita Anam, Aryya Gangopadhyay, and Nirmalya Roy. 2018. Evaluating Disaster Time-Line from Social Media with Wavelet Analysis. In *2018 IEEE International Conference on Smart Computing (SMARTCOMP)*. IEEE, 41–48.
[3] Grégoire Burel, Hassan Saif, Miriam Fernandez, and Harith Alani. 2017. On semantics and deep learning for event detection in crisis situations. In *Proceedings of the International Workshop on Semantic Deep Learning (SemDeep)*. ESWC 2017.
[4] Xingyu Chen, Lei Zou, and Bo Zhao. 2019. Detecting Climate Change Deniers on Twitter Using a Deep Neural Network. In *Proceedings of the 2019 11th International Conference on Machine Learning and Computing*. ACM, 204–210.
[5] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.
[6] Yury Kryvasheyeu, Haohui Chen, Nick Obradovich, Esteban Moro, Pascal Van Hentenryck, James Fowler, and Manuel Cebrian. 2016. Rapid assessment of disaster damage using social media activity. *Science advances* 2, 3 (2016), e1500779.
[7] Peter M Landwehr and Kathleen M Carley. 2014. Social media in disaster relief. In *Data mining and knowledge discovery for big data*. Springer, 225–257.
[8] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*. 3111–3119.
[9] Wanli Ouyang and Xiaogang Wang. 2013. Joint deep learning for pedestrian detection. In *Proceedings of the IEEE International Conference on Computer Vision*.
[10] Hongwei Qin, Junjie Yan, Xiu Li, and Xiaolin Hu. 2016. Joint training of cascaded cnn for face detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 3456–3465.
[11] Jishnu Ray Chowdhury, Cornelia Caragea, and Doina Caragea. 2019. Keyphrase Extraction from Disaster-related Tweets. In *The World Wide Web Conference*. ACM, 1555–1566.
[12] A Reese. 2016. How we'll predict the next natural disaster: Advances in natural hazard forecasting could help keep more people out of harmâĂŹs way. *Discover Magazine, Sep* (2016).
[13] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. 2010. Earthquake shakes Twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on World wide web*. ACM, 851–860.
[14] Prathusha K Sarma, YIngyu Liang, and William A Sethares. 2018. Domain Adapted Word Embeddings for Improved Sentiment Classification. *arXiv preprint arXiv:1805.04576* (2018).
[15] Robert Speer, Joshua Chin, and Catherine Havasi. 2017. Conceptnet 5.5: An open multilingual graph of general knowledge. In *Thirty-First AAAI Conference on Artificial Intelligence*.
[16] Jonathan J Tompson, Arjun Jain, Yann LeCun, and Christoph Bregler. 2014. Joint training of a convolutional network and a graphical model for human pose estimation. In *Advances in neural information processing systems*. 1799–1807.