Knowledge-aware Few Shot Learning for Event Detection from Short Texts

Zhihang Li
lizhihang19@outlook.com
Huazhong Agricultural University
Wuhan, China

Jinjin Guo *
UMguojinjin@outlook.com
JD Intelligent Cities Research
Beijing, China

Zhichao Huang
iceshzc@gmail.com
JD Intelligent Cities Research
Beijing, China

ABSTRACT

Event detection in a city is of great significance for the government to listen to the voice of the citizens, be aware of the real occurrences in a city, and then make wiser policies. However, in reality it often occurs some important events with few samples are easily to be overwhelmed by the massive information and hard to be recognized, and the limited word description from the short texts even makes the recognition harder. To address the problems, we propose a knowledge-aware event detector by incorporating the external knowledge to detect the events with few examples. The external knowledge incorporation with different semantic relation is capable to enrich the short texts. In addition, we leverage the representative few shot learning framework to formulate the event detection as the text classification problem. The proposed model is evaluated on two widely event-detection datasets. The experiments show a consistent accuracy improvement. The findings validates that our model with the knowledge infusion is effective to detect the few shot events from the short texts.

CCS CONCEPTS

• **Computing methodologies → Machine learning**; **Supervised learning**; **Semantic networks**; **Information extraction**.

KEYWORDS

Few Shot Learning, Event Detection, Prototypical network, External knowledge.

ACM Reference Format:
Zhihang Li, Jinjin Guo, and Zhichao Huang. 2022. Knowledge-aware Few Shot Learning for Event Detection from Short Texts. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 5 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

The task of event detection means to extract the important, newsworthy and real-world occurrences from the short texts posted by the citizens on the social media platform. Conducting event detection greatly benefits the government to understand the public concerns, be aware of the real occurrences in a city, and then make wiser policies. Additionally, the event detection task supports plentiful applications including product recommendation, trend prediction [9] and crisis management. Therefore , the study of event detection has been extensively explored as either an incremental clustering problem [4, 12], or a topic modeling problem [11, 9, 8]. However, in reality some critical events with few examples are hard to be detected. For instance, though only few texts report the positive confirmed cases of coronavirus at very early stage, they are crucial to be noticed by the authority yet easily to be overwhelmed by other massive information. In this situation, the conventional event detection methods may lead to the overfitting problem when dealing with events with few examples. Hence, recently researchers have turned to the framework of few shot learning for event detection.

To emulate few shot learning from few examples, N-way K-shot episodic training [5, 2] is often exploited by the existing studies. At each training iteration a small subset of N event types with K ($K \in [1, 10]$) examples, known as a support set, is used for learning, and further examples of the same class, known as a query set, is used to evaluate the performance. Based on the framework of few shot learning, the most recent work further advances the event detection performance by fully leveraging the internal factors. The work [15] exploits the relationship between training task to enforce event prediction consistency, the work [19] infuses the semantic meaning of labels into the framework, and the work [27] proposes a curriculum data augmentation technique to boost the performance. Even though these methods are successful in the event detection, they overlook the limited semantic content when dealing with the short texts from social media.

To address the limitation, we propose a new approach, knowledge-aware few shot learning for event detection by introducing external knowledge to enrich the semantics of short texts. In our approach, one of the key components is the Knowledge-aware Retriever, which borrows the background knowledge in external knowledge bases (KBs) to enrich the semantics of short texts. We resort to the ConceptNet [24] as the external KB, which contains abundant semantic knowledge of concepts. For example, in ConceptNet, the encoded knowledge associated with "Conan" includes "Conan is capable of inference", "Conan desires Sherlock Holmes", "Conan desires clues", "Conan desires expedition" and so on. Such knowledge incorporation greatly enriches the limited content from the short texts. To learn the representation of the enriched texts with knowledge incorporation, the Knowledge Encoder component firstly exploits the pretrained BERT [3] to encode discrete text with the knowledge incorporation as structured sequence. Followed by

*Corresponding author

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.
Conference'17, July 2017, Washington, DC, USA
© 2022 Association for Computing Machinery.
ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/XXXXXXX.XXXXXXX

the Multilayer Perceptron (MLP), the integrated dense representation of short text and knowledge is learned. The last key component is Few Shot Learning, formulating the event detection as the N-way K-shot text classification problem, and it classifies the integrated representations of short texts and knowledge into the corresponding events. In this key component, two representative few shot frameworks, Prototypical network [23] and Model-Agnostic Meta-Learning (MAML) [6] are applied to evaluate the performance. In the experiment, we evaluate the proposed model based on two widely used datasets. Experimental results show that our proposed model consistently improve the event detection results with the external knowledge incorporation.

2 RELATED WORK

The related work is divided into two parts, and one is event detection and the other is few shot learning, which are summarized in the following.

2.1 Event Detection

Event detection is to extract the real-world event occurrence from the texts, and it is one of the extensively studied topics in recent years. Technically, event detection methods fall into two categories when separating them by the techniques. One line is inherited from topic modeling. Among them, the studies [10, 30, 8] extend the statistical framework and propose to detect the events in a batch fashion by clustering short texts into events, while the work [9, 12, 29, 4, 11] solve the online event detection problem by dealing with continuously streaming data. With the great success of Neural Network, the other line takes advantage of the Graph Neural Network [17] to detect events. Authors in [31] propose the hierarchical and incremental mul-sources feature learning algorithm to forecast future events. The work [20] uses inductive Graph Convolutional Network as their event categorization model. The work [22] solves the multilingual event detection problem from the social media. Differs from them, authors in [1] induce a knowledge-preserving graph to address the continuous social media data. Though these methods are highly successful in detecting events in both batch and online manner, they are built on the assumption that the training text examples are abundant to infer their event types. However in the situation of few shot event detection, their performance would be degraded when dealing with few training examples.

2.2 Few Shot Learning

Few-shot learning: it aims to classify new data having seen only a few training examples. Within the few shot framework, a support set $\mathcal{S}$ contains pairs of text x and its corresponding label y: $\mathcal{S} = \{x_i, y_i | i \in \{1, 2, \ldots, N \times K\}\}$, where N is the number of label types and K is the number of examples per label. A query set $\mathcal{Q}$ consists of N texts to be classified: $\mathcal{Q} = \{q_j | j \in \{1, 2, \ldots, N\}\}$. Note that both $\mathcal{S}$ and $\mathcal{Q}$ have the same types of labels. The few shot problem is to identify the labels of texts in the query set based on the knowledge learned from the support set.

In summary, the few shot learning methods can be classified into two categories. One is metric-based methods, which conduct the classification based on the distance metric, e.g., the matching network based on cosine similarity [25], the Prototypical network based on Euclidean distance [23] and Graph Neural Networks [7]. According to this metric, the work [15] exploits the relationship between training task to enforce event prediction consistency, the work [19] infuses the semantic meaning of labels into the framework, and the work [27] proposes a curriculum data augmentation technique to boost the performance. The other line is meta-learning based methods, which *learn to learn* by learning the parameter generation [6], learning rates and parameter updates [16], and parameter updates using gradients [21].

In our work, we take each representative model from the above two categories, Prototypical network [23] and Model-Agnostic Meta-Learning (MAML) [6] to detect the few shot events.

3 METHODOLOGY

To identify the diverse events with few examples from the massive short texts, we propose a knowledge-aware few shot learning model for event detection. Formally, it formulates the event detection as the text classification via N-way K-shot framework. In our approach, the proposed model is composed of three key components. They are Knowledge-aware Retriever, Knowledge Encoder and Few Shot Learning, which are presented in Figure 1.

- **Knowledge-aware Retriever.** Given the short texts, it retrieves the commonsense knowledge from *ConceptNet* for each entity word, and integrates the knowledge with short texts as the input to the Encoder component.
- **Knowledge Encoder.** The encoder consists of a pretrained BERT [3] and Multiplayer Perceptron (MLP) to learn the dense representation of texts and knowledge infusion. The pretrained BERT is firstly exploited to encode the raw text and retrieved knowledge into the structured representation respectively, which are then concatenated as the input to the Multilayer Perceptron (MLP) to learn the dense representations.
- **Few Shot Learning.** According to the N-way K-shot framework, the input to this component are sampled as support set and query set respectively, which constitute an episodic training task. Both Prototypical network [23] and Model-Agnostic Meta-Learning (MAML) [6] are used to classify the query instances.

3.1 Knowledge Retriever

To enrich the semantics of short texts, we resort to the external knowledge base *ConceptNet*. It structures concept knowledge as a graph, where each node corresponds a concept, and each edge with a weight indicates the semantic relation between nodes. For each entity word in a short text, the Knowledge Retriever retrieves its relations as knowledge to enhance the semantics. In *ConceptNet*, each node is associated with multiple relations. Indicated by Figure 1, take the entity word of "Conan" from the text "Detective Conan is a Japanese manga and anime series" as the example, the Retriever returns different relations attached to "Conan", including *CapableOf*, *Causes*, *Desires*, *IsA*, *HasProperty* and so on. In our case, for each entity word, we only consider its top 10 relations with highest weights, which indicate the most representative knowledge of an entity word. For the entity word "Conan", its most representative 10 relations returned are “Conan is capable of inference”, “Conan

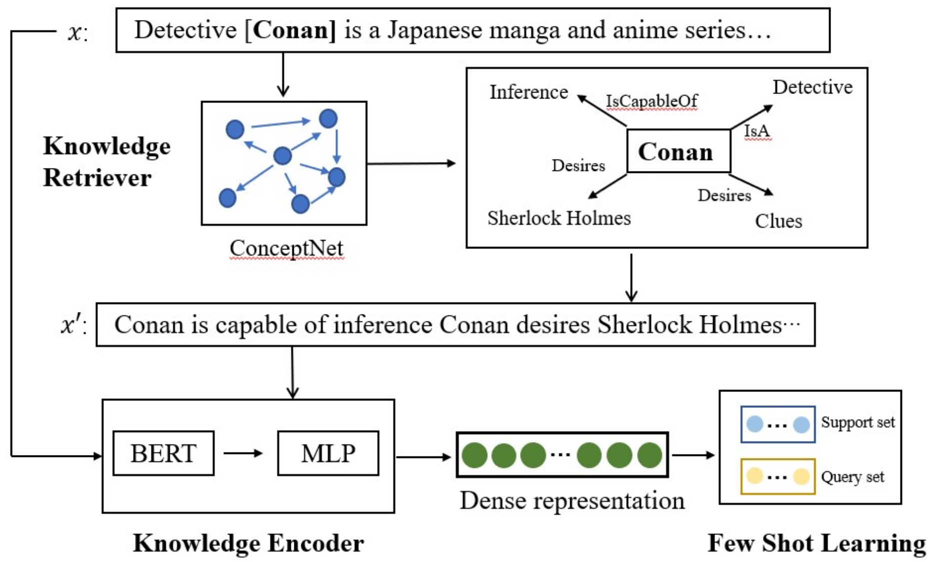

Figure 1: The structure of the purposed Knowledge-aware Few Shot Learning for Event Detection.

desires Sherlock Holmes", "Conan desires clues", and such external knowledge incorporation greatly enriches the semantics of the short text.

The original short text x concatenating the retrieved knowledge x' serves as the input to the following encoder.

3.2 Knowledge Encoder

The encoder $f(,\theta)$ encodes the raw text and knowledge incorporation into dense representation. Firstly, a pretrained BERT [3] is firstly exploited to transfer the discrete input into fixed dimensional representation, which is expressed as:

$$\begin{aligned} \mathbf{X} &= E(x) \in \mathcal{R}^d \\ \mathbf{K} &= E(x') \in \mathcal{R}^d \end{aligned} \tag{1}$$

where $E() \in \mathcal{R}^d$ denotes the pretrained BERT, and other text encoder (e.g. recurrent neural network [28] and convolutional neural network [13]) is also accepted, x denotes the raw text and x' is the retrieved knowledge, and $\mathcal{X}$ and $\mathcal{K}$ denote the corresponding representations from the pretrained BERT.

Multilayer Perceptron (MLP) is followed to learn the integrated dense representation.

$$\mathbf{H} = RELU([\mathbf{X};\mathbf{K}]\mathbf{W}_1 + \mathbf{b}_1)\mathbf{W}_2 + \mathbf{b}_2 \in \mathcal{R}^f \tag{2}$$

where $[;]$ is the concatenate operation, $\mathbf{W}_1$, $\mathbf{W}_2$, $\mathbf{b}_1$ and $\mathbf{b}_2$ are model parameters, and $\mathbf{H} \in \mathcal{R}^f$ is the learned dense representation of integrated short text and external knowledge.

3.3 Few Shot Learning

In this component, the event detection task is formulated as N-way K-shot text classification problem, which is a conventional episodic training strategy for few shot learning. Here we let D denote the whole dataset, D^{train} denote the training set sampled from D and the remaining is testing set D^{test}. At each training iteration, a subset $C \subset D^{train}$ with N classes is randomly sampled. For each class $c \in C$, $K+N$ samples are drawn with K samples as support set D_s and N samples as query set D_q, which constitute a train task T. At each training episode, the parameters of the model are updated on the query set, based on knowledge gained from its support set.

To validate the generality of knowledge-aware incorporation to the few shot learning , we leverage two representative models, Prototypical network and Model-Agnostic Meta-Learning (MAML) from metric-based and meta-learning based methods to classify short texts into different events.

Prototypical network. At each training episode, the classifier performs the classification task base on the distance between the representation of each query sample and the *prototype* of each class. A *prototype* c is a representation vector of the class it belongs to. It's defined as the average of all vectors derived of c-th this class [23] $\mathbf{p}_c = \frac{1}{K}\sum_{x \in D_s} f(x,\theta)$, where $f(x,\theta)$ denotes the encoded representation from the aforementioned knowledge encoder.

Hence, each query sample $x \in D_q$ is classified by the metric based distribution over all possible classes, which is determined by calculating the distance between the query sample and the *prototype* of each class,

$$P(y=k|x) = \frac{e^{-d(\mathbf{p}_k, f(x,\theta))}}{\sum_{x \in D_q} e^{-d(\mathbf{p}_i, f(x_q,\theta))}}, \tag{3}$$

where d denotes a distance function(e.g. Cosine Similarity [25], Euclidean Distance [23], Manhattan Distance).

Given this distribution, the negative log-probability loss function is defined as,

$$Loss = -\frac{1}{N}\sum_{i=1}^{N} logP(y=k|x_i). \tag{4}$$

Model-Agnostic Meta-Learning (MAML). When adapting to a new task T_i, the parameters θ of the model $f(,\theta)$ (i.e. the knowledge

Table 1: The accuracy results of 5-way 1-shot, 5-way 5-shot and 5-way 10-shot event detection on Twitter dataset.

Models	5-way 1-shot	5-way 5-shot	5-way 10-shot
Proto	51.2	63.1	74.7
KA-Proto	**52.3**	**64.0**	**77.5**
MAML	51.3	62.7	72.9
KA-MAML	**52.4**	**63.8**	**74.6**

Table 2: The accuracy results of 5-way 1-shot, 5-way 5-shot and 5-way 10-shot event detection on MAVEN dataset.

Models	5-way 1-shot	5-way 5-shot	5-way 10-shot
Proto	38.3	54.9	66.7
KA-Proto	**39.7**	**55.6**	**67.2**
MAML	37.5	52.1	63.4
KA-MAML	**38.6**	**53.0**	**64.1**

encoder) is computed using one gradient descent updates on tasks T_i, which is expressed as

$$\theta_i' = \theta - \alpha \Delta_\theta \mathcal{L}_{T_i}(f(,\theta)), \tag{5}$$

where α is the step size as a hyperparameter.

In effect, MAML aims to optimize the model parameters such that one or a small number of gradient steps on a new task would produce maximally effective behavior on that task [6]. During the meta-learning stage, the parameters θ are updated across the task via stochastic gradient descent (SGD), which is updated as:

$$\theta = \theta - \beta \Delta_\theta \sum_{T_i \sim P(T)} \mathcal{L}_{T_i}(f_{\theta_i'}), \tag{6}$$

where β is the meta step size.

4 EXPERIMENTS

In this section, we first present the experimental setup, then compare the proposed model with competitors over the short texts, and the evaluation results are followed.

4.1 Experimental Setup

We use two common event-detection datasets, Twitter data [18] and MAVEN dataset [26], to evaluate the performance of the proposed model. The Twitter dataset contains 68,841 labeled tweets associated with 503 event classes, while MAVEN dataset contains 10,242 messages related to 154 event classes.

The proposed model under the Prototypical network framework is referred to as *KA-Proto*, and the proposed under the Model-Agnostic Meta-Learning framework is referred to as *KA-MAML*. To evaluate them, two representative few shot learn framework, metric-based and meta-learning based models, Prototypical network [23] and Model-Agnostic Meta-Learning, are referred to as *Proto* and *MAML* respectively, and both of them are employed as strong baselines, which are also the state-of-the-art few shot classification models.

We train all models with 5-way 1-shot, 5-way 5-shot and 5-way 10-shot setting respectively. As an optimizer, Adam [14] is used to optimize the parameters with learning rate $1e^{-4}$. The training/evaluation/test are set to 6000, 1000 and 500 iterations respectively. The dimensions of knowledge encoder are set as [768, 200, 100] with dropout rate of 0.3 to prevent the overfitting problem.

4.2 Experimental Results

Table 1 and Table 2 report the accuracy results of the proposed models against the baselines in terms of event detection task on Twitter and MAVEN dataset respectively. According to the comparison results, we have the following remarks.

It's noted that with more shots involved in each class, the performance of all models on the event detection increases. In Twitter dataset, the accuracy result of Proto (Prototypical network) increases from 51.2% to 74.7% with 5-way 1-shot to 5-way 10-shot, which is true for the KA-Proto, MAML and KA-MAML. Most importantly, compared with the original Proto and MAML, the proposed knowledge-aware models KA-Proto and KA-MAML consistently acquire the higher accuracy results with more shots. On Twitter dataset, the proposed KA-Proto obtains distinctly higher accuracy results than the original Proto on 5-way 1-shot, 5-way 5-shot and 5-way 10-shot event detection task. Similarly, the proposed KA-MAML with knowledge incorporation also outperforms the original MAML. Once again, the performance on MAVAN dataset validates that the knowledge incorporation from CONCEPTNET into the few shot learning help boost the event detection task.

5 CONCLUSION AND FUTURE WORK

Event detection is indispensable for the government to learn the public concerns, be aware of the real occurrences in a city and thus make informed and wise policies. To better identify the events with few samples from the short texts, in this work we formulate it as a N-way 1 shot few shot learning problem and propose a knowledge-aware few shot learning for event detection. To mitigate the limited descriptions of short texts, we resort to the external knowledge base CONCEPTNET to incorporate the valuable semantics into the short texts. Two representative few shot learning models are leveraged to formulate the event detection as the text classification scenario. The external knowledge incorporation together with few shot learning framework accomplish the event detection task. The experimental results on two commonly used event-detection task shows that our proposed knowledge-aware model outperforms the strong baselines with an evident increase in the task of event detection.

Besides two representatives of few shot learning models, it's promising to explore other techniques to further boost the performance in future, e.g., the curriculum data augmentation technique in few shot [27] and exploiting cross-task relation [15].

ACKNOWLEDGMENTS

This work was supported by the Ministry of Industry and Information Technology Project, Beijing National Artificial Intelligence Innovation and Application Pilot Zone (2021-0166-1-1).

REFERENCES

[1] Yuwei Cao, Hao Peng, Jia Wu, Yingtong Dou, Jianxin Li, and Philip S Yu. 2021. Knowledge-preserving incremental social event detection via heterogeneous gnns. In *Proceedings of the Web Conference 2021*, 3383–3395.

[2] Shumin Deng, Ningyu Zhang, Jiaojian Kang, Yichi Zhang, Wei Zhang, and Huajun Chen. 2020. Meta-learning with dynamic-memory-based prototypical network for few-shot event detection. In *Proceedings of the 13th International Conference on Web Search and Data Mining*, 151–159.

[3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

[4] Mateusz Fedoryszak, Brent Frederick, Vijay Rajaram, and Changtao Zhong. 2019. Real-time event detection on social data streams. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2774–2782.

[5] Nanyi Fei, Zhiwu Lu, Tao Xiang, and Songfang Huang. 2020. Melr: meta-learning via modeling episode-level relationships for few-shot learning. In *International Conference on Learning Representations*.

[6] Chelsea Finn, Pieter Abbeel, and Sergey Levine. 2017. Model-agnostic meta-learning for fast adaptation of deep networks. In *International conference on machine learning*. PMLR, 1126–1135.

[7] Victor Garcia and Joan Bruna. 2017. Few-shot learning with graph neural networks. *arXiv preprint arXiv:1711.04043*.

[8] Jinjin Guo, Longbing Cao, and Zhiguo Gong. 2021. Recurrent coupled topic modeling over sequential documents. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 16, 1, 1–32.

[9] Jinjin Guo and Zhiguo Gong. 2017. A density-based nonparametric model for online event discovery from the social media data. In *IJCAI*, 1732–1738.

[10] Jinjin Guo and Zhiguo Gong. 2016. A nonparametric model for event discovery in the geospatial-temporal space. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, 499–508.

[11] Jinjin Guo, Zhiguo Gong, and Longbing Cao. 2021. Dhcm: dynamic and hierarchical event categorization and discovery for social media stream. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 12, 5, 1–25.

[12] Linmei Hu, Bin Zhang, Lei Hou, and Juanzi Li. 2017. Adaptive online event detection in news streams. *Knowledge-Based Systems*, 138, 105–112.

[13] Yoon Kim. 2014. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, Doha, Qatar, (Oct. 2014), 1746–1751. DOI: 10.3115/v1/D14-1181.

[14] Diederik P. Kingma and Jimmy Ba. 2015. Adam: a method for stochastic optimization. In 1–15.

[15] Viet Lai, Franck Dernoncourt, and Thien Huu Nguyen. 2021. Learning prototype representations across few-shot tasks for event detection. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 5270–5277.

[16] Zhenguo Li, Fengwei Zhou, Fei Chen, and Hang Li. 2017. Meta-sgd: learning to learn quickly for few-shot learning. *arXiv preprint arXiv:1707.09835*.

[17] Xiaoxiao Ma, Jia Wu, Shan Xue, Jian Yang, Chuan Zhou, Quan Z Sheng, Hui Xiong, and Leman Akoglu. 2021. A comprehensive survey on graph anomaly detection with deep learning. *IEEE Transactions on Knowledge and Data Engineering*.

[18] Andrew J. McMinn, Yashar Moshfeghi, and Joemon M. Jose. 2013. Building a large-scale corpus for evaluating event detection on twitter. In (CIKM '13). Association for Computing Machinery, San Francisco, California, USA, 409–418.

[19] Sora Ohashi, Junya Takayama, Tomoyuki Kajiwara, and Yuki Arase. 2021. Distinct label representations for few-shot text classification. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, 831–836.

[20] Hao Peng, Jianxin Li, Qiran Gong, Yangqiu Song, Yuanxing Ning, Kunfeng Lai, and Philip S Yu. 2019. Fine-grained event categorization with heterogeneous graph convolutional networks. *arXiv preprint arXiv:1906.04580*.

[21] Sachin Ravi and Hugo Larochelle. 2016. Optimization as a model for few-shot learning.

[22] Jiaqian Ren, Hao Peng, Lei Jiang, Jia Wu, Yongxin Tong, Lihong Wang, Xu Bai, Bo Wang, and Qiang Yang. 2021. Transferring knowledge distillation for multilingual social event detection. *arXiv preprint arXiv:2108.03084*.

[23] Jake Snell, Kevin Swersky, and Richard Zemel. 2017. Prototypical networks for few-shot learning. *Advances in neural information processing systems*, 30.

[24] Robyn 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*.

[25] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al. 2016. Matching networks for one shot learning. *Advances in neural information processing systems*, 29.

[26] Xiaozhi Wang et al. 2020. MAVEN: A Massive General Domain Event Detection Dataset. In (Apr. 2020).

[27] Jason Wei, Chengyu Huang, Soroush Vosoughi, Yu Cheng, and Shiqi Xu. 2021. Few-shot text classification with triplet networks, data augmentation, and curriculum learning.

[28] Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. 2016. Hierarchical attention networks for document classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, San Diego, California, (June 2016), 1480–1489. DOI: 10.18653/v1/N16-1174.

[29] Chao Zhang, Dongming Lei, Quan Yuan, Honglei Zhuang, Lance Kaplan, Shaowen Wang, and Jiawei Han. 2018. Geoburst+ effective and real-time local event detection in geo-tagged tweet streams. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 9, 3, 1–24.

[30] Chao Zhang, Liyuan Liu, Dongming Lei, Quan Yuan, Honglei Zhuang, Tim Hanratty, and Jiawei Han. 2017. Triovecevent: embedding-based online local event detection in geo-tagged tweet streams. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 595–604.

[31] Liang Zhao, Yuyang Gao, Jieping Ye, Feng Chen, Yanfang Ye, Chang-Tien Lu, and Naren Ramakrishnan. 2021. Spatio-temporal event forecasting using incremental multi-source feature learning. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 16, 2, 1–28.