Neurocomputing 578 (2024) 127438

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Towards low-resource rumor detection: Unified contrastive transfer with propagation structure

Hongzhan Lin [a], Jing Ma [a,*,1], Ruichao Yang [a], Zhiwei Yang [b,1], Mingfei Cheng [c]

[a] Hong Kong Baptist University, Hong Kong Special Administrative Region of China
[b] Guangdong Institute of Smart Education, Jinan University, China
[c] Singapore Management University, Singapore

ARTICLE INFO

Communicated by M. Gallo

Keywords:
Low resource
Rumor detection
Few-shot transfer
Contrastive learning
Propagation structure

ABSTRACT

The truth is significantly hampered by massive rumors that spread along with breaking news or popular topics. Since there is sufficient corpus gathered from the same domain for model training, existing rumor detection algorithms show promising performance on yesterday's news. However, due to a lack of substantial training data and prior expert knowledge, they are poor at spotting rumors concerning unforeseen events, especially those propagated in different languages (i.e., low-resource regimes). In this paper, we propose a simple yet effective framework with unified contrastive transfer learning, to detect rumors by adapting the features learned from well-resourced rumor data to that of the low-resourced with only few-shot annotations. More specifically, we first represent rumor circulated on social media as an undirected topology for enhancing the interaction of user opinions, and then train the propagation-structured model via a unified contrastive paradigm to mine effective clues simultaneously from both post semantics and propagation structure. Our model explicitly breaks the barriers of the domain and/or language issues, via language alignment and a novel domain-adaptive contrastive learning mechanism. To well-generalize the representation learning using a small set of annotated target events, we reveal that rumor-indicative signal is closely correlated with the uniformity of the distribution of these events. We design a target-wise contrastive training mechanism with three event-level data augmentation strategies, capable of unifying the representations by distinguishing target events. Extensive experiments conducted on four low-resource datasets collected from real-world microblog platforms demonstrate that our framework achieves much better performance than state-of-the-art methods and exhibits a superior capacity for detecting rumors at early stages.

1. Introduction

Rumor amid breaking news spreads like wildfire on social media, causing widespread confusion, fear, and distrust among individuals and society. However, rumor detection can be particularly challenging in low-resourced domains or languages due to the availability of domain expertise, nuances of language, cultural differences, etc. Taking the healthcare domain as an example, during the outbreak of COVID-19, a false rumor claimed "the vaccine has a chip in it which will control your mind".[2] The rumor was translated into various languages, enabling it to spread in different regions worldwide, including those with low levels of vaccine uptake or hesitant attitudes towards vaccination like Arabic, India, and other Muslim countries. Despite recent efforts to collect microblog posts related to COVID-19 [1–3], existing rumor detection methods are vulnerable to detecting such low-resource rumors without a substantial and suitable training corpus [4]. Therefore, to mitigate their harmful effects, it is crucial to develop robust methods to detect rumors in low-resource languages and domains during emerging news events.

A rumor is defined in social psychology literature as a narrative or a statement whose truth value is unconfirmed or intentionally untrue [5]. Recently, deep neural network (DNN) techniques [6–9] have shown considerable promise in recognizing rumors on microblogging services by extracting features indicative of rumors from a large corpus of labeled rumor data. Nevertheless, such DNN-based techniques are completely data-driven and have a significant drawback in detecting emergent events in low-resource domains, namely, the distinct subject coverage and word distribution [10] required for detecting

* Corresponding author.
E-mail addresses: majing@comp.hkbu.edu.hk (J. Ma), yangzw18@mails.jlu.edu.cn (Z. Yang).

1 Jing Ma and Zhiwei Yang contributed equally.
2 https://www.bbc.com/news/55768656

https://doi.org/10.1016/j.neucom.2024.127438
Received 16 October 2023; Received in revised form 27 December 2023; Accepted 17 February 2024
Available online 21 February 2024

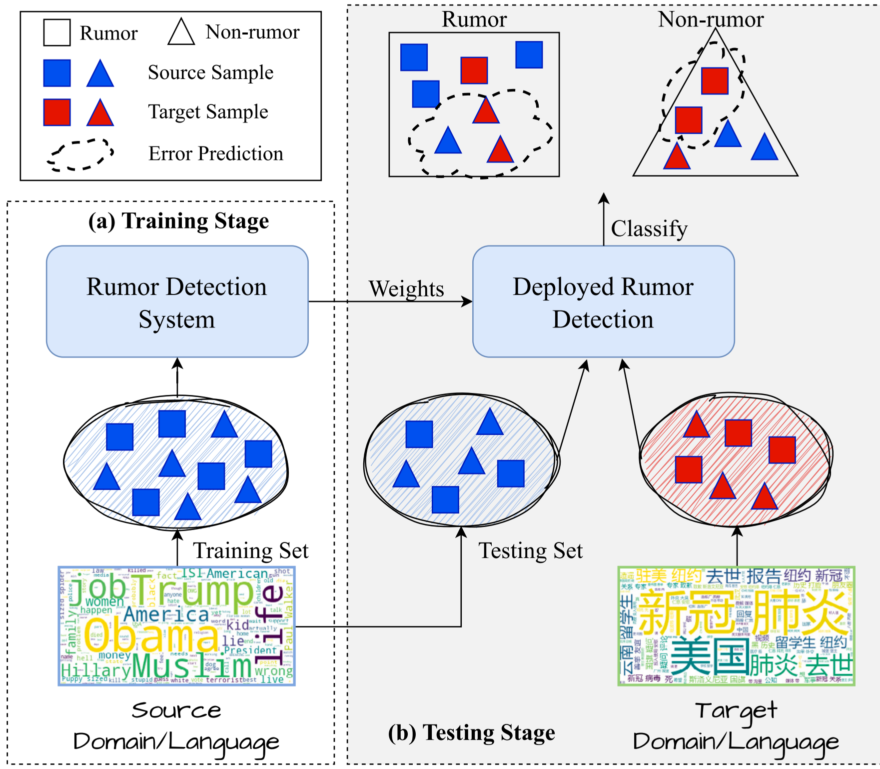
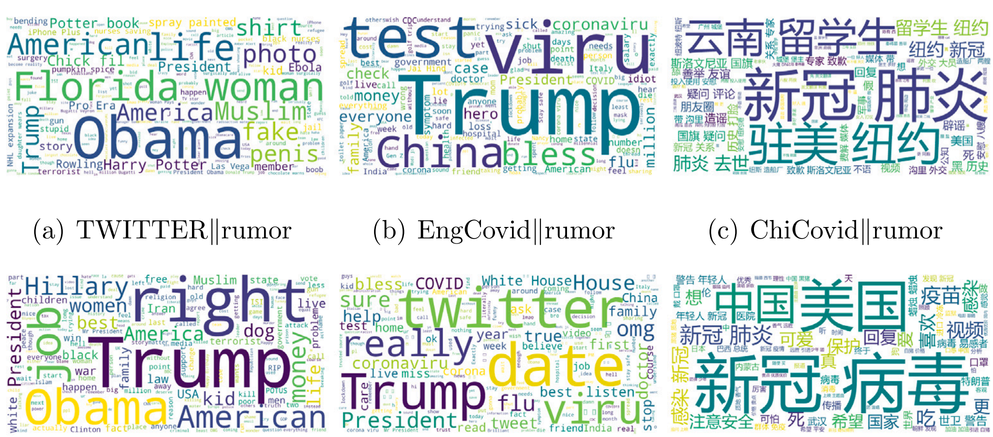


Fig. 1. A toy example to illustrate the gap between source data (blue) and target data (red). The rumor detection model trained on the source data may not adapt well to the target. This discrepancy becomes evident through increased error rates when directly inferring samples from the target data using the model pre-trained on the source data.



(a) TWITTER||rumor (b) EngCovid||rumor (c) ChiCovid||rumor

(d) TWITTER||non-rumor (e) EngCovid||non-rumor (f) ChiCovid||non-rumor

Fig. 2. Word clouds of rumor and non-rumor data generated from TWITTER, English-COVID19 (EngCovid), and Chinese-COVID19 (ChiCovid) datasets, where the size of terms corresponds to the word frequency. Both TWITTER and English-COVID19 are presented in English while Chinese-COVID19 in Chinese.

low-resource rumors are frequently not covered by the public benchmarks [11–13]. Existing monolingual techniques, on the other hand, are inapplicable for rumors disseminated in various languages due to the lack of sufficient open-domain data in the target language, which is necessary for effective monolingual model training. As exemplified in Fig. 1, the rumor detection model is initially trained on a source domain comprising English-language training samples. However, when confronted with newly emergent data that differs substantially in domain and/or language from the source, the model trained on this source data exhibits diminished performance within the target domain. Directly training the detection model on the source data and evaluating its performance in the newly emergent domain often yields unsatisfactory results.

In this study, we assume that establishing close correlations between well-resourced and low-resourced rumor data can help overcome domain and language barriers, thereby improving low-resource rumor detection within a more comprehensive framework. To illustrate our intuition, we collect rumorous and non-rumorous claims corresponding to COVID-19 with propagation threads from Twitter and Sina Weibo which are two popular social websites in English-spoken and Chinese-spoken communities, respectively. Fig. 2 illustrates the word clouds of rumor and non-rumor data from an open domain benchmark (i.e., TWITTER [13]) and two COVID-19 datasets [14] (i.e., English-COVID19 and Chinese-COVID19). It can be seen that both TWITTER and English-COVID19 contain denial opinions towards

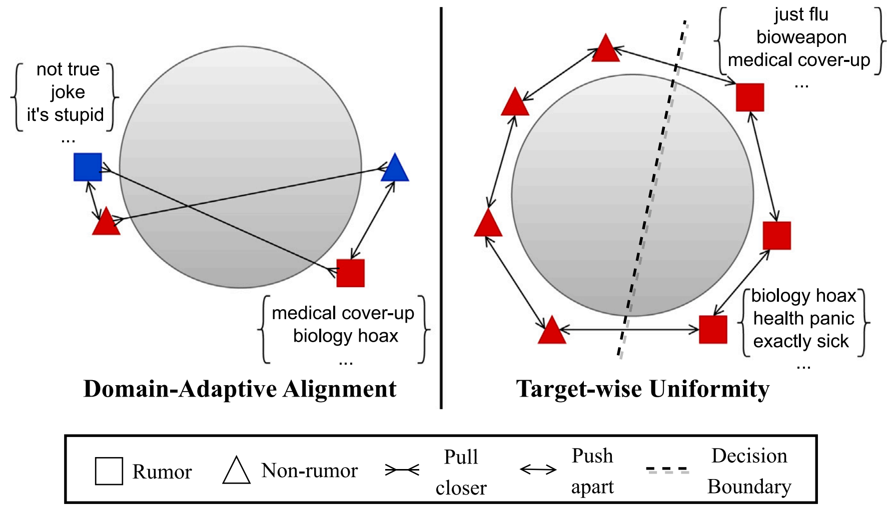


Fig. 3. The illustration of Alignment and Uniformity of features in source data (blue) and target data (red) on a hypersphere. The alignment dedicates identical rumor-indicative features from different domains closer, while the uniformity could help preserve maximal information of the features from target rumor data to capture nontrivial but more discriminative patterns for better generalization.

rumors, e.g., "fake", "joke", "stupid" in Fig. 2(a) and "wrong symptom", "exactly sick", "health panic" in Fig. 2(b). In contrast, supportive opinions towards non-rumors can be drawn from Figs. 2(d)–2(e). Moreover, considering that COVID-19 is a global disease, massive misinformation could be widely propagated in different languages such as Arabic [15], Indic [16], English [17] and Chinese [18]. Similar identical patterns can be observed in Chinese on Sina Weibo from Figs. 2(c) and 2(f). Despite the prevalence of domain-specific expertise jargon and language-related colloquialisms in COVID-19 data on social media, we contend that aligning the representation space of identical rumor-indicative patterns across various domains and/or languages can facilitate the adaptation of features extracted from well-resourced data to those of low-resourced data. Moreover, since rumor propagation generally reveals significant insight into how a claim is responded to by users irrespective of specific domains [6,19], we aim to develop an innovative domain and/or language transfer framework that is aware of such a structural social context.

To this end, inspired by contrastive learning [20–22], we proposed a domain-Adaptive Contrastive Learning approach for low-resource rumor detection, to encourage effective alignment of rumor-indicative features in the well-resourced and low-resource data. More specifically, we first transform each microblog post into a language-independent vector by semantically aligning the source and target language in a shared vector space. As the diffusion of rumors generally follows a propagation structure that provides valuable domain-invariant clues on how a claim is transmitted, we present the conversation propagation thread as an undirected topology, which allows full-duplex interactions between posts with responsive relationships so that the domain-invariant structural features can be fully aggregated and the interplay of user viewpoints can be enhanced. Thus we resort to a multi-scale Graph Convolutional mechanism to catch informative patterns fused from both claim semantics and event structure. Then, we propose a novel domain-adaptive contrastive learning paradigm to minimize the intra-class variance of source and target instances with same veracity, and maximize inter-class variance of instances with different veracity.

Previous literature reveals that two properties: (1) alignment and (2) uniformity on the unit hypersphere [23], are of great importance for representation learning in terms of contrastive paradigm. However, a problem of our domain-adaptive contrastive learning framework proposed in [14] is that it primarily emphasizes the alignment of different domains and/or languages, while largely disregarding the uniformity of target feature space that preserves distinctive rumor-indicative signals among target training samples [24]. The inductive bias of such an alignment-only paradigm could cause the representation degeneration issue [25] and limit the capacity of generalization to unlabeled low-resource rumor data. Fig. 3 illustrates the alignment between the source rumor data in open domains and a small amount of labeled target data in COVID-19 domain, and the ideal uniformity of target representation learning. The domain-adaptive contrastive learning can align the target rumor data containing target-specific patterns like "medical cover-up" or "biology hoax", with the well-resourced data containing general rumor-indicative patterns like "not true" or "it's stupid". So that almost all the few-shot target rumor data with the same veracity are shrinking towards the source data and thus degenerating into a narrow and anisotropic feature space, which however suppresses the uniformity of target representation learning. Generally, the more uniform the distribution of the small labeled target data, the more rumor-indicative information is retained, which can lead to better generalization to more target low-resource data [24]. As exemplified in Fig. 3, with an evenly distributed representation on the unit hypersphere [26], the target vector space is able to preserve enough discriminative information related to medicine or biology for each sample, which can be well-generalized on detecting COVID-19 rumors with nontrivial domain-specific patterns like "just flu" or "it's bioweapon".

In this work, we present a pioneering unified contrastive transfer framework that integrates a novel approach known as target-wise contrastive learning to augment our foundational domain-adaptive model, specifically tailored for low-resource rumor detection. The core objective of target-wise contrastive learning is to significantly enhance the generalization capability, particularly targeting the more intricate and challenging events within the target domain. Specifically, the incorporation of target-wise contrastive learning within our framework serves the purpose of unifying and refining the representations derived from limited samples in the target domain. By leveraging this technique, we aim to homogenize the few-shot target representations, thus enabling the model to glean more nuanced and discriminative features from the scarce labeled data available in the target domain. This enhancement is pivotal in bolstering the model's ability to discern and effectively detect rumors in scenarios where labeled data is notably scarce, thereby improving its performance on challenging target events.

Meanwhile, in target-wise contrastive learning, data augmentation is a crucial aspect as it enables the creation of new positive and negative pairs from a small amount of labeled target data. This increases the amount and diversity of target training data, improving the performance of detecting more challenging low-resource rumors. Therefore, we explore three event-level data augmentation strategies for target low-resource data (i.e., Adversarial Attack [27,28], Feature Dropout [29], and Graph Dropedge [30]) to effectively obtain pseudo diverse views of a target event in the latent space for the target-wise contrastive learning.

As there is no public benchmark available for detecting low-resource rumors with conversation threads, we collected the propagation structure for four rumor datasets corresponding to COVID-19 from social media in Mandarin, English, Cantonese and Arabic languages. We also instantiate our proposed unified contrastive transfer framework on three strong structure-based baseline methods that model the propagation thread of rumors.

Extensive experiments conducted on four real-world low-resource datasets confirm that (1) our model yields outstanding performances for detecting low-resource rumors over the state-of-the-art baselines with a large margin; (2) the unified contrastive transfer framework is more effective in contrastive representation learning for uniformity of target data distribution; and (3) our method performs particularly well on early rumor detection which is crucial for timely intervention and debunking especially for breaking events.

The main contributions of this paper are four-fold:

- To the best of our knowledge, we are the first to study rumor detection on social media from a fresh perspective of the low-resource regime, by presenting a novel simple yet effective framework via unified contrastive transfer integrated with propagation structure.
- We propose a contrastive learning framework for structural feature adaption between different domains and languages, which model domain-invariant similarities based on undirected propagation topology by pulling together events of the same veracity while pushing apart events of different veracity.
- Based on the domain-adaptive model previously proposed in our recent work [14], we further design a distinctive target-wise contrastive learning mechanism accompanied by three innovative event-level data augmentation strategies. These strategies converge to homogenize distributed event-level representations specifically tailored for target events, which enables our rumor detection model to discern and harness target-specific rumor signals.
- We collect four low-resource rumor benchmarks corresponding to COVID-19 domain with conversation threads, respectively represented in Chinese, English, Cantonese and Arabic languages. Experimental results on the four real-world benchmarks show that our model achieves superior performance for both rumor classification and early detection tasks under low-resource settings.

2. Related work

2.1. Sequence-based rumor detection

Initial research efforts in the field of automated rumor detection have primarily concentrated on developing supervised classifiers that leverage specifically designed features extracted from the content of posts, user profiles, and dissemination trends [31–33]. Subsequent research has introduced novel characteristics, including those that represent the dissemination and cascading effects of rumors [34–36]. [37] streamlined the engineering effort by employing a series of regular expressions to efficiently identify questioning and negating tweets. DNN-based models, including recurrent neural networks (RNNs) [12], convolutional neural networks (CNNs) [38], and attention mechanisms [39], are widely utilized to extract features from the continuous flow of social media content. Despite their efficacy, these techniques primarily treat the post structure as a linear sequence, oversimplifying the intricate propagation patterns inherent in social media interactions.

2.2. Structure-based rumor detection

To jointly derive valuable insights from content semantics and propagation structures, certain methodologies suggested employing kernel-learning models [13,40] for comparing different propagation trees. By utilizing tree-structured recursive neural networks (RvNN) [6] and transformer-based models [41,42], these approaches aimed to generate representations for individual posts within a propagation tree, guided by the tree's inherent structure. More recently, graph neural networks (GNN) [7,43] have been increasingly utilized for encoding propagation threads, leading to more sophisticated representations. However, these data-driven methodologies struggle to identify rumors in low-resource environments [44,45], as they typically necessitate substantial training data, which is often unavailable in less-researched domains or languages. To address this limitation, we present a novel plug-and-play framework designed to adapt pre-existing models by incorporating an efficient propagation structure, enabling the detection of rumors across diverse domains and/or languages.

2.3. Low-resource fact checking

To enable efficient fact-checking tasks in low-resource environments, domain adaptation techniques have been employed for fake news detection [10,46–48]. These methods leverage multi-modal data, such as text and images, to extract relevant features. In a related effort, [49] introduced a straightforward approach that utilizes perplexity scores obtained from pre-trained language models (PLMs) for few-shot fact-checking tasks. Our work, however, diverges from these multi-modal data adaptations and PLM transfer learning strategies. Instead, we concentrate on language and domain adaptation of structural features to identify rumors within low-resource microblog posts related to emerging events.

2.4. Applications of contrastive learning

Contrastive learning (CL) is designed to bolster representation learning by accentuating the concurrence between instances of the same category while simultaneously differentiating them from instances belonging to dissimilar categories [23]. In recent years, CL has achieved great success in unsupervised visual representation learning [20–22]. Besides computer vision, recent studies suggest that CL is promising in the semantic textual similarity [50,51], stance detection [52], short text clustering [53], unknown intent detection [54], and abstractive summarization [55], etc. Nevertheless, the aforementioned CL frameworks have been explicitly designed to enhance unstructured textual data, such as sentences and documents. Unfortunately, these frameworks do not adequately address the challenges posed by low-resource rumor detection tasks, which necessitate considering not only claims but also the intricate propagation patterns of community responses.

2.5. Contrastive learning for low-resource rumor detection

This work is a significant extension of the first supervised contrastive approach for low-resource rumor detection on social media using a structure-based framework in our recent work [14]. Since the publication of the conference version, several similar studies [56–58] with domain-adaptive contrastive learning on the fact-checking discipline have been conducted that build upon our findings. However, these studies have just focused on aligning different domains, while not adequately considering the distinctive informative features that may exist between different target samples and exploring effective augmentation strategies for structured data in the target-wise contrastive paradigm.

Specifically, although the domain-adaptive contrastive learning approach effectively aligns diverse domains and languages, it may neglect the crucial aspect of ensuring uniformity within the target feature

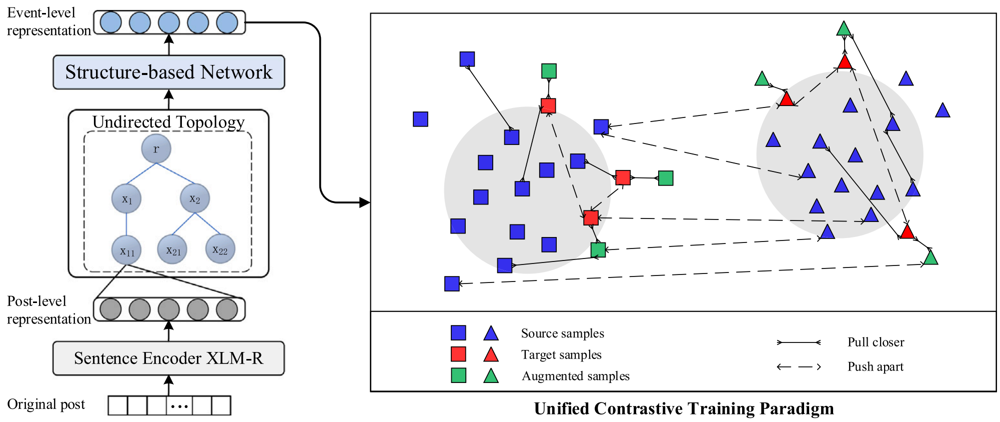


Fig. 4. The overall architecture of our proposed framework. For source and small target training data, we first obtain post-level representations after cross-lingual sentence encoding, then train the structure-based network (i.e., Multi-scale GCNs) with the proposed unified contrastive objective. For target test data, we extract the event-level representations to detect rumors.

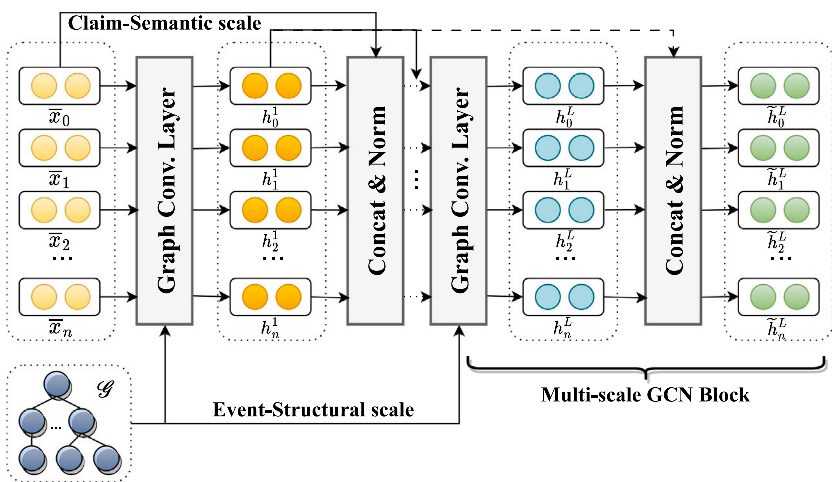


Fig. 5. The Multi-scale GCNs for propagation structure representations.

space. This oversight can lead to representation degradation, thereby constraining the model's ability to generalize effectively to unlabeled low-resource rumor data. In response, this work introduces a pioneering unified contrastive transfer framework that integrates target-wise contrastive learning. This augmentation of our foundational domain-adaptive model aims to bolster low-resource rumor detection by homogenizing the feature space of scarce target training samples, thereby enhancing their representation and generalization capabilities.

3. Problem statement

In the present study, we introduce the concept of low-resource rumor detection as follows: Given a well-resourced dataset as a source, the objective is to accurately categorize each event within a target low-resource dataset as either a rumor or a non-rumor. It is essential to note that the source and target datasets are derived from distinct domains and/or languages, thereby more realistic and challenging to the task. More specifically, a well-resourced source dataset for training is defined as a set of events $\mathcal{D}_s = \{C_1^s, C_2^s, \dots, C_M^s\}$, where M is the number of source events. Each event $C^s = (y, c, \mathcal{T}(c))$ is a triplet representing a given claim c which is associated with a veracity label $y \in \{\text{rumor}, \text{non-rumor}\}$, and ideally all its relevant responsive microblog post in chronological order, i.e., $\mathcal{T}(c) = \{c, x_1^s, x_2^s, \dots, x_{|C|}^s\}$,[3] where $|C|$ is the number of responsive tweets in the conversation structure. For the target dataset with low-resource domains and/or languages, a much smaller dataset is considered for training $\mathcal{D}_t = \{C_1^t, C_2^t, \dots, C_N^t\}$, where $N(N \ll M)$ is the number of target events and each $C^t = (y, c', \mathcal{T}(c'))$ has a similar composition structure to the source dataset.

We formulate the task of low-resource rumor detection as a supervised classification problem that trains a domain/language-agnostic classifier $f(\cdot)$ adapting the features learned from source datasets to that of the target events, that is, $f(C^t|\mathcal{D}_s) \rightarrow y$. Note that although the tweets are notated sequentially, there are connections among them based on their responsive relationships [6].

4. Our approach

In this section, we present our unified contrastive transfer framework with propagation structure to adapt the features extracted from the well-resourced data to detect rumors from low-resource events, which considers cross-lingual and cross-domain transfer. Figs. 4–5 show an overview of our backbone model and training paradigm, which will be depicted in the following subsections.

4.1. Cross-lingual sentence encoder

Given a post in an event that could be either from source or target data, to map it into a shared semantic space where the source and target languages are semantically aligned, we utilize XLM-RoBERTa [59]

[3] c is equivalent to x_0^s.

(XLM-R) to model the context interactions among tokens in the sequence for the sentence-level representation:

$$\bar{x} = XLM - R(x), \tag{1}$$

where x is the original post, and we obtain the post-level representation $\bar{x}$ using the output state of the $\langle s \rangle$ token in XLM-R. We thus denote the representation of posts in the source event C^s and the target event C^t as a matrix X^s and X^t respectively: $X^* = [\bar{x}_0^*, \bar{x}_1^*, \bar{x}_2^*, \dots, \bar{x}_{|X^*|-1}^*]^\top; * \in \{s, t\}$, where $X^s \in \mathbb{R}^{m \times d}$ and $X^t \in \mathbb{R}^{n \times d}$, d is the dimension of the output state of the sentence encoder.

4.2. Propagation structure representation

On top of the sentence encoder, different from the directed tree structure modeling in previous work [6,7], we first represent the propagation of each claim as an undirected propagation topology to explore the full-duplex interaction patterns between responsive nodes with the expressive capacity of graph neural networks [60]. To fully utilize the claim's abundant information while preventing off-topic coherence that strays from the claim's main point in the propagation structure, as illustrated in Fig. 5, we exploit a simple but effective Multi-scale Graph Convolutional Network to integrate both the claim semantics and the social context information for the subsequent contrastive training paradigm.

Given an event and its initialized embedding matrix $C^*, X^*; * \in \{s, t\}$, We model the propagation thread of the event as an undirected graph topology $\mathcal{G} = \langle V, E \rangle$, where V consists of the event claim and all its relevant responsive posts as nodes and E refers to a set of undirected edges corresponding to the response relation among the nodes in V. For example, for any $x_i, x_j \in V$, $x_i \to x_j$ and $x_j \to x_i$ exist if they have responsive relationships.

We transform the edge E into a symmetric adjacency matrix $\mathbf{A} \in \{0,1\}^{|V| \times |V|}$, where $\mathbf{A}_{i,j} = 1$ if $\mathbf{x}_i$ has a responsive relationship with $\mathbf{x}_j$ or $i = j$, else $\mathbf{A}_{i,j} = 0$. Then we utilize a layer-wise propagation rule to update the node vector at the lth layer:

$$H^{(l+1)} = \mathrm{ReLU}\left(\hat{\mathbf{A}} \cdot \tilde{H}^{(l)} \cdot W^{(l)}\right), \tag{2}$$

where $\hat{\mathbf{A}} = \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$ is the symmetric normalized adjacency matrix, $\mathbf{D}$ denotes the degree matrix of $\mathbf{A}$. $W^{(l)} \in \mathbb{R}^{d^{(l)} \times d^{(l+1)}}$ is a layer-specific trainable transformation matrix. In terms of $\tilde{H}^{(l)} = [\tilde{h}_0^{(l)}, \tilde{h}_1^{(l)}, \tilde{h}_2^{(l)}, \dots, \tilde{h}_{|X^*|-1}^{(l)}]^\top$, we employ a residual connection [61] around each graph convolutional layers for the multi-scale information fusion from both the claim-semantic scale and the event-structural scale, to obtain the refined representations:

$$\tilde{H}^{(l)} = \mathrm{LayerNorm}\left(H^{(l)} \| h_0^{(l-1)}\right), \tag{3}$$

where $h_0^{(l-1)} \in \mathbb{R}^{d^{(l-1)}}$ is the hidden representations of the claim at the $(l-1)$th layer, and $\|$ is the concatenation operation with the broadcast mechanism. $H^{(0)}$ and $\tilde{H}^{(0)}$ are both initialized as X^*.

For the Multi-scale GCN model with L-layers, we obtain the final node representation $\tilde{H}^{(L)}$ and jointly capture the opinions expressed in the propagation thread via mean-pooling:

$$o = \mathrm{mean{-}pooling}(\tilde{H}^{(L)}), \tag{4}$$

where $o \in \mathbb{R}^{d^{(L)}}$ is the event-level structural representation of the entire propagation thread, $d^{(L)}$ is the output dimension of GCN.

4.3. Domain-adaptive contrastive training

In order to align the representation space of rumor-relevant cues across various domains and/or languages, we introduce an innovative training framework that leverages labeled data, encompassing abundant source data and limited target data, to refine our model for target domains and languages. The key insight entails bringing the representations of source and target events belonging to the same category closer together, while maintaining a significant distance between representations of different categories, as shown in Fig. 4.

Given an event C_i^s from the source data, we firstly obtain the language-agnostic encoding for all the involved posts (see Eq. (1)) as well as the propagation structure representation o_i^s (see Eq. (4)) which is then fed into a *softmax* function to make rumor predictions. Then, we learn to minimize the cross-entropy loss between the prediction and the ground-truth label y_i^s:

$$\mathcal{L}_{CE}^s = -\frac{1}{N^s}\sum_{i=1}^{N^s} log(p_i), \tag{5}$$

where N^s is the total number of source examples in the batch, p_i is the probability of correct prediction. To improve the discrimination of rumor representation in source events, we propose a supervised contrastive learning aim to cluster the same class and separate various classes of samples:

$$\begin{aligned}\mathcal{L}_{SCL}^s = -\frac{1}{N^s}\sum_{i=1}^{N^s}\frac{1}{N_{y_i^s}-1}\sum_{j=1}^{N^s}\mathbb{1}_{[i\neq j]}\mathbb{1}_{[y_i^s=y_j^s]}\\ log\frac{exp(\mathrm{sim}(o_i^s,o_j^s))}{\sum_{k=1}^{N^s}\mathbb{1}_{[i\neq k]}exp(\mathrm{sim}(o_i^s,o_k^s))},\end{aligned} \tag{6}$$

where $N_{y_i^s}$ is the number of source examples with the same label y_i^s in the event C_i^s, and $\mathbb{1}$ is the indicator. $\mathrm{sim}(\cdot)$ is the normalized temperature-scaled cosine similarity function.

For an event C_i^t from the target data, we also compute the classification loss $\mathcal{L}_{CE}^t$ in the same manner as Eq. (5). In our study, we aligned both source and target languages to a shared semantic domain through sentence encoding. However, effective rumor detection depends not only on post-level lingual attributes but also on event-level context. In the absence of constraints, the structure-based network solely extracts event-level features from all samples according to their ultimate classification indicators. Nonetheless, these features might not hold significant relevance to the target domain and/or language. We make full use of the minor labels in the low-resource rumor data by parameterizing our model according to the contrastive objective between the source and target instances in the event-level representation space:

$$\begin{aligned}\mathcal{L}_{SCL}^t = -\frac{1}{N^t}\sum_{i=1}^{N^t}\frac{1}{N_{y_i^t}}\sum_{j=1}^{N^s}\mathbb{1}_{[y_i^t=y_j^s]}\\ log\frac{exp(\mathrm{sim}(o_i^t,o_j^s))}{\sum_{k=1}^{N^s}exp(\mathrm{sim}(o_i^t,o_k^s))},\end{aligned} \tag{7}$$

where N^t is the total number of target examples in the batch and $N_{y_i^t}$ is the number of source examples with the same label y_i^t in the event C_i^t. As a result, we project the source and target samples belonging to the same class closer than that of different categories, for feature alignment with minor annotation at the target domain and language.

4.4. Target-wise contrastive training

A critical defect of the domain-adaptive contrastive training is that almost all the target instances with the same veracity are mapped into a small projection space by simply maximizing inter-class variance and minimizing intra-class variance. Therefore, when there is a limited amount of labeled target samples, the structure-derived event-level representations could be somehow collapsed [62] and less discriminative to identify individual target samples. For instance, in the context of the COVID-19 domain, the learned rumor-indicative features for target low-resource data could just converge to the general patterns like "not true" or "joke" in well-resourced data of open domains, which may fail to detect unlabeled COVID-19 data with unseen domain-specific patterns like "just flu" or "it's bioweapon" - Denial opinions towards rumors that minimize the severity of COVID-19 or be fueled by conspiracy theories.

Thus it's important to make the target representation evenly distributed and discriminative with each other for more representative feature learning. To this end, we further exploit a target-wise contrastive learning to distinguish individual targets on top of the alignment between groups of samples with different veracity classes, which reinforces it to preserve maximal rumor-indicative information of the target events.

Given the event-level structural representation of a target sample o_i^t, we perform the target-wise contrastive objective based on the augmented event-level target data. As N^t events are randomly selected from D_t during each training stage to create a mini-batch, we first augment each target event to construct a pair of positive samples, leading to $2N^t$ event-level representations. To differentiate from other target samples, each target data sample is taught to identify its corresponding augmented sample from a batch of $2(N^t - 1)$ negative samples:

$$\mathcal{L}^t_{TCL} = -\frac{1}{N^t}\sum_{i=1}^{N^t} log \frac{exp(\mathrm{sim}(o_i^t, \overline{o}_i^t))}{\sum_{k=1}^{N^t} \mathbb{1}_{[i\neq k]}\left(exp(\mathrm{sim}(o_i^t, o_k^t)) + exp(\mathrm{sim}(o_i^t, \overline{o}_k^t))\right)}, \tag{8}$$

where $\overline{o}_i^t$ denotes the augmented event-level target representation of o_i^t, which is generated with the data augmentation strategies that would be depicted in the following subsection. In summary, as shown in Fig. 4, the target-wise contrastive objective focuses on distinguishing different target events for uniformly distributed event-level representations, and meanwhile the domain-adaptive contrastive objective identifies distinct rumor veracity from different domains and/or languages. As a result, the representations can be further enhanced by capturing more target-specific informative signals and well-generalized on diverse low-resource breaking events.

4.5. Data augmentation strategies

Data augmentation techniques were successfully utilized to enhance contrastive learning models [21], which involve creating different views or perspectives of the same data to be used as positive pairs in the target-wise contrastive learning process. In this study, we consider data augmentation from two perspectives: the encoding of the event-level representations and the modeling of the propagation structure. Thus we investigate three data augmentation techniques. The first two strategies (i.e., Adversarial Attack [27,28] and Feature Dropout [29]) are utilized to encode the event-level representations in our framework, which are widely utilized in recent studies [50,51]. To model the inherent complexity and dynamic nature of rumor dissemination [63], we attempt to augment data based on the propagation structure of target events by masking some sampling edges in the undirected propagation structure as shown in the third strategy.

Adversarial Attack. Adversarial training is commonly used to improve the robustness of a model. To create an adversarial example, we apply Fast Gradient Value [64] to approximate a worst-case perturbation at the event-level representations, where the gradient is normalized to represent the direction that significantly decreases the model's prediction performance. Then we obtain the pseudo adversarial sample by adding the perturbation to the event-level representations.

Feature Dropout. Dropout is a widely used regularization method that avoids overfitting. However, in this work, we also show its effectiveness as an augmentation strategy of event-level representations for contrastive learning. For this setting, we randomly drop elements in the event-level representations by a specific probability and set their values to zero.

Graph Dropedge. Different from Adversarial Attack and Feature Dropout directly applied to the encoding of event-level representations, we further explore an augmentation strategy based on the propagation structure. In particular, we randomly remove edges from the input undirected graphs throughout each training period to produce the deformed copies by a specific probability, which then be input into the structure-based network, i.e., Multi-scale GCNs, for the augmented event-level representations.

Algorithm 1 Unified Contrastive Learning

Require: A small set of events C_i^t in the target domain and language; A set of events C_i^s in the source domain and language.
Ensure: Assign rumor labels y to given unlabeled target data.
1: **for** each mini-batch N^t of the target events C_i^t **do**:
2: **for** each mini-batch N^s of the source events C_i^s **do**:
3: Pass C_i^* to the sentence encoder and then structure-based network to obtain its event-level feature o_i^*, where $* \in \{s, t\}$.
4: Compute the classification loss $\mathcal{L}^*_{CE}$ for source and target data, respectively.
5: Data augmentation for target data to compute the target-wise contrastive loss $\mathcal{L}^t_{TCL}$ and update $\mathcal{L}^t_{CE}$.
6: Compute the domain-adaptive contrastive loss $\mathcal{L}^*_{SCL}$.
7: Compute the joint loss $\mathcal{L}^*$ as Eq. (9).
8: Jointly optimize all parameters of the model using the average loss $\mathcal{L} = \mathrm{mean}(\mathcal{L}^s + \mathcal{L}^t)$.

4.6. Model training

We jointly train the model with the cross-entropy and contrastive objectives for the source and target training data:

$$\mathcal{L}^* = (1-\alpha)\mathcal{L}^*_{CE} + \alpha\left(\mathcal{L}^*_{SCL} + \mathbb{1}_{[*=t]}\mathcal{L}^*_{TCL}\right); * \in \{s, t\}, \tag{9}$$

where α is a trade-off parameter, which is set to 0.5 in our experiments. Algorithm 1 presents the training process of our approach. The framework is alternately trained using stochastic gradient descent with mini-batches [65]. For each mini-batch of target training data, we traverse the source data by repeating Step 3–8 in Algorithm 1. Firstly, We encode post-level representations, obtain the structure-derived event-level representations and compute traditional classification losses for source and target training data, respectively. After that, the data augmentation is conducted on target training data for the computation of the target-wise contrastive loss. And then the domain-adaptive contrastive loss is computed. In terms of Step 7, note that the training objective for the target data considers the target-wise contrastive loss in addition to the supervised contrastive loss and classification loss. We set the number L of the graph convolutional layer as 2. Parameters are updated through back-propagation [66] with the Adam optimizer [67]. The learning rate is initialized as 0.0001, and the dropout rate is 0.2. Early stopping [68] is applied to avoid overfitting.

5. Experiments

5.1. Datasets

The focus of this work, as well as in many previous studies [6,7,13, 41], is rumors on social media, instead of "fake news" strictly defined as a news article published by a news outlet that is verifiably false [19,69]. To the best of our understanding, no public benchmarks currently exist for identifying low-resource rumors featuring propagation tree structures within tweets. In this study, we focus on the COVID-19 breaking event as a representative of a low-resource domain and gather corresponding rumors and non-rumors from Twitter in English, Cantonese, and Arabic, as well as from Sina Weibo in Chinese. For the data from Twitter (English-COVID19, Cantonese-COVID19 and Arabic-COVID19), we resort two COVID-19 rumor datasets [15,70] of tweets, which only contains textual claims without propagation threads. We extend each claim by collecting its propagation thread via Twitter academic API with a twarc2 package[4] in python. For data from Sina Weibo (Chinese-COVID19), data annotation similar to [12], a set of rumorous claims is gathered from the Sina community management

[4] https://twarc-project.readthedocs.io/en/latest/twarc2_en_us/

Table 1
Statistics of datasets.

Dataset	Source		Target			
	TWITTER	WEIBO	English-COVID19	Chinese-COVID19	Cantonese-COVID19	Arabic-COVID19
# of events	1154	4649	400	399	1481	218
# of tree nodes	60 409	1 956 449	406 185	26 687	68 490	99 786
# of non-rumors	579	2336	148	146	920	78
# of rumors	575	2313	252	253	561	140
Avg. time/tree	389 h	1007 h	2497 h	248 h	668 h	2154 h
Avg. depth/tree	11.67	49.85	143.03	4.31	9.98	35.54
Language	English	Chinese	English	Chinese	Cantonese	Arabic

center[5] and non-rumorous claims by randomly filtering out the posts that are not reported as rumors. Weibo API is utilized to collect all the repost/reply messages towards each claim. All the datasets contain two binary labels: Rumor and Non-rumor. The statistics of the six datasets are illustrated in Table 1.

5.2. Experimental setup

We compare our model and several state-of-the-art baseline methods described below:

- (1) **CNN**: A CNN-based model for misinformation identification [38] by framing the relevant posts as a fixed-length sequence;
- (2) **RNN**: A RNN-based rumor detection model [12] with GRU for feature learning of relevant posts over time;
- (3) **RvNN**: A rumor detection approach based on tree-structured recursive neural networks [6] that learn rumor representations guided by the propagation structure;
- (4) **PLAN**: A transformer-based model [41] for rumor detection to capture long-distance interactions between any pair of involved tweets;
- (5) **BiGCN**: A GNN-based model [7] based on directed conversation trees to learn higher-level representations;
- (6) **DANN-***: We employ and extend an existing few-shot learning technique, domain-adversarial neural network [71], based on the structure-based model where * could be RvNN, PLAN, and BiGCN;
- (7) **UCLR-***: our proposed unified contrastive learning objectives on top of RvNN, PLAN, or BiGCN;
- (8) **UCLR**: our proposed unified propagation-aware contrastive transfer framework with multi-scale GCNs.

As the key insight to fill the low-resource gap is to relieve the limitation imposed by the specific language resource dependency besides the specific domain, in this work, we consider the most challenging case, i.e., detecting events (i.e., target) from a new domain and language. Specifically, we use TWITTER [13] and WEIBO [12] datasets as the source data; Chinese-COVID19, English-COVID19, Cantonese-COVID19 and Arabic-COVID19 datasets as the target. We use accuracy and macro-averaged F1, as well as class-specific F1 scores as the evaluation metrics.

5.3. Implementation details

In our study, all of the experiments are conducted using a solitary NVIDIA Tesla V100 GPU. The aggregate batch size is configured at 64, with equal batch sizes of 32 for both source and target samples. Within the structure-based network, each node possesses hidden and output dimensions set at 512 and 128, respectively. Since the focus in this paper is primarily on better leveraging the contrastive learning for domain and language adaptation on top of event-level representations, we choose the off-the-shelf multilingual PLM XLM-R$_{Base}$ (Layer number = 12, Hidden dimension = 768, Attention head = 12, 270M params) as our sentence encoder for language-agnostic representations at the post level. In order to perform five-fold cross-validation on the intended dataset within the context of limited target resources, we sequentially utilize each fold of the dataset for training, in conjunction with the entire source dataset, and subsequently evaluate the model on the remaining target dataset. On average, our method requires approximately 3 h to complete one iteration of five-fold cross-validation. Our model consists of 562,818 trainable parameters in total. The implementation of our model is carried out using the PyTorch framework. We also make our resources publicly available[6].

5.4. Rumor detection performance

Table 2 shows the performance of our proposed method versus all the compared methods on the Chinese-COVID19 and English-COVID19 test sets, respectively. And Table 3 further demonstrates the performance of all the compared models on the Cantonese-COVID19 and Arabic-COVID19 datasets. It is observed that the performances of the baselines in the first group are undoubtedly subpar as a result of neglecting inherent structural patterns. To make fair comparisons, all baselines are employed with the same cross-lingual sentence encoder of our framework as inputs. Other state-of-the-art baselines exploit the structural property of community wisdom on social media, which verifies the necessity of propagation structure representations aware in our framework.

Due to the expressive strength of message-passing architectures, PLAN and BiGCN beat RvNN among the structure-based baselines in the second group, though just trained with a small amount of labeled target data. The third group displays the outcomes for DANN-based methods with pre-determined training datasets TWITTER and WEIBO. Via generative adversarial nets [27] to capture cross-domain characteristics from the source and target datasets, it enhances the performance of structure-based baselines generally.

In contrast, our proposed UCLR-based framework on top of existing structure-based approaches in the fourth group achieves superior performance among all their counterparts ranging from 24.5% (13.6%) to 30.9% (18.0%) in terms of Macro F1 score on Chinese-COVID19 (English-COVID19) datasets in Table 2, and the similar phenomenon could be observed in Table 3, which suggests their strong judgment on low-resource rumors from different domains/languages. And the choice of propagation structure representation is orthogonal to our proposed framework that can be easily replaced with any existing structure-based models without any other change to our unified contrastive learning architecture. Meanwhile, it can be seen from Table 3 that, for the same target data, our framework performs generally better when utilizing WEIBO as the source data. The plausible reason might be that WEIBO has a relatively larger amount of training data than TWITTER so our domain-adaptive contrastive learning could make full use of the well-resourced data for few-shot transfer.

Our perfect model UCLR performs the best among all the baselines, even much better than the three UCLR-based variants, by mining effective clues simultaneously from the post semantics and the structural

[5] https://service.account.weibo.com/

[6] https://github.com/DanielLin97/ACLR4RUMOR-NAACL2022

Table 2
Rumor detection results on the target test datasets Chinese-COVID19 and English-COVID19.

Target (source)	Chinese-COVID19 (TWITTER)				English-COVID19 (WEIBO)			
Model	Acc.	Mac-F_1	Rumor F_1	Non-rumor F_1	Acc.	Mac-F_1	Rumor F_1	Non-rumor F_1
CNN	0.445	0.402	0.476	0.328	0.498	0.389	0.528	0.249
RNN	0.463	0.414	0.498	0.329	0.510	0.388	0.533	0.243
RvNN	0.514	0.482	0.538	0.426	0.540	0.391	0.534	0.247
PLAN	0.532	0.496	0.578	0.414	0.573	0.423	0.549	0.298
BiGCN	0.569	0.508	0.586	0.429	0.616	0.415	0.577	0.252
DANN-RvNN	0.583	0.498	0.591	0.405	0.577	0.482	0.648	0.317
DANN-PLAN	0.601	0.507	0.606	0.409	0.593	0.471	0.574	0.369
DANN-BiGCN	0.629	0.561	0.616	0.506	0.618	0.510	0.676	0.344
UCLR-RvNN	0.801	0.743	0.844	0.642	0.676	0.618	0.740	0.496
UCLR-PLAN	0.849	0.816	0.871	0.760	0.724	0.651	0.769	0.533
UCLR-BiGCN	0.885	0.867	0.898	0.835	0.769	0.687	0.815	**0.559**
UCLR	**0.895**	**0.883**	**0.916**	**0.851**	**0.773**	**0.692**	**0.827**	0.556

Table 3
Rumor detection results on the target test datasets Cantonese-COVID19 and Arabic-COVID19. The symbol ·|· for the transfer models denotes the different performance from the models trained on different source datasets, TWITTER and WEIBO, respectively.

Target	Cantonese-COVID19				Arabic-COVID19			
Model	Acc.	Mac-F_1	Rumor F_1	Non-rumor F_1	Acc.	Mac-F_1	Rumor F_1	Non-rumor F_1
CNN	0.508	0.347	0.272	0.422	0.556	0.430	0.632	0.227
RNN	0.488	0.371	0.341	0.401	0.560	0.463	0.687	0.238
RvNN	0.535	0.451	0.334	0.568	0.565	0.467	0.694	0.239
PLAN	0.544	0.459	0.289	0.629	0.573	0.470	0.641	0.298
BiGCN	0.538	0.504	0.383	0.625	0.586	0.487	0.698	0.276
DANN-RvNN	0.499|0.564	0.465|0.539	0.359|0.437	0.570|0.641	0.612|0.612	0.547|0.547	0.713|0.713	0.381|0.381
DANN-PLAN	0.531|0.572	0.473|0.522	0.339|0.370	0.607|0.673	0.636|0.631	0.568|0.555	0.717|0.736	0.419|0.374
DANN-BiGCN	0.591|0.631	0.575|0.587	0.539|0.454	0.611|0.720	0.642|0.665	0.563|0.563	0.744|0.773	0.381|0.353
UCLR-RvNN	0.571|0.670	0.501|0.627	0.323|0.499	0.679|0.754	0.659|0.678	0.611|0.636	0.732|0.756	0.489|0.516
UCLR-PLAN	0.650|0.703	0.652|0.644	0.599|0.509	0.704|0.779	0.686|0.690	0.587|0.643	0.780|0.773	0.393|0.512
UCLR-BiGCN	0.685|0.713	0.656|0.692	0.569|0.631	0.742|0.752	0.673|0.714	0.618|0.665	0.759|0.782	0.477|0.548
UCLR	0.730|**0.733**	**0.705**|0.701	**0.632**|0.612	0.777|**0.789**	0.713|**0.732**	0.670|**0.687**	0.786|**0.797**	0.554|**0.577**

property via multi-scale encoding for conversation threads. Furthermore, the structure-based counterparts generally have more parameters and complex structures (UCLR-BiGCN with total trainable parameters 1,117,954) than Multi-scale GCNs of UCLR framework with total trainable parameters 562,818. Although such complex structure-based networks like BiGCN may show promising performance on the mono-domain and mono-lingual training corpora, their generalization ability in cross-domain and cross-lingual settings may be compromised. This is because excessively complex models may overfit the training set data, leading to inaccurate generalization to new target data. This also justifies the complementary of our proposed Multi-scale GCNs backbone and the UCLR training paradigm. In summary, the main results indicate that the unified contrastive learning framework can effectively transfer knowledge from the source to target data at the event level, and substantiate our method is model-agnostic for different structure-based networks. For a more clear qualitative analysis of the effectiveness of the domain-adaptive contrastive learning [14] and the target-wise contrastive learning, we further provide the ablative test on the unified contrastive transfer framework UCLR in the following subsection Section 5.5.

5.5. Ablative study

We perform ablation studies based on our proposed approach UCLR, where the performance from models trained with TWITTER as the source data is shown in Table 4 and that with WEIBO as the source data is shown in Table 5. For the cross-domain and cross-lingual settings, we use Chinese-COVID19, Cantonese-COVID19, and Arabic-COVID19 as the target data when TWITTER is utilized as the source data, and we use English-COVID19, Cantonese-COVID19, and Arabic-COVID19 as the target data when WEIBO is utilized as the source data.

Table 4
Results of the ablation study of UCLR with TWITTER as the source data.

Source	TWITTER					
Target	Chinese-COVID19		Cantonese-COVID19		Arabic-COVID19	
Model	Acc.	Mac-F_1	Acc.	Mac-F_1	Acc.	Mac-F_1
BiGCN(T)	0.569	0.508	0.538	0.504	0.586	0.487
BiGCN(S)	0.578	0.463	0.562	0.541	0.631	0.536
BiGCN(S,T)	0.693	0.472	0.576	0.558	0.655	0.539
DANN-BiGCN	0.629	0.561	0.591	0.575	0.642	0.563
ACLR-BiGCN	0.873	0.861	0.653	0.617	0.671	0.579
UCLR-BiGCN	0.885	0.867	0.685	0.656	0.673	0.618
UCLR_{Adv}	0.890	0.871	0.718	0.697	0.709	0.616
$\text{UCLR}_{\text{Dropout}}$	0.888	0.869	0.721	0.702	0.699	0.580
$\text{UCLR}_{\text{DropEdge}}$	0.895	0.883	0.730	0.705	0.713	0.670

Effect of Well-resourced Data. As demonstrated in Tables 4 and 5, the first group shows the results for the best-performed data-driven baseline BiGCN. We observe that the model performs best if pre-trained on source data and then fine-tuned on target training data (i.e., BiGCN(S,T)), compared with the poor performance when trained on either minor labeled target data only (i.e., BiGCN(T)) or well-resourced source data (i.e., BiGCN(S)). This suggests that our hypothesis of leveraging well-resourced source data to improve the low-resource rumor detection on target data is feasible.

Effect of Feature Alignment. In the second group, the DANN-based model makes better use of the source data to extract domain-agnostic

Table 5
Results of the ablation study of UCLR with WEIBO as the source data.

Source	WEIBO					
Target	Chinese-COVID19		Cantonese-COVID19		Arabic-COVID19	
Model	Acc.	Mac-F_1	Acc.	Mac-F_1	Acc.	Mac-F_1
BiGCN(T)	0.616	0.415	0.538	0.504	0.586	0.487
BiGCN(S)	0.578	0.463	0.562	0.525	0.612	0.506
BiGCN(S,T)	0.693	0.472	0.581	0.538	0.633	0.519
DANN-BiGCN	0.618	0.510	0.631	0.587	0.665	0.563
ACLR-BiGCN	0.765	0.686	0.698	0.679	0.707	0.624
UCLR-BiGCN	0.769	0.687	0.713	0.692	0.714	0.665
$\mathrm{UCLR_{Adv}}$	0.768	0.691	0.729	0.697	0.710	0.658
$\mathrm{UCLR_{Dropout}}$	0.771	0.689	0.727	0.683	0.718	0.676
$\mathrm{UCLR_{DropEdge}}$	0.773	0.692	0.733	0.701	0.732	0.687

features, which further leads to performance improvement. Our proposed domain-Adaptive Contrastive Learning approach ACLR-BiGCN has already achieved outstanding performance compared with other baselines, which illustrates its effectiveness on domain and language adaptation.

Effect of Target-wise Uniform Distribution. We further notice that our UCLR-BiGCN consistently outperforms all baselines and improves the prediction performance of ACLR-BiGCN, suggesting that training model to preserve more rumor-indicative information on target data with more uniform distribution, could provide robust generalization for more accurate rumor predictions, especially in low-resource regimes.

Effect of Multi-scale GCNs. Our proposed UCLR frameworks with Multi-scale GCNs in the third group generally perform better than the UCLR-BiGCN, which indicates the potential of Multi-scale GCNs as the backbone of our few-shot transfer framework for propagation structure representation learning, complementary to the proposed unified contrastive training paradigm.

Effect of Data Augmentation Strategies. In the third group, we explore the effectiveness of different augmentation strategies to our proposed Target-wise Contrastive Learning. We can observe that the Graph Dropedge we employ as the data augmentation for the propagation structure is the most effective strategy, outperforming Adversarial Attacks and Feature Dropout. This is probably because Grpah Dropedge is more related to our propagation-aware framework since they are directly operated on the Event-structural scale of Multi-scale GCNs and change the structure of the propagation to produce hard examples.

5.6. Early detection

Early alerts of rumors are important to minimize their detrimental impact. By setting detection checkpoints of "delays" that can be either the count of reply posts or the time elapsed since the first posting, only contents posted within the specified checkpoint parameters are available for model evaluation. The efficacy is assessed by Macro F1 score attained at each respective checkpoint. To meet each checkpoint, we incrementally scan test data in order of time until the pre-determined time delay or post volume is reached.

Fig. 6 shows the performances of our approach versus ACLR-BiGCN [14], DANN-BiGCN [71], BiGCN [7], PLAN [41], and RvNN [6] at various deadlines.

We observe that the accuracies of all systems obviously increase with elapsed time or post counts, our proposed UCLR approach outperforms other counterparts and baselines throughout the whole lifecycle, which grows more quickly to supersede the other baselines and reaches a relatively high Macro F1 score at a very early period after the initial broadcast. One interesting phenomenon is that the early performance of some baselines may fluctuate more or less. It is because with the propagation of the claim, there is more semantic and structural information but the noisy information is increased simultaneously. However, our model has better steady rise of early detection performance than the baselines. We speculate the reason is that our proposed framework with multi-scale GCNs stands out for its simplicity and effectiveness in simultaneously leveraging the different scales from semantic and structural information, where integrating the claim information for the claim-semantic scale could not only guard the consistency of topics but also alleviate the potential noise resulting from the diffusion with the event-structural scale. This simple yet effective approach proves instrumental in the early stages of domain adaptation, complementary with the unified contrastive training paradigm for enhancing rumor detection. As a result, our method only needs about 50 posts and around 4 h with TWITTER and WEIBO as source data, respectively, to achieve the saturated performance, indicating the remarkably superior early detection performance of our method.

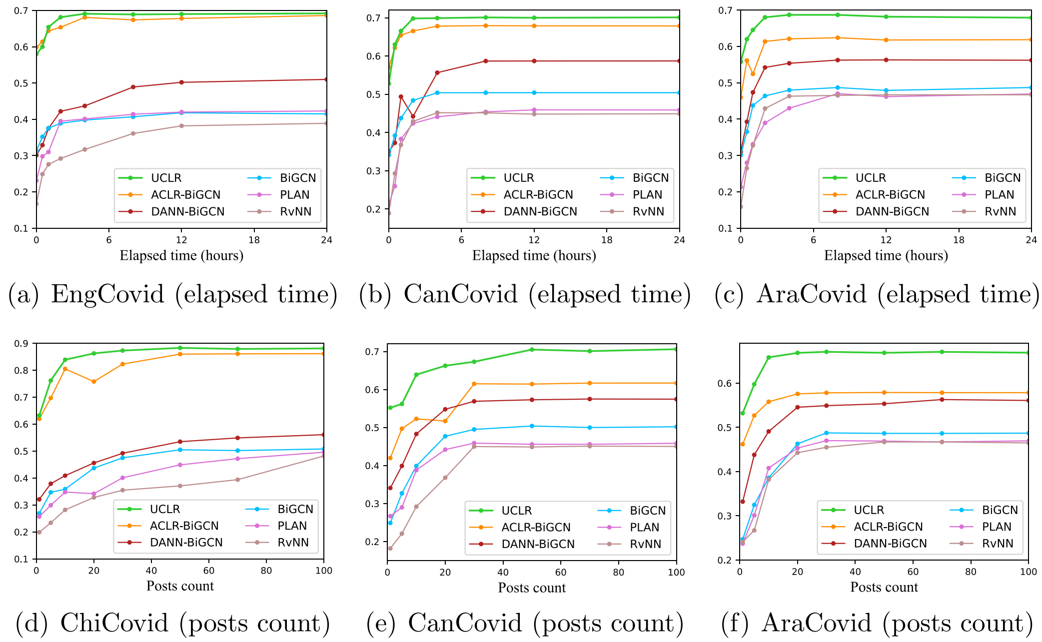
5.7. Feature visualization

Fig. 7 shows the PCA visualization of learned target event-level features obtained from traditional classification (left) and ACLR (right) paradigms on Chinese-COVID19 data for analysis. The left figure represents model training with only classification loss, and the right figure uses our proposed domain-Adaptive Contrastive Learning for training. We observe that (1) due to the lack of sufficient training data, the features extracted with the traditional training paradigm are entangled, making it difficult to detect rumors in low-resource regimes; and (2) our ACLR-based approach learns more discriminative representations to improve low-resource rumor classification, reaffirming that our training paradigm can effectively transfer knowledge to bridge the gap between source and target data distribution resulting from different domains and languages. Furthermore, Fig. 8 illustrates the difference in feature visualization obtained from ACLR (left) and UCLR (right) paradigms on Arabic-COVID19 data. It is observed that, besides the better decoupling for different rumor-related labels, the UCLR achieves a relatively more evenly distributed feature set for the target data compared with the ACLR, which indicates the effectiveness of Target-wise Contrastive Learning in contributing to the generalization ability of our framework in low-resource regimes.

5.8. Case study of propagation structure

For a more comprehensive analysis of the propagation structure, we present an example of correctly detected rumors with part of its propagation structure. The visualization of tweets in Fig. 9 shows that when a post challenges a rumor, it tends to elicit supportive or affirming replies that confirm the denial. Conversely, when a post endorses a rumor, it tends to trigger denials in its replies. Furthermore, it is observed that a reply typically responds to its immediate parent node rather than directly to the root claim. This observation aligns with our motivation to explore the propagation structure of rumors for representation learning. By adopting an undirected topology, the structure can be naturally modeled to capture the signals indicative of rumors and enhance representation by fully aggregating features from all informative neighbors. This enables the adaptive propagation of information association between nodes in the conversation thread along responsive parent–child relationships.

Furthermore, we can observe that the informative posts should be developed and extended around the content of the claim, i.e., the potential and implicit target to be checked. This highlights the significance of the claim content to catch informative posts. Our proposed multi-scale GCNs could integrate claim information from the claim-semantic scale with the propagation thread from the event-structural scale, to enrich the semantic context of replies and better guard the consistency of topics for the correct prediction.



(a) EngCovid (elapsed time) (b) CanCovid (elapsed time) (c) AraCovid (elapsed time)

(d) ChiCovid (posts count) (e) CanCovid (posts count) (f) AraCovid (posts count)

Fig. 6. Early detection performance Macro F1 at different checkpoints of elapsed time (or posts count), where EngCovid, ChiCovid, CanCovid and AraCovid denote the English-COVID19, Chinese-COVID19, Cantonese-COVID19 and Arabic-COVID19, respectively.

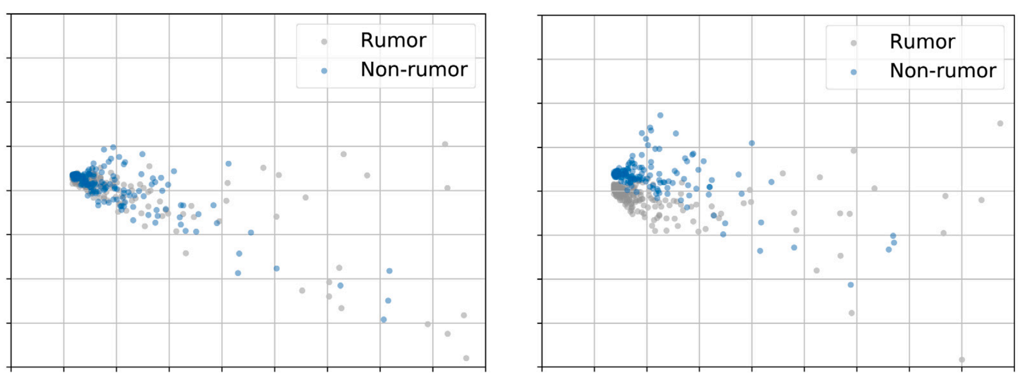


Fig. 7. Visualization of target event-level representation distribution for traditional classification (left) and ACLR (right) paradigms on Chinese-COVID19 data.

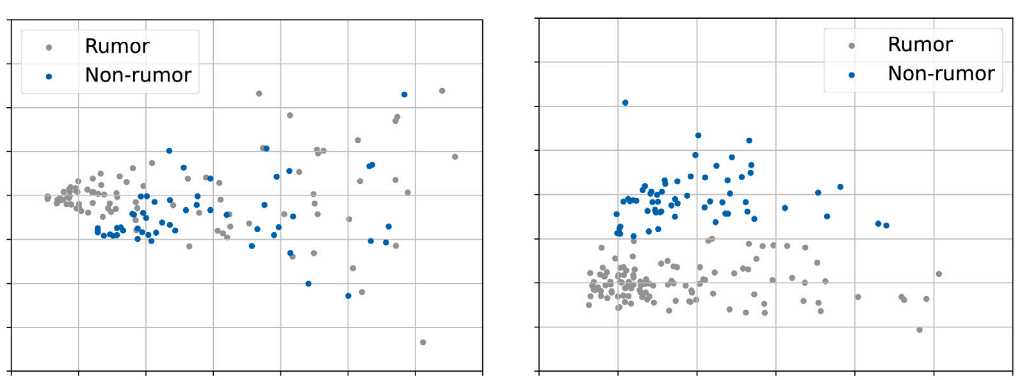


Fig. 8. Visualization of target event-level representation distribution for ACLR (left) and UCLR (right) paradigms on Arabic-COVID19 data.

5.9. Error analysis

In this section, to gain deeper insights into our model's behavior and to provide groundwork for future investigations, we perform an error analysis specifically focusing on the misclassified rumor examples by our proposed framework. This analysis aims to illuminate the nuances behind erroneous predictions, contributing to a more comprehensive understanding of the model's performance with the propagation structure.

We found that the major error exists in that our framework still cannot perfectly handle the instance in which few users' engagements are available. In real-world scenarios, certain users opt to reshare content without appending their own opinions or comments. This behavior presents a challenge for our model, particularly when engagements

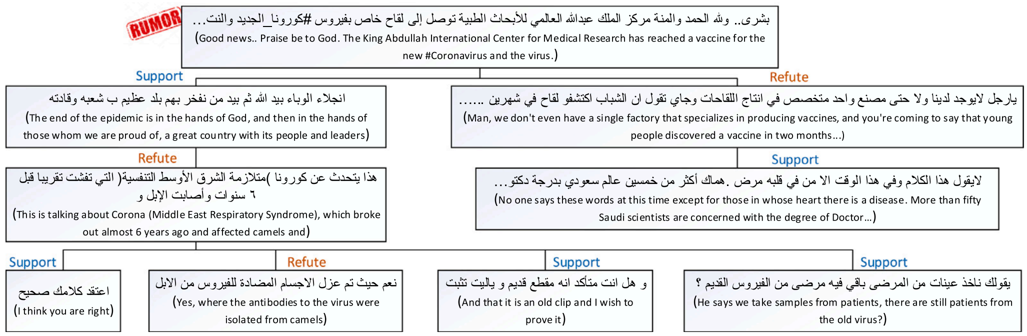


Fig. 9. A sample case of correctly detected rumors of our model. We show important tweets in the propagation structure and truncate others.

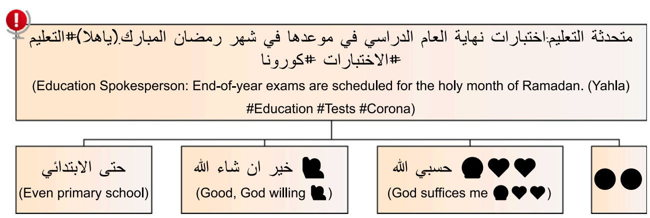


Fig. 10. A sample case of wrongly detected rumors of our model.

Table 6
Rumor detection results of our proposed framework in different low-resource settings. Cross-D&L denotes the cross-domain and cross-lingual settings and Cross-D denotes the cross-domain and mono-lingual settings.

Target	Chinese-COVID19		English-COVID19	
Settings	Acc.	Mac-F_1	Acc.	Mac-F_1
Cross-D&L	0.895	**0.883**	**0.773**	**0.692**
Cross-D	**0.899**	0.864	0.752	0.645

from a limited number of users are available. This scenario mirrors the complexity of early rumor detection. While our model demonstrates impressive performance in early rumor detection, it encounters inaccuracies stemming from situations where users predominantly retweet claims by attaching emojis with ambiguous meanings, but lack additional text of opinions, as shown in Fig. 10. Thus we plan to investigate the role of non-textual media such as images or emojis in the effectiveness of detecting rumors. Additionally, an intriguing observation is that users may reply to their own claims during information propagation. To enhance our modeling approach for novel social networks, accounting for these distinct behaviors (e.g., retweets or replies originating from the node posting the claim itself) is crucial for more heuristic rumor propagation analysis.

5.10. Effect of trade-off hyper-parameter

To study the effects on performance (Macro F1 score) of the trade-off hyper-parameter in our training paradigm, we conduct qualitative analysis under UCLR architecture (Fig. 11). For the target data Chinese-COVID19 and Cantonese-COVID19, we use TWITTER as the source data; in terms of English-COVID19 and Arabic-COVID19, we use WEIBO as the source data. Since the platform for collecting Chinese-COVID19 data is Sina Weibo while the platform for the other three datasets is Twitter, there will be a large gap between the model's performance on Chinese-COVID19 data and its performance on the other three datasets. We can see that $\alpha = 0.5$ achieves the best performance while the point where $\alpha = 0.3$ also has good performance. Looking at the overall trend, the performance fluctuates more or less as the value of α grows. We conjecture that this is because the unified contrastive objective, while optimizing the representation distribution, compromises the mapping relationship with labels. Such a multi-task paradigm means optimizing the traditional classification loss and the unified contrastive loss simultaneously. This setting leads to mutual interference between two tasks, which affects the convergence effect. This phenomenon points out the direction for our further research in the future.

5.11. Discussion about low-resource settings

To highlight the low-resource settings in our experiments, we analyze our proposed framework in this section using mono-lingual and cross-lingual source datasets. Considering the cross-domain and cross-lingual settings in Table 2 of the main experiments, we also conduct an experiment in cross-domain and mono-lingual settings. Specifically, for the Chinese-COVID19 as the target data, we utilize the WEIBO dataset as the source data with rich annotations. In terms of English-COVID19, we set the TWITTER dataset as the source data. Table 6 depicted the results in different low-resource settings. The results show that our model generally performs better in both cross-domain and cross-lingual settings compared to just cross-domain, revealing that relieving language resource dependency is crucial in addressing the low-resource gap. Our proposed framework, the unified propagation-aware contrastive transfer, can alleviate the low-resource issue in rumor detection and reduce dependence on domain and language-specific annotated datasets.

6. Conclusion and future work

In this work, we propose a unified contrastive transfer framework with propagation structure to bridge low-resource gaps for rumor detection on social media by adapting social contextual features learned from well-resourced data to that of the low-resource breaking events. For the novel Low-Resource Rumor Detection task, our domain-adaptive contrastive learning aligns identical features from different domains and/or languages. Furthermore, we propose target-wise contrastive learning with three data augmentation strategies to optimize representations of target data more uniformly by distinguishing individual target training samples, for better generalization to unseen target data. Results on four real-world datasets show that: (1) our method is more effective and robust compared with state-of-the-art baselines; and (2) our extended unified contrastive transfer framework with target-wise contrastive learning makes further improvements over the original domain-adaptive contrastive model. We also compare different data augmentation strategies for target-wise contrastive learning and provide comprehensive analysis for interpreting how our approach works.

For future work, we will explore the following directions: (1) We will investigate the pre-training paradigm in contrastive manners and then fine-tune the model using classification loss, which may further enhance the model's performance and stability; (2) Besides the textual information of the relevant posts, we will incorporate more information types (e.g., user profiles, post time, etc.) for improving our unified

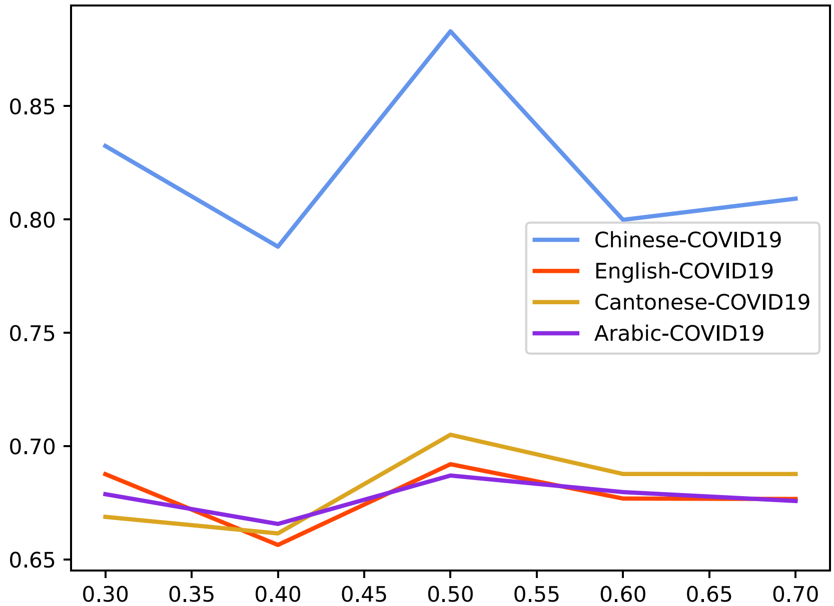


Fig. 11. Effect of trade-off hyper-parameter α between Classification and Contrastive Objectives.

contrastive training paradigm; (3) To deepen the interpretability of our framework, we aim to further extract explicit token-level insights by elucidating the specific textual features learned. This will involve enhancing the cross-lingual sentence encoder within our approach; (4) We intend to leverage more datasets with abundant annotation and adapt our model to other domains and minority languages since our model has explicitly conquered the restriction of both domain and language usage in distinct datasets.

Acknowledgements

This work is partially supported by National Natural Science Foundation of China Young Scientists Fund (No. 62206233) and Hong Kong RGC ECS (No. 22200722).

CRediT authorship contribution statement

Hongzhan Lin: Data curation, Methodology, Writing – original draft. **Jing Ma:** Supervision, Writing – review & editing, Project administration. **Ruichao Yang:** Visualization, Writing – review & editing. **Zhiwei Yang:** Validation, Writing – review & editing. **Mingfei Cheng:** Formal analysis, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

We have shared the link for some codes and data in the manuscript.

References

[1] E. Chen, K. Lerman, E. Ferrara, et al., Tracking social media discourse about the covid-19 pandemic: Development of a public coronavirus twitter data set, JMIR Public Health Surveil. 6 (2) (2020) e19273.

[2] K. Zarei, R. Farahbakhsh, N. Crespi, G. Tyson, A first instagram dataset on covid-19, 2020, arXiv preprint arXiv:2004.12226.

[3] S. Alqurashi, A. Alhindi, E. Alanazi, Large arabic twitter dataset on covid-19, 2020, arXiv preprint arXiv:2004.04315.

[4] M.A. Hedderich, L. Lange, H. Adel, J. Strötgen, D. Klakow, A survey on recent approaches for natural language processing in low-resource scenarios, in: NAACL, 2021, pp. 2545–2568.

[5] G.W. Allport, L.J. Postman, Section of psychology: The basic psychology of rumor, Trans.New York Acad. Sci. 8 (2 Series II) (1945) 61–81.

[6] J. Ma, W. Gao, K.-F. Wong, Rumor detection on Twitter with tree-structured recursive neural networks, in: ACL, 2018, pp. 1980–1989.

[7] T. Bian, X. Xiao, T. Xu, P. Zhao, W. Huang, Y. Rong, J. Huang, Rumor detection on social media with bi-directional graph convolutional networks, in: AAAI, 2020, pp. 549–556.

[8] Z. Chen, L. Wang, X. Zhu, S. Dietze, TSNN: A topic and structure aware neural network for rumor detection, Neurocomputing 531 (2023) 114–124.

[9] Y. Yan, Y. Wang, P. Zheng, Rumor detection on social networks focusing on endogenous psychological motivation, Neurocomputing 552 (2023) 126548.

[10] A. Silva, L. Luo, S. Karunasekera, C. Leckie, Embracing domain differences in fake news: Cross-domain fake news detection using multi-modal data, in: AAAI, Vol. 35, No. 1, 2021, pp. 557–565.

[11] A. Zubiaga, M. Liakata, R. Procter, Learning reporting dynamics during breaking news for rumour detection in social media, 2016, arXiv preprint arXiv:1610.07363.

[12] J. Ma, W. Gao, P. Mitra, S. Kwon, B.J. Jansen, K.-F. Wong, M. Cha, Detecting rumors from microblogs with recurrent neural networks, in: IJCAI, 2016, pp. 3818–3824.

[13] J. Ma, W. Gao, K.-F. Wong, Detect rumors in microblog posts using propagation structure via kernel learning, in: ACL, 2017, pp. 708–717.

[14] H. Lin, J. Ma, L. Chen, Z. Yang, M. Cheng, G. Chen, Detect rumors in microblog posts for low-resource domains via adversarial contrastive learning, in: NAACL, Association for Computational Linguistics, 2022, pp. 2543–2556.

[15] F. Alam, S. Shaar, F. Dalvi, H. Sajjad, A. Nikolov, H. Mubarak, G. Da San Martino, A. Abdelali, N. Durrani, K. Darwish, et al., Fighting the COVID-19 infodemic: Modeling the perspective of journalists, fact-checkers, social media platforms, policy makers, and the society, in: EMNLP, 2021, pp. 611–649.

[16] D. Kar, M. Bhardwaj, S. Samanta, A.P. Azad, No rumours please! a multi-indic-lingual approach for covid fake-tweet detection, in: GHCI, IEEE, 2021, pp. 1–5.

[17] L. Cui, D. Lee, Coaid: Covid-19 healthcare misinformation dataset, 2020, arXiv preprint arXiv:2006.00885.

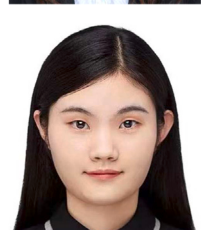
[18] Y. Hu, H.-Y. Huang, A. Chen, X.-L. Mao, Weibo-COV: A large-scale COVID-19 social media dataset from weibo, in: EMNLP Workshops, 2020.

[19] A. Zubiaga, A. Aker, K. Bontcheva, M. Liakata, R. Procter, Detection and resolution of rumours in social media: A survey, ACM Comput. Surv. 51 (2) (2018) 1–36.

[20] K. He, H. Fan, Y. Wu, S. Xie, R. Girshick, Momentum contrast for unsupervised visual representation learning, in: CVPR, 2020, pp. 9729–9738.
[21] T. Chen, S. Kornblith, M. Norouzi, G. Hinton, A simple framework for contrastive learning of visual representations, in: ICML, PMLR, 2020, pp. 1597–1607.
[22] T. Chen, S. Kornblith, K. Swersky, M. Norouzi, G. Hinton, Big self-supervised models are strong semi-supervised learners, in: NeurIPS, 2020, pp. 22243–22255.
[23] T. Wang, P. Isola, Understanding contrastive representation learning through alignment and uniformity on the hypersphere, in: ICML, PMLR, 2020, pp. 9929–9939.
[24] P. Bachman, R.D. Hjelm, W. Buchwalter, Learning representations by maximizing mutual information across views, in: NeurIPS, 2019, pp. 15535–15545.
[25] J. Gao, D. He, X. Tan, T. Qin, L. Wang, T. Liu, Representation degeneration problem in training natural language generation models, in: ICLR, 2018.
[26] Y. Tian, D. Krishnan, P. Isola, Contrastive multiview coding, in: ECCV, Springer, 2020, pp. 776–794.
[27] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: NeurIPS, 27, 2014.
[28] A. Kurakin, I.J. Goodfellow, S. Bengio, Adversarial examples in the physical world, in: Artificial Intelligence Safety and Security, Chapman and Hall/CRC, 2018, pp. 99–112.
[29] G.E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, R.R. Salakhutdinov, Improving neural networks by preventing co-adaptation of feature detectors, 2012, arXiv preprint arXiv:1207.0580.
[30] Y. Rong, W. Huang, T. Xu, J. Huang, DropEdge: Towards deep graph convolutional networks on node classification, in: ICLR, 2019.
[31] C. Castillo, M. Mendoza, B. Poblete, Information credibility on twitter, in: WWW, 2011, pp. 675–684.
[32] F. Yang, Y. Liu, X. Yu, M. Yang, Automatic detection of rumor on sina weibo, in: KDD Workshops, 2012, pp. 1–7.
[33] X. Liu, A. Nourbakhsh, Q. Li, R. Fang, S. Shah, Real-time rumor debunking on twitter, in: CIKM, 2015, pp. 1867–1870.
[34] S. Kwon, M. Cha, K. Jung, W. Chen, Y. Wang, Prominent features of rumor propagation in online social media, in: ICDM, IEEE, 2013, pp. 1103–1108.
[35] A. Friggeri, L. Adamic, D. Eckles, J. Cheng, Rumor cascades, in: Eighth International AAAI Conference on Weblogs and Social Media, 2014.
[36] A. Hannak, D. Margolin, B. Keegan, I. Weber, Get back! you don't know me like that: The social mediation of fact checking interventions in Twitter conversations., in: ICWSM, 2014.
[37] Z. Zhao, P. Resnick, Q. Mei, Enquiring minds: Early detection of rumors in social media from enquiry posts, in: WWW, 2015, pp. 1395–1405.
[38] F. Yu, Q. Liu, S. Wu, L. Wang, T. Tan, A convolutional approach for misinformation identification, in: IJCAI, 2017.
[39] H. Guo, J. Cao, Y. Zhang, J. Guo, J. Li, Rumor detection with hierarchical social attention network, in: CIKM, 2018, pp. 943–951.
[40] K. Wu, S. Yang, K.Q. Zhu, False rumors detection on sina weibo by propagation structures, in: ICDE, IEEE, 2015, pp. 651–662.
[41] L.M.S. Khoo, H.L. Chieu, Z. Qian, J. Jiang, Interpretable rumor detection in microblogs by attending to user interactions, in: AAAI, 2020, pp. 8783–8790.
[42] J. Ma, W. Gao, Debunking rumors on Twitter with tree transformer, in: COLING, 2020, pp. 5455–5466.
[43] H. Lin, J. Ma, M. Cheng, Z. Yang, L. Chen, G. Chen, Rumor detection on twitter with claim-guided hierarchical graph attention networks, in: EMNLP, 2021.
[44] M. Janicka, M. Pszona, A. Wawer, Cross-domain failures of fake news detection, Comput. Sist. 23 (3) (2019).
[45] H. Lin, P. Yi, J. Ma, H. Jiang, Z. Luo, S. Shi, R. Liu, Zero-shot rumor detection with propagation structure via prompt learning, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 37, No. 4, 2023, pp. 5213–5221.
[46] Y. Wang, F. Ma, Z. Jin, Y. Yuan, G. Xun, K. Jha, L. Su, J. Gao, Eann: Event adversarial neural networks for multi-modal fake news detection, in: KDD, 2018, pp. 849–857.
[47] H. Yuan, J. Zheng, Q. Ye, Y. Qian, Y. Zhang, Improving fake news detection with domain-adversarial and graph-attention neural network, Decis. Support Syst. (2021) 113633.
[48] T. Zhang, D. Wang, H. Chen, Z. Zeng, W. Guo, C. Miao, L. Cui, BDANN: BERT-based domain adaptation neural network for multi-modal fake news detection, in: IJCNN, IEEE, 2020, pp. 1–8.
[49] N. Lee, Y. Bang, A. Madotto, P. Fung, Towards few-shot fact-checking via perplexity, in: NAACL, 2021, pp. 1971–1981.
[50] T. Gao, X. Yao, D. Chen, Simcse: Simple contrastive learning of sentence embeddings, in: EMNLP, 2021, pp. 6894–6910.
[51] Y. Yan, R. Li, S. Wang, F. Zhang, W. Wu, W. Xu, Consert: A contrastive framework for self-supervised sentence representation transfer, in: ACL-IJCNLP, 2021, pp. 5065–5075.
[52] M. Mohtarami, J. Glass, P. Nakov, Contrastive language adaptation for cross-lingual stance detection, in: EMNLP-IJCNLP, 2019, pp. 4442–4452.
[53] D. Zhang, F. Nan, X. Wei, S.-W. Li, H. Zhu, K. McKeown, R. Nallapati, A.O. Arnold, B. Xiang, Supporting clustering with contrastive learning, in: NAACL, 2021, pp. 5419–5430.
[54] H. Lin, Y. Yan, G. Chen, Boosting low-resource intent detection with in-scope prototypical networks, in: ICASSP, IEEE, 2021, pp. 7623–7627.
[55] Y. Liu, P. Liu, Simcls: A simple framework for contrastive learning of abstractive summarization, in: ACL-IJCNLP, 2021, pp. 1065–1072.
[56] Z. Yue, H. Zeng, Z. Kou, L. Shang, D. Wang, Contrastive domain adaptation for early misinformation detection: A case study on covid-19, in: CIKM, 2022, pp. 2423–2433.
[57] J. Ma, Y. Liu, M. Liu, M. Han, Curriculum contrastive learning for fake news detection, in: CIKM, 2022, pp. 4309–4313.
[58] J. Wu, W. Xu, Q. Liu, S. Wu, L. Wang, Adversarial contrastive learning for evidence-aware fake news detection with graph neural networks, 2022, arXiv preprint arXiv:2210.05498.
[59] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, É. Grave, M. Ott, L. Zettlemoyer, V. Stoyanov, Unsupervised cross-lingual representation learning at scale, in: ACL, 2020, pp. 8440–8451.
[60] T.N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, 2016, arXiv preprint arXiv:1609.02907.
[61] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: CVPR, 2016, pp. 770–778.
[62] X. Chen, K. He, Exploring simple siamese representation learning, in: CVPR, 2021, pp. 15750–15758.
[63] Y. Rong, W. Huang, T. Xu, J. Huang, DropEdge: Towards deep graph convolutional networks on node classification, in: ICLR, 2019.
[64] A. Rozsa, E.M. Rudd, T.E. Boult, Adversarial diversity and hard positive generation, in: CVPR Workshops, 2016, pp. 25–32.
[65] A. Bordes, L. Bottou, P. Gallinari, SGD-QN: Careful quasi-Newton stochastic gradient descent, J. Mach. Learn. Res. 10 (2009) 1737–1754.
[66] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, P. Kuksa, Natural language processing (almost) from scratch, J. Mach. Learn. Res. 12 (ARTICLE) (2011) 2493–2537.
[67] I. Loshchilov, F. Hutter, Decoupled weight decay regularization, in: ICLR, 2018.
[68] Y. Yao, L. Rosasco, A. Caponnetto, On early stopping in gradient descent learning, Constr. Approx. 26 (2) (2007) 289–315.
[69] K. Shu, A. Sliva, S. Wang, J. Tang, H. Liu, Fake news detection on social media: A data mining perspective, ACM SIGKDD Explor. Newslett. 19 (1) (2017) 22–36.
[70] L. Ke, X. Chen, Z. Lu, H. Su, H. Wang, A novel approach for cantonese rumor detection based on deep neural network, in: SMC, IEEE, 2020, pp. 1610–1615.
[71] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, V. Lempitsky, Domain-adversarial training of neural networks, J. Mach. Learn. Res. 17 (59) (2016) 1–35.

Hongzhan Lin is currently pursuing a Ph.D. degree at the Department of Computer Science, Hong Kong Baptist University (HKBU). He received a master's degree in Information and Communication Engineering at the Pattern Recognition and Intelligent System Laboratory, Beijing University of Posts and Telecommunications (BUPT). His research interests include social media analytics, rumor detection, and natural language processing. He has published and served at several international conferences including AAAI, WWW, EMNLP, ACL, NAACL, SIGIR etc.

Jing Ma received the Ph.D. degree from The Chinese University of Hong Kong (CUHK) in 2020. She is currently an Assistant Professor at the department of Computer Science, Hong Kong Baptist University (HKBU). Her current research interests include Natural Language Processing, Information Verification, and Social Media Analytics. She has been serving on the program committee of several international conferences, including: IJCAI, AAAI, WWW, CIKM, ACL and EMNLP.



Ruichao Yang received the bachelor's degree from Jinlin University (JLU) in 2015 and master's degree from Peking University (PKU) in 2018 respectively. She is currently pursuing the Ph.D. degree at the department of Computer Science, Hong Kong Baptist University (HKBU). Her current research interests include Natural Language Processing, Rumor Verification, Fake News Detection, Misinformation Detection and Social Media Analytics.

Zhiwei Yang is currently an assistant professor at Jinan University. Before that, he was a Ph.D. student at the College of Computer Science and Technology, Jilin University (JLU), Changchun, Jilin Province, China, and a full-time exchange ph.D. student in Department of Computer Science, Hong Kong Baptist University (HKBU), Hong Kong. His research interests include information extraction, rumor detection, and artificial intelligence. His publications include AAAI, IJCAI, EMNLP, COLING, TNNLS, IP&M, Neurocomputing and et al. He has been serving as a reviewer for Neural Networks, Neurocomputing, KSEM, and et al.

Mingfei Cheng is currently pursuing a Ph.D. degree in Computer Science, Singapore Management University (SMU). He received a master's degree in Information and Communication Engineering at the Pattern Recognition and Intelligent System Laboratory, Beijing University of Posts and Telecommunications (BUPT). His research interests include trustworthy AI, autonomous driving, and computer vision. His publications include ICCV, EMNLP, NAACL, ACCV, etc.