Reinforcement Tuning for Detecting Stances and Debunking Rumors Jointly with Large Language Models

Ruichao Yang[1], **Wei Gao**[2], **Jing Ma**[1*], **Hongzhan Lin**[1], **Bo Wang**[3]
[1]Hong Kong Baptist University, Hong Kong SAR, China
[2]Singapore Management University, Singapore
[3]Jilin University, Changchun, Jilin, China
{csrcyang,majing,cshzlin}@comp.hkbu.edu.hk,
weigao@smu.edu.sg, wangbo21@mails.jlu.edu.cn

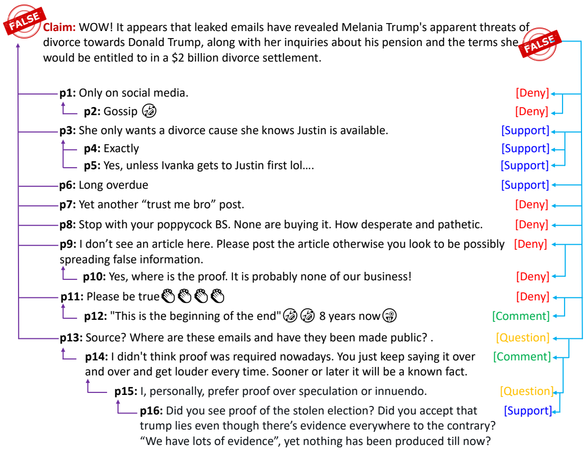
Abstract

Learning multi-task models for jointly detecting stance and verifying rumors poses challenges due to the need for training data of stance at post level and rumor veracity at claim level, which are difficult to obtain. To address this issue, we leverage large language models (LLMs) as the foundation annotators for the joint stance detection (SD) and rumor verification (RV) tasks, dubbed as JSDRV. We introduce a novel reinforcement tuning framework to enhance the joint predictive capabilities of LLM-based SD and RV components. Specifically, we devise a policy for selecting LLM-annotated data at the two levels, employing a hybrid reward mechanism to choose high-quality labels for effective LLM fine-tuning on both tasks. Results demonstrate that JSDRV improves the capabilities of LLMs in the joint tasks, not only outperforming state-of-the-art methods but also generalizing to non-LLMs accommodated as task models.

1 Introduction

Social media has transformed the ways people access information by facilitating rapid information sharing. However, it has become a fertile ground for nurturing rumors and misinformation due to its lack of systematic moderation (Vosoughi et al., 2018; Cheng et al., 2021). Their rampant spread has become a considerable global societal issue that can profoundly impact people's beliefs and normal life (Roozenbeek and van der Linden, 2019).

In general, stance provides insights into the attitudes, opinions, and beliefs of individuals regarding a specific target (Küçük and Can, 2020; AL-Dayel and Magdy, 2021). Stance and rumor are deeply coupled since the stance expressed by social media posts toward a rumorous event, i.e., rumor stance, offers valuable cues for assessing the overall credibility of the target claim. Figure 1 shows that the rumor stance helps in understanding the context and the way the rumor is perceived by different users in an intuitive and explainable manner.



Claim: WOW! It appears that leaked emails have revealed Melania Trump's apparent threats of divorce towards Donald Trump, along with her inquiries about his pension and the terms she would be entitled to in a $2 billion divorce settlement.

p1: Only on social media. [Deny]
p2: Gossip 😂 [Deny]
p3: She only wants a divorce cause she knows Justin is available. [Support]
p4: Exactly [Support]
p5: Yes, unless Ivanka gets to Justin first lol…. [Support]
p6: Long overdue [Support]
p7: Yet another "trust me bro" post. [Deny]
p8: Stop with your poppycock BS. None are buying it. How desperate and pathetic. [Deny]
p9: I don't see an article here. Please post the article otherwise you look to be possibly spreading false information. [Deny]
p10: Yes, where is the proof. It is probably none of our business! [Deny]
p11: Please be true👏👏👏👏 [Deny]
p12: "This is the beginning of the end"😂😂 8 years now🤪 [Comment]
p13: Source? Where are these emails and have they been made public? . [Question]
p14: I didn't think proof was required nowadays. You just keep saying it over and over and get louder every time. Sooner or later it will be a known fact. [Comment]
p15: I, personally, prefer proof over speculation or innuendo. [Question]
p16: Did you see proof of the stolen election? Did you accept that trump lies even though there's evidence everywhere to the contrary? "We have lots of evidence", yet nothing has been produced till now? [Support]

Figure 1: A false rumor claim and the stances of responding posts. The arrow lines indicate the direction of inference for claim veracity and posts stances.

Early research has found that detecting stances expressed in related conversation threads on social media is beneficial for rumor detection and monitoring (Qazvinian et al., 2011; Zhao et al., 2015; Zubiaga et al., 2016b). Nowadays, researchers are increasingly focusing on leveraging related stances for improving rumor detection and verification effectiveness (Dungs et al., 2018; Ma et al., 2020; Li et al., 2020; Tian et al., 2020; Yuan et al., 2021) as well as performing stance-rumor joint detection tasks through multi-task learning (Ma et al., 2018; Kochkina et al., 2018; Wei et al., 2019b; Li et al., 2019; Yu et al., 2020; Yang et al., 2022).

However, most existing approaches require extensive post-level stance labels and claim-level veracity labels for training stance detection and rumor verification models, respectively, which are very expensive to obtain. While some "for free" unsupervised (Kobbe et al., 2020; Allaway and McKeown, 2020a; Pick et al., 2022; Li et al., 2023) meth-

[*] Jing Ma is the corresponding author.

Findings of the Association for Computational Linguistics: ACL 2024, pages 13423–13439
August 11-16, 2024 ©2024 Association for Computational Linguistics

ods were developed, they are only for the stance detection task and suffer from poor generalizability due to reliance on crafted features or specific models pre-trained on the data (e.g., online debate) unmatched with rumor stances from social media. More recently, a weakly supervised method (Yang et al., 2022) based on multiple instance learning (Dietterich et al., 1997; James Richard Foulds, 2010) is proposed to predict the stance of an individual post and the rumor veracity of the claim only being supervised with claim veracity labels. However, determining post stances via dispersing the veracity of a source claim down to stances of many individual posts, as illustrated by Figure 1, is intuitively much more challenging than inferring a source claim's veracity via aggregating post stances in an opposite direction.

The emergence of disruptive technologies for Large Language Models (LLMs) such as ChatGPT (Brown et al., 2020) and smaller variants such as Llama (Touvron et al., 2023) and ALPACA (Taori et al., 2023) have shown performance and explainability on a par with human or even better in various NLP tasks. However, their abilities in rumor-stance related tasks within social media context are still understudied, especially for the joint prediction for detecting posts stances and debunking rumors from claims at the same time. On one hand, LLMs might suit such tasks thanks to their rich pre-trained knowledge and strong zero- and few-shot capabilities; on the other hand, their predictive power based on a large number of rumor-related social media posts might be compromised by noise, unreliability, and lack of supervision.

To study and unleash the potential of LLMs for this joint task, we propose a reinforcement tuning framework for joint stance detection and rumor verification (**JSDRV**) based on LLMs[1]. The framework contains three complementary parts: the LLM stance detection (SD) network, the reinforcement label selector, and the LLM rumor verification (RV) network. Specifically, assuming merely a small set of seeding veracity labels at claim level, the reinforcement selector chooses high-quality examples for fine-tuning the SD and RV LLMs based on their generated labels and explanations. We present an end-to-end joint optimization mechanism to boost the integrated framework. Our contributions are summarized as follows:

- We propose a novel LLM-based reinforcement tuning framework to detect stance and verify rumor veracity jointly starting off with a small set of seed instances labeled by humans.

- Our JSDRV framework is generic, which can not only accommodate open or closed LLMs, but also non-LLM-based models as stance detection and rumor verification networks.

- Extensive experiments on multiple benchmark rumor stance datasets demonstrate that JSDRV outperforms a range of strong baselines, including pretrained language models and fully supervised models on both tasks.

2 Related Work

Rumor Verification. Early studies on rumor verification train supervised classifiers by utilizing content (Yang et al., 2012; Liu et al., 2015) or contextual features (Zhao et al., 2015; Zubiaga et al., 2017) extracted from claim and related posts from social media. Nowadays, most methods predominately focus on utilizing neural networks such as RNN (Ma et al., 2016), CNNs (Yu et al., 2017), tree-/graph-based (Lu and Li, 2020; Ma et al., 2020; Rosenfeld et al., 2020; Lin et al., 2021), transformer-based (Khoo et al., 2020; Ma and Gao, 2020; Yu et al., 2020) and adversarial contrastive (Ma et al., 2021; Lin et al., 2022) models. Researchers also find that stances towards the specific claim shared among social media users, can assist rumor verification by revealing crucial cues and dissemination patterns (Qazvinian et al., 2011; Zhao et al., 2015; Zubiaga et al., 2016b), leading to more recent stance-aware approaches for rumor verification (Dungs et al., 2018; Ma et al., 2020; Li et al., 2020; Tian et al., 2020; Yuan et al., 2021).

Recently, research has leveraged pre-trained language models (PLMs) including LLMs for misinformation-related tasks such as fact checking (Lee et al., 2021; Pan et al., 2023; Zeng and Gao, 2023; Zhang and Gao, 2023, 2024). Lin et al. (2023) proposed zero-shot prompt learning for rumor detection using a multilingual PLM addressing diverse languages and domains on social media. Little has been done on using LLMs for rumor stance detection and verification in social media contexts.

Stance Detection. Stance detection, initially relies on hand-crafted features (Lukasik et al., 2016; Zubiaga et al., 2018), and later has advanced to use

[1] Code is released at https://anonymous.4open.science/r/JSDRV-F3CE/

deep learning (Augenstein et al., 2016; Zhang et al., 2019; Liang et al., 2021). Subsequent research has explored incorporating propagation structure (Zubiaga et al., 2016a; Kochkina et al., 2017). Recent approaches delve into reinforcement learning (Wei et al., 2019a), contrastive learning (Liang et al., 2022) and teacher-student models (Li et al., 2023). Yet these methods require large annotated corpora for training. To address this limitation, unsupervised (Kobbe et al., 2020; Allaway and McKeown, 2020a; Pick et al., 2022; Ran and Jia, 2023) and weakly supervised model (Yang et al., 2022) have emerged, albeit with concerns about their weak detection efficacy and generalizability. PLMs elevate stance detection by utilizing variants of BERT (Devlin et al., 2019), setting new standards this task (Allaway and McKeown, 2020b; Li et al., 2021a). ChatGPT demonstrates its accuracy in stance detection (Aiyappa et al., 2023) via zero-shot and few-shot prompt engineering.

Rumor-Stance Dual Task. The rumor-stance dual task commenced from RumorEval shared task series (Derczynski et al., 2017; Gorrell et al., 2019) as a two-step pipeline, where rumor verification (subtask B) performs veracity prediction based on the claim and SDQC stances of the posts classified in stance detection (subtask A). Then, joint detection has been studied through multi-task learning (Ma et al., 2018; Kochkina et al., 2018; Wei et al., 2019b; Li et al., 2019; Yu et al., 2020). However, such approach is fully supervised by large training sets labeled for claim veracity and posts stance. Yang et al. (2022) proposed a weakly supervised neural model with multiple instance learning only using a full set of veracity-labeled claims for training. We assume only a small set of training examples at claim level as seeds for LLM-based annotation, and learn a policy to select high-quality annotations for fine-tuning SD and RV LLMs.

3 Problem Statement

We define a rumor dataset as $\mathcal{C} = \{(c_i, X_i)\}_{i=1}^{|\mathcal{C}|}$, where each instance (c_i, X_i) is a tuple consisting of a source claim c_i and a conversation thread of posts responding to c_i denoted as $X_i = \{x_{i,1}, x_{i,2}, \cdots, x_{i,T}\}$. The posts are presented in a *chronological* order while reply structure may exist via '@user' symbol in the text. We define the dual tasks as follows:

- **Stance Detection:** The task is to determine the stance $y_{i,j}$ for each post $x_{i,j} \in X_i$ under claim c_i. That is, $f : x_{i,1}x_{i,2} \cdots x_{i,T} \rightarrow y_{i,1}y_{i,2} \cdots y_{i,T}$, where $y_{i,j}$ takes one of the Support (S), Deny (D), Question (Q) or Comment (C) stance labels.

- **Rumor Verification:** The task is to classify each claim c_i together with the responding posts into one of the four veracity classes Y_i: Non-Rumor (N), True Rumor (T), False Rumor (F), or Unverified Rumor (U). That is, $g : (c_i, X_i) \rightarrow Y_i$.

Traditionally, the ground-truth of $y_{i,1}y_{i,2} \cdots y_{i,T}$ and Y_i of all training instances are assumed available for full supervision (Ma et al., 2018; Kochkina et al., 2018; Wei et al., 2019b; Li et al., 2019; Yu et al., 2020), or only Y_i of each training instance is available for weak supervision (Yang et al., 2022). In contrast, we target a more challenging setting, where only a small set of seeding claims $\mathcal{C}' \in \mathcal{C}$ ($|\mathcal{C}'| \ll |\mathcal{C}|$) are provided with veracity labels, while no post stance is provided for training.

4 Our JSDRV Framework

Using zero-/few-shot prompting (Radford et al., 2019; Brown et al., 2020) and parameter-efficient fine-tuning (PEFT) (Hu et al., 2021; Houlsby et al., 2019; Lester et al., 2021) techniques, LLMs can perform on a par with or even better than traditional supervised models. However, using LLMs for the stance-rumor joint tasks faces two major issues: 1) Human labels, especially stance labels of posts in social conversation about a claim, are difficult to obtain. Even if they were abundant, it could be hard to fully utilize them for fine-tuning due to the limit of computing resources; 2) While one could use LLMs to annotate data in scale, LLM-provided labels may be unreliable, subject to further refinement for quality. Thus, how to label and select high-quality instances becomes paramount for effective prompting or fine-tuning.

We propose a reinforcement tuning framework to perform data annotation, selection, and model fine-tuning for the dual tasks, based on a small set of seeding claims with veracity labeled manually. By iteratively refining the selection policy with LLM's annotations, the model prioritizes instances that align with the desired task objectives. With heterogeneous reward functions based on limited ground-truth data, JSDRV enhances the instance selection process, ultimately boosting the overall performance of both tasks.

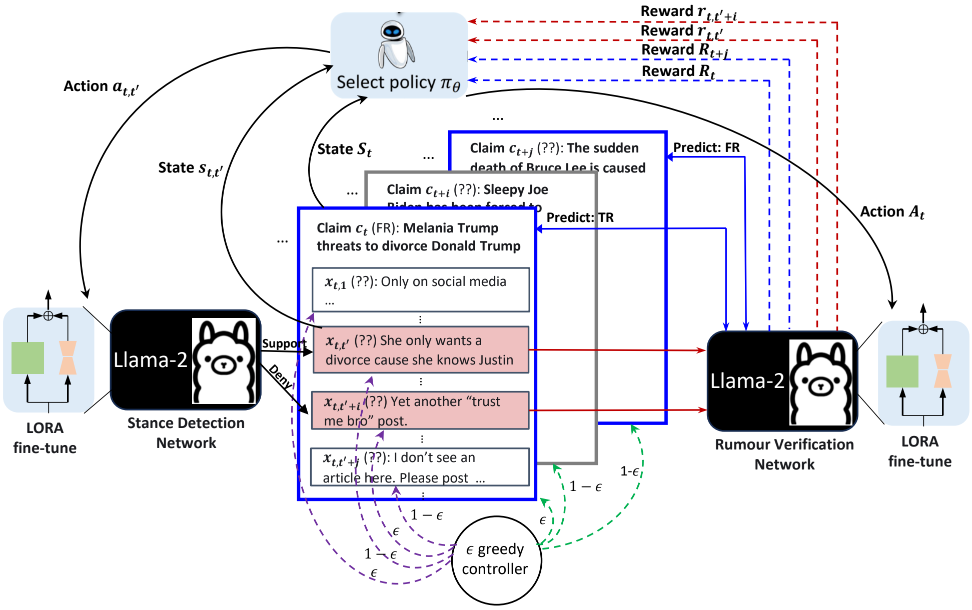

Reward $r_{t,t'+i}$
Reward $r_{t,t'}$
Reward R_{t+j}
Reward R_t
Select policy π_θ
Action $a_{t,t'}$
State $s_{t,t'}$
State S_t
Action A_t
Claim c_{t+j} (??): The sudden death of Bruce Lee is caused
Predict: FR
Claim c_{t+i} (??): Sleepy Joe Biden has been forced to
Predict: TR
Claim c_t (FR): Melania Trump threats to divorce Donald Trump
$x_{t,1}$ (??): Only on social media …
$x_{t,t'}$ (??) She only wants a divorce cause she knows Justin
$x_{t,t'+i}$ (??) Yet another "trust me bro" post.
$x_{t,t'+j}$ (??): I don't see an article here. Please post …
Support
Deny
Llama-2
Stance Detection Network
LORA fine-tune
Llama-2
Rumour Verification Network
LORA fine-tune
ϵ greedy controller
$1-\epsilon$
ϵ
1-ϵ

Figure 2: Our reinforcement fine-tuning framework for joint stance detection and rumor verification.

Figure 2 illustrates our end-to-end framework, encompassing an SD network, a reinforcement selection policy model, and an RV network. We adopt Llama-2 7B version[2] (Touvron et al., 2023), an open-source LLM for both SD and RV networks, which can be fine-tuned with PEFT techniques such as LoRA (Hu et al., 2021).

4.1 LLM Stance Detection Network

The purpose of the LLM-based SD network is to provide a stance label to any post in X_t of a given claim c_t. Since the size of X_t can be very large, it is neither necessary nor feasible to use all the posts due to the potential high costs of interaction with LLMs or fine-tuning. We use an ϵ-greedy strategy to pre-select a subset of posts for LLM to annotate before learning our selector policy model (§ 4.3), aiming to facilitate the continuous iterative optimization of the stance LLM.

ϵ-greedy Controller for Post Pre-Selection We propose an ϵ greedy controller to judiciously regulate the quantity of posts submitted for LLM prediction, thereby balancing the exploitation, i.e., choosing posts from X_t by following the time order of posts, and exploration, i.e., selecting posts in X_t randomly. The intuition is that the model strikes to balance the earliness and time span for getting useful posts. Thus, a post $\tilde{x}_{t,t'}$ is sampled at any step t' based on the following trade-off:

$$\tilde{x}_{t,t'} \sim \begin{cases} \text{Next}(X_t) & \text{with prob } \epsilon \\ \text{Uniform}(X_t) & \text{with prob 1- } \epsilon \end{cases} \tag{1}$$

where X_t is a temporally ordered list of posts as aforementioned, Next$(.)$ is a function of choosing the next post in the list, and Uniform$(.)$ is a uniform distribution over all posts in X_t. At each step t', we sample a post without replacement from one of the two distributions and feed it into SD LLM for prompting and label generation.

Stance Prompt Learning We design a prompt as the instruction to make the SD LLM generate a stance label and a brief explanation for each sampled post, following a required format. An example prompt is shown in Figure 4 in Appendix A.1.

Pretraining For a better initial quality of labeling, we pre-train the SD LLM with P-stance dataset[3] (Li et al., 2021b). The objective is to minimize the negative conditional language modeling for generating correct stance labels.

4.2 LLM Rumor Verification Network

Posts after pre-selection and annotation by SD network will be further selected by the policy model (§ 4.3) for determining whether they should be retained or discarded, considering how useful they

[2] While we opt to host local, open-sourced LLMs, our framework is general and can accommodate closed LLMs like GPT-3.5/4 from OpenAI.

[3] P-stance is a general stance dataset with 3-category stances. We find it still helps SD LLM in 4-class stance classification in our task.

are to claim veracity prediction by the RV network. The RV network also needs to augment labeled instances at claim level, for which a pre-selection of claims is conducted using ϵ-greedy method. Since we have some human labeled claims, the sampling of a claim $\tilde{c}_t$ at step t strives to balance exploiting human labeled claims and exploring unlabeled claims both uniformly, which yields:

$$\tilde{c}_t \sim \begin{cases} \text{Uniform}(\mathcal{C}') & \text{with prob } \epsilon \\ \text{Uniform}(\mathcal{C} - \mathcal{C}') & \text{with prob 1- } \epsilon \end{cases} \quad (2)$$

where $\mathcal{C}'$ denotes the labeled claim set.

Given each pre-selected claim and its related posts that are retained by the selection policy model, RV LLM network is then prompted to generate veracity label for the claim and a brief explanation of the decision, considering the claim and posts content and stance. An example prompt is shown in Appendix A.2.

Pretraining Similarly, we also pretrain the RV LLM for better initial quality using the small manually labeled claim set $\mathcal{C}'$. Without post stance labels, we just feed posts content with the claim into the RV LLM, which is trained to minimize the negative likelihood of predicted labels and ground truth.

4.3 Selector Policy

We design a selector policy to transform input states, i.e., annotated instances that are pre-selected, to their corresponding actions, i.e., decisions to discard or retain an instance. For optimizing the selector's policy π_θ with parameter θ, each step corresponds to sampling an annotated claim followed by a sequence of sub-steps sampling annotated posts, and receiving rewards from the RV network.

We formulate a two-level Markov Decision Process (MDP) with the following elements: (1) $\{S_t\}$ and $\{s_{t,t'}\}$ correspond to a sequence of states at claim and post level, respectively; (2) $\{A_t\}$ and $\{a_{t,t'}\}$ correspond to a sequence of actions for sampled claims and posts, respectively, for deciding whether to keep the current instance; (3) $\{R_t\}$ and $\{r_{t,t'}\}$ are rewards received after taking an action for the respective level. This reward incorporates predictions based on instances with and without human labels. For easing presentation, we will use a unified notation to describe these elements at both levels, that is, ς, α, γ, and τ denote state, action, reward, and time step, respectively, but note that in practice each of them is separated into two versions corresponding to claim and post as mentioned.

State State ς_τ denotes the current status at step τ after the previous instances, i.e., claims or posts, sampled up to $\tau - 1$. It contains the representations of three parts: the current claim $\tilde{c}_\tau$, the selected instances thus far, and the prediction explanation for the current instance. This yields $\varsigma_\tau = [\tilde{\mathbf{c}}_\tau, \mathbf{C}_{\tau-1}, \mathbf{E}_\tau]$, where $\tilde{\mathbf{c}}_\tau$, $\mathbf{C}_{\tau-1}$ and $\mathbf{E}_\tau$ denote the embedding of $\tilde{c}_\tau$, the context by averaging the embeddings of selected instances up to $\tau - 1$, and the embedding of explanation for the current instance, respectively. The embeddings are obtained through RoBERTa (Liu et al., 2019).

Action An action α_τ is sampled from π_θ stochastically given the state ς_τ. Specifically, the policy network will output a probability distribution over action space {*discard*, *retain*}, which yields:

$$\pi_\theta(\alpha_\tau, \varsigma_\tau) = \sigma(w_2 \cdot \text{ReLU}(w_1 \cdot \varsigma_\tau)) \quad (3)$$

where $\theta = \{w_1, w_2\}$ are the weights of policy network and σ is the sigmoid activation function. Then, the action α_τ is sampled according to the output probability: $\alpha_\tau \sim \pi_\theta(\alpha_\tau, \varsigma_\tau)$.

Reward For a selected claim with human label, the reward considers congruence between the RV network prediction and its ground truth, since the contribution of each selected post under the claim can be reflected by claim veracity prediction; For a selected claim without ground truth, the reward considers how well the stance distribution of its sampled posts conforms to the posts stance distribution of those claims in $\mathcal{C}'$ that have the same veracity label as the predicted label of the selected claim. This yields:

$$\gamma_\tau = \begin{cases} \mathbb{E}(\cos(\hat{\mathbf{Y}}_{\tilde{c}_\tau}, \mathbf{Y}_{\tilde{c}_\tau})) & \text{if } \tilde{c}_\tau \in \mathcal{C}' \\ \mathbb{E}(\cos(\overline{\hat{\mathbf{y}}_{\tilde{c}_\tau, \tilde{x}_{\tau,*}}}, \overline{\hat{\mathbf{y}}_{c \in \mathcal{C}', x_{c,*}}} | \hat{Y}_{\tilde{c}_\tau} = Y_c)) & \text{otherwise} \end{cases} \quad (4)$$

Here, $\hat{\mathbf{Y}}_{\tilde{c}_\tau}$ and $\mathbf{Y}_{\tilde{c}_\tau}$ are respectively the predicted and ground-truth veracity distributions of sampled claim $\tilde{c}_\tau$; $\overline{\hat{\mathbf{y}}_{\tilde{c}_\tau, \tilde{x}_{\tau,*}}}$ is the mean of stance distributions predicted on the selected posts $\tilde{x}_{\tau,*}$ under $\tilde{c}_\tau$; $\overline{\hat{\mathbf{y}}_{c \in \mathcal{C}', x_{c,*}}}$ is the mean of stance distributions predicted on all the posts $x_{c,*}$ over all the claims $c \in \mathcal{C}'$, of which the veracity label is same as the predicted veracity of $\tilde{c}_\tau$; and $\mathbb{E}(.)$ is a sign function that turns the cosine similarity of two distributions to -1, 0 or 1 depending on the sign of similarity. The reward encourages the model to retain the instances (i.e., claims and posts) that can help veracity prediction keep close to the human-labeled claims providing similar stance distribution.

4.4 Model Training

Our training is an end-to-end joint optimization process, which involves alternating training on policy network and LLM-based SD and RV networks in each epoch.

Policy Network We employ the widely used offline optimization method (Sutton and Barto, 2018) to maximize expected accumulative reward $\mathcal{R}_t$:

$$\mathcal{R}_t = \frac{1}{t}\sum_{i=1}^{t}\Bigg(R_i \log(\pi_\theta(A_i, S_i)) + \frac{1}{t'}\sum_{j=1}^{t'} r_{i,j}\log(\pi_\theta(a_{i,j}, s_{i,j}))\Bigg) \tag{5}$$

where R_i and $r_{i,j}$ are calculated by Equation 4, and t and t' respectively denote the current time step at claim and post levels. The policy network is updated after a claim and its posts are selected.

LLMs We fine-tune the LLM-based SD and RV networks using the annotations (human or machine labeled) of selected instances by minimizing standard Negative Log Likelihood loss in language model training (Kanamori, 2010). In each epoch, the two LLMs are fine-tuned only once following the selection process. Note that we can skip fine-tuning them but use the selected instances for few-shot in-context learning.

Training Procedure The training detail is depicted as Algorithm 1. JSDRV is trained with an end-to-end fashion by optimizing the two-level selector policy model for claim selection and post selection and the SD and RV LLMs.

Termination condition The ϵ greedy process can terminate automatically. For the termination condition, we utilize the reward function described in Equation 4. The selection process stops if the model receives reward $\gamma_\tau = 1$ for N continuous steps, for which we set N as 100 tuned on the validation set.

5 Experiments and Results

5.1 Datasets and Setup

Datasets For model training, we utilize three public rumor verification benchmark datasets, Twitter15/16 (T15/16) (Ma et al., 2017) and PHEME (PH)[4], where only claim veracity labels are provided. Since both stance label and rumor veracity are required for testing, we resort to RumorEval-S[5] (Yang et al., 2022) and SemEval-8 (Derczynski et al., 2017) datasets with labeled claim veracity and posts stance. Details of the datasets are provided in Table 4 and Table 5 in Appendix B.

Setup We hold out 20% instances of test sets as validation sets (Val), for training models that require stance labels and tuning hyper-parameters. We use micro-averaged, macro-averaged F1 score, and class-specific F-measure as evaluation metrics considering imbalanced class distributions (Zubiaga et al., 2016a). We use public codes for baselines or re-implement them if codes are not released. We use 50% claims in the training set as seeding claims for JSDRV for the main results and examine how the percentage affects its performance (see Appendix E.2). We set $\epsilon = 0.3$ with validation data. The detail of other hyper-parameters setup is provided in Appendix C.

5.2 Stance Detection Performance

Since our SD network does not need post stance label, we choose the following unsupervised, supervised, and weakly supervised baselines, which are detailed in Appendix D.1: (1) **TGA** (Allaway and McKeown, 2020a); (2) **BerTweet** (Nguyen et al., 2020); (3) **Llama 2-ST** (Touvron et al., 2023); (4) **Llama 2-MT** (Touvron et al., 2023); (5) **BiGRU** (Augenstein et al., 2016); (6) **BrLSTM** (Kochkina et al., 2017); (7) **MT-GRU** (Ma et al., 2018); (8) **JointCL** (Liang et al., 2022); (9) **SRLF** (Yuan et al., 2021); (10) **TD-MIL** (Yang et al., 2022). We use **model (DATASET)** to denote **model** trained on DATASET[6].

For the stance detection baselines, we use the original source code of TGA. The first group refers to unsupervised baselines, while BiGRU, BrLSTM, MT-GRU, and JointCL in the second group are four popular supervised stance detection baselines. SRLF and TD-MIL are weakly supervised models that do not need stance annotation. We use the validation dataset to train the JSDRV (Val) variant for fair comparison with the supervised models. This is because there is no stance annotations in the training set.

[4] https://figshare.com/articles/PHEME_dataset_of_rumours_and_non-rumours/4010619.

[5] https://github.com/2302Jerry/Data-Repo

[6] We also train JSDRV on the validation set for fairly compared with supervised methods that need stance labels.

Algorithm 1 The JSDRV Model Training

```
Input: A rumor dataset $\mathcal{C}$, seeding claims $\mathcal{C}' \in \mathcal{C}$, P-stance dataset, SD network, RV network, $\epsilon$, $\pi_\theta$ policy network.
Pretrain SD network with P-stance dataset according to § 4.1, and RV network with seeding claims according to § 4.2.
for $t \leftarrow 1$ to $|\mathcal{C}|$ do
    if termination condition meets then
        break.
    end if
    Sample a claim $\tilde{c}_t$ according to Equation 2.
    for $t' \leftarrow 1$ to $|X_{\tilde{c}_t}|$ do
        if termination condition meets then
            break
        end if
        Sample a post according to Equation 1.
        Predict post stance according to § 4.1.
        Calculate post-level reward according to Equation 4.
    end for
    Predict claim veracity according to § 4.2.
    Calculate claim-level reward according to Equation 4.
    Update parameters of $\pi_\theta$ based on Equation 5.
    Fine-tune SD network (w/ LoRA).
    Fine-tune RV network (w/ LoRA).
end for
return SD network, $\pi_\theta$, RV network.
```

Dataset	RumorEval-S						SemEval-8					
Method	MicF	MacF	S	D	Q	C	MicF	MacF	S	D	Q	C
			F1	F1	F1	F1			F1	F1	F1	F1
TGA	-	0.324	0.301	0.168	0.342	0.486	0.383	0.344	0.278	0.162	0.480	0.456
BerTweet	0.619	0.492	0.497	0.203	0.513	0.753	0.611	0.428	0.512	0.131	0.326	0.742
Llama 2-ST	0.630	0.500	0.501	0.203	0.532	0.763	0.631	0.471	0.533	0.138	0.472	0.740
Llama 2-MT	0.632	0.500	0.502	0.199	0.533	0.766	0.630	0.473	0.534	0.142	0.471	0.742
BiGRU (Val)	0.630	0.417	0.392	0.162	0.360	0.754	0.633	0.416	0.460	0.168	0.328	0.708
BrLSTM (Val)	0.660	0.420	0.460	0.000	0.391	0.758	0.665	0.401	0.493	0.000	0.381	0.730
MT-GRU (Val)	0.636	0.432	0.313	0.156	0.506	0.748	0.630	0.413	0.498	0.116	0.312	0.729
JointCL (Val)	0.639	0.505	0.532	0.210	0.516	0.760	0.640	0.475	0.536	0.136	0.478	0.751
SRLF (PH)	0.606	0.479	0.492	0.280	0.468	0.676	0.510	0.393	0.328	0.205	0.420	0.619
TD-MIL (PH)	0.691	0.434	0.344	0.179	0.467	0.767	0.651	0.426	0.335	0.175	0.430	0.763
JSDRV-Bert (Val)	0.668	0.541	0.531	0.316	0.562	0.755	0.658	0.489	0.540	0.173	0.482	0.761
JSDRV-Bert (T15)	0.680	0.562	0.534	0.380	0.570	0.764	0.671	0.498	0.549	0.169	0.490	0.784
JSDRV-Bert (T16)	0.681	0.560	0.527	0.381	0.573	0.759	0.673	0.500	0.550	0.170	0.490	0.790
JSDRV-Bert (PH)	0.683	0.565	0.535	0.383	0.573	0.765	0.780	0.502	0.556	0.170	0.493	0.789
JSDRV (Val)	0.672	0.550	0.536	0.310	0.576	0.779	0.673	0.496	0.542	0.168	0.483	0.790
JSDRV (T15)	0.696	0.576	0.535	0.383	0.586	**0.801**	0.693	0.506	0.558	0.170	0.496	0.798
JSDRV (T16)	0.697	0.574	0.536	0.380	0.580	0.798	0.696	0.507	0.560	0.173	0.498	0.796
JSDRV (PH)	**0.723**	**0.605**	**0.546**	**0.476**	**0.595**	**0.801**	**0.705**	**0.522**	**0.563**	**0.216**	**0.506**	**0.801**

Table 1: Stance detection results. JSDRV models use 50% claims in training sets as seeding claims.

From Table 1, we observe that: 1) In zero-shot models, TGA performs worst as it is pretrained on specific topics and cannot generalize well to Twitter data; BerTweet, which is fine-tuned on enormous

Twitter datasets, outperforms TGA; Llama 2-ST and -MT outperform BerTweet, indicating that the LLM has promising zero-shot capability to detect stances. 2) In fully supervised baselines, Bi-GRU based on a sequential architecture performs worst; BrLSTM, benefiting from propagation structure, makes improvements, but is unable to detect deny stance as it is sensitive to such infrequent class in the training data; JointCL leveraging both context- and target-aware features outperforms MT-GRU which is sequential. 3) While both TD-MIL and SRL are weakly supervised only using claim labels, TD-MIL benefits from propagation information while SRLF cannot. 4) Trained on the validation set, JSDRV (Val) and its BERT-based variant are comparable to TD-MIL trained on the full training set, indicating JSDRV is less demanding on labeled data.

JSDRV outperforms all the baselines on the corresponding datasets. When LLM is replaced by BERT in JSDRV, there is a performance drop but JSDRV-Bert is still superior to baselines, suggesting it can be generalized to non-LLM task models. JSDRV (PH) performs the best, beating its counterparts trained on T15/16 datasets with large margins due to the much larger size of PH dataset.

5.3 Rumor Verification Performance

We compare to unsupervised, supervised, weakly supervised, and multi-task rumor verification baselines: (1) **BerTweet** (Nguyen et al., 2020); (2) **Llama 2-ST** (Touvron et al., 2023); (3) **Llama 2-MT** (Touvron et al., 2023); (4) **GCAN** (Lu and Li, 2020); (5) **TD-RvNN** (Ma et al., 2020); (6) **PLAN** (Khoo et al., 2020); (7) **DDGCN** (Sun et al., 2022); (8) **SRLF** (Yuan et al., 2021); (9) **TD-MIL** (Yang et al., 2022); (10) **MTL2** (Kochkina et al., 2018); (11) **MT-GRU** (Ma et al., 2018). The details are depicted in Appendix D.2. In Table 2, we report the best rumor verification results obtained across different training and test datasets. MT-GRU and MTL2 require both rumor and stance labels for training. So, we train them on the validation set, which has both rumor and stance labels. We also compare JSDRV (Val) with MT-GRU and MTL2.

From Table 2, we observe a similar trend as the stance detection task. While Llama 2-ST and -MT outperform BerTweet, their performance is still on a par, indicating directly prompting Llama 2 to perform both tasks together is similar to doing them separately. Among supervised baselines, GCAN performs worst because it only considers local structure while structure-aware models such as TD-RvNN, PLAN and DDGCN appear much better. MTL2 and MT-GRU are multi-task frameworks that require both labels, so we trained them on validation sets. JSDRV (Val), trained on the same dataset only using a small set of veracity-labeled claims, outperforms MT-GRU on MacF and class-level F1, indicating that JSDRV is effective in multi-task learning.

Similarly, JSDRV outperforms all the baselines on the corresponding datasets for the RV task, and is still well generalized to the non-LLM-based task models with the LLMs replaced by BERT. JSDRV (PH) outperforms all the baselines regardless of datasets used and gets 3.2%/17.1% MicF/MacF improvement over the best baseline TD-MIL (PH) on RumorEval-S.

5.4 Ablation Study

For the ablation, we separate the components in JSDRV. (**RSS**): reinforced stance selector; (**RVS**): reinforced veracity selector; (**FTSD**): fine-tuning SD; (**FTRV**): fine-tuning RV; (**PTSD**): pretraining SD; (**PTRV**): pretraining RV; (**epSD**): ϵ greedy control for SD; (**epRV**): ϵ greedy control for RV.

Table 3 shows that except removing **epSD** and **epRV**, removing other components decreases the performance, indicating they are useful and the pre-selection with ϵ greedy method can maintain performance with a little cost. Removing **RSS** and **RVS** drops the most, meaning that the reinforced selector is the most vial. Removing **FTRV** and **FTSD** drops more than removing **PTSD** and **PTRV** respectively, indicating that joint fine-tuning is more important.

5.5 Analysis

We plot posts with their stances on RumorEval-S dataset with tSNE (Van der Maaten and Hinton, 2008) to show the effect of post stance selection. We input each post as "[CLS] post [SEP] stance [SEP] reason" into RoBERTa-base and take [CLS] token representations to plot Figure 3. We observe that the selected stances are separated much better, indicating that JSDRV can differentiate stances and may help rumor verification.

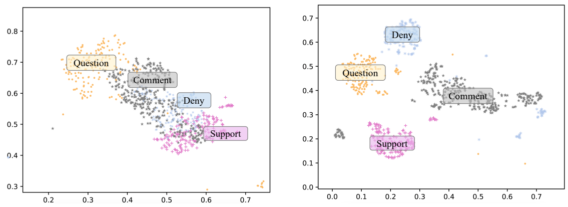
We also plot example outputs of JSDRV in Figure 6 in Appendix E.1 and find that the selector policy can choose posts with high-quality stances annotated by stance LLM, which provides useful

Dataset	RumorEval-S						SemEval-8				
Method	MicF	MacF	T	F	U	N	MicF	MacF	T	F	U
			F1	F1	F1	F1			F1	F1	F1
BerTweet	0.760	0.452	0.641	0.293	0.367	0.460	0.755	0.427	0.630	0.256	0.395
Llama 2-ST	0.754	0.450	0.660	0.271	0.400	0.469	0.746	0.424	0.632	0.260	0.380
Llama 2-MT	0.758	0.465	0.678	0.301	0.403	0.478	0.756	0.427	0.635	0.263	0.382
GCAN (PH)	0.645	0.253	0.249	0.310	0.113	0.339	0.645	0.255	0.241	0.326	0.198
TD-RvNN (PH)	0.753	0.677	0.755	0.666	0.673	0.615	0.748	0.694	0.712	0.617	0.753
PLAN (PH)	0.800	0.743	0.819	0.760	0.780	0.612	0.794	0.720	0.741	0.694	0.726
DDGCN (PH)	0.759	0.663	0.713	0.663	0.669	0.607	0.755	0.685	0.709	0.624	0.723
SRLF (PH)	0.742	0.447	0.667	0.290	0.381	0.452	0.742	0.423	0.635	0.249	0.386
TD-MIL (PH)	0.809	0.776	0.826	0.659	0.669	0.852	0.798	0.741	0.741	0.672	0.810
MTL2 (Val)	0.653	0.430	0.622	0.279	0.352	0.457	0.651	0.433	0.640	0.289	0.372
MT-GRU (Val)	0.768	0.452	0.662	0.298	0.373	0.457	0.761	0.428	0.639	0.254	0.391
JSDRV-Bert (Val)	0.752	0.580	0.758	0.488	0.500	0.574	0.750	0.577	0.746	0.493	0.492
JSDRV-Bert (T15)	0.803	0.770	0.800	0.750	0.753	0.777	0.783	0.730	0.758	0.701	0.731
JSDRV-Bert (T16)	0.810	0.776	0.802	0.751	0.751	0.800	0.785	0.732	0.760	0.710	0.726
JSDRV-Bert (PH)	0.813	0.780	0.810	0.762	0.756	0.792	0.796	0.746	0.779	0.712	0.747
JSDRV (Val)	0.763	0.592	0.765	0.486	0.510	0.612	0.756	0.579	0.749	0.492	0.495
JSDRV (T15)	0.828	0.786	**0.830**	0.759	0.788	0.762	0.829	0.755	0.769	0.731	0.766
JSDRV (T16)	0.838	0.786	0.829	**0.770**	0.782	0.765	0.830	0.768	0.800	0.734	0.770
JSDRV (PH)	**0.842**	**0.804**	0.829	**0.774**	**0.824**	**0.787**	**0.834**	**0.784**	**0.820**	**0.741**	**0.792**

Table 2: Rumor verification results. JSDRV models use 50% claims in training sets as seeding claims.

Method	Rumor Verification		Stance Detection	
	MicF	MacF	MicF	MacF
JSDRV	0.842	0.804	0.723	0.605
- RSS	0.820	0.783	0.700	0.571
- RVS	0.826	0.790	0.705	0.573
- FTSD	0.828	0.790	0.709	0.578
- FTRV	0.830	0.793	0.710	0.583
- PTSD	0.831	0.794	0.712	0.589
- PTRV	0.833	0.795	0.714	0.600
- epSD	0.844	0.805	0.725	0.606
- epRV	0.843	0.806	0.723	0.605

Table 3: Ablation study on RumorEval-S dataset.



(a) Original (b) Selected by JSDRV

Figure 3: Visualization of stance distributions of original and selected posts.

cues to rumor verification. More analyses on the impact of seeding claims and ϵ are in Appendix E.2.

6 Conclusion and Future Work

We introduced a novel reinforcement tuning framework, JSDRV, for joint stance detection and rumor verification tasks based on LLMs. By leveraging LLMs as annotators and employing a reinforcement label selector, we effectively fine-tune the LLMs for both tasks supervised only with minimal human-labeled claims. Our experiments on benchmark datasets demonstrated the superiority of JSDRV over existing methods, suggesting its promising potential for addressing the challenges of joint stance detection and rumor verification in social media environments. Our future work aims to develop unsupervised methods for both tasks.

Limitations

JSDRV relies on the SD and RV networks to provide class probability distributions for calculating rewards. While some LLMs like Llama 2 can return word distribution, converting it to the distribution of a limited number of classes may compromise accuracy. Moreover, LLMs like ChatGPT, which do not return such distributions, could hinder the applicability and practical use of our method.

In JSDRV, frequent interactions with SD and RV networks, particularly when they rely on closed LLM services, can introduce overhead during training. This overhead may impact training efficiency.

Despite demanding much less labeled data, JS-DRV necessitates human-labeled claims, posing challenges in obtaining them, especially in certain domains such as healthcare and science. This constraint may limit its applicability in scenarios where labeled claims are scarce. Therefore, our future work aims to develop unsupervised methods for both tasks.

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).

Ethical Considerations

While our research utilizes publicly accessible datasets, using social media conversations for debunking rumors and detecting user stances may raise privacy concerns. While our approach does not necessitate access to sensitive user data, we anonymized all social media posts, ensuring user information is invisible and unusable by others.

Both tasks studied in this paper hold social implications. Key considerations include system reliability and the potential for mislabeling information and misleading users. To address this, we will establish responsible policies for code dissemination, aligning with ethical standards.

References

Rachith Aiyappa, Jisun An, Haewoon Kwak, and Yong-yeol Ahn. 2023. Can we trust the evaluation on ChatGPT? In Proceedings of the 3rd Workshop on Trustworthy Natural Language Processing (TrustNLP 2023), pages 47–54, Toronto, Canada. Association for Computational Linguistics.

Abeer ALDayel and Walid Magdy. 2021. Stance detection on social media: State of the art and trends. Information Processing & Management, 58(4):102597.

Emily Allaway and Kathleen McKeown. 2020a. Zero-shot stance detection: A dataset and model using generalized topic representations. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 8913–8931.

Emily Allaway and Kathleen McKeown. 2020b. Zero-Shot Stance Detection: A Dataset and Model using Generalized Topic Representations. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 8913–8931, Online. Association for Computational Linguistics.

Isabelle Augenstein, Tim Rocktäschel, Andreas Vlachos, and Kalina Bontcheva. 2016. Stance detection with bidirectional conditional encoding. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pages 876–885.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. In Advances in Neural Information Processing Systems, volume 33, pages 1877–1901. Curran Associates, Inc.

Lu Cheng, Ruocheng Guo, Kai Shu, and Huan Liu. 2021. Causal understanding of fake news dissemination on social media. In KDD.

Leon Derczynski, Kalina Bontcheva, Maria Liakata, Rob Procter, Geraldine Wong Sak Hoi, and Arkaitz Zubiaga. 2017. SemEval-2017 task 8: RumourEval: Determining rumour veracity and support for rumours. In Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017), pages 69–76, Vancouver, Canada. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Thomas G. Dietterich, Richard H. Lathrop, and Tomás Lozano-Pérez. 1997. Solving the multiple instance problem with axis-parallel rectangles. Artificial Intelligence, 89(1):31–71.

Sebastian Dungs, Ahmet Aker, Norbert Fuhr, and Kalina Bontcheva. 2018. Can rumour stance alone predict veracity? In Proceedings of the 27th International Conference on Computational Linguistics, pages 3360–3370, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Genevieve Gorrell, Elena Kochkina, Maria Liakata, Ahmet Aker, Arkaitz Zubiaga, Kalina Bontcheva, and Leon Derczynski. 2019. SemEval-2019 task 7: RumourEval, determining rumour veracity and support for rumours. In Proceedings of the 13th International Workshop on Semantic Evaluation, pages 845–854, Minneapolis, Minnesota, USA. Association for Computational Linguistics.

Neil Houlsby, Andrei Giurgiu, Stanislaw Jastrzebski, Bruna Morrone, Quentin De Laroussilhe, Andrea Gesmundo, Mona Attariyan, and Sylvain Gelly. 2019. Parameter-efficient transfer learning for NLP. In Proceedings of the 36th International Conference on Machine Learning, volume 97 of Proceedings of Machine Learning Research, pages 2790–2799. PMLR.

Edward J Hu, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen, et al. 2021. Lora: Low-rank adaptation of large language models. In International Conference on Learning Representations.

Eibe Frank James Richard Foulds. 2010. A review of multi-instance learning assumptions. The Knowledge Engineering Review, 25(1):1–25.

Ankur Joshi, Saket Kale, Satish Chandel, and D Kumar Pal. 2015. Likert scale: Explored and explained. British journal of applied science & technology, 7(4):396.

Takafumi Kanamori. 2010. Deformation of log-likelihood loss function for multiclass boosting. Neural Networks, 23(7):843–864.

Ling Min Serena Khoo, Hai Leong Chieu, Zhong Qian, and Jing Jiang. 2020. Interpretable rumor detection in microblogs by attending to user interactions. In AAAI.

Jonathan Kobbe, Ioana Hulpuș, and Heiner Stuckenschmidt. 2020. Unsupervised stance detection for arguments from consequences. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 50–60.

Elena Kochkina, Maria Liakata, and Isabelle Augenstein. 2017. Turing at semeval-2017 task 8: Sequential approach to rumour stance classification with branch-lstm. In Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017), pages 475–480.

Elena Kochkina, Maria Liakata, and Arkaitz Zubiaga. 2018. All-in-one: Multi-task learning for rumour verification. In Proceedings of the 26th International Conference on Computational Linguistics, pages 3402–3413.

Dilek Küçük and Fazli Can. 2020. Stance detection: A survey. ACM Comput. Surv., 53(1).

Nayeon Lee, Yejin Bang, Andrea Madotto, and Pascale Fung. 2021. Towards few-shot fact-checking via perplexity. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 1971–1981, Online. Association for Computational Linguistics.

Brian Lester, Rami Al-Rfou, and Noah Constant. 2021. The power of scale for parameter-efficient prompt tuning. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, pages 3045–3059, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Jiawen Li, Yudianto Sujana, and Hung-Yu Kao. 2020. Exploiting microblog conversation structures to detect rumors. In Proceedings of the 28th International Conference on Computational Linguistics, pages 5420–5429, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Quanzhi Li, Qiong Zhang, and Luo Si. 2019. Rumor detection by exploiting user credibility information, attention and multi-task learning. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 1173–1179, Florence, Italy. Association for Computational Linguistics.

Yingjie Li, Tiberiu Sosea, Aditya Sawant, Ajith Jayaraman Nair, Diana Inkpen, and Cornelia Caragea. 2021a. P-stance: A large dataset for stance detection in political domain. In Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, pages 2355–2365, Online. Association for Computational Linguistics.

Yingjie Li, Tiberiu Sosea, Aditya Sawant, Ajith Jayaraman Nair, Diana Inkpen, and Cornelia Caragea. 2021b. P-stance: A large dataset for stance detection in political domain. In Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, pages 2355–2365.

Yingjie Li, Chenye Zhao, and Cornelia Caragea. 2023. Tts: A target-based teacher-student framework for zero-shot stance detection. In Proceedings of the ACM Web Conference 2023, pages 1500–1509.

Bin Liang, Yonghao Fu, Lin Gui, Min Yang, Jiachen Du, Yulan He, and Ruifeng Xu. 2021. Target-adaptive graph for cross-target stance detection. In Proceedings of the Web Conference 2021, pages 3453–3464.

Bin Liang, Qinglin Zhu, Xiang Li, Min Yang, Lin Gui, Yulan He, and Ruifeng Xu. 2022. Jointcl: A joint contrastive learning framework for zero-shot stance detection. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 81–91.

Hongzhan Lin, Jing Ma, Liangliang Chen, Zhiwei Yang, Mingfei Cheng, and Chen Guang. 2022. Detect rumors in microblog posts for low-resource domains via adversarial contrastive learning. In Findings of the Association for Computational Linguistics: NAACL 2022, pages 2543–2556.

Hongzhan Lin, Jing Ma, Mingfei Cheng, Zhiwei Yang, Liangliang Chen, and Guang Chen. 2021. Rumor detection on twitter with claim-guided hierarchical graph attention networks. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, pages 10035–10047.

Hongzhan Lin, Pengyao Yi, Jing Ma, Haiyun Jiang, Ziyang Luo, Shuming Shi, and Ruifang Liu. 2023. Zero-shot rumor detection with propagation structure via prompt learning. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 37, pages 5213–5221.

Xiaomo Liu, Armineh Nourbakhsh, Quanzhi Li, Rui Fang, and Sameena Shah. 2015. Real-time rumor debunking on twitter. In Proceedings of the 24th ACM international on conference on information and knowledge management, pages 1867–1870.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692.

Yi-Ju Lu and Cheng-Te Li. 2020. Gcan: Graph-aware co-attention networks for explainable fake news detection on social media. In ACL.

Michal Lukasik, PK Srijith, Duy Vu, Kalina Bontcheva, Arkaitz Zubiaga, and Trevor Cohn. 2016. Hawkes processes for continuous time sequence classification: an application to rumour stance classification in twitter. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pages 393–398.

Jing Ma and Wei Gao. 2020. Debunking rumors on twitter with tree transformer. In Proceedings of the 28th International Conference on Computational Linguistics, pages 5455–5466.

Jing Ma, Wei Gao, Shafiq Joty, and Kam-Fai Wong. 2020. An attention-based rumor detection model with tree-structured recursive neural networks. ACM Transactions on Intelligent Systems and Technology (TIST), 11(4):1–28.

Jing Ma, Wei Gao, Prasenjit Mitra, Sejeong Kwon, Bernard J Jansen, Kam-Fai Wong, and Meeyoung Cha. 2016. Detecting rumors from microblogs with recurrent neural networks. In IJCAI.

Jing Ma, Wei Gao, and Kam-Fai Wong. 2017. Detect rumors in microblog posts using propagation structure via kernel learning. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 708–717.

Jing Ma, Wei Gao, and Kam-Fai Wong. 2018. Detect rumor and stance jointly by neural multi-task learning. In Companion proceedings of the the web conference 2018, pages 585–593.

Jing Ma, Jun Li, Wei Gao, Yang Yang, and Kam-Fai Wong. 2021. Improving rumor detection by promoting information campaigns with transformer-based generative adversarial learning. IEEE Transactions on Knowledge and Data Engineering.

Dat Quoc Nguyen, Thanh Vu, and Anh Tuan Nguyen. 2020. Bertweet: A pre-trained language model for english tweets. In EMNLP.

Liangming Pan, Xiaobao Wu, Xinyuan Lu, Anh Tuan Luu, William Yang Wang, Min-Yen Kan, and Preslav Nakov. 2023. Fact-checking complex claims with program-guided reasoning. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 6981–7004, Toronto, Canada. Association for Computational Linguistics.

Ron Korenblum Pick, Vladyslav Kozhukhov, Dan Vilenchik, and Oren Tsur. 2022. Stem: unsupervised structural embedding for stance detection. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 36, pages 11174–11182.

Vahed Qazvinian, Emily Rosengren, Dragomir R. Radev, and Qiaozhu Mei. 2011. Rumor has it: Identifying misinformation in microblogs. In Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, pages 1589–1599, Edinburgh, Scotland, UK. Association for Computational Linguistics.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners. OpenAI blog, 1(8):9.

Hongyan Ran and Caiyan Jia. 2023. Unsupervised cross-domain rumor detection with contrastive learning and cross-attention. In Thirty-Seventh AAAI Conference on Artificial Intelligence, AAAI 2023, pages 13510–13518. AAAI Press.

Jon Roozenbeek and Sander van der Linden. 2019. Fake news game confers psychological resistance against online misinformation. Palgrave Communications, 5(1):1–10.

Nir Rosenfeld, Aron Szanto, and David C Parkes. 2020. A kernel of truth: Determining rumor veracity on twitter by diffusion pattern alone. In Proceedings of The Web Conference 2020, pages 1018–1028.

Mengzhu Sun, Xi Zhang, Jiaqi Zheng, and Guixiang Ma. 2022. Ddgcn: Dual dynamic graph convolutional networks for rumor detection on social media. In Proceedings of the AAAI conference on artificial intelligence, volume 36, pages 4611–4619.

Richard S Sutton and Andrew G Barto. 2018. Reinforcement learning: An introduction. MIT press.

Rohan Taori, Ishaan Gulrajani, Tianyi Zhang, Yann Dubois, Xuechen Li, Carlos Guestrin, Percy Liang, and Tatsunori B Hashimoto. 2023. Alpaca: A strong, replicable instruction-following model. Stanford Center for Research on Foundation Models.

https://crfm. stanford. edu/2023/03/13/alpaca. html, 3(6):7.

Lin Tian, Xiuzhen Zhang, Yan Wang, and Huan Liu. 2020. Early detection of rumours on twitter via stance transfer learning. In Advances in Information Retrieval, pages 575–588, Cham. Springer International Publishing.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. 2023. Llama 2: Open foundation and fine-tuned chat models. arXiv preprint arXiv:2307.09288.

Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-sne. Journal of machine learning research, 9(11).

Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. Science, 359(6380):1146–1151.

Penghui Wei, Wenji Mao, and Guandan Chen. 2019a. A topic-aware reinforced model for weakly supervised stance detection. In Proceedings of the aaai conference on artificial intelligence, volume 33, pages 7249–7256.

Penghui Wei, Nan Xu, and Wenji Mao. 2019b. Modeling conversation structure and temporal dynamics for jointly predicting rumor stance and veracity. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 4787–4798.

Fan Yang, Yang Liu, Xiaohui Yu, and Min Yang. 2012. Automatic detection of rumor on sina weibo. In KDD.

Ruichao Yang, Jing Ma, Hongzhan Lin, and Wei Gao. 2022. A weakly supervised propagation model for rumor verification and stance detection with multiple instance learning. In Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 1761–1772.

Feng Yu, Qiang Liu, Shu Wu, Liang Wang, and Tieniu Tan. 2017. A convolutional approach for misinformation identification. In Proceedings of IJCAI, pages 3901–3907.

Jianfei Yu, Jing Jiang, Ling Min Serena Khoo, Hai Leong Chieu, and Rui Xia. 2020. Coupled hierarchical transformer for stance-aware rumor verification in social media conversations. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1392–1401, Online. Association for Computational Linguistics.

Chunyuan Yuan, Wanhui Qian, Qianwen Ma, Wei Zhou, and Songlin Hu. 2021. Srlf: a stance-aware reinforcement learning framework for content-based rumor detection on social media. In 2021 International Joint Conference on Neural Networks (IJCNN), pages 1–8. IEEE.

Fengzhu Zeng and Wei Gao. 2023. Prompt to be consistent is better than self-consistent? few-shot and zero-shot fact verification with pre-trained language models. In Findings of the Association for Computational Linguistics: ACL 2023, pages 4555–4569, Toronto, Canada. Association for Computational Linguistics.

Qiang Zhang, Shangsong Liang, Aldo Lipani, Zhaochun Ren, and Emine Yilmaz. 2019. From stances' imbalance to their hierarchical representation and detection. In The World Wide Web Conference, pages 2323–2332.

Xuan Zhang and Wei Gao. 2023. Towards llm-based fact verification on news claims with a hierarchical step-by-step prompting method. In Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics, pages 996–1011, Nusa Dua, Bali. Association for Computational Linguistics.

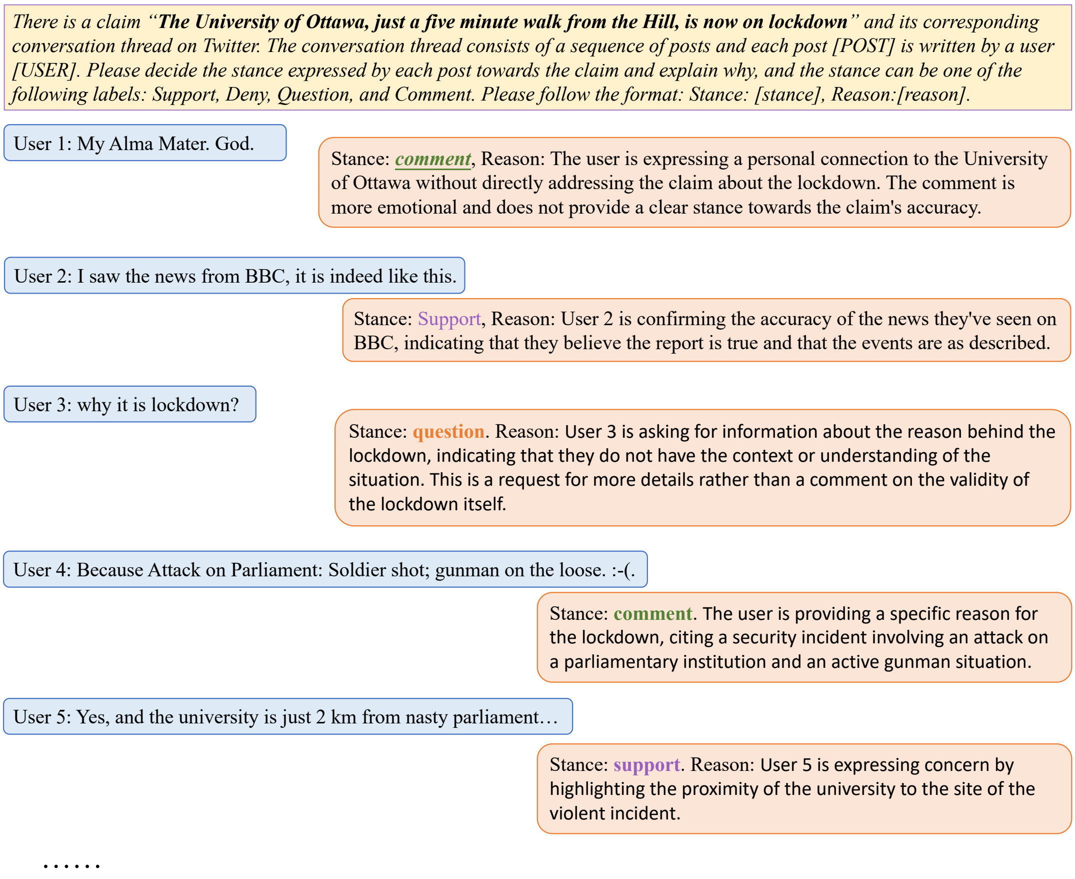
Xuan Zhang and Wei Gao. 2024. Reinforcement retrieval leveraging fine-grained feedback for fact checking news claims with black-box LLM. In Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), pages 13861–13873, Torino, Italia. ELRA and ICCL.

Zhe Zhao, Paul Resnick, and Qiaozhu Mei. 2015. Enquiring minds: Early detection of rumors in social media from enquiry posts. In Proceedings of the 24th international conference on world wide web, pages 1395–1405.

Arkaitz Zubiaga, Elena Kochkina, Maria Liakata, Rob Procter, and Michal Lukasik. 2016a. Stance classification in rumours as a sequential task exploiting the tree structure of social media conversations. In Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, pages 2438–2448.

Arkaitz Zubiaga, Elena Kochkina, Maria Liakata, Rob Procter, Michal Lukasik, Kalina Bontcheva, Trevor Cohn, and Isabelle Augenstein. 2018. Discourse-aware rumour stance classification in social media using sequential classifiers. Information Processing & Management, 54(2):273–290.

Arkaitz Zubiaga, Maria Liakata, Rob Procter, Geraldine Wong Sak Hoi, and Peter Tolmie. 2016b. Analysing how people orient to and spread rumours in social media by looking at conversational threads. PLOS ONE, 11(3):1–29.



*There is a claim "**The University of Ottawa, just a five minute walk from the Hill, is now on lockdown**" and its corresponding conversation thread on Twitter. The conversation thread consists of a sequence of posts and each post [POST] is written by a user [USER]. Please decide the stance expressed by each post towards the claim and explain why, and the stance can be one of the following labels: Support, Deny, Question, and Comment. Please follow the format: Stance: [stance], Reason:[reason].*

User 1: My Alma Mater. God.

Stance: ***comment***, Reason: The user is expressing a personal connection to the University of Ottawa without directly addressing the claim about the lockdown. The comment is more emotional and does not provide a clear stance towards the claim's accuracy.

User 2: I saw the news from BBC, it is indeed like this.

Stance: Support, Reason: User 2 is confirming the accuracy of the news they've seen on BBC, indicating that they believe the report is true and that the events are as described.

User 3: why it is lockdown?

Stance: **question**. Reason: User 3 is asking for information about the reason behind the lockdown, indicating that they do not have the context or understanding of the situation. This is a request for more details rather than a comment on the validity of the lockdown itself.

User 4: Because Attack on Parliament: Soldier shot; gunman on the loose. :-(.

Stance: **comment**. The user is providing a specific reason for the lockdown, citing a security incident involving an attack on a parliamentary institution and an active gunman situation.

User 5: Yes, and the university is just 2 km from nasty parliament…

Stance: **support**. Reason: User 5 is expressing concern by highlighting the proximity of the university to the site of the violent incident.

……

Figure 4: An example of stance prompt.

Arkaitz Zubiaga, Alex Voss, Rob Procter, Maria Liakata, Bo Wang, and Adam Tsakalidis. 2017. Towards real-time, country-level location classification of worldwide tweets. IEEE Transactions on Knowledge and Data Engineering, 29(9):2053–2066.

A Prompt Design

A.1 Stance Detection Prompt

For stance labeling, we use the following prompt:

There is a claim [CLAIM] and I will give you its corresponding conversation thread on Twitter. The conversation thread consists of a sequence of posts and each post [POST] is written by a user [USER]. Please decide the stance expressed by each post towards the claim and explain why, and the stance can be one of the following labels: Support, Deny, Question, and Comment. Please follow the format: Stance: [stance], Reason:[reason].

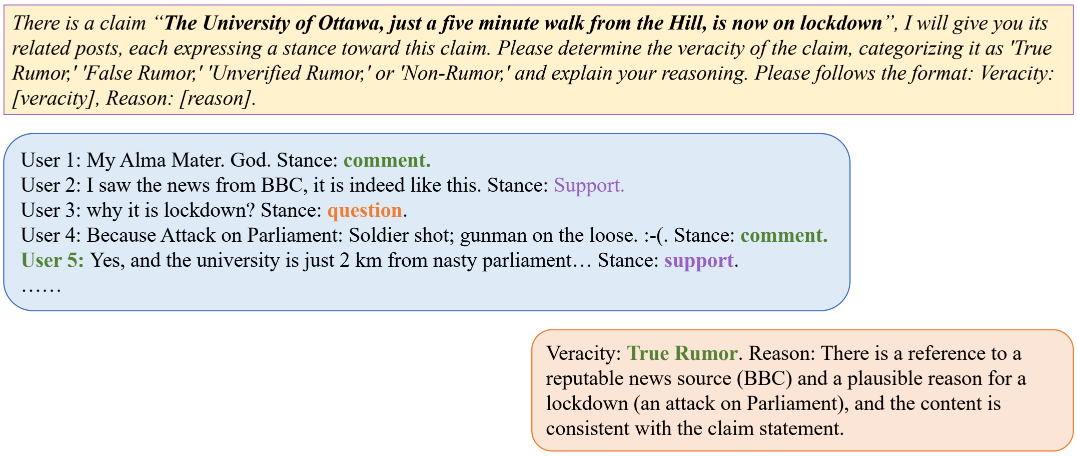
The SD LLM will output a stance label and an explanation of labeling rationale for each input post. We show the details of stance detection prompt in Figure 4. The blue box represents all the context, the text in the orange box represents the stance and reason output from the stance LLM.

A.2 Rumor Verification Prompt

Given a claim and its related posts that are retained by the selection policy model, RV LLM network is prompted to generate veracity label for the claim and a brief explanation of the decision, considering the claim and posts content and stance. The prompt is constructed as follows:

There is a claim [CLAIM], I will give you its related posts, each expressing a stance toward this claim. Please determine the veracity of the claim, categorizing it as 'True Rumor,' 'False Rumor,' 'Unverified Rumor,' or 'Non-Rumor,' and explain your reasoning. Please follows the format: Veracity: [veracity], Reason: [reason].

The RV LLM will output a veracity label and an explanation of labeling rationale for any input claim. We show the details of claim veracity classification prompt in Figure 5. The blue box represents all the context, the text in the orange box represents the outputs of claim veracity and the reason.



*There is a claim "**The University of Ottawa, just a five minute walk from the Hill, is now on lockdown**", I will give you its related posts, each expressing a stance toward this claim. Please determine the veracity of the claim, categorizing it as 'True Rumor,' 'False Rumor,' 'Unverified Rumor,' or 'Non-Rumor,' and explain your reasoning. Please follows the format: Veracity: [veracity], Reason: [reason].*

User 1: My Alma Mater. God. Stance: **comment.**
User 2: I saw the news from BBC, it is indeed like this. Stance: Support.
User 3: why it is lockdown? Stance: **question**.
User 4: Because Attack on Parliament: Soldier shot; gunman on the loose. :-(. Stance: **comment.**
User 5: Yes, and the university is just 2 km from nasty parliament… Stance: **support**.
……

Veracity: **True Rumor**. Reason: There is a reference to a reputable news source (BBC) and a plausible reason for a lockdown (an attack on Parliament), and the content is consistent with the claim statement.

Figure 5: An example of rumor prompt.

B Dataset Statistics

We provide the statistics of training and testing dataset in Table 4 and Table 5, respectively.

Statistics	Twitter15	Twitter16	PHEME
Total claims	1,308	818	6,425
Non-rumor	374 (28.6%)	205 (25.1%)	4,023 (62.6%)
False-rumor	370 (28.3%)	207 (25.3%)	638 (9.9%)
True-rumor	190 (14.5%)	205 (25.1%)	1,067 (16.6%)
Unverified-rumor	374 (28.6%)	201 (24.5%)	697 (10.8%)
Total posts	68,026	40,867	383,569
Avg. posts/claim	52	50	6
Max. posts/claim	814	757	228
Min. posts/claim	1	1	3

Table 4: Statistics of datasets used for training.

Statistics	RumorEval-S	SemEval-8
Total claims	425	297
Non-rumor	100 (23.53%)	——
False-rumor	74 (17.41%)	62 (20.8%)
True-rumor	145 (34.12%)	137 (46.1%)
Unverified-rumor	106 (24.94%)	98 (33.0%)
Posts of Support	1,320 (19.65%)	910 (20.1%)
Posts of Deny	522 (7.77%)	334 (7.6%)
Posts of Question	531 (7.90%)	358 (7.9%)
Posts of Comment	4,345 (64.68%)	2,907 (64.3%)
Total posts	6,718	4,519
Avg. posts/claim	16	15
Max. posts/claim	249	228
Min. posts/claim	2	3

Table 5: Statistics of the datasets used for testing.

C Hyper-parameter Setting

As for Llama 2 model, We download the version with 7 billion parameters(Llama-2-7b) on August, 20, 2023 to obtain its responses. we then employ HuggingFace AutoTokenizer[7] and classes to run the model. The following arguments are setup during fine-tuning stage:

"model_name": "Llama-2-7b";
"learning_rate": 1e-4;
"num_train_epochs": 6;
"max_seq_length": 4096;
"load_in_4bit": True;
"lr_scheduler_type": "linear";
"temperature": 0

As for selector policy model, we set the hyper-parameters as the following:

"learning rate": 5e-5
"optimizer": Adam
"warm-up rate": 0.1
"batch size": 4
"maximum epoch": 50

D Details of Baseline Models

D.1 Stance Detection Baselines

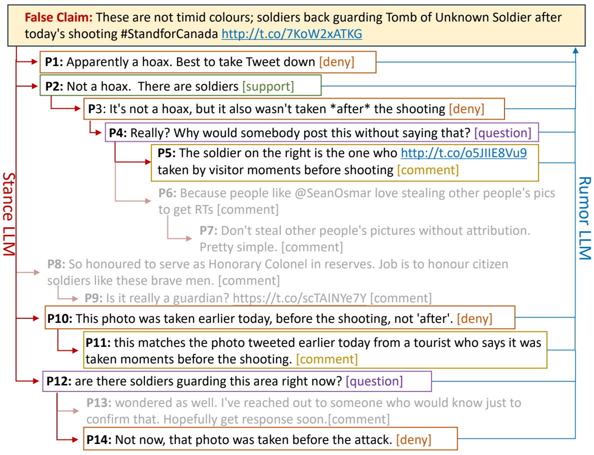
Since our SD network only requires claim veracity label but not post stance label, we choose unsupervised, supervised, and weakly supervised baselines: (1) **TGA** (Allaway and McKeown, 2020a): A topic-grouped attention network for zero-shot stance detection. (2) **BerTweet** (Nguyen et al., 2020): A language model pre-trained on 850M tweets, which is fine-tuned on validataion dataset to adapt to stance detection task. (3) **Llama 2-ST** (Touvron et al., 2023): A pre-trained large language model developed by Meta for only stance detection task. (4) **Llama 2-MT** (Touvron et al., 2023): A pre-trained

[7] https://huggingface.co/docs/transformers/v4.33.0/en/model_doc/auto#transformers.AutoTokenizer

large language model prompted to perform multitask of stance detection and rumor verification together. (5) **BiGRU** (Augenstein et al., 2016): A bidirectional GRU-based stance detection model. (6) **BrLSTM** (Kochkina et al., 2017): An LSTM-based model that models the structured conversational thread to detect stance. (7) **MT-GRU** (Ma et al., 2018): A RNN-based multi-task learning model to jointly detect rumors and stances. (8) **JointCL** (Liang et al., 2022): A zero-shot stance detection model based on contrastive learning. (9) **SRLF** (Yuan et al., 2021): A stance-aware reinforced framework for stance detection.[8] (10) **TD-MIL** (Yang et al., 2022): A weakly supervised model for stance detection and rumor verification based on top-down tree structure. Additionally, **JS-DRV (DATASET)** is our LLM-based reinforcement tuning method applied for stance detection.

D.2 Rumor Verification Baselines

We collect unsupervised, supervised, weakly supervised baselines for rumor verification. Since multi-task learning also can enhance rumor detection, we also introduce two multi-task baselines: (1) **BerTweet** (Nguyen et al., 2020): A pre-trained language model with 850M tweets, and we fine-tune it for rumor verification here. (2) **Llama 2-ST** (Touvron et al., 2023): A large language model pre-trained on various domains myriad corpus, we use it for single rumor verification task. (3) **Llama 2-MT** (Touvron et al., 2023): A pre-trained large language model prompted to perform multi-task of stance detection and rumor verification together. (4) **GCAN** (Lu and Li, 2020): A graph-aware co-attention model utilizing retweet to detect rumor veracity. (5) **TD-RvNN** (Ma et al., 2020): A top-down tree-structured attention networks for rumor verification. (6) **PLAN** (Khoo et al., 2020): A transformer based rumor verification model utilizing interactions between users. (7) **DDGCN** (Sun et al., 2022): A rumor verification model utilizing the dynamic propagation texts and external knowledge. (8) **SRLF** (Yuan et al., 2021): A stance-aware refinforced framework for rumor verification. (9) **TD-MIL** (Yang et al., 2022): A top-down tree propagation model for joint detection tasks on stance type and rumor veracity. (10) **MTL2** (Kochkina et al., 2018): A sequential model utilizing a number of task-specific layers for stance detection and rumor verification. (11) **MT-GRU** (Ma et al., 2018): A multi-task learning approach to jointly detect rumors and stances by capturing both shared and task-specific features. Additionally, **JSDRV(DATASET)** is our reinforcement tuning method applied for rumor verification.



False Claim: These are not timid colours; soldiers back guarding Tomb of Unknown Soldier after today's shooting #StandforCanada http://t.co/7KoW2xATKG

P1: Apparently a hoax. Best to take Tweet down [deny]

P2: Not a hoax. There are soldiers [support]

P3: It's not a hoax, but it also wasn't taken *after* the shooting [deny]

P4: Really? Why would somebody post this without saying that? [question]

P5: The soldier on the right is the one who http://t.co/o5JIIE8Vu9 taken by visitor moments before shooting [comment]

P6: Because people like @SeanOsmar love stealing other people's pics to get RTs [comment]

P7: Don't steal other people's pictures without attribution. Pretty simple. [comment]

P8: So honoured to serve as Honorary Colonel in reserves. Job is to honour citizen soldiers like these brave men. [comment]

P9: Is it really a guardian? https://t.co/scTAINYe7Y [comment]

P10: This photo was taken earlier today, before the shooting, not 'after'. [deny]

P11: this matches the photo tweeted earlier today from a tourist who says it was taken moments before the shooting. [comment]

P12: are there soldiers guarding this area right now? [question]

P13: wondered as well. I've reached out to someone who would know just to confirm that. Hopefully get response soon.[comment]

P14: Not now, that photo was taken before the attack. [deny]

Figure 6: Case study. SD LLM predicts posts stance (left red line). The selector policy chooses posts stances (color boxes) for RV LLM to predict rumor veracity (right blue line).

E Analysis

E.1 Case Study

We also plot example outputs of JSDRV in Figure 6, where the red line represents the stance detection process, the gray texts denote discarded posts, and the blue line indicates the rumor verification process during inference. We observe that the selector policy can choose posts with high-quality stances annotated by stance LLM, which provides useful cues to rumor verification LLM.

E.2 Sensitivity Study

We conduct a sensitivity study to see the impact of the size of seeding claims. Using JSDRV (PH) model, we show the variation of micF score on stance detection and rumor verification tasks with the increase of the proportion of seeding claims. We also show the performance of the strongest baseline TD-MIL trained on the same sets of seeding claims. Figure 7 indicates that (1) With more seeding claims, the performance of TD-MIL (PH), JSDRV (PH), and RSDRV-Bert (PH) all gets improved; (2) The performance of JSDRV tends to stabilize more quickly than TD-MIL. When the ratio reaches 50%, the performance of JSDRV becomes

[8] The original method just outputs rumor veracity result, we put a Softmax layer behind stance embeddings to get stance output here.

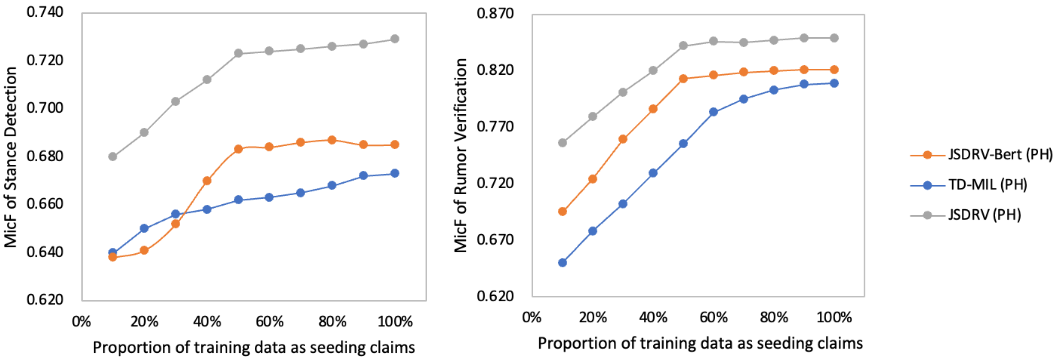


Figure 7: Impact of the size of seeding claims.

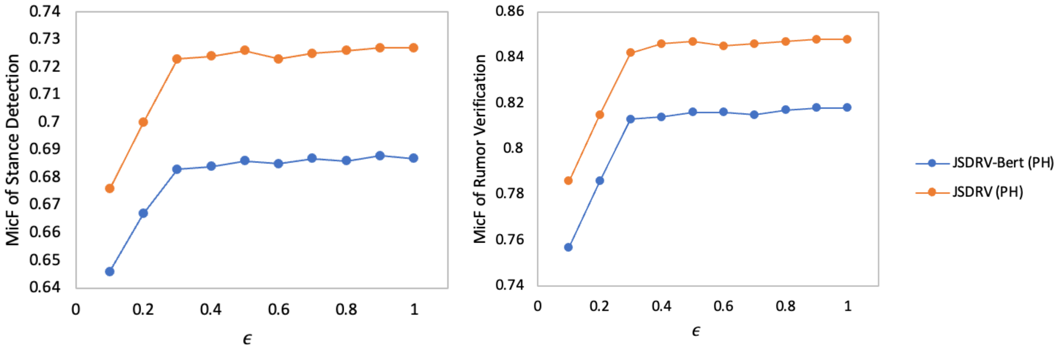


Figure 8: Impact of ϵ.

saturated at a much higher level, while TD-MIL cannot catch up even using up all the training data. (3) JSDRV performs much better than TD-MIL using a same proportion of training data.

To get an intuitive understanding of the greedy controller, we conduct a study to assess the performance of JSDRV-Bert (PH) and JSDRV (PH) across varying ϵ rates. We use the same ϵ to control both post and claim sampling. The MicF scores for stance detection and rumor verification on the RumorEval-S dataset are shown in Figure 8. Our observations include: (1) The performance of both models improves as ϵ increases. (2) Model performance stabilizes as ϵ reaches 0.3, indicating that our greedy controller can achieve comparable results with limited instances labels compared to abundant instances labels.

E.3 User Study

We conduct a user study to evaluate the quality of the model output. We sample 120 samples from RumorEval-S and present them in two forms: Baseline (claim, posts) and JSDRV (claim, selected post-stance pairs, reasons). We then ask 6 users to label the articles and give their confidence in a 5-point Likert Scale (Joshi et al., 2015), and each person is given only one form to avoid cross influence.

	F1	Acc	Confidence	Avg. Time/news
Baseline	0.693	0.713	1.017	25 sec
JSDRV	0.961	0.990	4.165	5 sec

Table 6: User study results.

Table 6 shows that 1) users determine the rumors more accurately with JSDRV, 2) users spent 75% less time identifying rumors, and 3) users show higher confidence with the results of JSDRV, suggesting that users tend to be more sure about their decision when stance and related reasons have been provided.