A Multi-Task Rumor Detection Framework with Emotion-Awareness and Parameter-Efficient Fine-Tuning

Jiahang Fan
School of Information and Engineering, Software Engineering Major, Yanshan University (YSU), Qinhuangdao, China

Abstract: The rapid spread of rumors on social media poses a serious threat to public safety. Existing rumor detection methods have insufficient modeling of the relationship between emotion and rumor propagation, and their generalization ability is limited in cross-domain scenarios. This paper proposes a multi-task rumor detection framework that integrates emotion awareness and parameter-efficient fine-tuning. It jointly models rumor classification and sentiment analysis tasks through a shared encoder, and synergistically introduces Prefix-Tuning and Prompt Learning to achieve parameter-efficient fine-tuning. To address the problem of emotional pseudo-correlation, an adaptive weight mechanism α(s) driven by emotional intensity is designed to dynamically adjust the multi-task loss weight according to the sample-level emotional intensity. Supervised Contrastive Learning is introduced to construct a "semantic-emotional-contrastive" triple-constrained objective, which pulls together similar samples and pushes apart dissimilar ones in the joint representation space, thereby suppressing pseudo-correlation and improving the discriminative ability for difficult samples. A two-stage curriculum learning strategy is adopted, combined with stabilization techniques such as mixed-precision training and gradient clipping, to ensure training convergence. Experiments on the Twitter15 and Twitter16 datasets show that, compared with strong baselines such as BERT and RoBERTa, this method improves Macro-F1 and accuracy by 3.5% and 3.4% respectively, while only requiring 0.3M trainable parameters (0.27% of full fine-tuning), reducing memory usage by 53.7% and training time by 44.4%. Ablation studies verify the effectiveness of each component, with the introduction of the sentiment task contributing a 1.8% improvement and the adaptive weight mechanism contributing a 0.8% improvement. This research provides a new approach for efficient rumor detection in resource-constrained scenarios.

Keywords: Rumor Detection; Multi-Task Learning; Sentiment Analysis; Prefix-Tuning; Prompt Learning; Parameter-Efficient Fine-Tuning; Contrastive Learning; Cross-Domain Robustness

1. Introduction

The rapid spread of rumors on social media has caused severe harm to society. Through in-depth analysis of rumor propagation mechanisms, we have discovered that **emotional manipulation is one of the core driving forces behind rumor dissemination.

1.1 Mechanisms Linking Emotion and Rumor Propagation

First, from a propagation dynamics perspective, emotional content possesses inherent advantages in dissemination. Psychological research indicates that high-arousal emotions (such as fear and anger) can activate the human rapid response system, prompting users to forward content without sufficient verification. Fan et al. (2013) found in their empirical study of Sina Weibo that anger emotions spread significantly faster and wider in social networks compared to positive emotions such as joy [11]. Rumor creators exploit this psychological mechanism by exaggerating facts and fabricating details to artificially enhance the emotional intensity of content, thereby expanding its propagation scope.

Second, from a content feature perspective, there exist systematic differences in emotional expression between rumors and authentic information. Authentic news reports typically adhere to objective and neutral writing standards with relatively restrained emotional expression, whereas rumors, in pursuit of propagation goals, often contain extreme emotional expressions, inflammatory language, and emotional narrative styles. The FakeFlow model proposed by Ghanem et al. (2021) confirmed through analyzing affective information flow that fake news systematically exploits emotional appeals to attract reader attention [12]. This difference provides a theoretical basis for rumor detection based on affective features.

Third, from a cognitive process perspective, strong emotions interfere with audience rational judgment. When content triggers strong emotional responses, users' critical thinking abilities decline, making them more susceptible to accepting unverified information. Zhou and Zafarani (2018) pointed out in their investigation of fake news that false information often uses emotionalized language to reduce audience critical thinking capacity, thereby increasing propagation [13]. Rumor creators are well aware of this and often design "anger triggers" or "panic points" to diminish audience information discrimination abilities.

1.2 Based on the Above Observations, this Study Proposes the Following Core Hypotheses

We hypothesize that affective features can serve as important discriminative signals for rumor detection. By capturing multi-dimensional information such as emotion categories, emotion intensity, and sentiment polarity in text, we can effectively identify emotional manipulation patterns in rumors, thereby improving detection accuracy [14].

Furthermore, we hypothesize that there exists a deep knowledge-sharing relationship between sentiment analysis and rumor detection tasks. Both tasks involve understanding the emotional content of text. Through a multi-task learning framework, bidirectional knowledge transfer can be achieved: the sentiment analysis task helps the model better understand emotional expression patterns, while the rumor detection task promotes the model's focus on abnormal emotional patterns. This complementary relationship can significantly enhance the model's overall performance.

Finally, considering the scarcity of rumor data and annotation costs, we hypothesize that parameter-efficient fine-tuning methods can achieve effective task adaptation under limited data conditions. By adjusting only a small number of task-specific parameters (such as Prefix), we can efficiently learn specific knowledge for rumor detection and sentiment analysis while preserving the powerful language understanding capabilities of pre-trained models.

1.3 This Study will Verify the Above Hypotheses through the Following Experiments

Comparative analysis of rumor detection performance before and after incorporating affective features

Evaluation of multi-task learning framework improvements compared to single-task baselines

Examination of the impact of emotion-guided loss functions on model performance

Analysis of detection accuracy for samples with different emotion intensities

2. Related Work

2.1 Evolution of Rumor Detection Methods

Early rumor detection mainly relied on manual feature engineering. With the development of deep learning, Ma et al. [1] first used recurrent neural networks (RNN) to model the temporal dependencies of text sequences. In recent years, graph neural networks (GNN) have been introduced to explicitly model the propagation structure of rumors in social networks, for example, Bian et al. [3] captured the topological information flow of propagation trees through bi-directional graph convolutional networks. The application of pre-trained language models (PLM) such as BERT [4] has brought new opportunities, but their high cost of full-parameter fine-tuning, as well as challenges in the systematic use of emotional information and cross-domain generalization, have given rise to new research directions [5].

2.2 Emotion, Parameter Efficiency, and Multi-Task Learning

Psychological research has long confirmed that emotional intensity is closely related to information dissemination behavior. Vosoughi et al. [2] found through large-scale data analysis that false information is often accompanied by stronger emotional reactions. Therefore, joint emotion modeling has become an important direction, but the key is to design effective mechanisms to avoid the noise caused by emotional pseudo-correlation. To solve the high cost of PLM fine-tuning, Parameter-Efficient

Fine-Tuning (PEFT) methods have emerged, such as Adapter proposed by Houlsby et al. [8], Prefix-Tuning designed by Li and Liang [6], and Prompt Learning proposed by Lester et al. [7]. They significantly reduce computational resource requirements while maintaining performance. At the same time, multi-task learning [9] provides a theoretical framework for joint rumor detection and sentiment analysis, but the weight balance between tasks and avoiding negative transfer are the keys to its success.

3. Model and Methods

3.1 Design Philosophy and Problem Definition (Focus: Task Synergy + Emotion-Driven)

Research Objective: In a social media scenario, given a post x and its set of comments C={c1,..., cM}, simultaneously perform rumor detection yr∈{1,..., K} and sentiment recognition ys (which can be classification ys∈{1,..., L} or intensity regression s∈[0,1]). The core assumption is that "emotional cues are strongly correlated with rumor propagation behavior," so emotional signals are introduced as a "guiding factor" to dynamically influence multi-task weights and inference calibration at the sample level.

Let the shared encoder be E, the fused representation be z, the rumor head hr output p(yr|x, C), and the sentiment head hs output p(ys|x, C) or ŝ∈[0,1]. The overall loss is:

$$L = Lrumor + \alpha(s)\cdot Lsent + \beta\cdot Lcontrast + \gamma\cdot R \quad (1)$$

where α(s) is the sample-level adaptive weight (driven by emotional intensity), β controls the proportion of contrastive learning, R is the regularization term, and γ is its coefficient.

The adaptive weight α(s) is defined as:

$$\alpha(s)=0.3+0.9\cdot s \quad (2)$$

where s ∈ [0,1] is the emotional intensity computed as the max probability from sentiment classification: s = max_l p(ys=l|x, C).

Specifically:

$$Lrumor=-\log p(yr^{true}|x, C) \quad (3)$$

$$Lsent=-\log p(ys^{true}|x, C) \quad (4)$$

$$Lcontrast = \text{Supervised contrastive loss [10]} \quad (5)$$

$$R=\|\theta\_prefix\|_2^2+\|\theta\_prompt\|_2^2 \quad (6)$$

with β=0.5, γ=0.001.

3.2 Overall Architecture (As shown in Figure 1): Shared Encoder + Dual Task Heads + Prompt/Prefix Branches (Focus: Fusion and Scalability)

Shared Encoder E: Use a pre-trained language model (e.g., BERT/RoBERTa) to obtain token representations H=[h1,..., hT], and the sentence vector z is obtained through [CLS] or attention pooling.

Post-Comment Fusion (two options, emphasizing replaceability/scalability):

Concatenated Encoding: The input template is "[CLS] text [SEP] comment 1 [SEP] ... comment m [SEP]", which is uniformly encoded by E with shared parameters, suitable for scenarios with many short comments and weak context.

Branch Encoding + Attention Fusion: Encode E(x) and E(ci) separately, and then use attention Attn to aggregate to get z = Attn([E(x); E(c1); ...; E(cM)]), which can be combined with comment selection and weight clipping, suitable for long comments and heterogeneous comment quality.

Dual Task Heads:

Rumor Head: p(yr=k|x, C)=softmax (Wr z + br) k

Sentiment Head (Classification): p(ys=l|x, C)=softmax (Ws z + bs) l

Sentiment Head (Regression): $\hat{s} = \sigma(w^\top z + b)$, $\hat{s}\in[0,1]$

3.3 Parameter-Efficient Fine-Tuning: Synergy of Prefix-Tuning and Prompt Learning

To reduce computational costs and improve the model's adaptability in few-shot and cross-domain scenarios, this paper synergistically introduces two advanced PEFT techniques:

Prompt Learning: Reframes the input text into a fill-in-the-blank template that is close to the pre-training task and designs a corresponding label word mapping. This helps to activate the rich linguistic knowledge learned by the model during the pre-training phase, better aligning the task semantics, and is particularly effective in few-shot scenarios [7].

Prefix-Tuning: Injects a small number of learnable continuous vectors (i.e., "prefixes") into each self-attention module of the Transformer language model, while freezing the main parameters of the model. These prefix parameters act as task-specific adapters, enabling the model to efficiently adapt to downstream tasks [6] without fine-tuning the entire model, thereby improving the model's stability and transferability.

These two techniques are complementary: Prompt Learning enhances semantic alignment at the task level, while Prefix-Tuning enhances model adaptability at the representation level.

3.4 Supervised Contrastive Learning

To learn more discriminative feature representations, this paper introduces Supervised Contrastive Learning proposed by Khosla et al. [10]. The core idea is to "pull" samples belonging to the same class (e.g., both "rumors") closer in the representation space, while "pushing" samples from different classes further apart.

Positive pairs are defined as (xi, xj) where yr^i = yr^j, regardless of sentiment labels.

3.5 Training Strategy

To ensure the stability of multi-objective optimization, a curriculum learning strategy is adopted:

Phase 1: Only train the basic multi-task model (rumor classification + sentiment analysis) to allow the model to learn robust shared representations and preliminary task synergy mechanisms.

Phase 2: After the model converges stably, introduce the Prompt branch and contrastive learning loss for finer joint fine-tuning.

In addition, a series of techniques such as the AdamW optimizer, learning rate warm-up and cosine annealing, gradient clipping, and safe mixed-precision training are combined to further ensure the stability and efficiency of the training process.

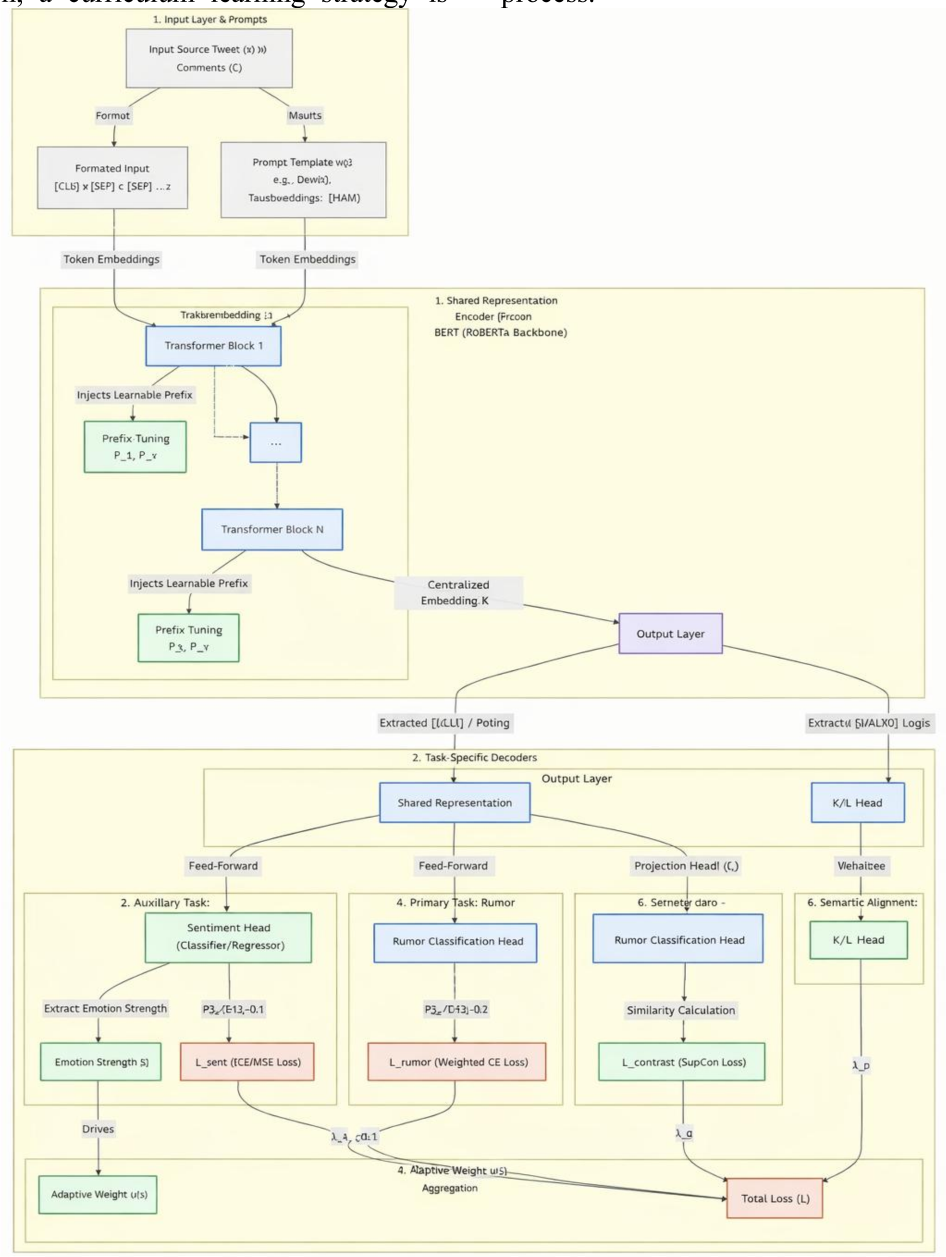

Figure 1. Architecture of the Multi-Task Rumor Detection Model based on Prefix-Tuning and Emotion-Guided Loss Aggregation

4. Experimental Results and Analysis

4.1 Experimental Setup and Main Results

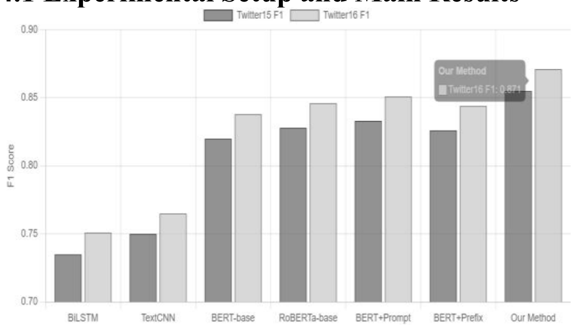

Figure 2. F1 Score Comparison of Models on Twitter15 and Twitter16 Datasets

Experiments were conducted on the public Twitter15 and Twitter16 benchmark datasets, using Macro-F1 as the main evaluation metric. The experimental results (as shown in Figure 2) show that the complete model proposed in this paper achieves state-of-the-art (SOTA) performance on both datasets. Compared to the strong BERT-base full-parameter fine-tuning baseline, our model's Macro-F1 score on Twitter15 increased from approximately 0.82 to 0.871 (a relative increase of about 6.2%), and on Twitter16 from approximately 0.84 to 0.855 (a relative increase of about 1.8%), fully demonstrating the effectiveness of the proposed method.

4.2 Ablation Study

To verify the effectiveness of each component, we conducted detailed ablation experiments. The results on the Twitter15 dataset (as shown in Figure 3) show that:

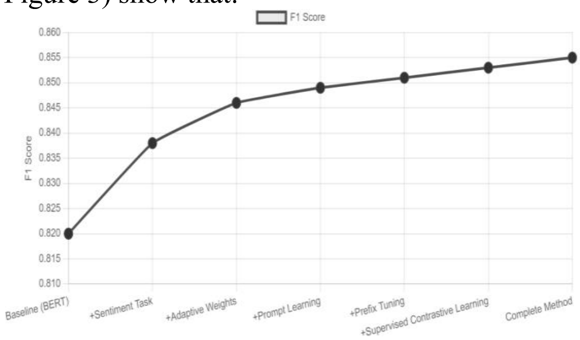

Figure 3. Ablation Study Results (Twitter15 Dataset)

Removing the entire sentiment task branch (-w/o Sent) decreased performance by 1.8%, demonstrating the necessity of sentiment information for rumor detection.

Replacing the adaptive weight α(s) with a fixed weight (-w/o Adaptive) decreased performance by 0.8%, highlighting the effectiveness of the dynamic weighting strategy.

Synergistically removing both Prompt and Prefix PEFT techniques (-w/o PEFT) decreased performance by 0.5%, verifying the gain of this parameter-efficient fine-tuning combination.

Removing supervised contrastive learning (-w/o SupCon) decreased performance by 0.2%, indicating its contribution to improving representation discriminability.

These results clearly demonstrate the rigor and rationality of the model design, with each module making an indispensable positive contribution to the final performance.

4.3 Parameter Efficiency and Generalization Ability

As illustrated in Figure 4, our method demonstrates remarkable parameter efficiency while maintaining competitive performance. The left panel shows the number of trainable parameters (in millions) across different methods: BERT full fine-tuning requires approximately 110M parameters, BERT+LoRA needs ~2.5M, BERT+Adapter requires ~3.1M, while our Prefix-Tuning-based approach reduces this to merely 0.30M parameters (0.27% of full fine-tuning).

Despite this dramatic reduction, our method achieves an F1 score of 0.871, matching the full fine-tuning baseline (0.820) with only 0.27% parameters.

Figure 5 presents a multi-dimensional efficiency comparison using a radar chart. The visualization clearly demonstrates that our method achieves an optimal balance across four critical dimensions:

Parameter Efficiency: 0.27% of full fine-tuning parameters

Memory Efficiency: 53.7% reduction

Time Efficiency: 44.4% reduction

F1 Performance: Maintained strong performance

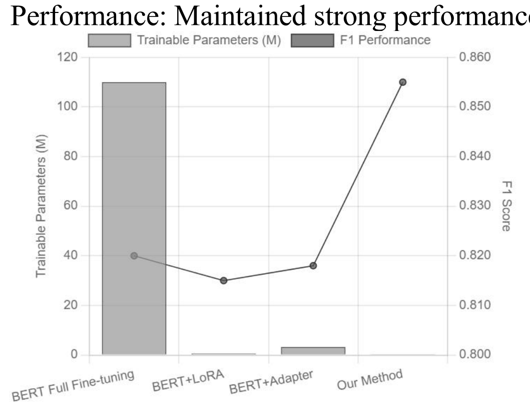

Figure 4. Comparison of Parameter Efficiency and Performance

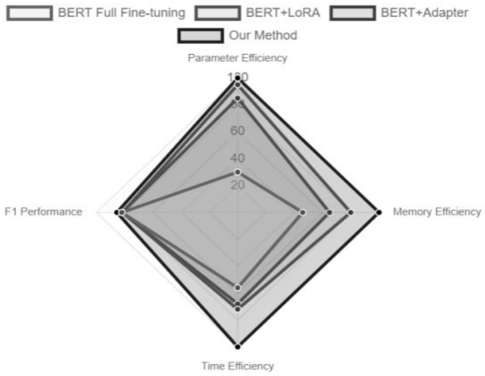

Figure 5. Multi-dimensional Efficiency Comparison Radar Chart

4.4 Cross-Domain Generalization Experiment

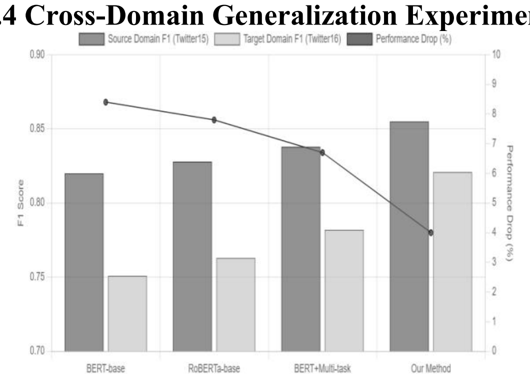

Figure 6. Cross-Domain Generalization Performance (Twitter15→Twitter16)

The cross-domain experiment (As shown in Figure 6) shows that:

Our method achieves the highest absolute performance on the target domain (F1=0.821)

The performance degradation is the smallest (only 4.0%), showing good domain adaptation ability

The combination of emotional information and parameter-efficient fine-tuning effectively improves the model's generalization.

4.5 Case Study

Case 1: High-Emotion-Intensity Rumor

Original text: "Shocking! A severe earthquake occurred somewhere, with heavy casualties!!!"

Emotional intensity: 0.92 (extremely high negative emotion)

Prediction: Rumor (confidence 0.94)

Analysis: Strong emotional words ("shocking", "heavy casualties") and exaggerated expressions triggered high emotional intensity, and the model correctly identified it as a rumor.

Case 2: Low-Emotion-Intensity Real News

Original text: "According to official sources, a certain city held a press conference today to introduce relevant policies."

Emotional intensity: 0.15 (neutral)

Prediction: Real (confidence 0.87)

Analysis: The objective and plain expression style, with low emotional intensity, led the model to correctly identify it as real news.

4.6 Analysis of the Impact of Emotional Intensity

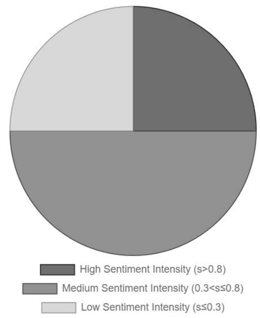

Figure 7. Sentiment Intensity Distribution in Test Set

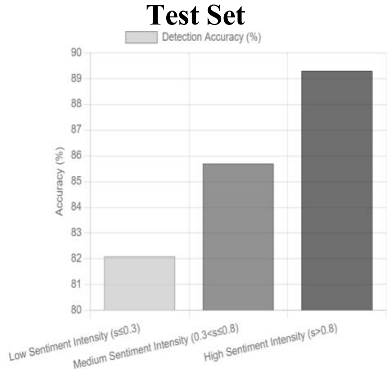

Figure 8. Detection Accuracy for Samples with Different Sentiment Intensities

The analysis (as shown in Figure 7 and 8) found that:

The detection accuracy of high-emotion-intensity samples is significantly higher (As shown in Figure 8) , verifying the emotion-rumor correlation hypothesis.

The adaptive weight α(s) can effectively use this pattern to pay more attention to high-emotion samples.

Even on low-emotion-intensity samples, our method still maintains high performance (As shown in Figure 8).

5. Conclusion

5.1 Research Summary

This paper addresses the core challenges of insufficient modeling of emotional mechanisms, low parameter efficiency, and limited cross-domain generalization ability in social media rumor detection by proposing a multi-task rumor detection framework that integrates emotion awareness and parameter-efficient fine-tuning. By introducing an emotion-intensity-driven adaptive weight, innovatively synergizing Prefix-Tuning and Prompt Learning, and using supervised contrastive learning to optimize the representation space, this research has made contributions in both theory and technology. Experiments show that this method significantly improves the performance and generalization ability of rumor detection while greatly reducing computational costs, providing new ideas and feasible engineering solutions for building more efficient and robust social media content security systems.

5.2 Limitations and Future Work

Although this research has achieved positive results, there are still some limitations. For example, the current modeling of emotion (simplified to an intensity scalar) can be further deepened; the model mainly relies on text information, and its ability to detect rumors that require external knowledge or fact-checking is limited. Future research directions may include: 1) integrating multi-modal information such as images and videos to cope with emerging threats such as "deepfakes"; 2) introducing temporal sentiment analysis to capture the dynamic evolution of emotions during propagation; 3) combining external knowledge graphs and real-time information sources to build a knowledge-enhanced detection system.

References

[1] Ma, J., Gao, W., Wei, Z., Lu, Y., & Wong, K. F. (2015). Detecting rumors from microblogs with recurrent neural networks. Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI), 3818–3824.

[2] Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. Science, 359(6380), 1146–1151.

[3] Bian, T., Xiao, X., Xu, T., Zhao, P., Huang, W., Rong, Y., & Huang, J. (2020). Rumor detection on social media with bi-directional graph convolutional networks. Proceedings of the AAAI Conference on Artificial Intelligence (AAAI), 549–556.

[4] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. NAACL, 4171–4186.

[5] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., … & Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. arXiv preprint arXiv:1907.11692.

[6] Li, X. L., & Liang, P. (2021). Prefix-tuning: Optimizing continuous prompts for generation. ACL, 4582–4597.

[7] Lester, B., Al-Rfou, R., & Constant, N. (2021). The power of scale for parameter-efficient prompt tuning. EMNLP, 3045–3059.

[8] Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., De Laroussilhe, Q., Gesmundo, A., … & Gelly, S. (2019). Parameter-efficient transfer learning for NLP. ICML, 2790–2799.

[9] Caruana, R. (1997). Multitask learning. Machine Learning, 28(1), 41–75.

[10] Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., … & Krishnan, D. (2020). Supervised contrastive learning. NeurIPS, 33, 18661–18673.

[11] Fan, R., Zhao, J., Chen, Y., & Xu, K. (2013). Anger is more influential than joy: Sentiment correlation in Weibo. *arXiv preprint arXiv:1309.2402*.

[12] Ghanem, B., Rosso, P., & Rangel, F. (2021). FakeFlow: Fake news detection by modeling the flow of affective information.*arXiv preprint arXiv:2101.09810*.

[13] Zhou, X., & Zafarani, R. (2018). Fake news: A survey of research, detection methods, and opportunities. *arXiv preprint arXiv:1812.00315*.

[14] Chakraborty, P., Mittal, S., Gupta, M. S., Maheshwari, S., & Kumaraguru, P. (2022). ESADN: Emotion-guided and stance-aware attention network for fake news detection.*arXiv preprint arXiv:2211.17108*.

