Stance-Conditioned Modeling for Rumor Verification

Gibson Nkhata, Susan Gauch

The Seventeenth International Conference on Information, Process, and Knowledge Management

eKNOW 2025

Presenter: Gibson Nkhata

University of Arkansas, USA

gnkhata@uark.edu

Presenter Biography

- **Gibson Nkhata** received a Bachelor’s degree in Information and Communication Technology from Mzuzu University, Malawi, in 2018 and a Master’s degree in Computer Science from the University of Arkansas (UARK), USA, in 2022. He is currently a doctoral student in Computer Science at the university. He is also a Graduate Research Assistant at UARK in the Electrical Engineering and Computer Science (EECS) department under the DART project.

Research Interests

- Our research interest lies in the application of Deep Learning techniques and pre-trained language models in Natural Language Processing (NLP) tasks like Sentiment Analysis, Stance Detection, and Rumor Verification on social media. Additionally, the research interest lies in leveraging contextual word embeddings for Personalized Information Retrieval.

Outline

- Motivation
- Our approach
- Experiments
- Discussion
- Conclusion
- Limitations and Future Work

Motivation

- A rumor may be defined as a statement whose truth value is unverified or deliberately false.
- The exponential rise of social media platforms has fueled the rapid spread of misinformation and rumors, making Rumor Verification (RV) a critical challenge.
- Traditional RV approaches primarily rely on transformer-based language models, such as BERT, to analyze textual representations of posts.
- However, these methods often truncate conversational threads due to sequence length constraints and overlook valuable discourse signals, such as stance labels, that reflect user perspectives on rumors.
- We propose a sequential stance aggregation mechanism that accounts for the temporal ordering of replies and chronologically embedding stance labels using a Bidirectional LSTM (BiLSTM).

Our approach

- **Task formulation**

- Given a conversational thread consisting of a source post p and a set of reply posts $R = \{r_1, r_2, ..., r_n\}$, where n is the total number of reply posts, the goal of RV is to classify the source post p into one of three categories: $y_c \in$ {*true rumor*, *false rumor*, *unverified rumor*}.
- Each post (both p and r_i) is associated with a stance label s_i, where $s_i \in$ {*support*, *deny*, *query*, *comment*}.

A sample online discourse

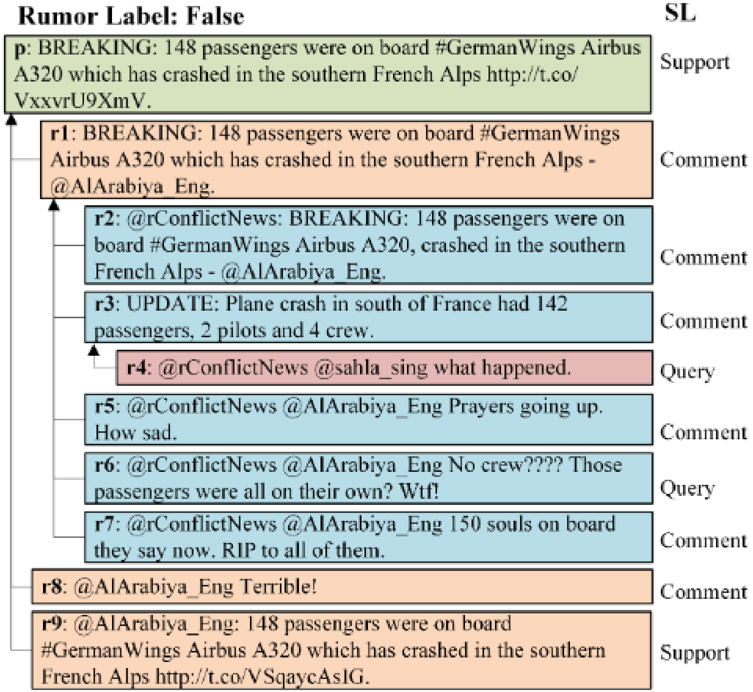
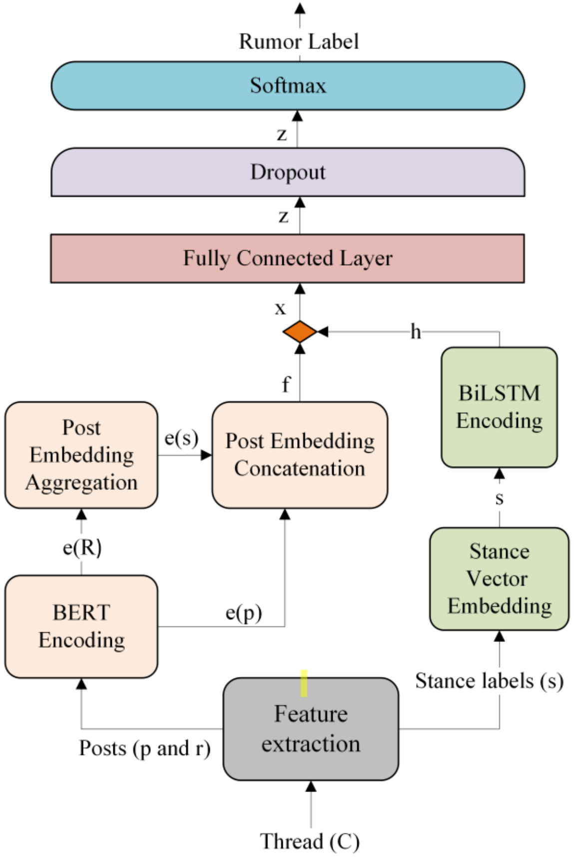


Figure 1. A sample thread C with a false veracity label. SL stands for Stance Labels.

Our approach

- The approach consists of three main components:

1) **Post embedding representation and aggregation**: BERT extracts contextual embeddings for the source and reply posts.
   - Reply posts are aggregated based on stance label and concatenated with source post embedding.
2) **Stance-aware sequence encoding**: A BiLSTM encodes the sequence of stance labels in temporal order.
3) **Unified feed-forward layer**: The post embeddings and stance representations are concatenated and fed into a classifier.



Experiments

TABLE I
DETAILED STATISTICS OF THE DATASETS.

Dataset	#Threads	#Tweets	Stance Distribution				Rumor Veracity Labels		
			#Support	#Deny	#Query	#Comment	#True	#False	#Unverified
SemEval-17	325	5,568	1,004	415	468	3,685	145	74	106
RumorEval-19	446	8,574	1,184	606	608	6,176	185	138	123
PHEME	2,402	105,354	-	-	-	-	1,067	638	697

Baselines:

1. **eventAI**: ensemble learning.
2. **Longformer**: multi-task learning (SD and RV).
3. **Coupled Hierarchical Transformer** (CHT): Global transformer
4. **Joint Rumor and Stance Model** (JRSM): Graph transformer.
5. **SAMGAT**: Uses Graph Attention Networks

TABLE II
COMPARISON OF OUR RESULTS WITH BASELINE MODELS.

Model	SemEval-2017		RumorEval-2019		PHEME	
	Macro-F1	Acc	Macro-F1	Acc	Macro-F1	Acc
eventAI	0.618	0.629	0.577	0.591	0.342	0.357
Longformer	0.662	0.673	0.672	0.684	0.452	0.469
CHT	0.680	0.678	0.579	0.611	0.396	0.466
SAMGAT	0.702	0.709	0.542	0.562	0.409	0.418
JRSM	0.754	0.767	0.598	0.623	0.448	0.479
Ours	**0.774**	**0.781**	**0.636**	**0.648**	**0.641**	**0.643**

Experiments

TABLE III
ABLATION STUDY RESULTS.

Model	RumorEval-2017		RumorEval-2019	
	Macro-F1	Acc	Macro-F1	Acc
-Replies	0.624	0.632	0.540	0.566
-Emb agg	0.642	0.649	0.552	0.579
-Stance-aware	0.647	0.651	0.548	0.564
Ours-whole	**0.774**	**0.781**	**0.636**	**0.648**

Ablation study:

- ***-Replies***: Excludes reaction posts *R*, encoding only the source post *p*.
- ***-Emb agg***: Discards stance-conditioned embedding aggregation, instead encoding the entire rumor event as a single BERT embedding.
- ***-Stance-aware***: Omits the sequential modeling of stance labels using BiLSTM.
- ***Ours-whole:*** the complete model.

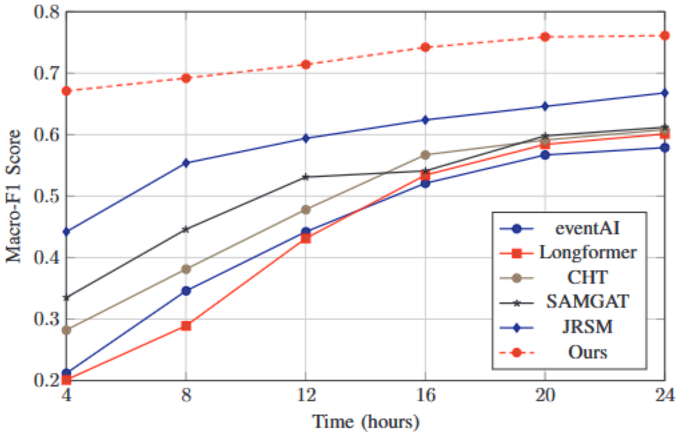


Figure 3. Early Rumor Detection Performance on SemEval-2017.

To assess early detection capabilities, we define detection checkpoints based on the elapsed time, spanning 24 hours, since the initial post. At each checkpoint, only replies accumulated up to that point are considered for model evaluation.

Discussion

- Unlike baseline models, such as *CHT*, that process hierarchical groups separately, our aggregation strategy preserves stance distribution and reduces information loss.
- Our model aggregates reply post embeddings grouped by stance type, ensuring that stance-conditioned representations provide a holistic view of the conversation.
- Aggregation also mitigates the sequence length limitation of BERT by summarizing the impact of all replies in a stance-specific manner.
- While baselines implicitly incorporate stance, this work explicitly embeds and encodes stance labels with BiLSTM, preserving the chronological order of stance evolution.

Conclusion

- We introduced a novel stance-conditioned rumor verification model that integrates BERT-based source post embeddings and reply post embedding aggregation and BiLSTM encoding of stance labels, to enhance the detection of rumors in online discourse.
- Our results show that integrating source post embeddings with aggregated reply features enhances classification performance compared to models that treat them separately.
- Processing stance sequences chronologically using BiLSTM preserves the natural evolution of discussions, leading to more context-aware representations.
- Early rumor detection analysis demonstrates that our model achieves faster and more accurate misinformation detection than competing methods, underscoring its practical utility in real-world rumor detection.

Limitations and Future work

- **Limitations**
- Heavy reliance on accurate stance annotations, that may not always be available or reliable in real-world scenarios.
- The model is trained and evaluated on datasets that may not fully capture real-world misinformation trends across different social media platforms.
- The study focuses exclusively on textual content, ignoring visual information (images, memes, videos) that often accompany social media rumors.

- **Future work could**
- Explore weakly supervised and self-supervised learning to reduce reliance on manually labeled stance data.
- Cross-platform adaptation to generalize rumor detection across different social media environments.
- Incorporation of multi-modal data (text + images + videos) to enhance rumor verification in diverse contexts.
- Explore extra structural dynamics, e.g., stance distribution and hierarchical levels encoding, and attention mechanisms.