Collaboration and Controversy Among Experters:
Rumor Early Detection by Tuning a Comment Generator

Authors: **Bing Wang** (wangbing1416@gmail.com),

Bingrui Zhao, Ximing Li*, Changchun Li, Renchu Guan, Shengsheng Wang

SIGIR 2025
Padova
ITALY

Project page **My homepage**

Background

Social media platforms are inevitably full of rumors, causing lots of damage

Over-the-counter cold and cough medications are being pulled from drugstore shelves in an effort to start the "next plandemic" or force people to get the COVID-19 vaccine.

COVID-19 vaccines are safe for people who have existing health conditions, including conditions that have a higher risk of getting serious illness with COVID-19.

RUMOR

Rumor Detection

Background

Fully-Supervised RD

A world without chocolate?! Two of the world's biggest chocolate makers could face a shortage

Well, there goes your sweet tooth

If you like your chocolate you can keep your chocolate.

Thankfully, this wouldn't bother me one bit. I love caramel.

×K

Rumor Early Detection

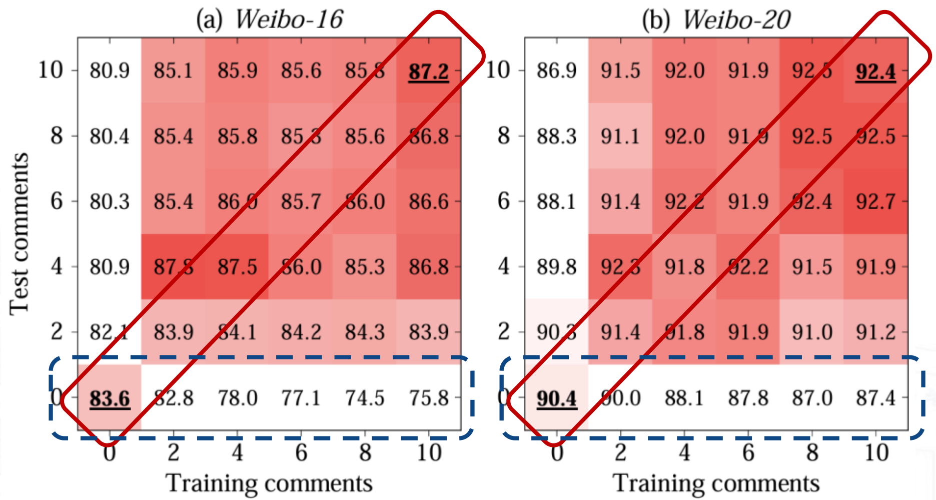
A world without chocolate?! Two of the world's biggest chocolate makers could face a shortage

…

- Previous RD models always assume that user comments are sufficient to support the detection
- User engagement is limited, resulting in few or even no comments available

Preliminary Analysis

How do the dynamics of comments affect model performance?



C2

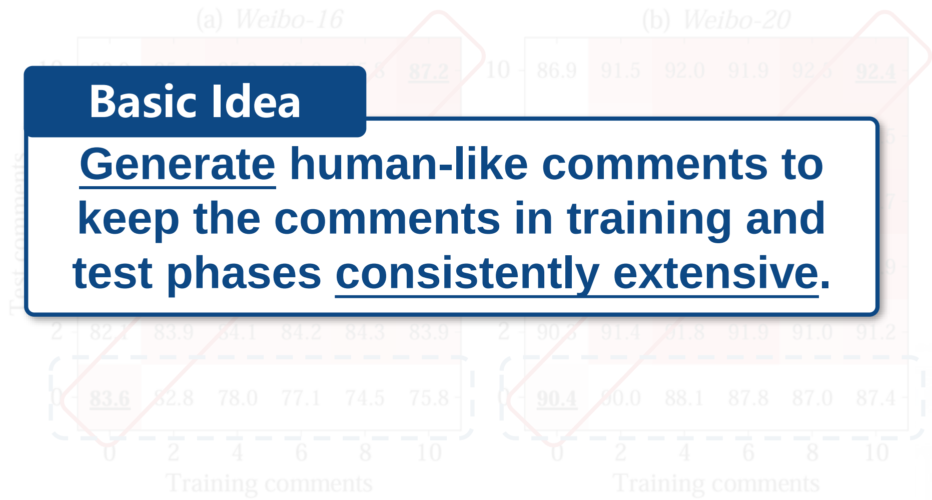
The model performs best when the training and test comments is consistently extensive

C1

Limited comments in early scenarios significantly reduce model performance

Preliminary Analysis

How the dynamics of comments affect model performance?



Basic Idea

Generate human-like comments to keep the comments in training and test phases consistently extensive.

C2

The model performs best when the training and test comments is consistently extensive

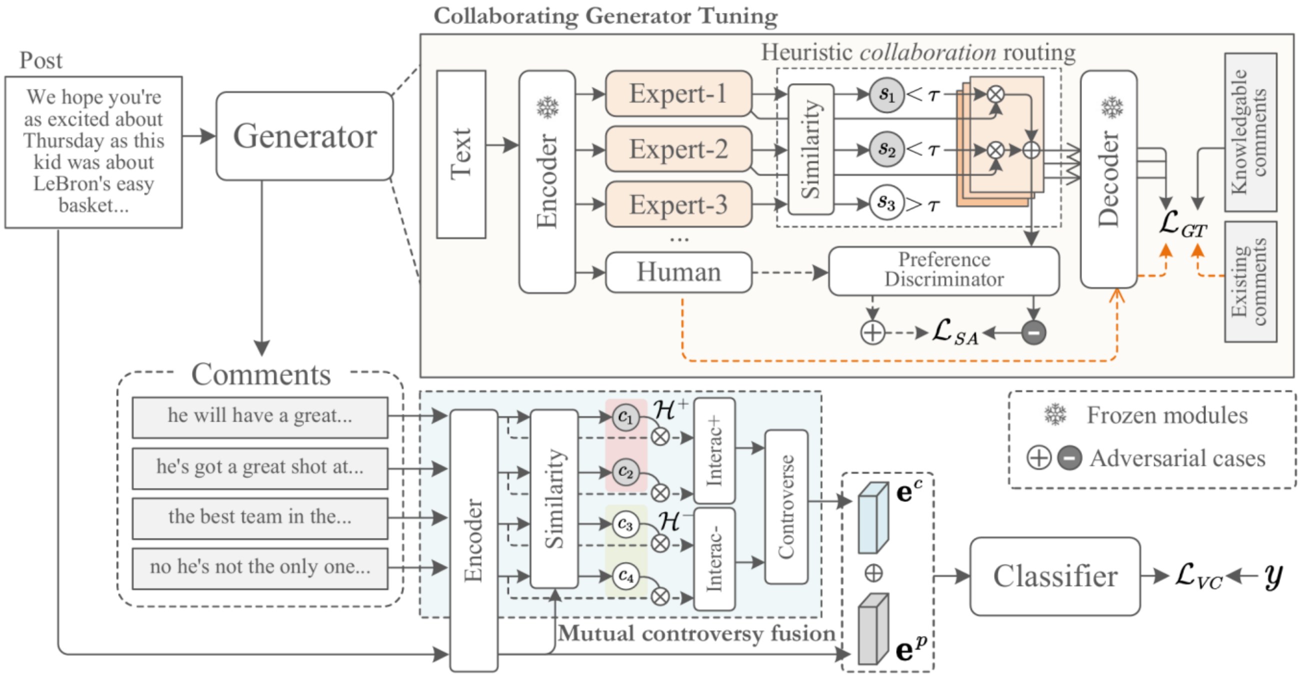
C1

Limited comments in early scenarios significantly reduce model performance

Overview of CAMERED

Basic Idea

Generate human-like comments to keep the training and test comments consistently extensive.



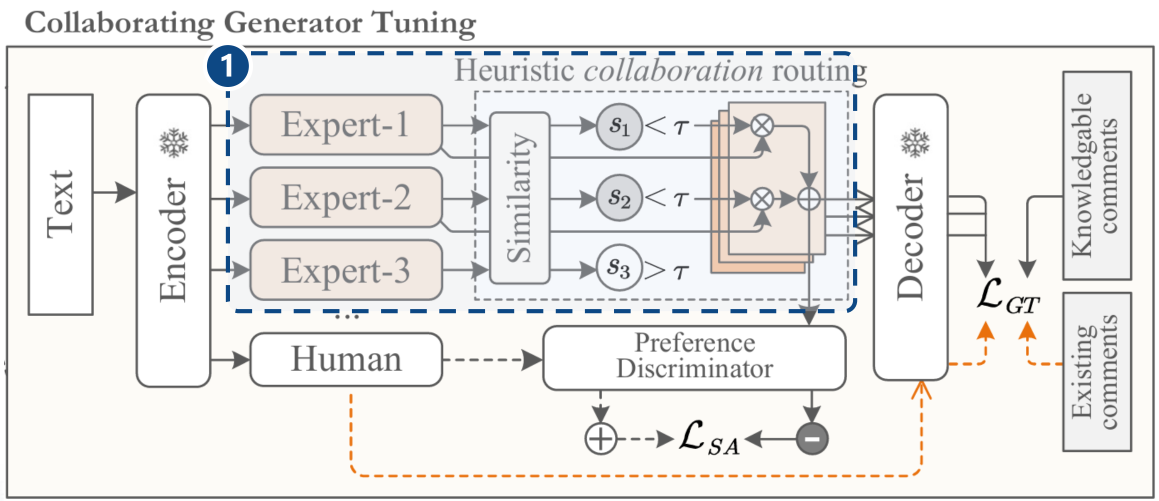
1 Collaborating Generator Tuning

Tuning a generator to produce **diverse**, **knowledgeable**, and **human-like** comments

2 Mutual Controversy Fusion

Integrating generated and original comments by grouping comments with their **stances**

Collaborating Generator Tuning



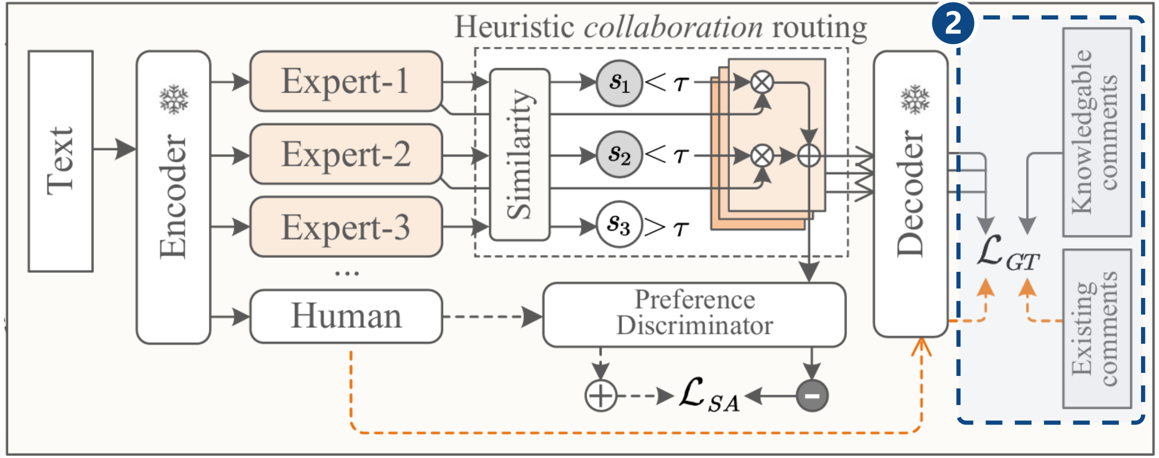
1 Multiple Experts Structure

- Injecting **tunable MoE** into frozen pre-trained language models
- Grouping experts with **similar semantics** for heuristic routing

$$\mathbf{A}_i = \{a_{ilm}\}_{l,m \in \{1,\cdots,L\}}, \quad a_{ilm} = \frac{\mathbf{h}_{il} \cdot \mathbf{h}_{im}}{\|\mathbf{h}_{il}\| \times \|\mathbf{h}_{im}\|}.$$

Collaborating Generator Tuning

Collaborating Generator Tuning

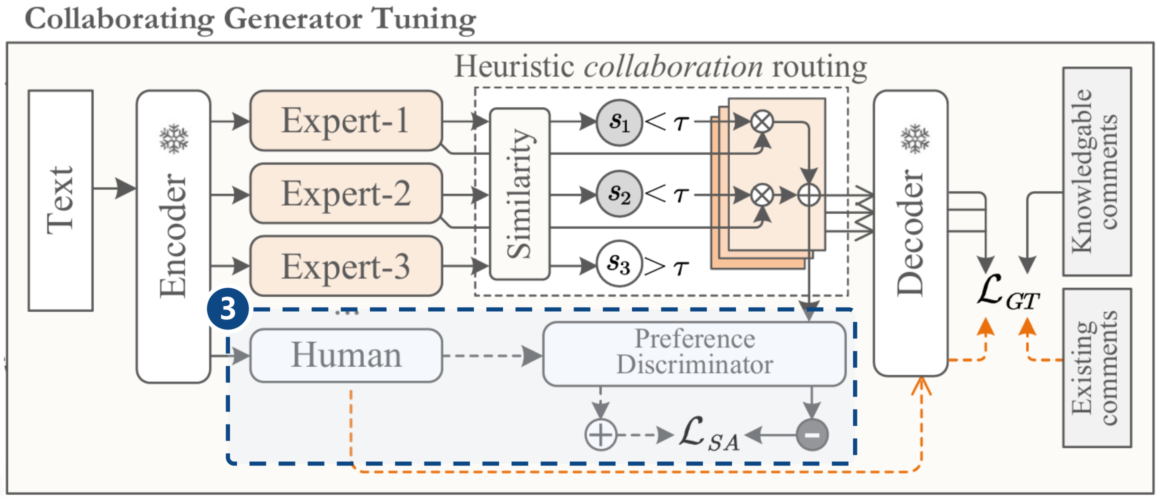


2 Knowledgeable Data Synthesis

- **To make generated comments knowledge-able, synthesizing** training comments through entity descriptions.

$$\min_{\phi_{1:L}} \mathcal{L}_{GT} = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\frac{1}{|\mathbf{c}_{ij}|}\sum_{k=1}^{|\mathbf{c}_{ij}|}\ell_{CE}\Big(\mathcal{G}_{\pi}\big(\mathbf{x}_i,\mathbf{c}_{ij<k}\big),c_{ijk}\Big) + \frac{1}{|\mathcal{D}_\kappa|}\sum_{i=1}^{|\mathcal{D}_\kappa|}\frac{1}{|\mathbf{c}_i^\kappa|}\sum_{k=1}^{|\mathbf{c}_i^\kappa|}\ell_{CE}\Big(\mathcal{G}_{\pi}\big(\mathbf{x}_i,\mathbf{c}^\kappa_{i<k}\big),c^\kappa_{ik}\Big).$$

Collaborating Generator Tuning

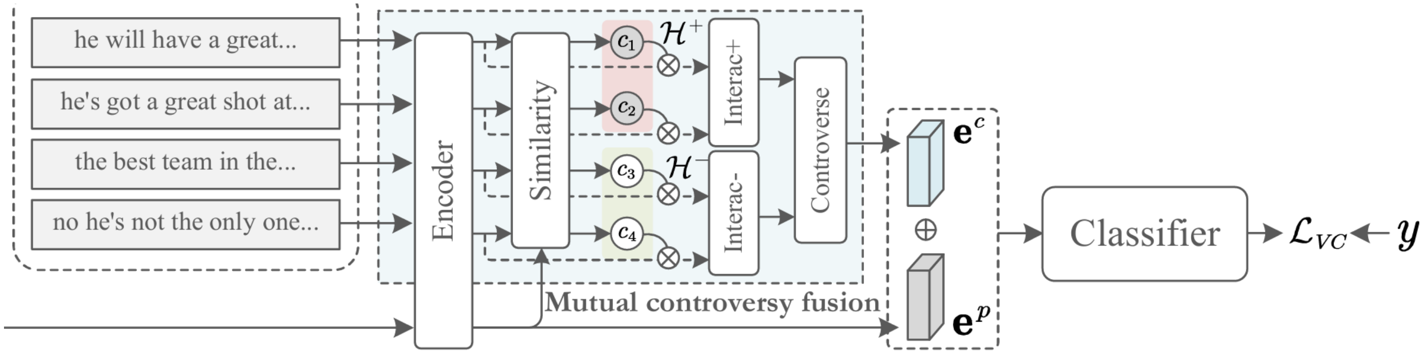


3 Adversarial Style Alignment

- Training an additional expert model to simulate the **language style of humans**
- **Adversarially fooling** the style discriminator to confuse styles from experts and humans

$$\max_{\phi_{1:L}} \min_{\mathbf{W}_S} \mathcal{L}_{SA} = \frac{1}{N + |\mathcal{D}_\kappa|} \sum_{i=1}^{N+|\mathcal{D}_\kappa|} \ell_{CE}(\mathbf{o}_i \mathbf{W}_S, 0) + \ell_{CE}(\mathbf{h}_i^H \mathbf{W}_S, 1).$$

Mutual Controversy Fusion



I. Grouping generated and original comments into **two subsets** by semantic similarities

II. Extracting **features** of comments in two subsets, respectively

III. **Fusing** them into one comment feature and feed it into the classifier

$$\begin{cases} \mathcal{H}_i^+ \leftarrow \mathbf{h}_{ij}^c, & \xi_{ij} > \tau, \\ \mathcal{H}_i^- \leftarrow \mathbf{h}_{ij}^c, & \text{otherwise.} \end{cases} \qquad \xi_{ij} = 1 - \frac{\mathbf{h}_{ij}^c \cdot \mathbf{e}_i^p}{\|\mathbf{h}_{ij}^c\| \times \|\mathbf{e}_i^p\|},$$

Experimental Results on 5 Baselines and 4 Datasets

JILIN UNIVERSITY

Training: 16 original comments = **Test: 2 original and 14 generated comments**

Model	Dataset: *Twitter15* [25] Acc.	F1	AUC	P.	R.	Avg. Δ	Dataset: *Weibo16* [24] Acc.	F1	AUC	P.	R.	Avg. Δ
cBERT [7]	76.07±1.3	75.49±1.2	91.72±0.3	76.25±2.0	76.12±1.3	-	82.96±0.7	81.85±0.5	81.84±0.3	81.98±1.0	81.84±0.3	-
+ CGT (ours)	79.29±1.8*	78.91±1.7*	93.24±0.3*	79.85±1.4*	79.35±1.8*	**+3.00**	84.94±0.6*	83.87±0.7*	83.67±1.0*	84.18±0.7*	83.67±1.0*	**+1.97**
dEFEND [33]	75.72±2.0	75.17±2.2	91.44±0.7	75.62±2.0	75.82±1.6	-	82.80±1.0	81.98±0.8	82.60±0.8	81.87±0.8	82.60±0.8	-
+ CGT (ours)	79.82±1.5*	79.27±1.8*	92.42±0.9*	79.97±1.9*	79.80±1.5*	**+3.50**	84.22±0.9*	83.35±0.8*	83.63±0.8*	83.26±1.0*	83.63±0.8*	**+1.25**
BERTEmo [58]	75.90±1.9	75.39±2.0	91.63±0.4	76.02±1.6	75.87±1.5	-	82.75±0.9	81.72±0.9	81.85±0.8	81.62±0.9	81.85±0.8	-
+ CGT (ours)	78.57±1.9*	77.73±2.3*	93.25±0.7*	79.63±1.7*	78.67±1.8*	**+2.61**	84.72±1.1*	83.68±0.7*	83.52±0.7*	83.87±0.8*	83.52±0.7*	**+1.90**
KAHAN [38]	75.89±2.0	75.70±1.9	92.58±0.3	76.21±1.8	75.91±2.0	-	82.93±1.1	81.83±0.8	81.87±1.1	81.95±1.0	81.87±1.1	-
+ CGT (ours)	78.57±1.0*	78.15±1.0*	92.11±0.5	79.77±1.7*	78.57±1.0*	**+2.18**	84.38±1.1*	83.45±1.0*	83.57±1.1*	83.46±0.9*	83.57±1.1*	**+1.60**
CAS-FEND [29]	75.18±1.2	74.99±1.2	91.56±0.6	75.13±1.1	75.20±1.3	-	83.25±0.6	81.69±0.6	80.99±0.8	83.00±1.1	80.99±0.8	-
+ CGT (ours)	78.93±1.8*	78.86±1.8*	92.31±0.5*	79.39±1.5*	78.89±1.3*	**+3.26**	84.54±0.7*	83.33±0.8*	82.93±0.9*	83.95±0.9*	82.93±0.9*	**+1.55**
CAMERED	76.43±1.6	76.18±1.4	91.92±0.3	77.12±1.5	76.47±1.7	-	83.72±0.3	82.66±0.4	82.60±0.4	82.72±0.4	82.60±0.4	-
+ CGT (ours)	80.36±1.3*	80.11±1.4*	93.54±1.0*	80.68±1.6*	80.28±1.5*	**+3.37**	86.21±1.0*	85.32±1.0*	85.29±1.1*	85.48±1.2*	85.29±1.1*	**+2.66**

Model	Dataset: *Twitter16* [25] Acc.	F1	AUC	P.	R.	Avg. Δ	Dataset: *Weibo20* [58] Acc.	F1	AUC	P.	R.	Avg. Δ
cBERT [7]	74.54±2.5	74.21±2.0	92.27±1.8	75.50±2.7	74.54±2.5	-	86.16±0.6	86.14±0.6	86.19±0.6	86.47±0.6	86.19±0.6	-
+ CGT (ours)	78.44±1.6*	78.40±1.6*	93.77±0.7*	79.52±1.2*	78.85±1.7*	**+3.58**	88.33±0.8*	88.32±0.8*	88.32±0.8*	88.38±0.8*	88.32±0.8*	**+2.10**
dEFEND [33]	72.98±1.9	72.82±2.1	92.50±0.6	75.25±1.9	72.96±2.0	-	86.28±0.5	86.26±0.5	86.27±0.5	86.36±0.4	86.27±0.5	-
+ CGT (ours)	77.66±1.5*	77.65±1.6*	93.72±0.3*	79.00±1.5*	78.15±1.4*	**+3.93**	88.38±0.8*	88.38±0.8*	88.39±0.8*	88.41±0.8*	88.39±0.8*	**+2.10**
BERTEmo [58]	74.02±2.4	73.83±2.4	92.23±2.1	75.16±1.9	74.26±2.6	-	86.03±0.9	86.00±0.9	86.05±0.9	86.33±0.8	86.05±0.9	-
+ CGT (ours)	77.14±1.7*	77.12±1.7*	92.87±1.8	77.87±2.2*	77.35±1.8*	**+2.57**	88.08±0.7*	88.08±0.7*	88.09±0.7*	88.13±0.7*	88.09±0.7*	**+2.00**
KAHAN [38]	74.80±1.9	74.89±2.0	91.04±1.2	75.41±2.0	74.93±2.0	-	86.13±0.3	86.13±0.3	86.14±0.3	86.20±0.3	86.14±0.3	-
+ CGT (ours)	77.92±0.9*	77.98±0.9*	92.28±0.3*	78.21±1.0*	78.11±0.9*	**+2.68**	88.25±0.5*	88.25±0.5*	88.26±0.5*	88.27±0.5*	88.26±0.5*	**+2.11**
CAS-FEND [29]	73.76±1.8											
+ CGT (ours)	78.44±1.6*											
CAMERED	75.06±1.9											

+ our generated comments

+ our comment fusion module

Our method consistently and significantly improves the performance of baseline models

Experimental Results on 5 Baselines and 4 Datasets

Train: 2 original and 14 generated comments = **Test: 2 original and 14 generated comments**

Model	Dataset: *Twitter15* [25]						Dataset: *Weibo16* [24]					
	Acc.	F1	AUC	P.	R.	Avg. Δ	Acc.	F1	AUC	P.	R.	Avg. Δ
cBERT [7]	76.61±1.4	76.02±1.5	91.86±1.0	76.83±1.3	76.64±1.5	-	82.56±0.5	81.23±0.7	80.97±1.2	81.79±0.6	80.97±1.2	-
+ CGT (ours)	78.22±1.5*	77.81±1.2*	91.88±0.3	78.99±1.2*	78.25±1.6*	**+1.44**	84.88±1.0*	83.91±0.9*	83.91±1.1*	84.21±1.3*	83.91±1.1*	**+2.66**
dEFEND [33]	76.79±1.2	75.77±1.3	92.12±0.8	78.39±1.0	76.96±1.1	-	82.58±1.2	81.31±1.1	81.03±1.2	81.75±0.8	81.03±1.2	-
+ CGT (ours)	79.42±0.9*	79.08±1.2*	93.01±0.5*	80.34±1.1*	79.56±0.9*	**+2.28**	84.91±0.4*	83.66±0.3*	83.13±0.4*	84.50±0.8*	83.13±0.4*	**+2.33**
BERTEmo [58]	76.43±1.0	75.70±1.2	92.54±0.6	77.76±1.4	76.61±1.5	-	82.69±1.3	81.60±1.1	81.64±0.8	81.79±1.1	81.64±0.8	-
+ CGT (ours)	78.57±0.9*	77.73±0.9*	93.25±0.7*	79.63±1.2*	78.67±1.3*	**+1.76**	84.38±0.6*	83.28±0.8*	83.14±0.9*	83.52±0.7*	83.14±0.9*	**+1.62**
KAHAN [38]	76.43±1.2	75.89±1.2	92.56±0.7	76.40±1.1	76.42±1.2	-	82.51±1.1	81.24±1.0	81.01±0.8	81.64±1.0	81.01±0.8	-
+ CGT (ours)	79.29±1.1*	78.66±0.9*	93.17±0.6*	80.02±1.2*	79.38±1.1*	**+2.56**	84.41±0.7*	83.46±0.7*	83.55±0.9*	83.46±0.8*	83.55±0.9*	**+2.20**
CAS-FEND [29]	75.54±0.8	75.33±0.8	91.39±0.7	75.68±0.8	75.45±0.7	-	83.19±1.1	81.99±1.0	81.77±0.7	82.37±1.6	81.77±0.7	-
+ CGT (ours)	78.93±0.9*	78.66±0.9*	92.66±0.7*	79.16±1.0*	78.90±0.8*	**+2.98**	84.51±0.8*	82.95±1.0*	82.64±1.0*	83.90±0.9*	82.64±1.0*	**+1.11**
CAMERED	76.79±1.5	76.30±1.2	91.79±0.0	77.98±1.8	76.93±1.1	-	82.23±0.7	82.00±0.5	81.76±0.7	82.60±1.1	81.76±0.7	-
+ CGT (ours)	80.00±1.3*	79.46±1.3*	92.75±0.8*	81.30±1.2*	79.98±1.2*	**+2.74**	86.00±1.1*	85.13±1.4*	85.18±1.5*	85.08±1.4*	85.18±1.5*	**+3.24**

Model	Dataset: *Twitter16* [25]						Dataset: *Weibo20* [58]					
	Acc.	F1	AUC	P.	R.	Avg. Δ	Acc.	F1	AUC	P.	R.	Avg. Δ
cBERT [7]	75.32±2.4	75.07±2.5	93.18±0.9	76.90±2.1	75.67±2.4	-	85.84±0.4	85.82±0.4	85.86±0.4	86.01±0.3	85.86±0.4	-
+ CGT (ours)	77.92±2.0*	77.96±2.1*	94.07±0.3*	78.53±2.2*	78.26±2.1*	**+2.12**	87.78±0.3*	87.78±0.3*	87.79±0.3*	87.85±0.3*	87.79±0.3*	**+1.92**
dEFEND [33]	75.06±2.1	75.09±2.0	93.31±0.5	76.49±1.4	75.34±2.0	-	85.87±0.7	85.86±0.7	85.88±0.7	85.97±0.7	85.88±0.7	-
+ CGT (ours)	77.92±2.4*	77.86±2.4*	94.32±1.0*	79.65±1.9*	78.34±2.0*	**+2.56**	87.94±0.7*	87.93±0.7*	87.95±0.7*	88.03±0.8*	87.95±0.7*	**+2.07**
BERTEmo [58]	74.80±1.9	74.78±1.9	92.22±1.3	75.83±2.3	74.95±2.0	-	85.62±1.0	85.60±1.0	85.64±1.0	85.80±0.9	85.64±1.0	-
+ CGT (ours)	77.40±1.8*	77.48±1.8*	94.11±0.4*	78.85±1.7*	77.66±1.5*	**+2.58**	87.88±0.3*	87.87±0.3*	87.88±0.3*	87.92±0.3*	87.88±0.3*	**+2.23**
KAHAN [38]	74.80±2.3	74.85±2.2	92.32±0.6	75.22±2.0	74.89±2.3	-	85.90±0.7	85.89±0.7	85.92±0.7	86.06±0.7	85.92±0.7	-
+ CGT (ours)	78.25±1.6*	78.37±1.7*	92.44±0.4	78.55±1.5*	78.37±1.8*	**+2.78**	88.00±0.6*	87.99±0.6*	88.02±0.6*	88.18±0.5*	88.02±0.6*	**+2.10**
CAS-FEND [29]	74.55±1.4	74.59±1.3	92.34±1.0	75.45±1.5	74.85±1.4	-	85.74±0.5	85.73±0.5	85.76±0.5	85.85±0.5	85.76±0.5	
+ CGT (ours)	78.70±0.7*	78.71±0.7*	93.60±0.7*	78.78±0.6*	78.96±0.6*	**+3.39**	87.72±0.7*	87.72±0.7*	87.73±0.7*	87.80±0.6*	87.73±0.7*	**+1.89**
CAMERED	75.84±1.9	75.93±2.0	93.30±1.1	77.05±1.5	76.33±1.8	-	85.90±0.6	85.90±0.6	85.91±0.6	85.93±0.6	85.91±0.6	-
+ CGT (ours)	79.48±2.1*	79.52±2.1*	94.64±0.7*	80.77±2.0*	79.86±2.1*	**+3.16**	88.14±0.5*	88.13±0.5*	88.16±0.5*	88.25±0.5*	88.16±0.5*	**+2.26**

+ our generated comments

+ our comment fusion module

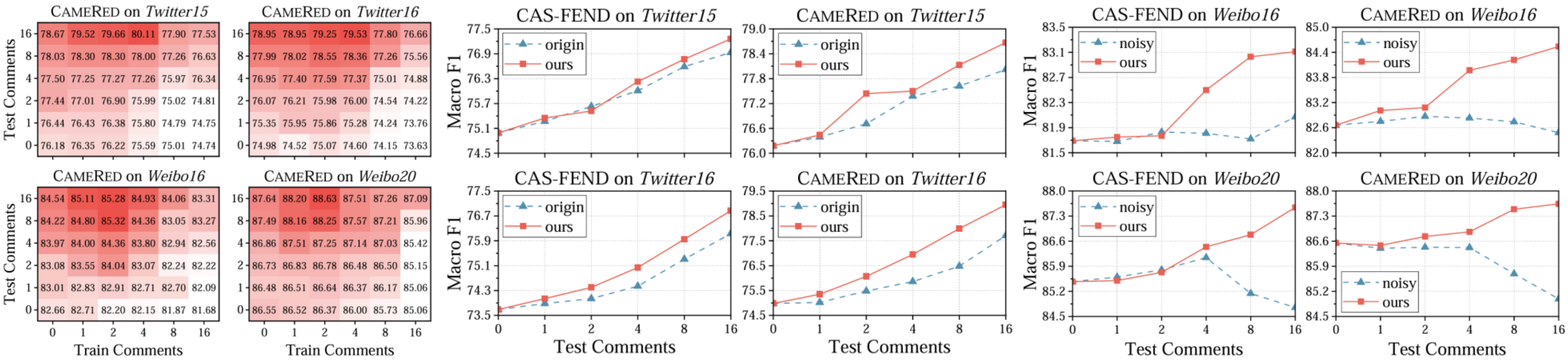
Compared with Generation Methods

Model	Dataset: *Twitter15* [25]			Dataset: *Twitter16* [24]			Dataset: *Weibo16* [25]			Dataset: *Weibo20* [24]			Avg. Δ
	Acc.	F1	AUC	Acc.	F1	AUC	Acc.	F1	AUC	Acc.	F1	AUC	
CAS-FEND [29]	71.18	74.99	91.56	73.76	73.69	91.44	83.25	81.69	80.99	85.51	85.47	85.54	-
+ CGT *w/* T5 [6]	77.92	78.09	92.79	78.04	77.43	92.51	83.75	82.47	82.08	86.86	86.86	86.86	+2.22
+ CGT *w/* Llama [37]	78.93	78.86	92.31	78.44	78.32	93.90	84.54	83.33	82.93	87.92	87.92	87.93	+3.02
+ DELL *w/* Llama [41]	77.80	77.64	92.72	77.86	77.26	91.20	83.85	82.48	81.92	86.84	86.83	86.86	+2.02
+ GenFEND *w/* Llama [28]	77.92	77.88	92.76	77.86	77.25	91.83	83.86	82.55	82.13	87.03	87.03	87.03	+2.17
CAMERED *w/o* CGT	76.43	76.18	91.92	75.06	74.98	92.92	83.72	82.66	82.60	86.56	86.55	86.56	-
+ CGT *w/* T5 [6]	79.02	79.29	93.02	78.75	78.30	93.02	85.07	84.12	84.13	87.85	87.84	87.85	**+1.84**
+ CGT *w/* Llama [37]	80.36	80.11	93.54	79.55	79.53	93.69	86.21	85.32	85.29	88.63	88.63	88.65	**+2.78**
+ DELL *w/* Llama [41]	78.90	79.07	92.45	78.68	78.45	92.86	85.01	84.00	83.87	87.65	87.65	87.65	**+1.68**
+ GenFEND *w/* Llama [28]	79.22	79.30	92.62	78.83	78.46	92.51	85.09	84.10	84.00	87.80	87.80	87.80	**+1.78**

Our generation method outperforms SOTA generation methods DELL and GenFEND

Our T5(220M)-based generator performs consistently Llama(7B)-based SOTA generator

Sensitivity Analysis



1 The model performs best when the total number of comments Is balanced between training and testing

2 Our generated comments even outperform human-written comments in the original dataset

3 Comments with noisy writing styles consistently decrease the model performance

Conclusion

- We first empirically reveal a conclusion: **the model performs best when the training and test comments is consistently extensive**
- We tune a comment generator to produce **diverse, knowledgeable, and human-like** comments to keep the comments in training and test phases consistently extensive
- We integrate original and generated comments by designing a **mutual controversy fusion** module
- **Extensive experiments** are conducted to demonstrate the performance of our generated comments and comment fusion method

Thanks.

Collaboration and Controversy Among Experts:
Rumor Early Detection by Tuning a Comment Generator

Authors: **Bing Wang** (wangbing1416@gmail.com),

Bingrui Zhao, Ximing Li*, Changchun Li, Renchu Guan, Shengsheng Wang

SIGIR 2025 Padova ITALY

Project page **My homepage**