Integrating Pattern- and Fact-based Fake News Detection via Model Preference Learning
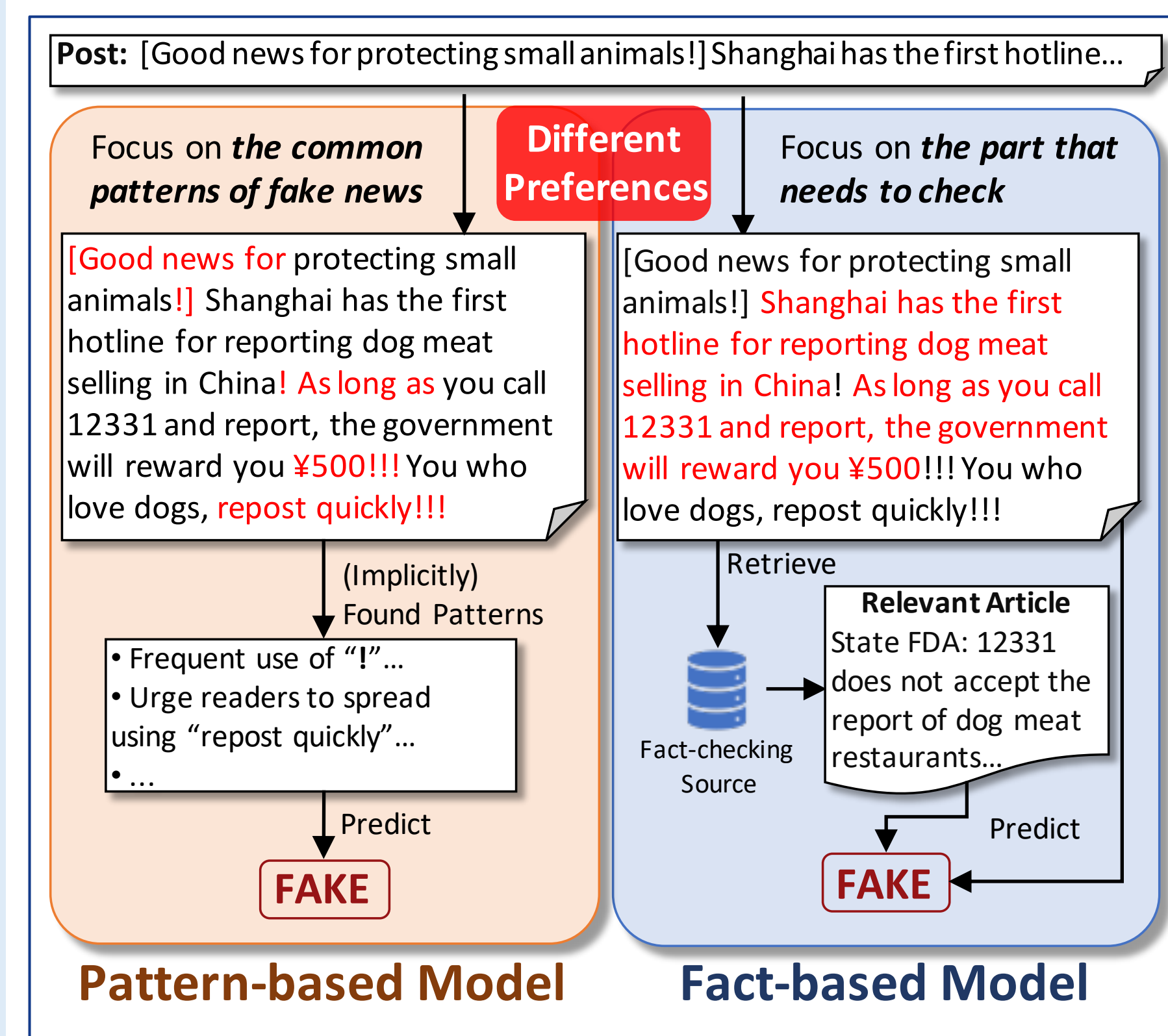
Qiang Sheng*, Xueyao Zhang*, Juan Cao, Lei Zhong   *: Equal Contribution

Key Lab of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences & University of Chinese Academy of Sciences

Introduction

- There are two schools of text-based fake news detection methods.
- They have **different preferences** of textual clues, playing **complementary** roles.

Post: [Good news for protecting small animals!] Shanghai has the first hotline...

Focus on **the common patterns of fake news**   **Different Preferences**   Focus on **the part that needs to check**

[Good news for protecting small animals!] Shanghai has the first hotline for reporting dog meat selling in China! As long as you call 12331 and report, the government will reward you ¥500!!! You who love dogs, repost quickly!!!

[Good news for protecting small animals!] Shanghai has the first hotline for reporting dog meat selling in China! As long as you call 12331 and report, the government will reward you ¥500!!! You who love dogs, repost quickly!!!

(Implicitly) Found Patterns
- Frequent use of "I"...
- Urge readers to spread using "repost quickly"...
- ...

Retrieve

Relevant Article
State FDA: 12331 does not accept the report of dog meat restaurants...

Fact-checking Source

Predict → **FAKE**   Predict → **FAKE**

Pattern-based Model   **Fact-based Model**

Few works consider their integration.
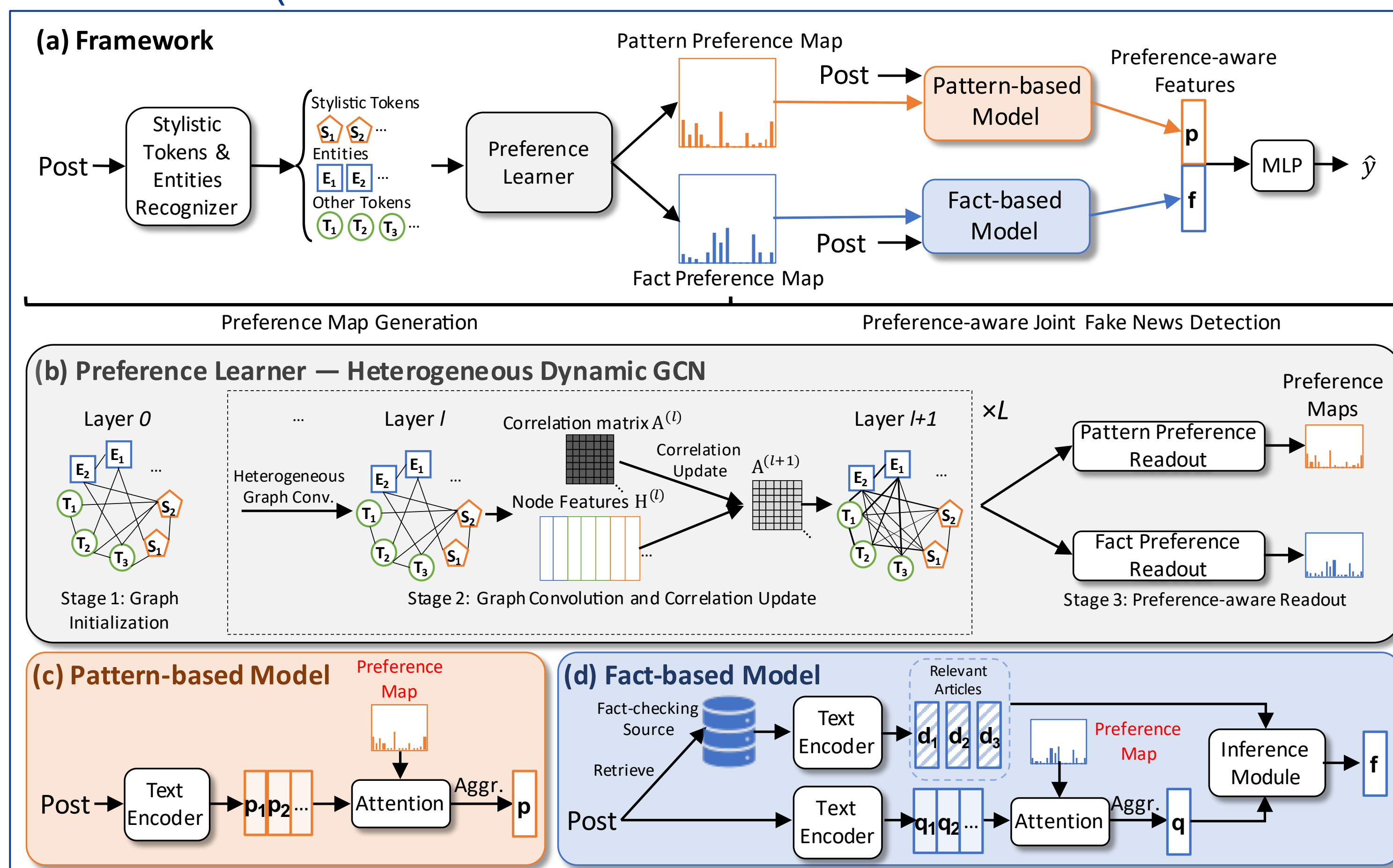
Challenge of Integration

The real models are generally not that ideal...
- A pattern-based model may *memorize* some event-specific words.
- A fact-based model may prefer similar-format evidence, distracting from fact descriptions.

[Good news for protecting small animals!] Shanghai has the first hotline for reporting dog meat selling in China! As long as you call 12331 and report, the government will reward you ¥500!!! You who love dogs, repost quickly!!!

I see many real news samples containing **Shanghai** during training. This post might be real.

I find many seemingly reliable relevant articles with **a title in []**. This post might be real.

$M_{pattern}$ ✗ ✗ M_{fact}

Our Proposed Framework

Pref-FEND (Preference-aware Fake News Detection Framework)

(a) Framework

Preference Map Generation   Preference-aware Joint Fake News Detection

(b) Preference Learner — Heterogeneous Dynamic GCN

Layer 0 ... Layer l Correlation matrix $A^{(l)}$ Correlation Update $A^{(l+1)}$ Layer $l+1$ ×L

Heterogeneous Graph Conv.   Node Features $H^{(l)}$

Stage 1: Graph Initialization   Stage 2: Graph Convolution and Correlation Update   Stage 3: Preference-aware Readout

Pattern Preference Readout   Fact Preference Readout   Preference Maps

(c) Pattern-based Model   Preference Map
Post → Text Encoder → $p_1 p_2$ ... → Attention → Aggr. → p

(d) Fact-based Model   Relevant Articles   Preference Map
Fact-checking Source   Retrieve → $d_1 d_2 d_3$
Post → Text Encoder / Text Encoder → $q_1 q_2$ ... → Attention → Aggr. → q → Inference Module → f

CORE IDEA: Model preferences of the pattern- and the fact-based model and make the learning process more focused.

Step 1: Preference Map Generation *Left part of (a)*

Preference Map: A score distribution of length n where the i-th score represents to what extent the i-th token is *preferred* *by the model.*

① **Stylistic Tokens & Entities Recognition**
   Use *prior knowledge* to guide the map generation.
   ➤ **Pattern Map:** Recognize **Stylistic Tokens** (e.g., emotional words, punctuations)
   ➤ **Fact Map:** Recognize **Entities** (as a verifiable claim generally has >=1 entity)
② **Dynamic Preference Learning**
   Preference Learner: **Heterogeneous Dynamic GCN (HetDGCN), shown in (b)**
   ➤ **Init:** Tokens as nodes with preserving their types *(heterogeneous)* and cosine similarity between each token pair as the initial edge weight.
   ➤ **Graph Conv & Correlation Update:** After standard heterogeneous graph convolution, use the latest representation to update edge weights *(dynamic)*

➤ **Preference-aware Readout:** Use the final correlation matrix, sum all edge weight starting from one node *excluding* those connected with non-preferred nodes.

Step 2: Preference-aware Joint Fake News Detection *Right part of (a)*

- **Attention-based guidance** to Pattern- and Fact-based models to obtain the preference-aware features (shown in (c) and (d)), which are further concatenated for final prediction.
- Besides normal cross-entropy, **two new losses** are used to make the models supervise *mutually*.
  ➤ Similarity Loss. Cosine Similarity between the two Maps.
  ➤ Reverse Cross-entropy Loss. Pattern Map to guide fact-based model; and vice versa. Expect the model to output the wrong prediction.

Evaluation

Table 6: Frequently preferred tokens (separated by "|") in the pattern-preferred and the fact-preferred token set.

Set	Category	Token (Translated into English)
Pattern-preferred	Punctuation	. , . ! ! : ? [["	[(	. . .]] @	<	)	#	~	>	:	/
	Negation	not	no	don't
	Pronouns	we	they	you all
	Others	find	such	think	may	certainly	so	release	some	as	careful	become	focus	loving heart	but	kind of
Fact-preferred	Evidence-related	claim	video	link	webpage	full text	say	picture	investigation	according to	uncover
	Entity-related	China	Beijing	police	place	car	Shanghai	official
	Pronouns	he	its	it	you
	Others	's	done	just	also	and	already	will	wait	go to	do	female	too	want	what	certain	die	pass	death	when	in	second	more	make	suffer	night	society

Table 7: Three examples of fake news posts. Red represents pattern-preferred tokens and blue represents fact-preferred tokens. Darker color indicates a higher preference score.

#	Post (Translated into English)
1	A group of city administration officials in Sishui , Shandong , chased an old man until all his eggs were broken on the ground . The old man sat there helplessly . The officials ran away after hitting . The white-haired man should be about 80 years old , and he can't make much money by selling eggs . So why be aggressive ? Is there no time for the officials to be alone ? If the officials only oppresses citizens , what's the good of having these officials ? You will be punished sooner or later for bullying the underprivileged .
	Ground Truth: Fake **Judgment:** Bi-LSTM (Fake), DeClarE (Real), Pref-FEND (Fake)
2	I A student of ZJU jumping to the West Lake for a crazy graduation photo drowned] On June 29 , Xin (not his real name) from ZJU and his classmates went to the waters near the scenic spot of Konggu Chuanyin " in Gushan , Beili Lake , West Lake in Hangzhou , Xin asked his classmates to take pictures of his swimming underwater , He jumped into the West Lake from the side of Xiling brige on Beishan Road and swam to the lotus pool of Gushan park on the other side , He drowned when swimming to the center of the lake , Recently , he has received a full PhD scholarship from a U.S. university .
	Ground Truth: Fake **Judgment:** Bi-LSTM (Real), DeClarE (Fake), Pref-FEND (Fake)
3	Is anyone in Shanghai interested in raising a dog ? No Charge , Golden Retriever , Poodle , Samoyed and other more kinds There are dog-killing farms being destroyed , If no one adopts , they will be euthanized . Let these little cute lives accompany you If you are really not able to raise them , please forward this message
	Ground Truth: Fake **Judgment:** Bi-LSTM (Real), DeClarE (Real), Pref-FEND (Fake)

- Our Pref-FEND can improve the detection performance of models that prefer patterns, facts, or both. (See in paper)

↖ **Frequently preferred tokens** in the two kinds of maps show their differences, which is captured by Pref-FEND.

← **Cases with preferred tokens highlighted.** With effective guidance, Pref-FEND made correct predictions when the single-preference models did not.

Conclusion

- **Method:** We propose **Pref-FEND** to integrate pattern-based and fact-based fake news detection models in a *preference-aware* fashion.
- **Evaluation:** Experiments on two new datasets show that Pref-FEND outperforms the existing detection models. Further analysis shows that learned preference maps help models be more focused.
- **Data:** We construct two new datasets (Chinese & English) based on the existing ones and crawled web articles.

Our Project