Zoom Out and Observe: News Environment Perception for Fake News Detection

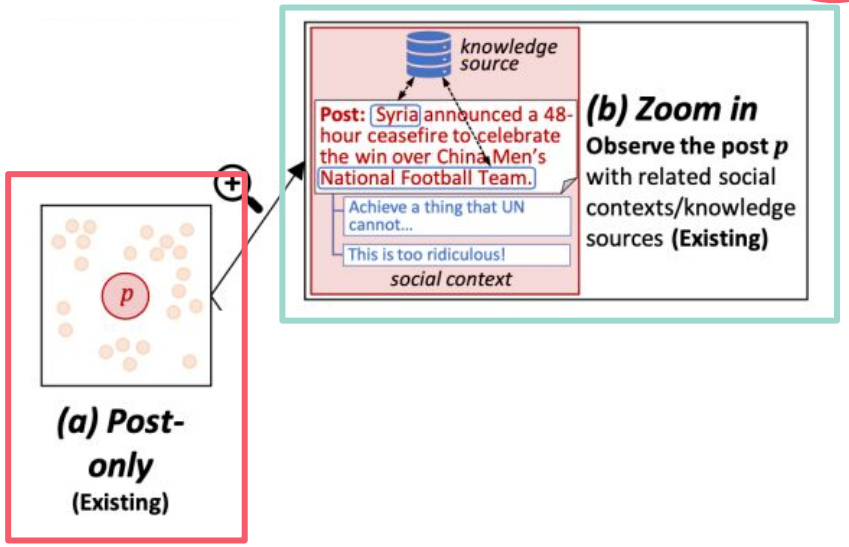
ADVISOR: Jia-Ling Koh
PRESENTER: Xiao-Yuan Hung
SOURCE: ACL' 22
DATE: 2023/01/31

Outline

1. Introduction
   - Problem
   - Solution
   - Example
   - Observation
2. Method
3. Experiment
4. Conclusion

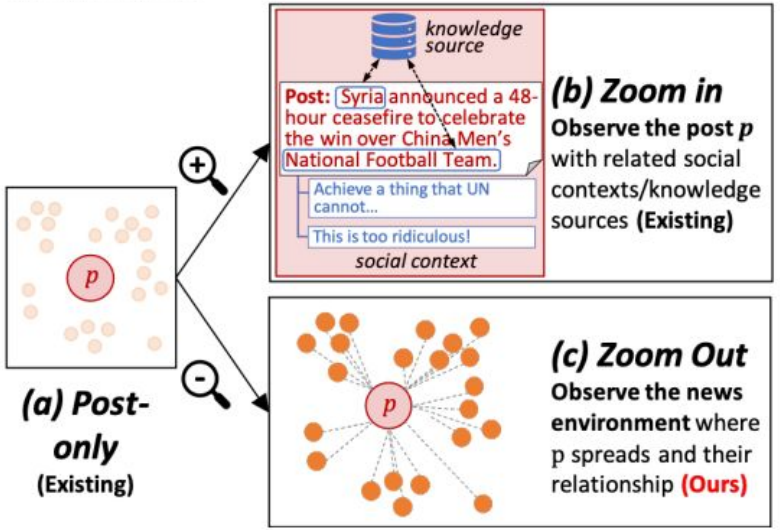
Problem

- **Existing method usually rely on**
  - **Post content**
    - **Pros**
      - **Detect immediately**
    - **Cons**
      - **Single source**
  - **Related post-level signals**
    - **Pros**
      - **Explainable**
    - **Cons**
      - **Social Context**
        - **Need sufficient social context**
      - **Knowledge**
        - **Emergency is difficult to check**

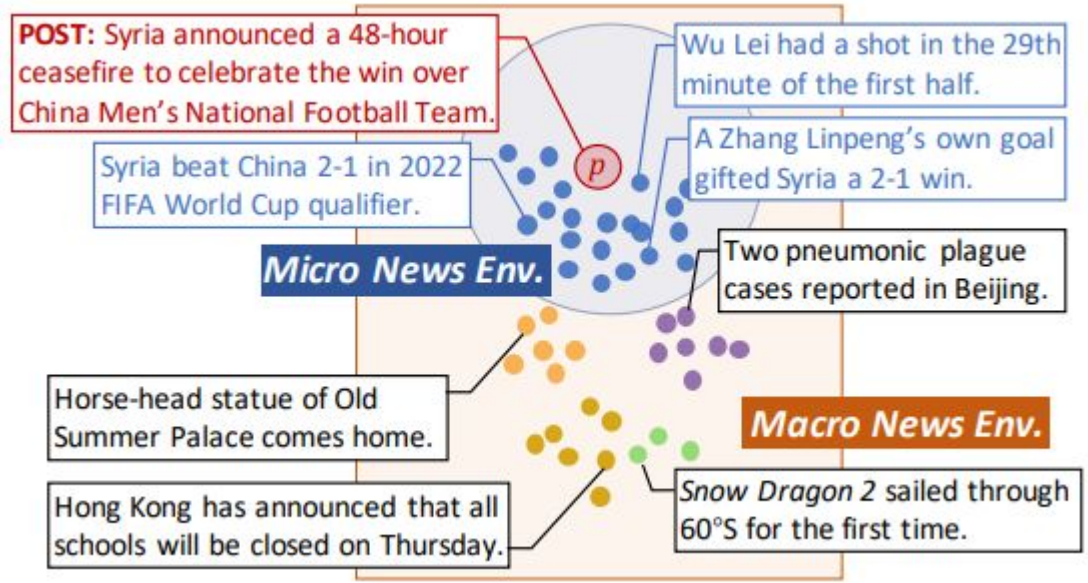


Solution

- Observe the external news environment where a fake news post is created and disseminated

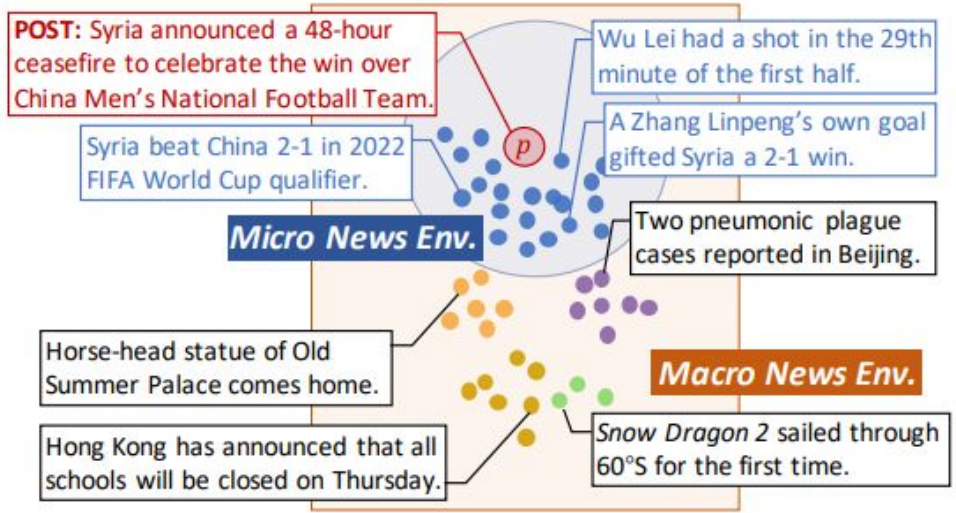


Example


POST: Syria announced a 48-hour ceasefire to celebrate the win over China Men's National Football Team.
Wu Lei had a shot in the 29th minute of the first half.
A Zhang Linpeng's own goal gifted Syria a 2-1 win.
Syria beat China 2-1 in 2022 FIFA World Cup qualifier.
p
Micro News Env.
Two pneumonic plague cases reported in Beijing.
Horse-head statue of Old Summer Palace comes home.
Macro News Env.
Hong Kong has announced that all schools will be closed on Thursday.
Snow Dragon 2 sailed through 60°S for the first time.

Observation

- **Popularity(Macro)**
  - Popular event might spread widely
- **Novelty(Micro)**
  - Catch the audience's attention



Outline

1 Introduction

2 Method

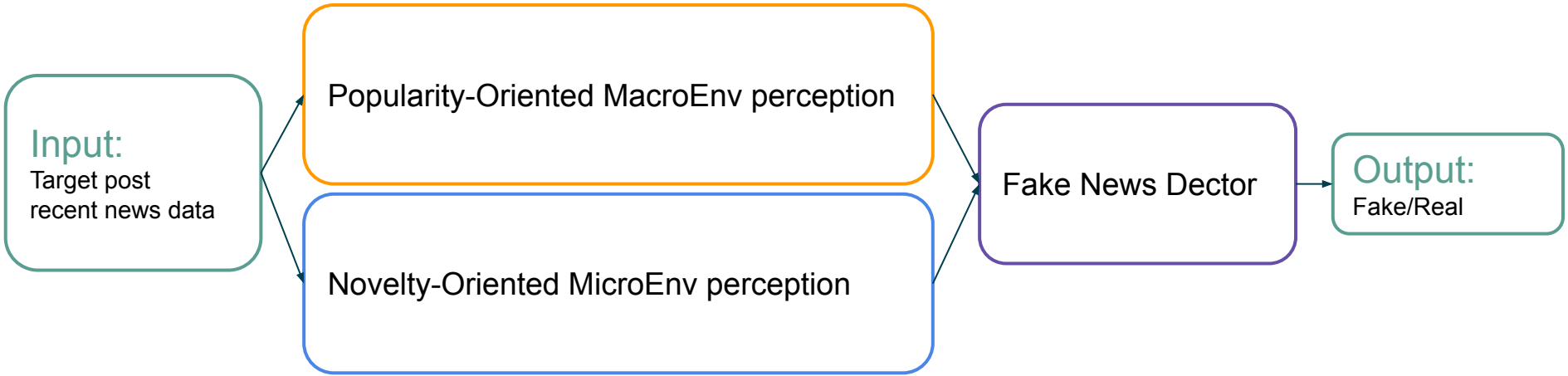
- Architecture
- (a)Constuction
- (b)Perception
- (c)Prediction

3 Experiment

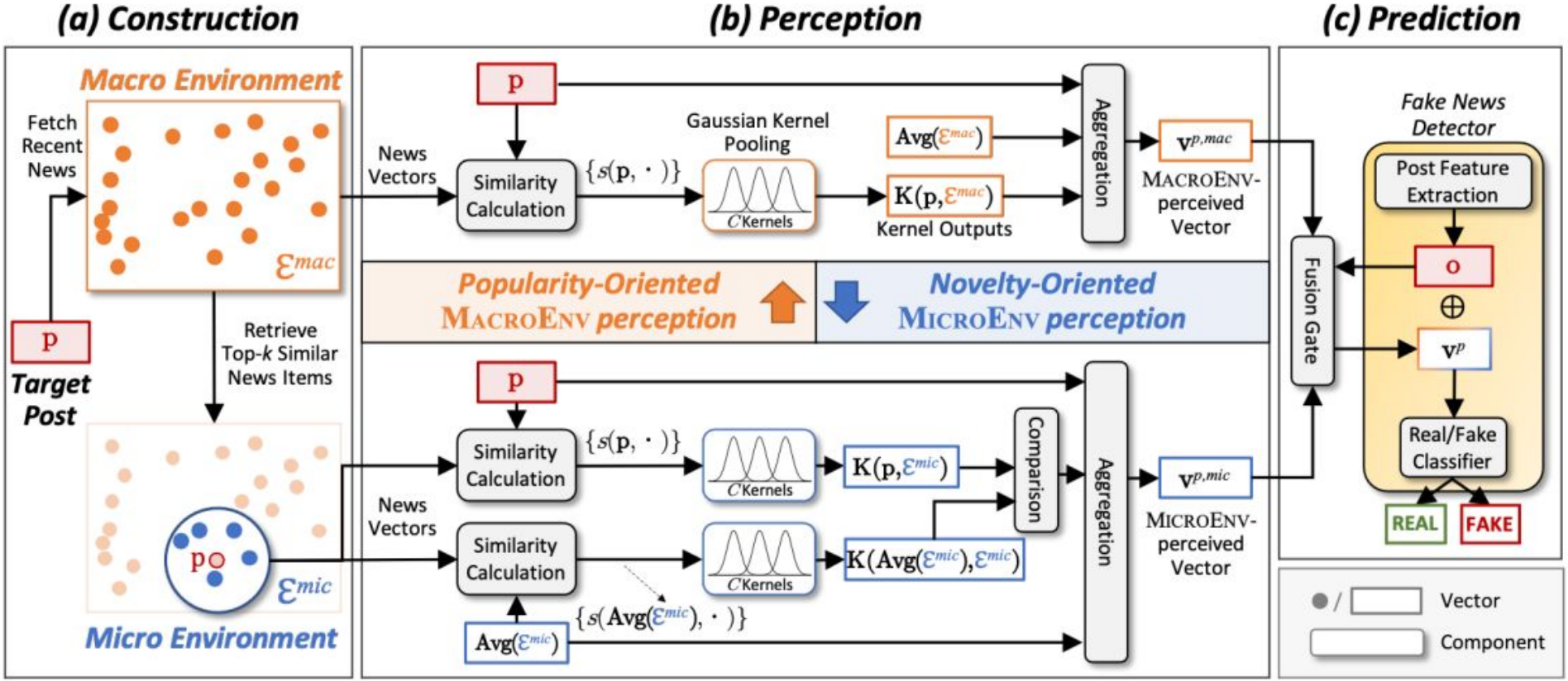
4 Conclusion

News Environment Perception(NEP)

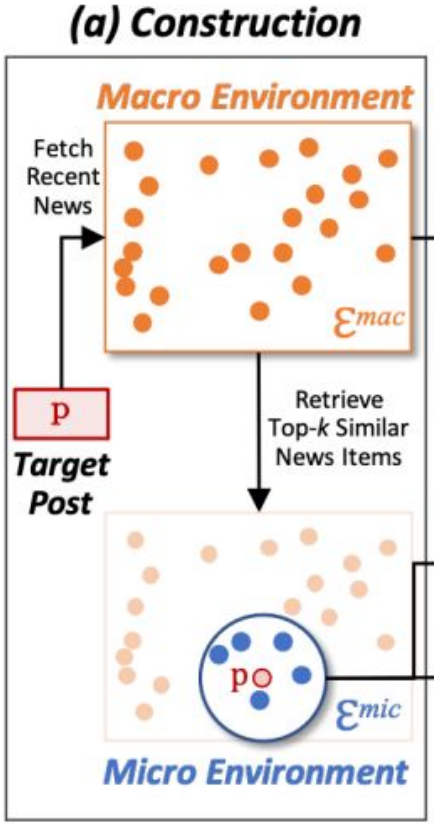
- **Goal**
  - Empower fake news detectors with the effective perception of news environments



Architecture



(a)Constuction



- **ε**
  - Set of all collected news items published earlier than p
- **MarcoEnv**

$$\mathcal{E}^{mac} = \{e : e \in \mathcal{E}, 0 < t_p - t_e \leq T\}$$

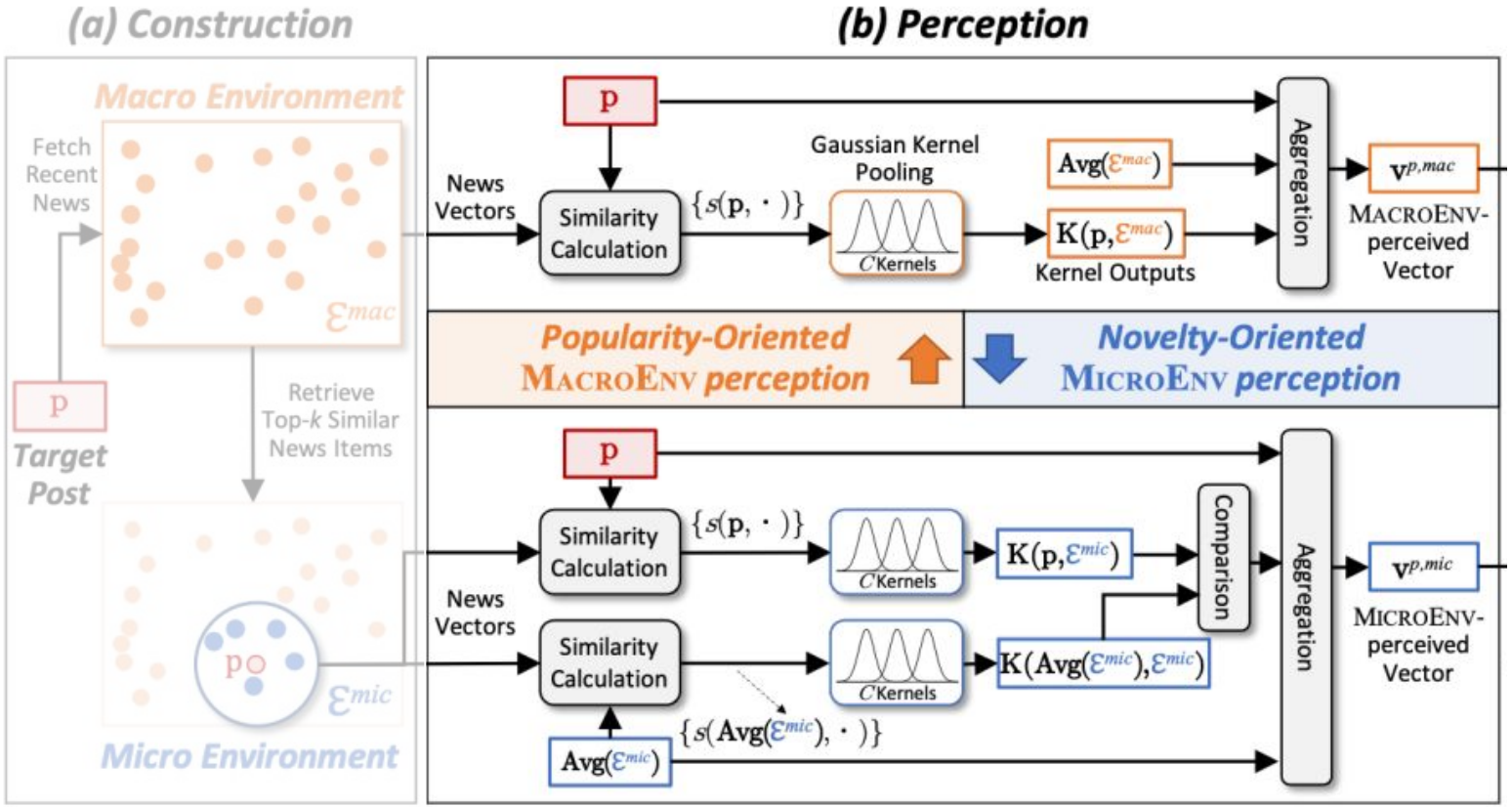
- **MircoEnv**

$$\mathcal{E}^{mic} = \{e : e \in \text{Topk}(p, \mathcal{E}^{mac})\}$$

$$k = \lceil r|\mathcal{E}^{mac}| \rceil \text{ and } r \in (0, 1)$$

T : days, hyperparameter
t_p : publication date of p
t_e : publication date of news item e
r : proportion, hyperparameter

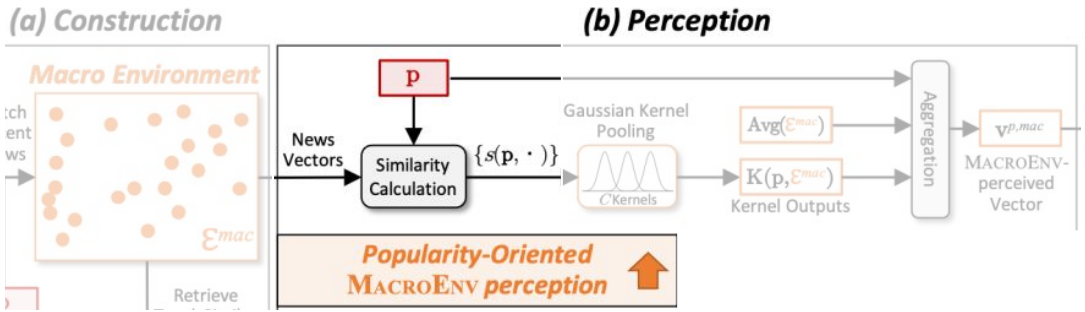
(b)Perception



Popularity-Oriented MACROENV Perception

- **Cosine Similarity**
  - If many items in the MACROENV are similar to p, then p might be also popular in such an environment.

$$s(\mathbf{p}, \mathbf{e}_i) = \frac{\mathbf{p} \cdot \mathbf{e}_i}{\|\mathbf{p}\| \|\mathbf{e}_i\|}.$$



Popularity-Oriented MACROENV Perception

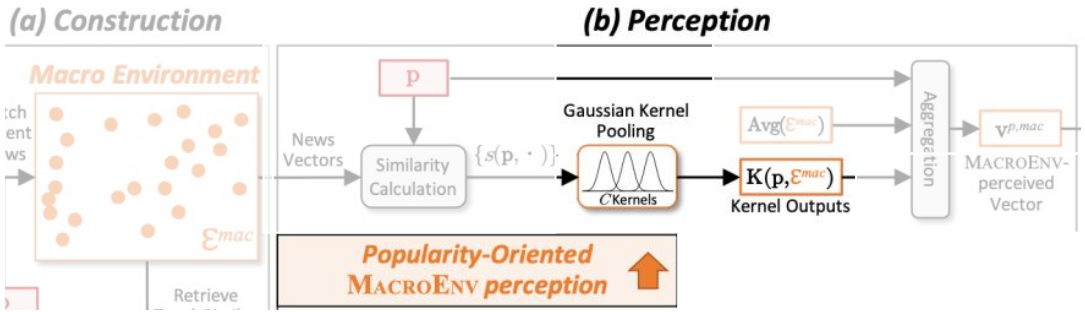
- **Gaussian Kernel Pooling**
  - Transform environment perceived vector into fixed-dimensional vectors

use C kernels $\{\mathbf{K}_i\}_{i=1}^{C}$

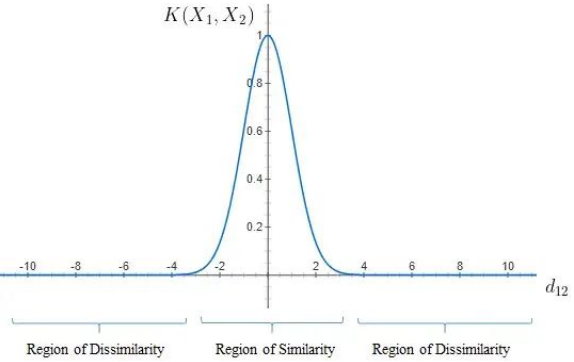
Scatter in [−1, 1] completely and evenly cover the range of cosine similarity

$$\mathbf{K}_k^i = \exp\left(-\frac{(s(\mathbf{p},\mathbf{e}_i)-\mu_k)^2}{2\sigma_k^2}\right), \quad (5)$$

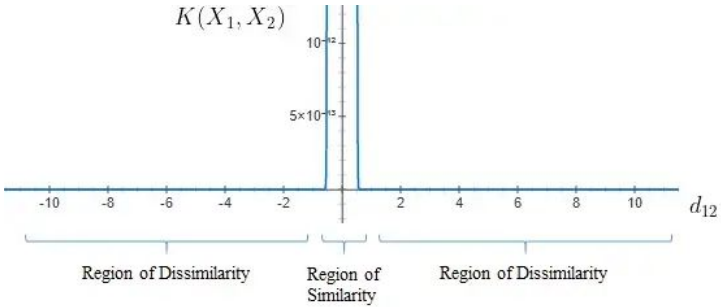
$$\mathbf{K}_k(\mathbf{p},\mathcal{E}^{mac}) = \sum_{i=1}^{|\mathcal{E}^{mac}|} \mathbf{K}_k^i, \quad (6)$$



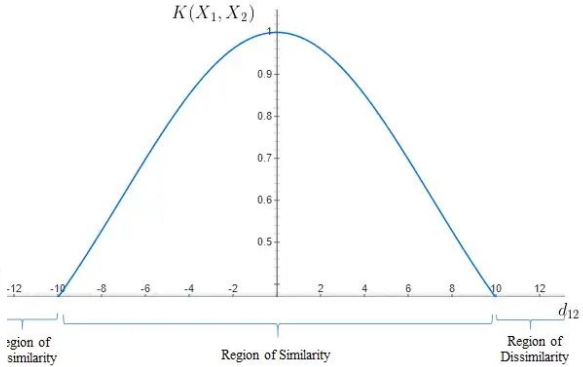
Different Kernels Example



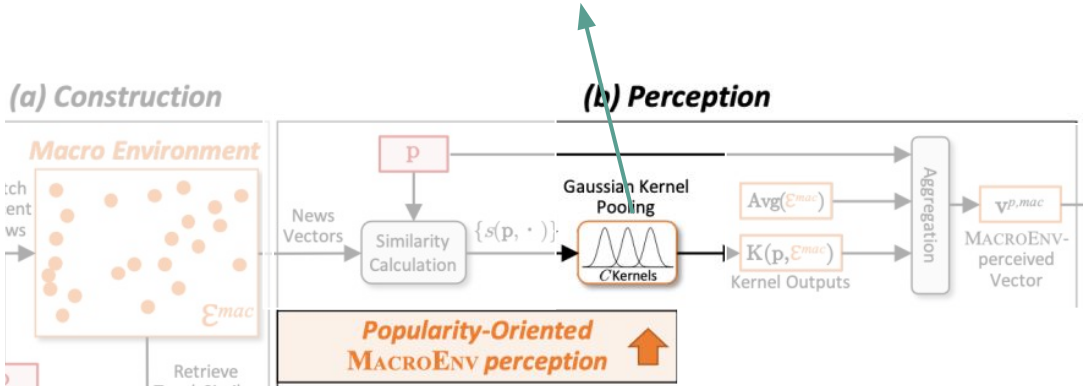
σ = 1



σ = 0.01



σ = 10



Popularity-Oriented MACROENV Perception

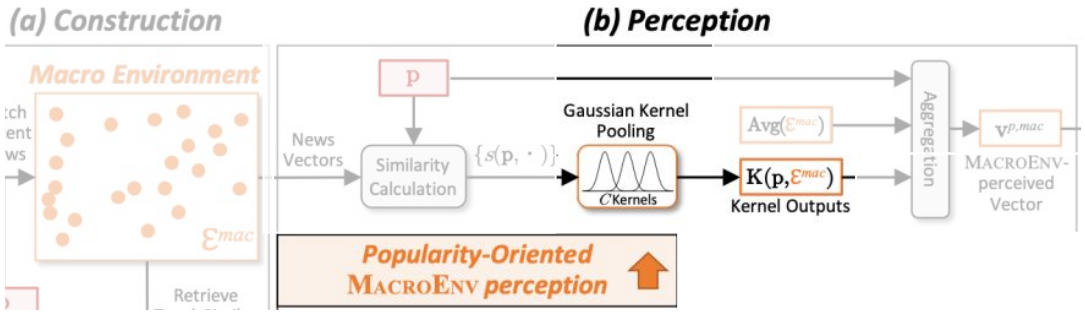
- **Gaussian Kernel Pooling**
  - Transform environment perceived vector into fixed-dimensional vectors

use C kernels $\{\mathbf{K}_i\}_{i=1}^{C}$

$$\mathbf{K}_k^i = \exp\left(-\frac{(s(\mathbf{p},\mathbf{e}_i)-\mu_k)^2}{2\sigma_k^2}\right), \quad (5)$$

$$\mathbf{K}_k(\mathbf{p},\mathcal{E}^{mac}) = \sum_{i=1}^{|\mathcal{E}^{mac}|} \mathbf{K}_k^i, \quad (6)$$

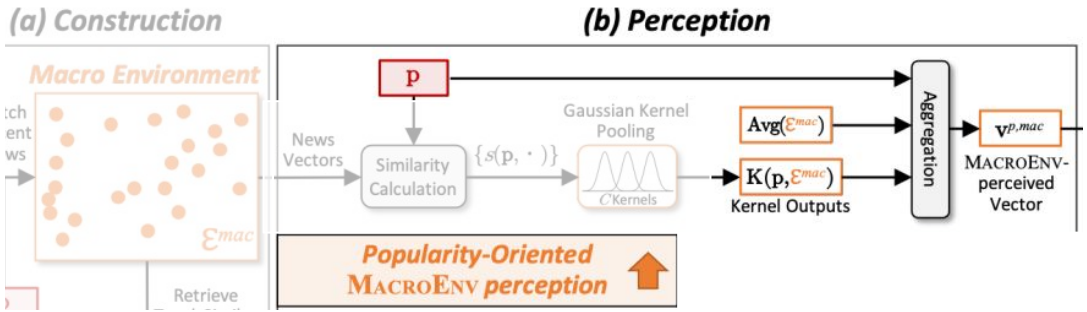
$$\mathbf{K}(\mathbf{p},\mathcal{E}^{mac}) = \text{Norm}\left(\bigoplus_{k=1}^{C} \mathbf{K}_k(\mathbf{p},\mathcal{E}^{mac})\right), \quad (7)$$



Popularity-Oriented MACROENV Perception

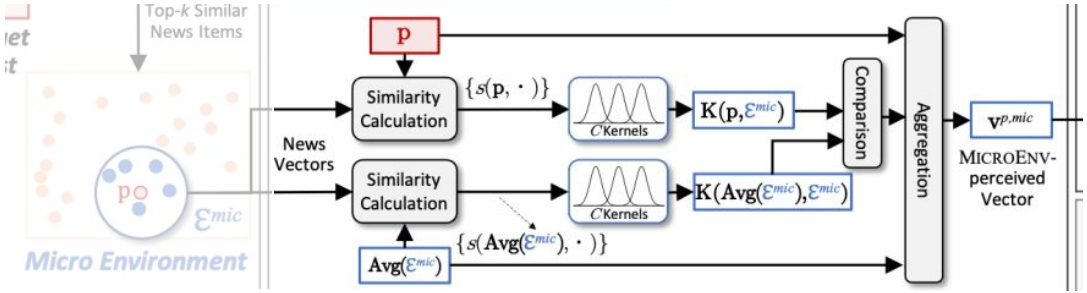
- **MACROENV-perceived vector** $\mathbf{v}^{p,mac}$
  - $\mathbf{m}(\mathcal{E}^{mac})$
    - center vector of the MACROENV
    - averaging all vectors

$$\mathbf{v}^{p,mac} = \mathrm{MLP}(\mathbf{p} \oplus \mathbf{m}(\mathcal{E}^{mac}) \oplus \mathbf{K}(\mathbf{p}, \mathcal{E}^{mac})).$$



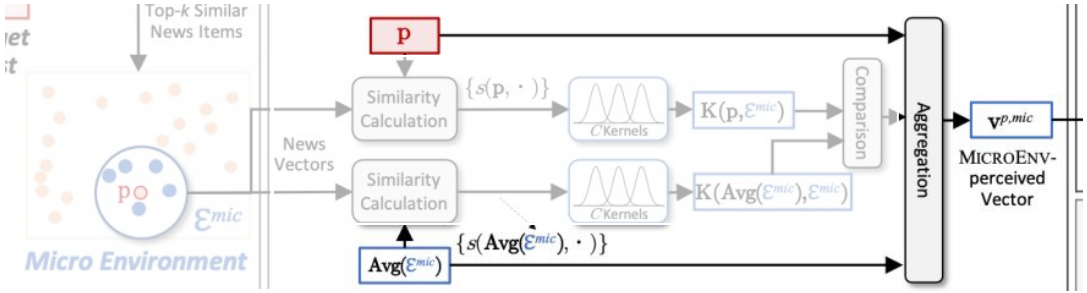
Novelty-Oriented MICROENV Perception

- **If the content of a post is novel, it is expected to be an outlier in such an event**
- **use Eqs. (5) to (7) again for K(·,·)**
- $\mathbf{m}(\mathcal{E}^{mic})/\mathbf{Avg}(\mathcal{E}^{mic})$
  - center vector of MICROENV



Novelty-Oriented MICROENV Perception

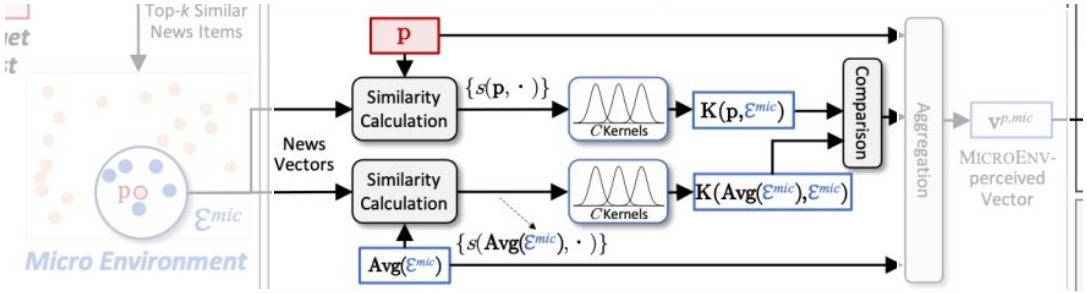
$$\mathbf{u}^{sem} = \mathrm{MLP}(\mathbf{p} \oplus \mathbf{m}(\mathcal{E}^{mic}))$$



Novelty-Oriented MICROENV Perception

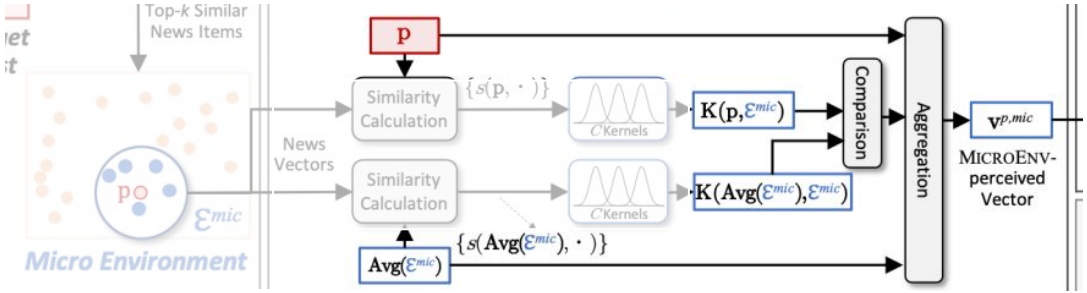
$$\mathbf{u}^{sim} = \mathrm{MLP}(\mathrm{g}(\mathbf{K}(\mathbf{p}, \mathcal{E}^{mic}), \mathbf{K}(\mathbf{m}(\mathcal{E}^{mic}), \mathcal{E}^{mic})))$$

where $\mathrm{g}(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \odot \mathbf{y}) \oplus (\mathbf{x} - \mathbf{y})$



Novelty-Oriented MICROENV Perception

$$\mathbf{v}^{p,mic} = \mathrm{MLP}(\mathbf{u}^{sem} \oplus \mathbf{u}^{sim})$$



(c)Prediction

- **Fusion Gate**
  - For a good compatibility with various fake news detectors

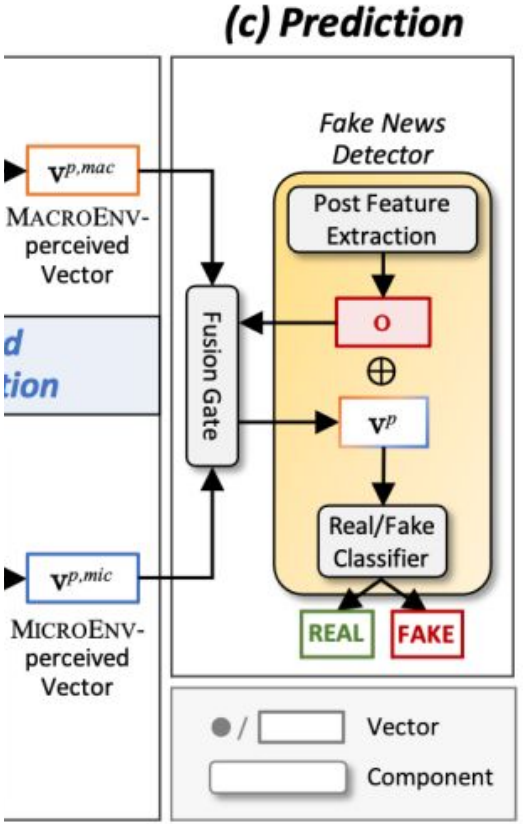
$$\mathbf{v}^p = \mathbf{g} \odot \mathbf{v}^{p,mac} + (\mathbf{1} - \mathbf{g}) \odot \mathbf{v}^{p,mic},$$

$$\mathbf{g} = \text{sigmoid}\left(\text{Linear}\left(\mathbf{o} \oplus \mathbf{v}^{p,mac}\right)\right)$$

o : the last-layer feature from a Fake News detector

$$\hat{\mathbf{y}} = \text{softmax}(\text{MLP}(\mathbf{o} \oplus \mathbf{v}^p)).$$

LOSS : cross-entropy loss



Outline

1 Introduction

2 Method

3 Experiment

4 Conclusion

- Dataset
- Performance
- Ablation
- Categorization of macro- and micro-preferred samples
- Case Study
- Case Weakly Related to News Env.

Dataset

Table 1: Statistics of the datasets.

Dataset	Chinese			English		
	Train	Val	Test	Train	Val	Test
#Real	8,787	5,131	5,625	1,976	656	661
#Fake	8,992	4,923	5,608	1,924	638	628
Total	17,779	10,054	11,233	3,900	1,294	1,289
#News Items	583,208			1,003,646		
Min/Avg/Max of $\lvert\mathcal{E}^{mac}\rvert$ in 3 days	41 / 505 / 1,563			308 / 1,614 / 2,211		

News Outlet
Chinese
People's Daily
Xinhua Agency
Xinhua Net
CCTV News
The Paper
Toutiao News

News Outlet
English
Huffington Post
NPR
Daily Mail

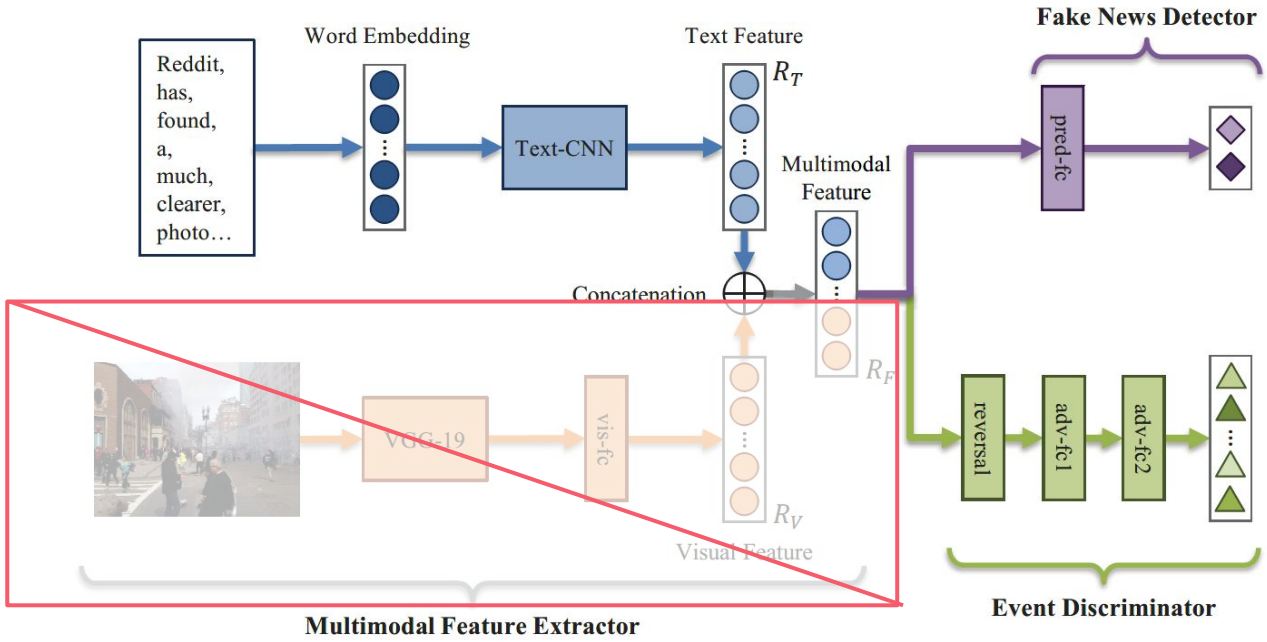
- **Chinese(2010-2021)**
  - Post
    - Merged the non-overlapping parts of 4 Weibo datasets
  - News Env.
    - Collected from the official accounts of six representative mainstream news outlets
- **English(2014-2018)**
  - Post
    - Merged the non-overlapping parts of 3 English datasets
  - News Env.
    - use news headlines (plus short descriptions if any) from Huffington Post, NPR, and Daily Mail

Performance

	Model	Chinese				English			
		Acc.	macF1	$F1_{fake}$	$F1_{real}$	Acc.	macF1	$F1_{fake}$	$F1_{real}$
Post-Only	Bi-LSTM	0.727	0.713	0.652	0.775	0.705	0.704	0.689	**0.719**
	+*NEP*	**0.776**	**0.771**	**0.739**	**0.803**	**0.718**	**0.718**	**0.720**	0.716
	$EANN_T$	0.732	0.718	0.657	0.780	0.700	0.699	0.683	0.714
	+*NEP*	**0.776**	**0.770**	**0.733**	**0.807**	**0.722**	**0.722**	**0.722**	**0.722**
	BERT	0.792	0.785	0.744	0.825	0.709	0.709	0.701	**0.716**
	+*NEP*	**0.810**	**0.805**	**0.772**	**0.837**	**0.718**	**0.718**	**0.720**	0.715
	BERT-Emo	0.812	0.807	0.776	0.838	0.718	0.718	0.719	0.718
	+*NEP*	**0.831**	**0.829**	**0.808**	**0.850**	**0.728**	**0.728**	**0.728**	**0.728**
"Zoom-In"	DeClarE	0.764	0.758	0.720	0.795	0.714	0.714	0.709	**0.718**
	+*NEP*	**0.800**	**0.797**	**0.773**	**0.822**	**0.717**	**0.716**	**0.718**	0.714
	MAC	0.755	0.751	0.717	0.784	0.706	0.705	0.708	0.701
	+*NEP*	**0.764**	**0.760**	**0.732**	**0.789**	**0.716**	**0.716**	**0.716**	**0.716**

EANN(Event Adversarial Neural Networks)

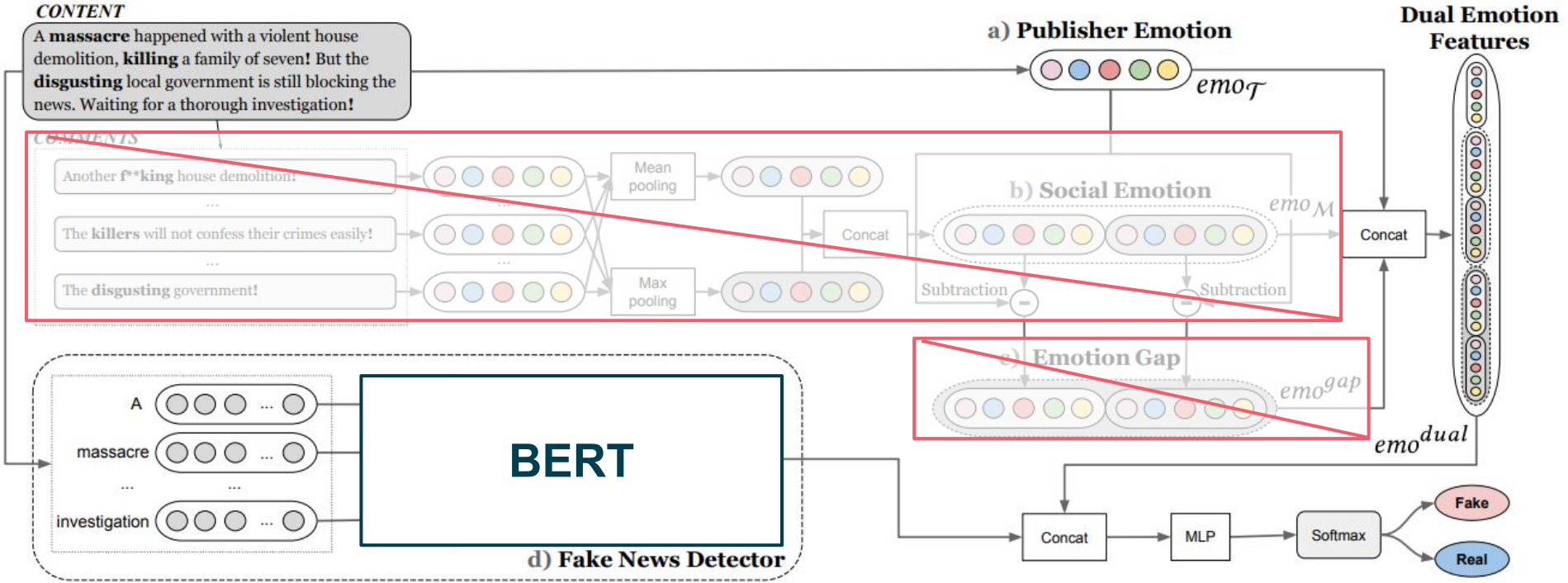
Identify whether a specific post is fake news or not



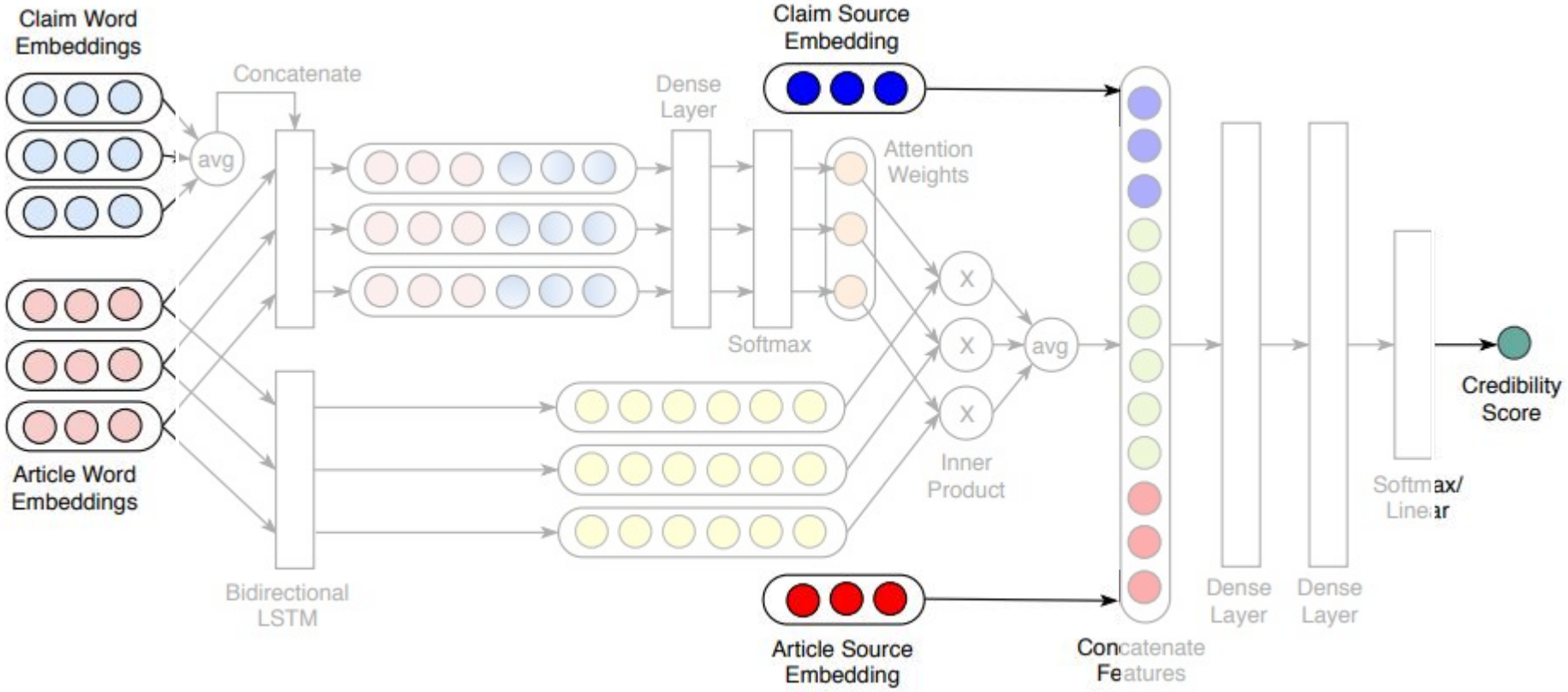
Extract the informative features from textual content

Discover the event-specific information included in the feature representations to recognize the event

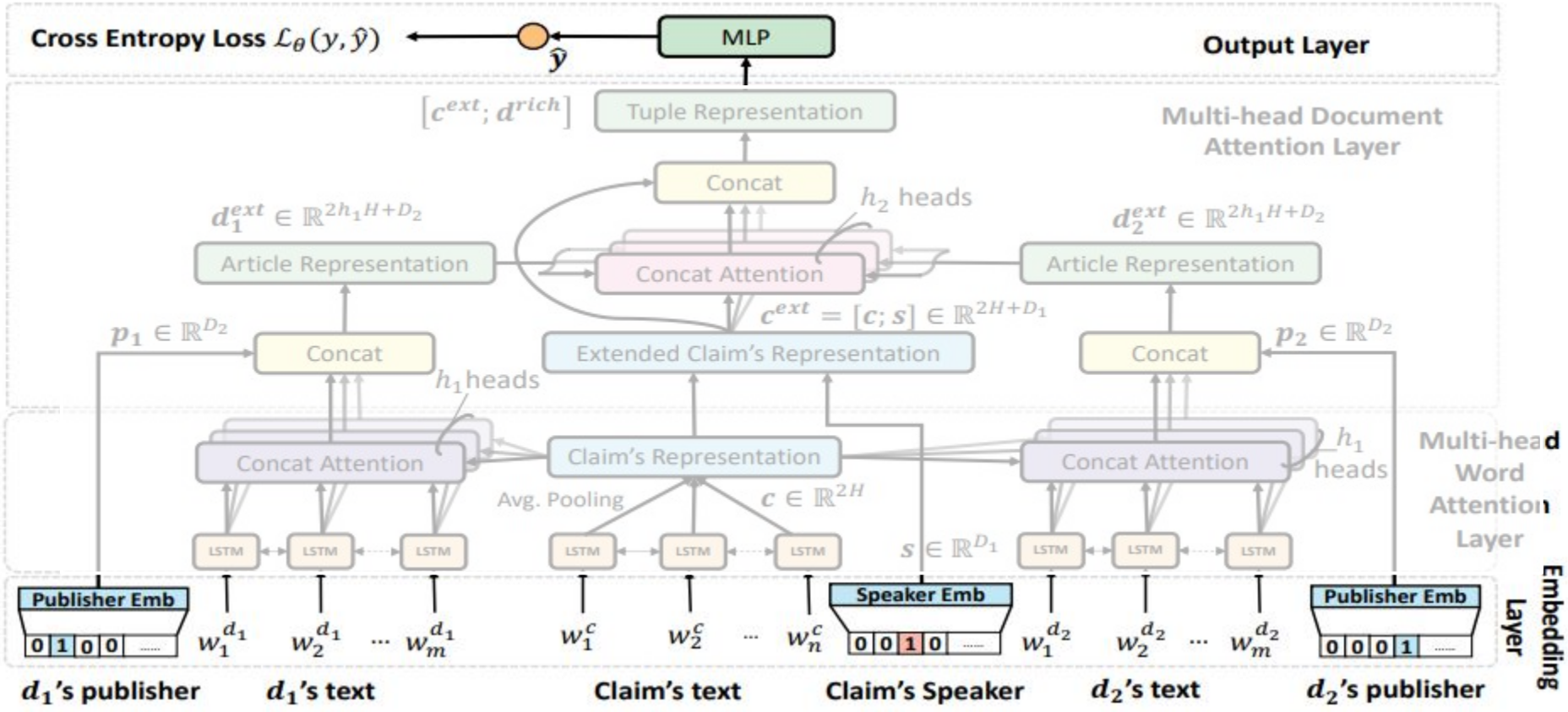
BERT-Emo


CONTENT
A massacre happened with a violent house demolition, killing a family of seven! But the disgusting local government is still blocking the news. Waiting for a thorough investigation!
COMMENTS
Another f**king house demolition!
The killers will not confess their crimes easily!
The disgusting government!
Mean pooling
Max pooling
Concat
a) Publisher Emotion
emo$_{\mathcal{T}}$
b) Social Emotion
emo$_{\mathcal{M}}$
Subtraction
Subtraction
c) Emotion Gap
emo^{gap}
Concat
Dual Emotion Features
emo^{dual}
A
massacre
investigation
BERT
d) Fake News Detector
Concat
MLP
Softmax
Fake
Real

DeClarE (Debunking Claims with Interpretable Evidence)



MAC(Multihead Attentive Network for Fact-Checking)

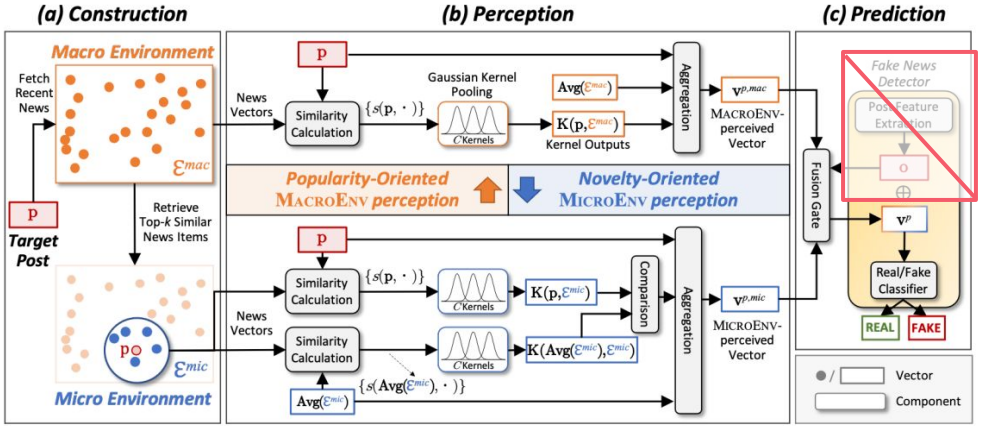


Performance

	Model	Chinese				English			
		Acc.	macF1	$F1_{fake}$	$F1_{real}$	Acc.	macF1	$F1_{fake}$	$F1_{real}$
Post-Only	Bi-LSTM	0.727	0.713	0.652	0.775	0.705	0.704	0.689	**0.719**
	+*NEP*	**0.776**	**0.771**	**0.739**	**0.803**	**0.718**	**0.718**	**0.720**	0.716
	$EANN_T$	0.732	0.718	0.657	0.780	0.700	0.699	0.683	0.714
	+*NEP*	**0.776**	**0.770**	**0.733**	**0.807**	**0.722**	**0.722**	**0.722**	**0.722**
	BERT	0.792	0.785	0.744	0.825	0.709	0.709	0.701	**0.716**
	+*NEP*	**0.810**	**0.805**	**0.772**	**0.837**	**0.718**	**0.718**	**0.720**	0.715
	BERT-Emo	0.812	0.807	0.776	0.838	0.718	0.718	0.719	0.718
	+*NEP*	**0.831**	**0.829**	**0.808**	**0.850**	**0.728**	**0.728**	**0.728**	**0.728**
"Zoom-In"	DeClarE	0.764	0.758	0.720	0.795	0.714	0.714	0.709	**0.718**
	+*NEP*	**0.800**	**0.797**	**0.773**	**0.822**	**0.717**	**0.716**	**0.718**	0.714
	MAC	0.755	0.751	0.717	0.784	0.706	0.705	0.708	0.701
	+*NEP*	**0.764**	**0.760**	**0.732**	**0.789**	**0.716**	**0.716**	**0.716**	**0.716**

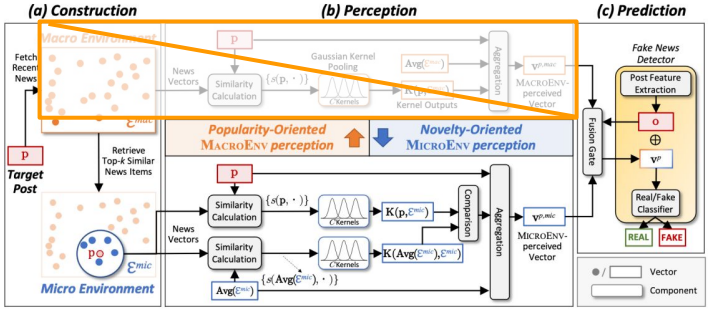
Ablation Study(w/o Fake News Detector)

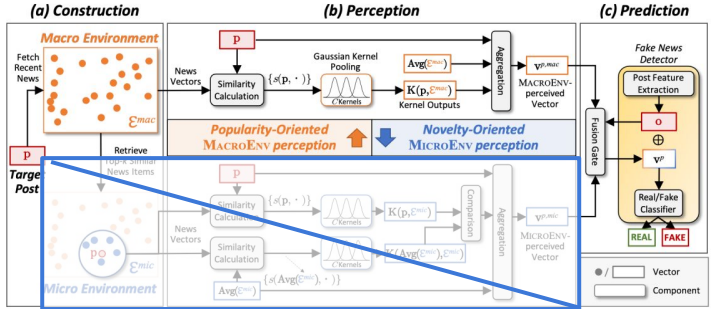
Model	Chinese				English			
	Acc.	macF1	$F1_{fake}$	$F1_{real}$	Acc.	macF1	$F1_{fake}$	$F1_{real}$
MACROENV	0.689	0.659	0.557	0.761	0.693	0.693	**0.696**	0.689
MICROENV	0.666	0.626	0.503	0.748	0.695	0.695	0.694	0.696
MACROENV+MICROENV	**0.694**	**0.666**	**0.569**	**0.763**	**0.696**	**0.696**	0.694	**0.697**



Ablation Study(w/o Env. Perception Modules)

Model	Chinese				English			
	Acc.	macF1	$F1_{fake}$	$F1_{real}$	Acc.	macF1	$F1_{fake}$	$F1_{real}$
BERT-Emo + *NEP*	**0.831**	**0.829**	**0.808**	**0.850**	**0.728**	**0.728**	**0.728**	0.728
w/o MACROENV	0.822	0.819	0.794	0.843	0.726	0.726	0.726	0.725
w/o MICROENV	0.824	0.820	0.795	0.845	0.723	0.723	0.715	**0.731**
DeClarE + *NEP*	**0.797**	**0.800**	**0.773**	**0.822**	**0.717**	**0.716**	0.718	**0.714**
w/o MACROENV	0.776	0.771	0.735	0.806	0.712	0.711	0.709	0.713
w/o MICROENV	0.778	0.773	0.736	0.809	0.709	0.709	**0.719**	0.698





Categorization of macro- and micro-preferred samples

- **Selected the top 1% of Chinese fake news samples**

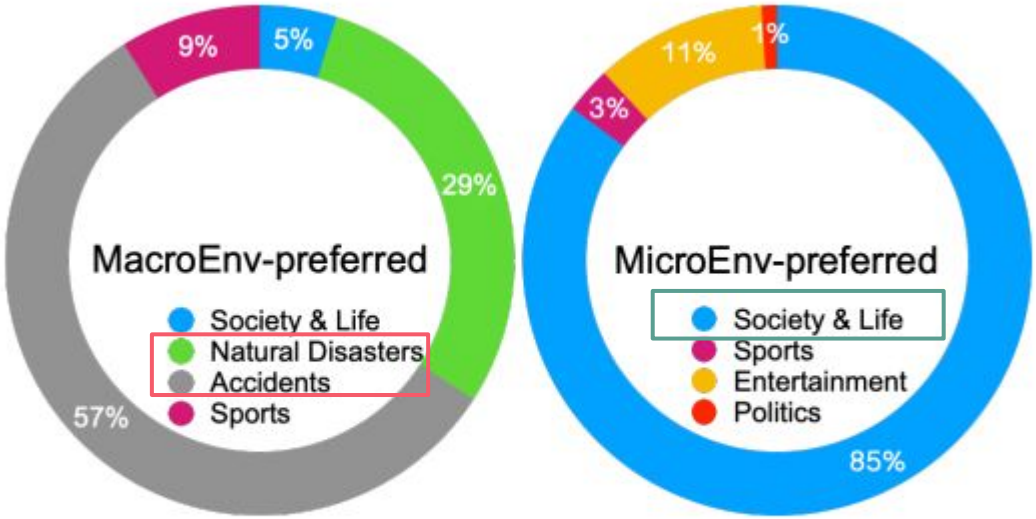
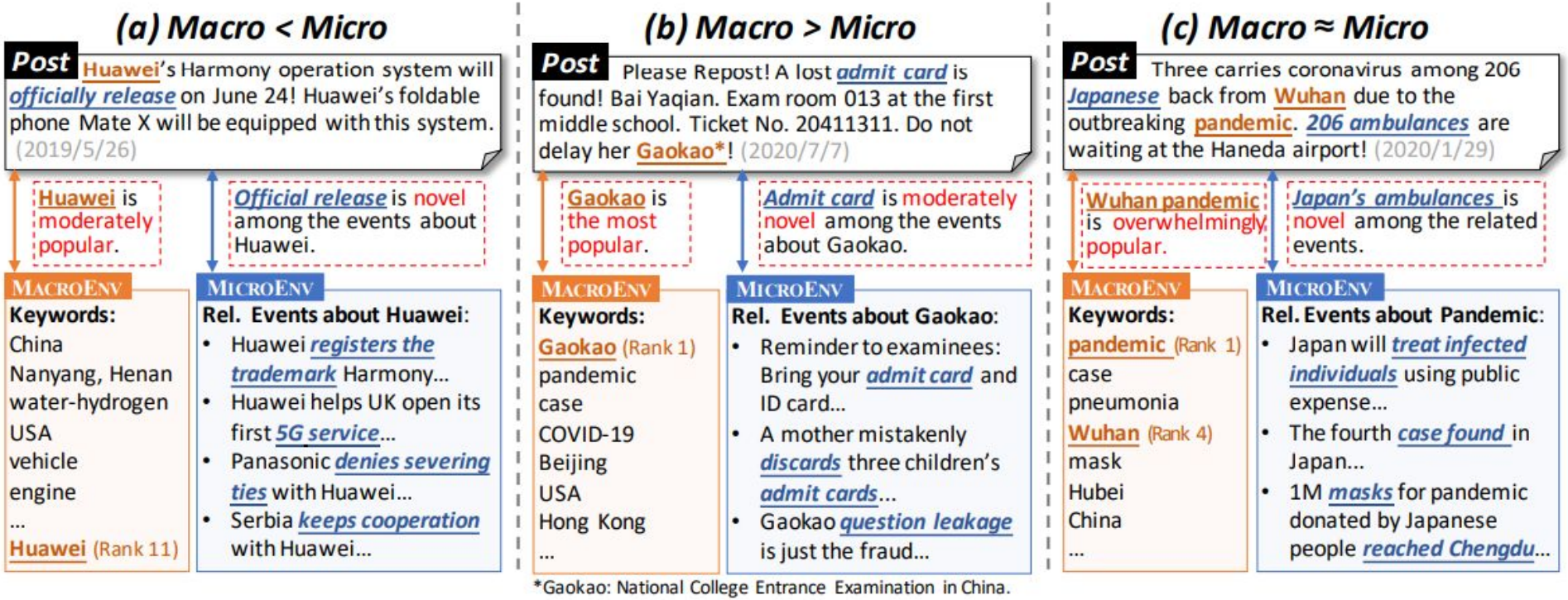


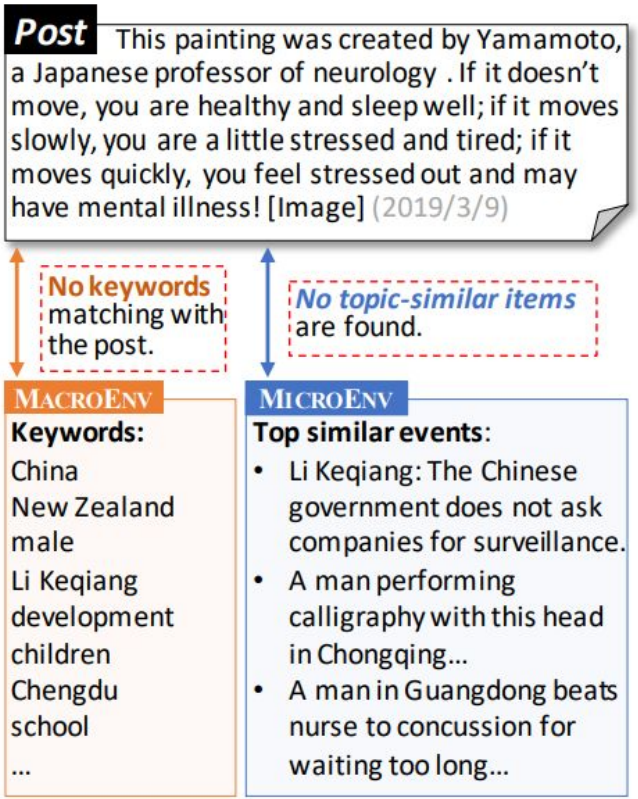
Figure 5: Categories of MACROENV- and MICROENV-preferred samples.

Case Study



intuitively show how NEP handles different scenarios

Case Weakly Related to News Env.



Outline

1 Introduction

2 Method

3 Experiment

4 Conclusion

Conclusion

- **It designed popularity- and novelty-oriented perception modules to assist fake news detectors**
- **Pros**
  - Timeliness
    - Only requires the post and mainstream news published a few days before.
  - Compatibility
    - Can be integrated with existing methods
  - Data Accessibility
    - The data to construct news environments is easy to access
- **Cons**
  - Works ineffective if involved with a long-lasting discussed topic(e.g.mental health)