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

Qiang Sheng, Juan Cao*, Xueyao Zhang, Rundong Li, Danding Wang, Yongchun Zhu
Key Lab of Intelligent Information Processing of Chinese Academy of Sciences,
Institute of Computing Technology, Chinese Academy of Sciences
University of Chinese Academy of Sciences
{shengqiang18z,caojuan,zhangxueyao19s}@ict.ac.cn
{lirundong20s,wangdanding,zhuyongchun18s}@ict.ac.cn

code:https://github.com/ICTMCG/News-Environment-Perception

ACL 2022

Reported by Minqin Li

Introduction

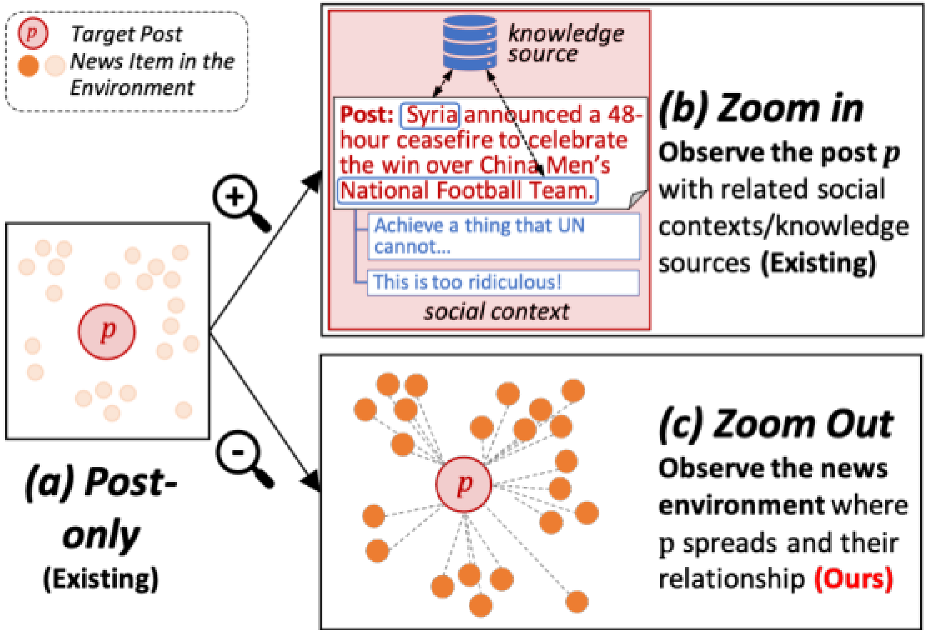


Figure 1: Existing methods for fake news detection rely on (a) the post content itself and (b) related post-level signals like social context and knowledge. Unlike (a) and (b), our method captures (c) signals from *news environments*.

Introduction

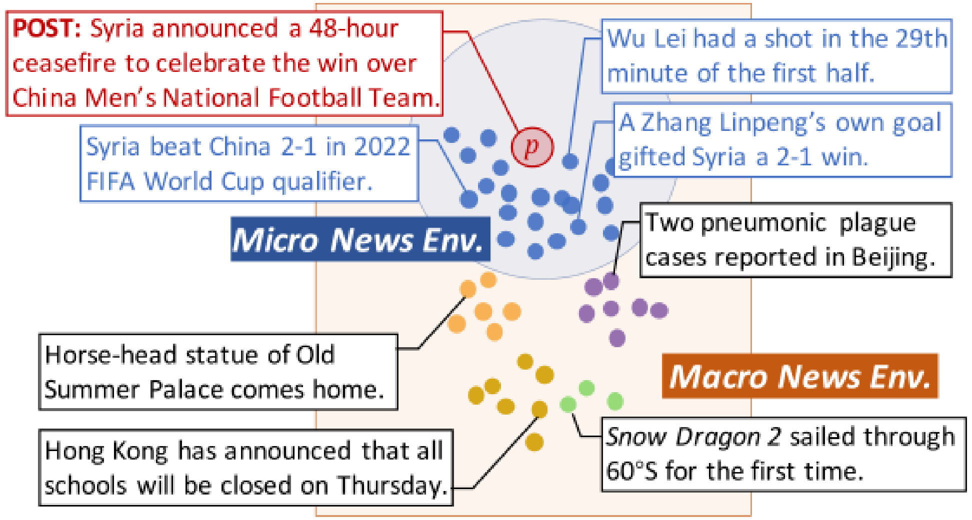


Figure 2: A fake news post p and its news environment containing recent news items in three days (2019/11/12 to 2019/11/14). Only the items in events that are reported multiple times (differentiated by dot colors) are displayed for brevity. We can see that p falls in a *popular* event on a Syria-China World Cup qualifier compared with other events and focuses on a *novel* aspect (unusual celebration in Syria).

events. We use a pretrained language model $\mathcal{M}$ (e.g., BERT (Devlin et al., 2019)) to obtain the post/news representation. For p or each item in the macro/micro environment e, the initial representation is the output of $\mathcal{M}$ for the [CLS] token:

$$\mathbf{p} = \mathcal{M}(p), \quad \mathbf{e} = \mathcal{M}(e). \tag{3}$$

Method

$$\mathbf{v}^p = \mathbf{g} \odot \mathbf{v}^{p,mac} + (\mathbf{1} - \mathbf{g}) \odot \mathbf{v}^{p,mic}, \tag{12}$$

where the gating vector $\mathbf{g} = \mathrm{sigmoid}(\mathrm{Linear}(\mathbf{o} \oplus \mathbf{v}^{p,mac}))$, sigmoid is to constrain the value of each element in $[0, 1]$, and $\mathbf{o}$ denotes the last-layer feature from a post-only detector.[3] $\mathbf{o}$ and $\mathbf{v}^p$ are further fed into an MLP and a softmax layer for final prediction:

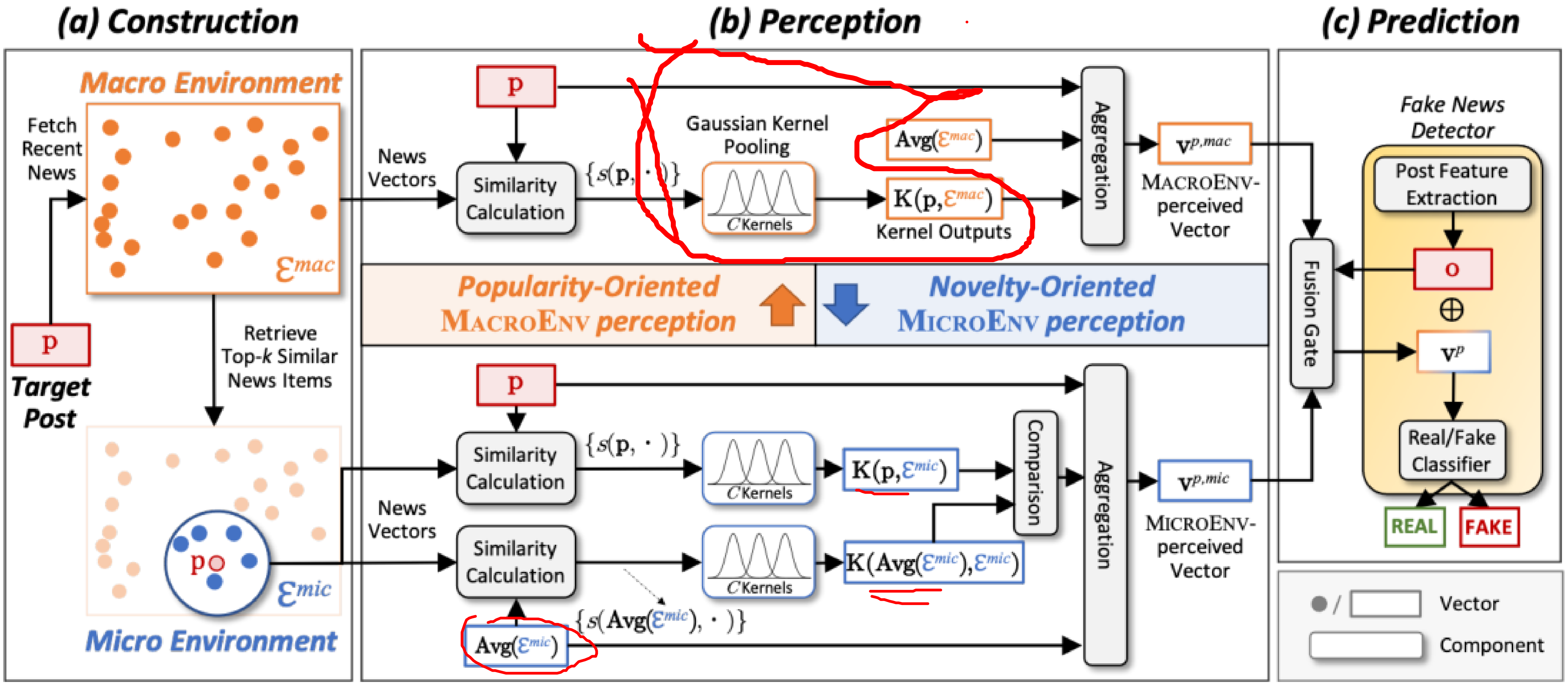
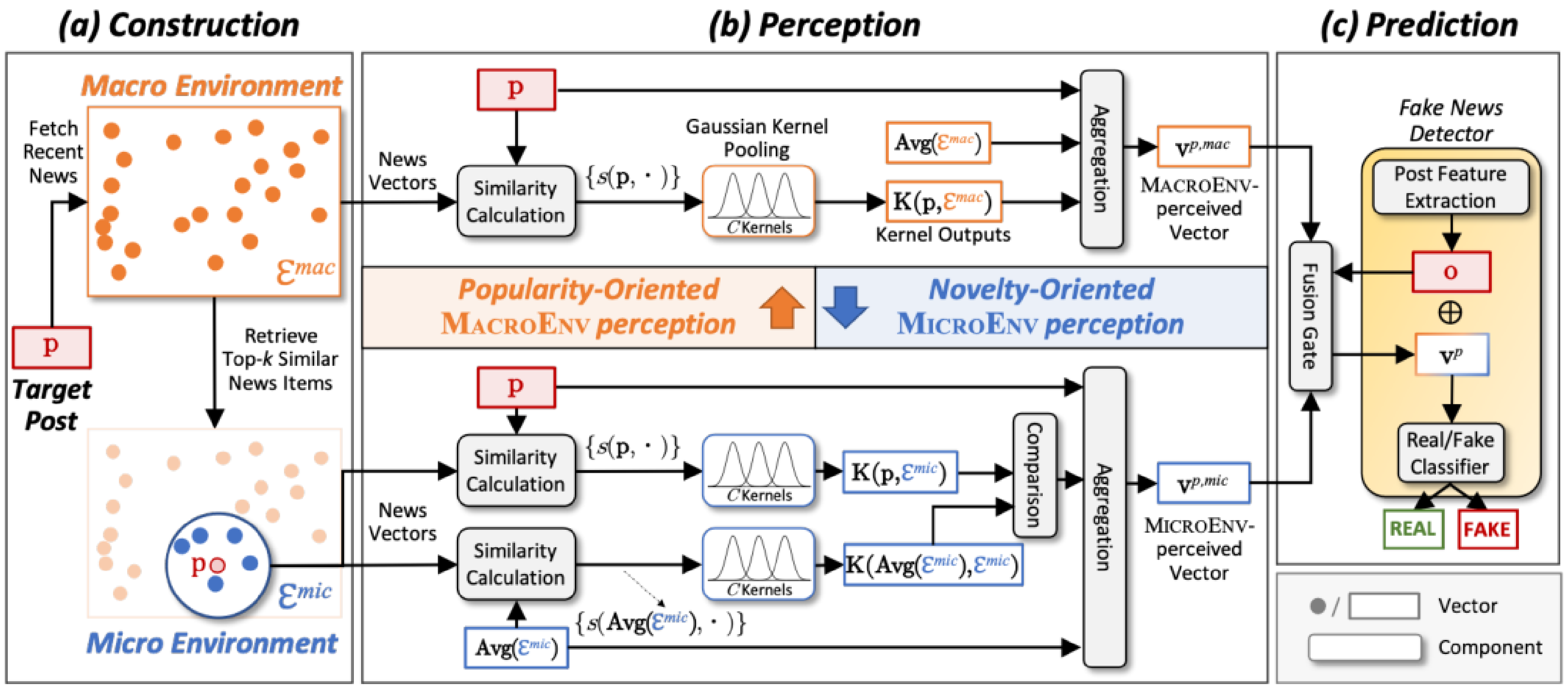


Figure 3: Architecture of the News Environment Perception Framework (NEP).

Method



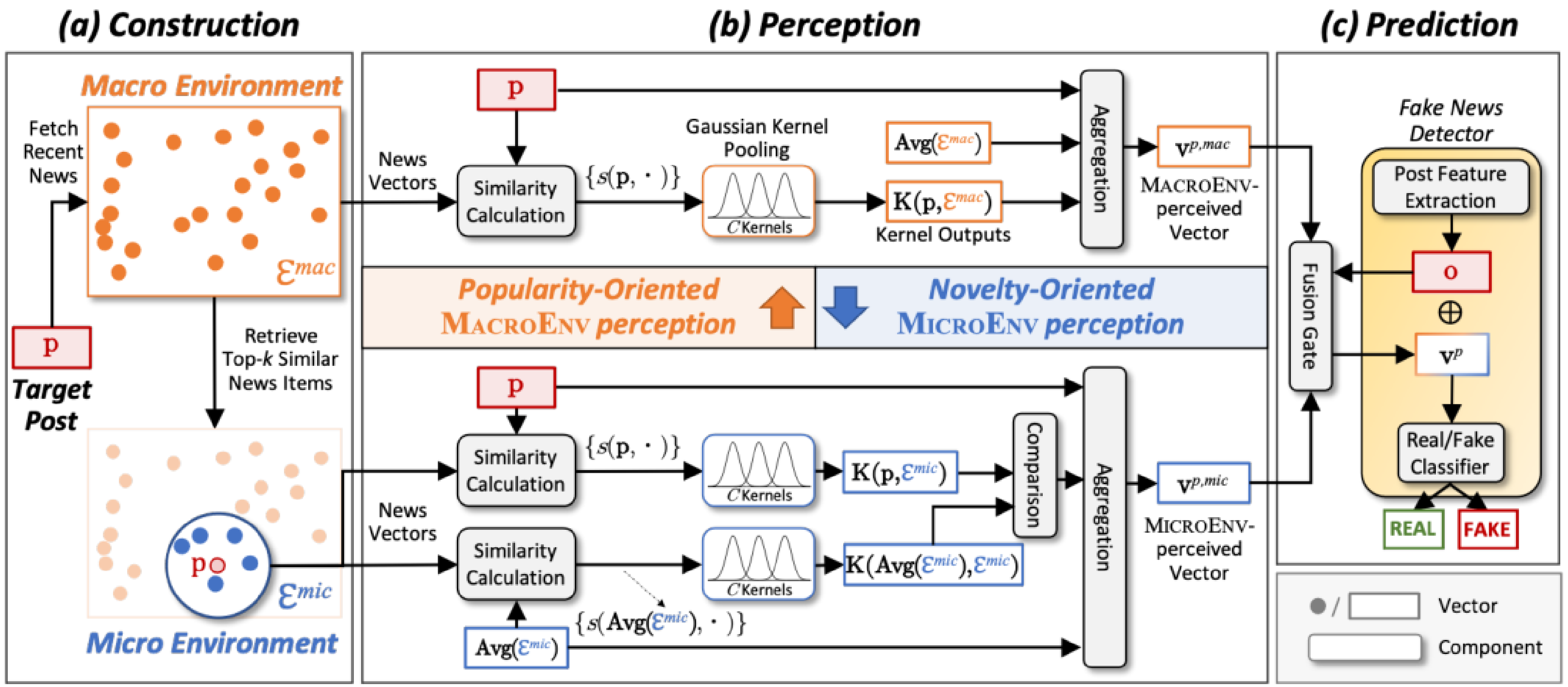
$$\mathcal{E}^{mac} = \{e : e \in \mathcal{E}, 0 < t_p - t_e \le T\} \quad (1)$$

$$\mathcal{E}^{mic} = \{e : e \in \mathrm{Topk}(p, \mathcal{E}^{mac})\} \quad (2)$$

$$\mathbf{p} = \mathcal{M}(p), \ \ \mathbf{e} = \mathcal{M}(e) \quad (3)$$

$$s(\mathbf{p}, \mathbf{e}_i) = \frac{\mathbf{p} \cdot \mathbf{e}_i}{\|\mathbf{p}\| \|\mathbf{e}_i\|} \quad (4)$$

Method



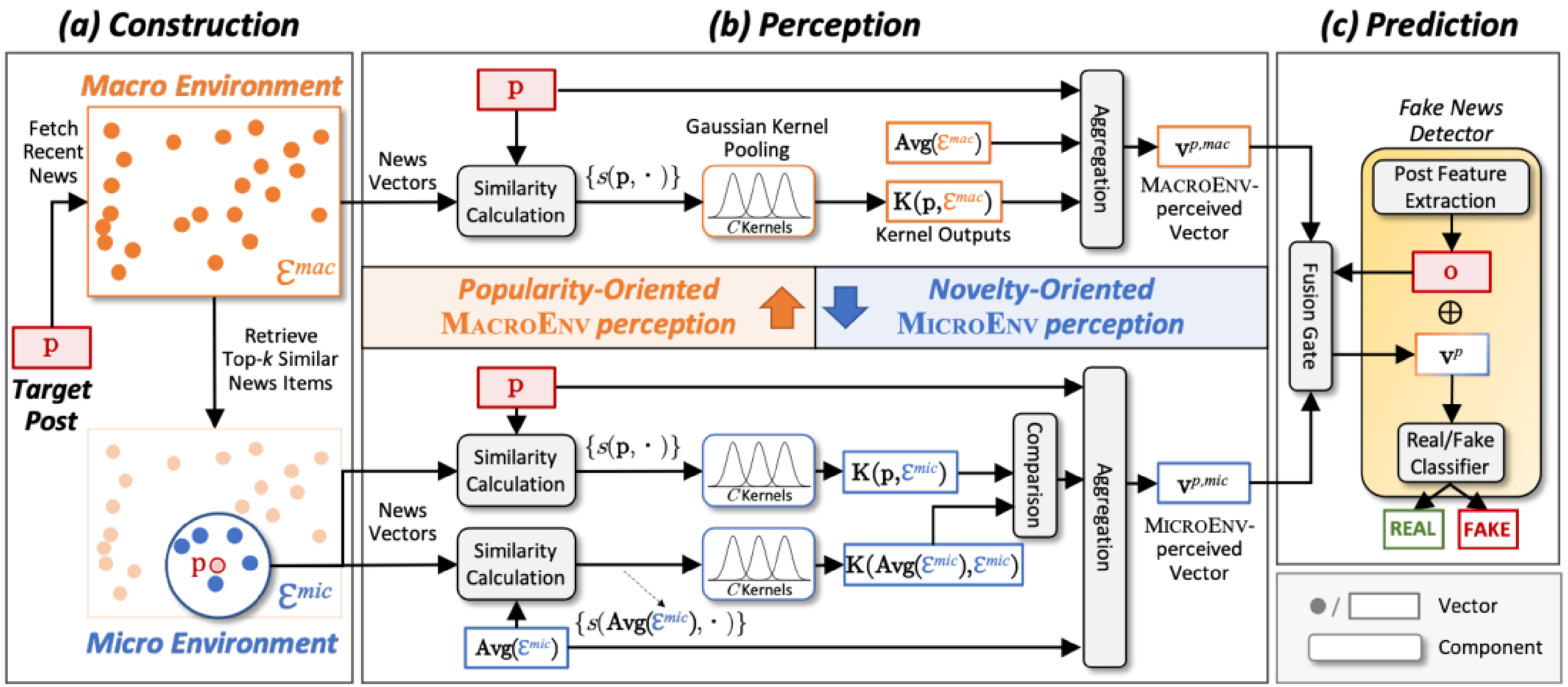
$$\mathbf{K}_k^i = \exp\left(-\frac{(s(\mathbf{p},\mathbf{e}_i)-\mu_k)^2}{2\sigma_k^2}\right) \quad (5)$$

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

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

$$\mathbf{v}^{p,mac} = \text{MLP}(\mathbf{p}\oplus\mathbf{m}(\mathcal{E}^{mac})\oplus\mathbf{K}(\mathbf{p},\mathcal{E}^{mac})) \quad (8)$$

Method



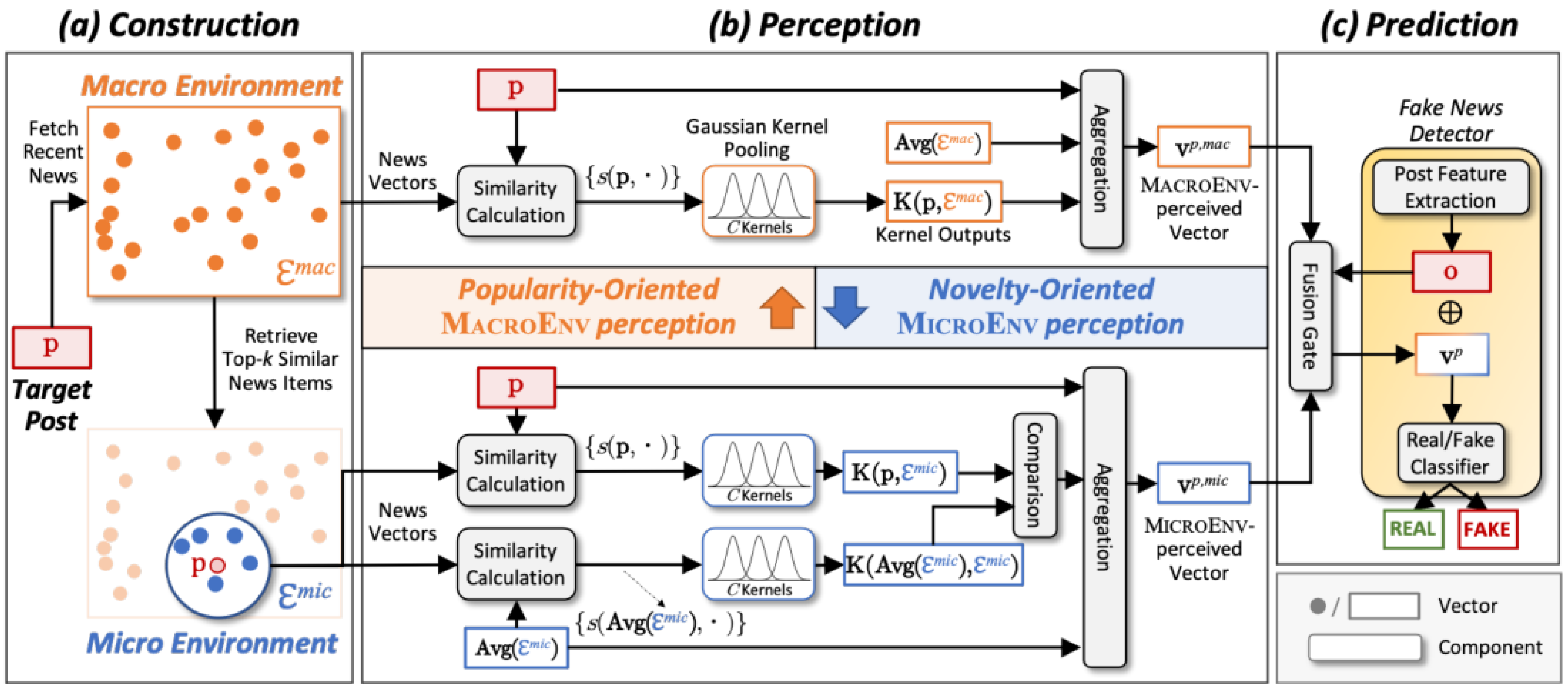
$$\mathbf{u}^{sem} = \mathrm{MLP}(\mathbf{p} \oplus \mathbf{m}(\mathcal{E}^{mic})) \quad (9)$$

$$\mathbf{v}^{p,mic} = \mathrm{MLP}(\mathbf{u}^{sem} \oplus \mathbf{u}^{sim}) \quad (11)$$

$$\mathbf{u}^{sim} = \mathrm{MLP}(\mathrm{g}(\mathbf{K}(\mathbf{p}, \mathcal{E}^{mic}), \mathbf{K}(\mathbf{m}(\mathcal{E}^{mic}), \mathcal{E}^{mic}))) \quad (10)$$

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

Method



$$\mathbf{v}^p = \mathbf{g} \odot \mathbf{v}^{p,mac} + (\mathbf{1} - \mathbf{g}) \odot \mathbf{v}^{p,mic} \quad (12)$$

$$\hat{\mathbf{y}} = \mathrm{softmax}(\mathrm{MLP}(\mathbf{o} \oplus \mathbf{v}^p)) \quad (13)$$

where the gating vector $\mathbf{g} = \mathrm{sigmoid}(\mathrm{Linear}(\mathbf{o} \oplus \mathbf{v}^{p,mac}))$, sigmoid is to constrain the value of each

Experiments

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		

Experiments

Table 2: Performance comparison of base models with and without the NEP. The better result in each group using the same base model are in **boldface**.

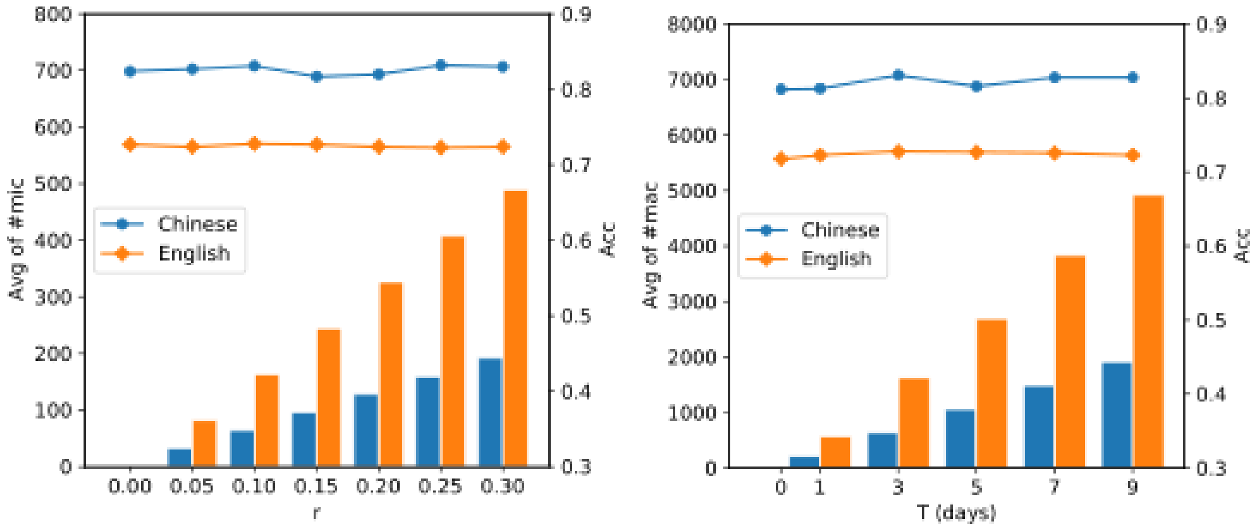
Model		Chinese Acc.	Chinese macF1	Chinese $F1_{fake}$	Chinese $F1_{real}$	English Acc.	English macF1	English $F1_{fake}$	English $F1_{real}$
Post-Only	Bi-LSTM	0.727	0.713	0.652	0.775	0.705	0.704	0.689	**0.719**
	+*NEP*	**0.776**	**0.771**	**0.739**	**0.803**	**0.718**	**0.718**	**0.720**	0.716
	$EANN_T$	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**

Experiments

Table 3: Performance comparison of the NEP and its variants without the fake news detector or without the environment perception module. The best result in each group is in **boldface**.

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**
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

Experiments



(a) Proportion Factor r (b) Day Difference T

Figure 4: Effects of (a) the proportion factor r and (b) the day difference T. Lines show the accuracies and bars show the average numbers of news items in the micro/macro environments.

Experiments

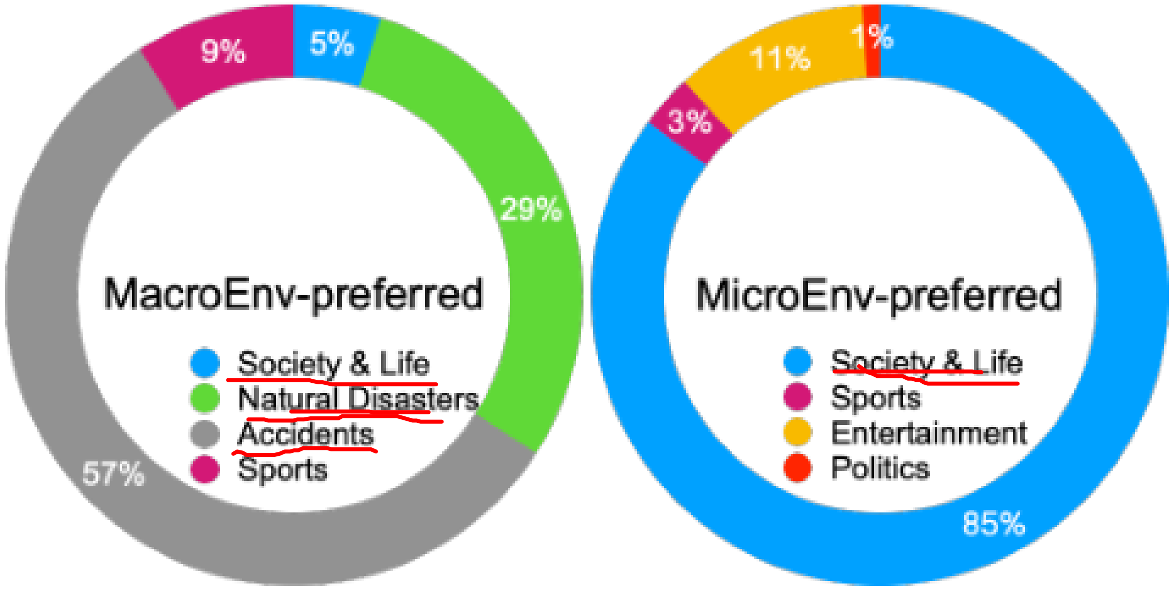


Figure 5: Categories of MACROENV- and MICROENV-preferred samples.

Experiments

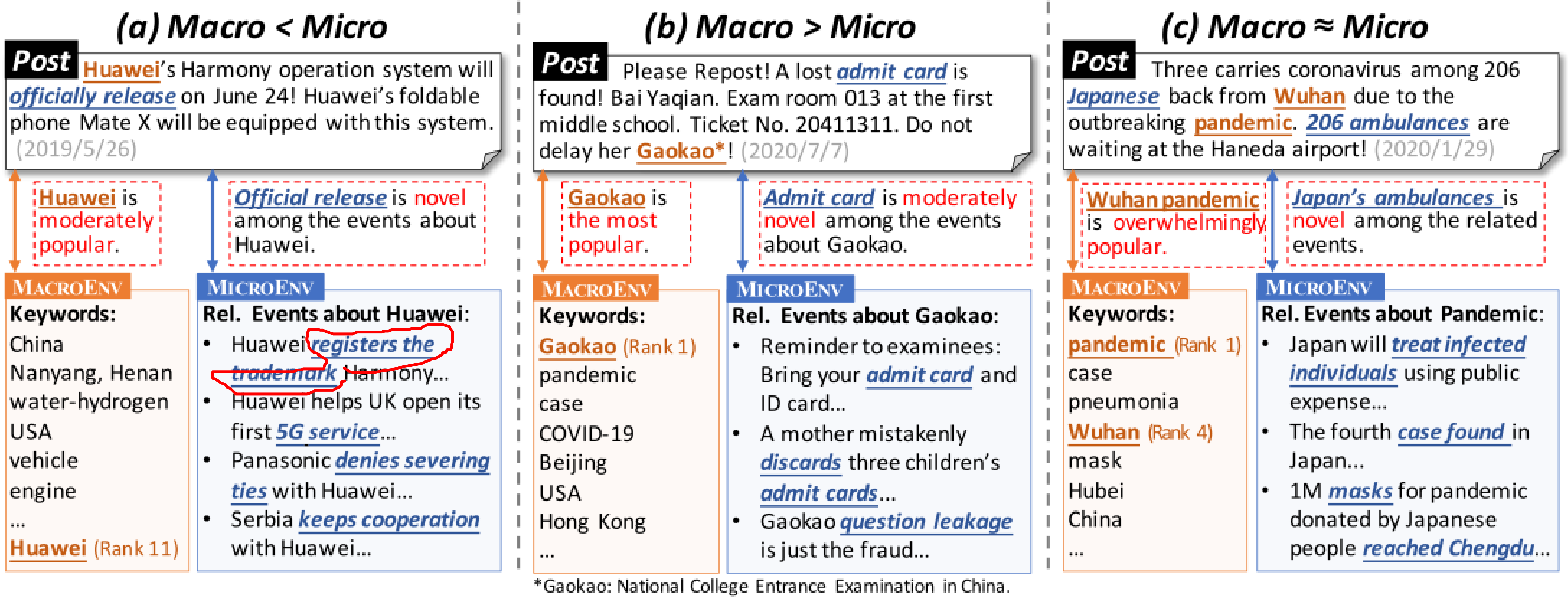


Figure 6: Three fake news cases with different preferences on environmental information. Underlined regular words hit the keywords in the MACROENV and *underlined italic* words are related to the MICROENV. Keywords are extracted using TextRank (Mihalcea and Tarau, 2004).

Experiments

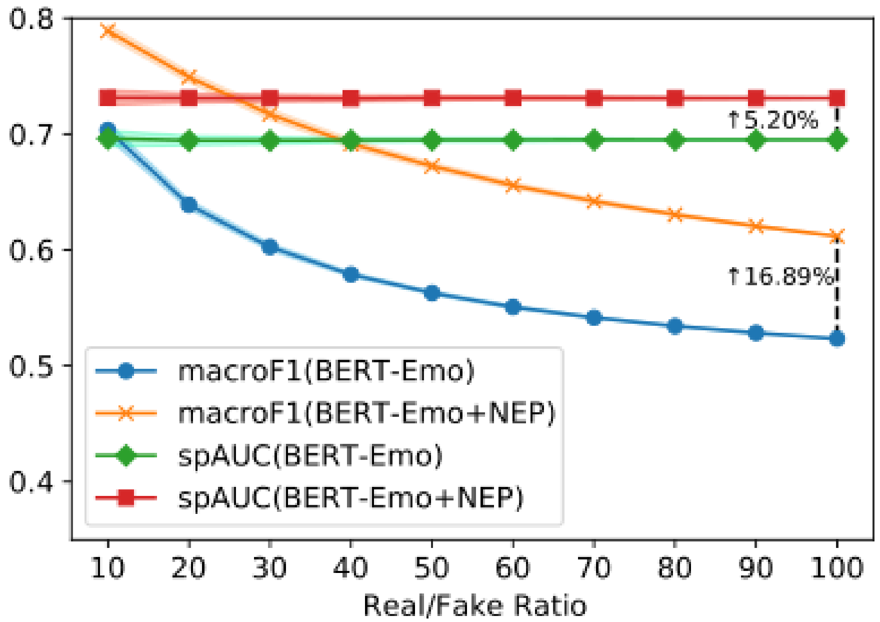


Figure 7: Macro F1s and spAUCs on the online data in different real/fake ratios. We sampled 100 times from the 100:1 set for each fo the first nine ratios. Shadows show the standard deviations. The percentages denote relative improvements using the NEP.

Thanks