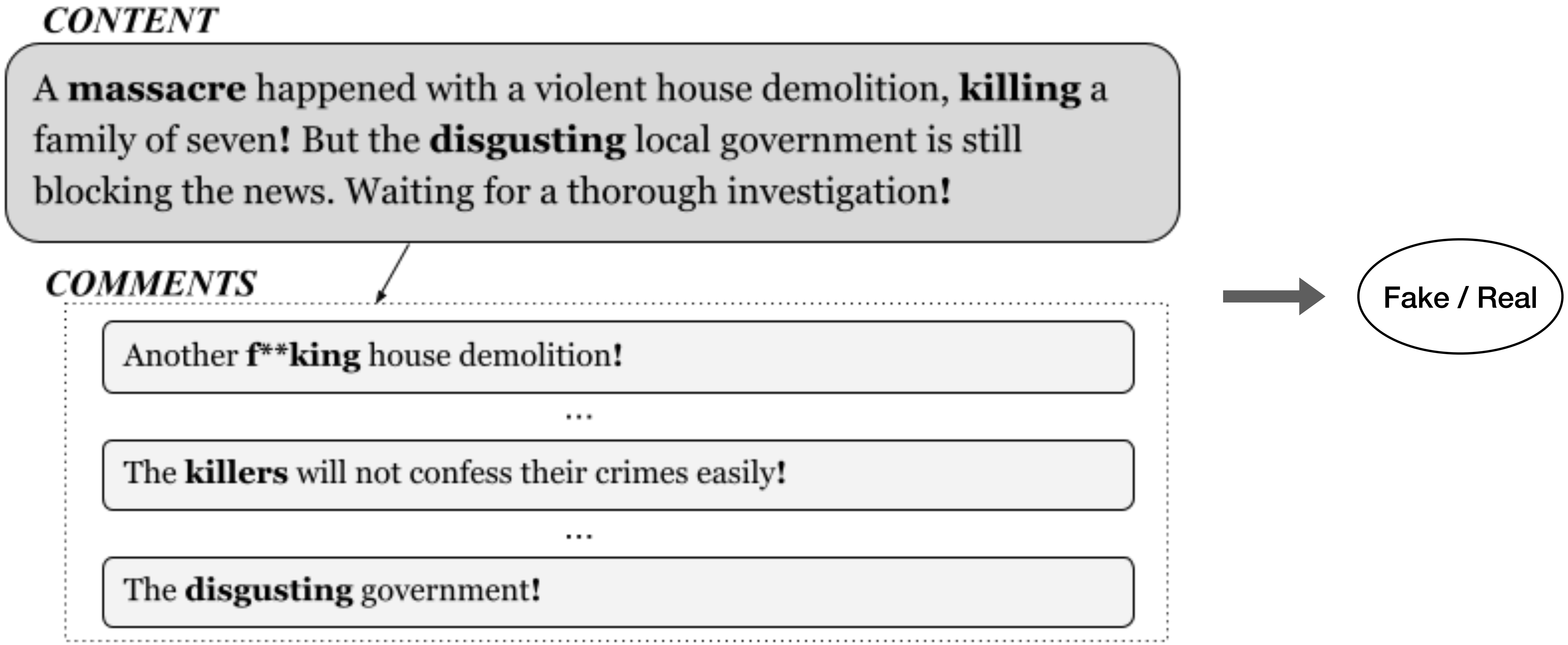
Mining Dual Emotion for Fake News Detection

Xueyao Zhang[1,2], Juan Cao[1,2], Xirong Li[3], Qiang Sheng[1,2], Lei Zhong[1,2], Kai Shu[4]

[1] Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China
[2] University of Chinese Academy of Sciences, Beijing, China
[3] Renmin University of China, Beijing, China
[4] Illinois Institute of Technology Chicago, Illinois, USA

TASK DESCRIPTION

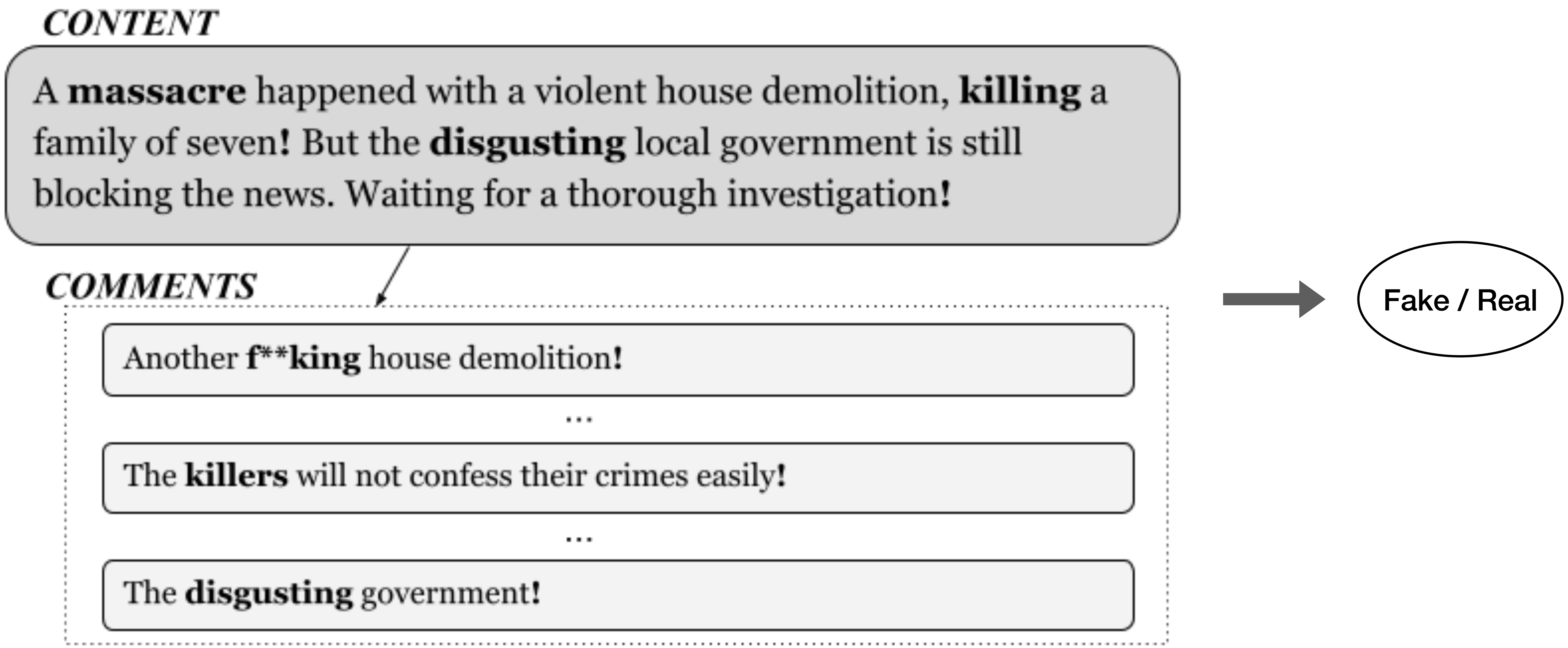
Fake News Detection



Design a classifier to judge a given news piece as fake or real

TASK DESCRIPTION

Fake News Detection



To detect fake news, we can leverage the signals from:

- content
- comments
- images
- social contexts (eg: the platform, the publisher, the crowd, the propagation structure, etc.)

Design a classifier to judge a given news piece as fake or real

THE WEB CONFERENCE

TASK DESCRIPTION

Fake News Detection

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!

→ Fake / Real

To detect fake news, we can leverage the signals from:

- content
- comments
- images
- social contexts (eg: the platform, the publisher, the crowd, the propagation structure, etc.)

Text-based Fake News Detection

Design a classifier to judge a given news piece as fake or real

MOTIVATION

Why emotion?

Quantitative Analysis:

- There are distinct distributions of *emotional category*, *emotional intensity* and *emotional expression* between fake and real news [1].
- Fake and real news inspire readers' different *emotion categories* [2].

The role of emotional signals in the existing techniques:

- In the earliest study on information credibility evaluation: *sentiment-based features* like the fraction of sentimental words and exclamation marks were proven to be effective [3].
- In 2019, Ajao et al. [4] designed a novel emotion feature (the ratio of the count of *negative and positive words* in news contents). And Giachanou et al. [5] extract emotion features based on *emotional lexicons* from news contents for fake news detection.

[1] Chuan Guo, Juan Cao, Xueyao Zhang, Kai Shu, Miao Yu. Exploiting Emotions for Fake News Detection on Social Media[J]. arXiv:1903.01728, 2019.
[2] Soroush Vosoughi, Deb Roy, Sinan Aral. The spread of true and false news online[J]. Science, 2018, 359(6380): 1146-1151.
[3] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. Information credibility on twitter. In WWW 2011. 675–684.
[4] Oluwaseun Ajao, Deepayan Bhowmik, and Shahrzad Zargari. Sentiment Aware Fake News Detection on Online Social Networks. In IEEE ICASSP 2019. 2507–2511.
[5] Anastasia Giachanou, Paolo Rosso, and Fabio Crestani. 2019. Leveraging Emotional Signals for Credibility Detection. In ACM SIGIR 2019. 877–880.

MOTIVATION

Why emotion?

Quantitative Analysis:

- There are distinct distributions of *emotional category*, *emotional intensity* and *emotional expression* between fake and real news [1].
- Fake and real news inspire readers' different *emotion categories* [2].

The role of emotional signals in the existing techniques:

Only focus on the emotions from news contents (i.e. *publisher emotion*)

- In the earliest study on information credibility evaluation: *sentiment-based features* like the fraction of sentimental words and exclamation marks were proven to be effective [3].
- In 2019, Ajao et al. [4] designed a novel emotion feature (the ratio of the count of *negative and positive words* in news contents). And Giachanou et al. [5] extract emotion features based on *emotional lexicons* from news contents for fake news detection.

[1] Chuan Guo, Juan Cao, Xueyao Zhang, Kai Shu, Miao Yu. Exploiting Emotions for Fake News Detection on Social Media[J]. arXiv:1903.01728, 2019.
[2] Soroush Vosoughi, Deb Roy, Sinan Aral. The spread of true and false news online[J]. Science, 2018, 359(6380): 1146-1151.
[3] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. Information credibility on twitter. In WWW 2011. 675–684.
[4] Oluwaseun Ajao, Deepayan Bhowmik, and Shahrzad Zargari. Sentiment Aware Fake News Detection on Online Social Networks. In IEEE ICASSP 2019. 2507–2511.
[5] Anastasia Giachanou, Paolo Rosso, and Fabio Crestani. 2019. Leveraging Emotional Signals for Credibility Detection. In ACM SIGIR 2019. 877–880.

MOTIVATION

Beyond publisher emotion

For spreading in the crowd virally, fake news often evokes high-arousal or activating emotions of the crowd [6], whose activated emotions are different from those of real news [2].

[6] Ralph L Rosnow. Inside rumor: A personal journey. American psychologist 46, 5 (1991), 484.

MOTIVATION

Beyond publisher emotion

For spreading in the crowd virally, fake news often evokes high-arousal or activating emotions of the crowd [6], whose activated emotions are different from those of real news [2].

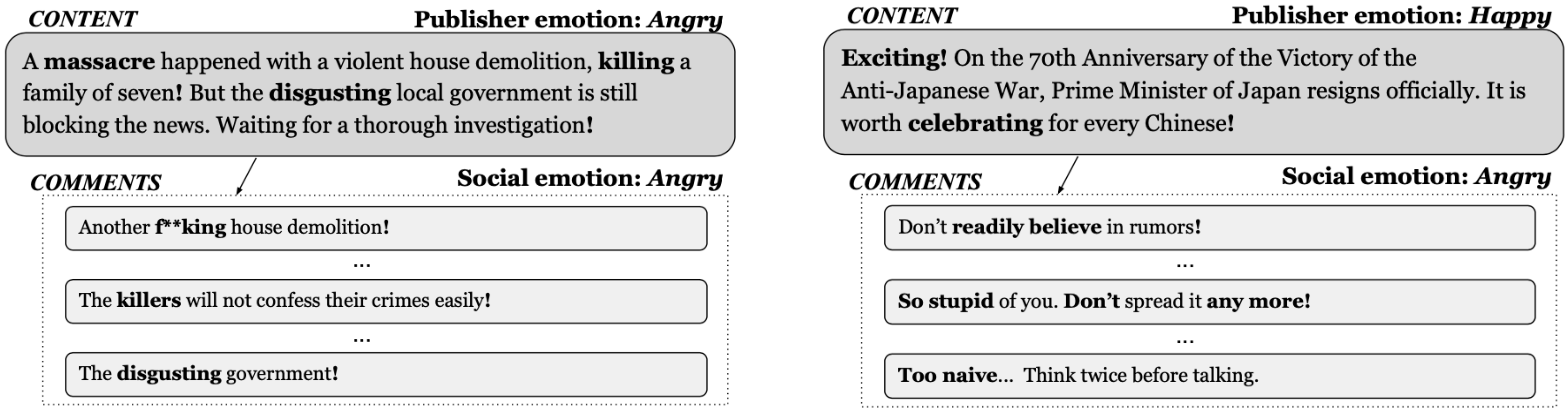
Definition:

- ***publisher emotion***: the emotions conveyed by publishers of the news pieces.
- ***social emotion***: the emotions aroused in the crowd facing to the news pieces.
- ***dual emotion***: a general term of these two emotions.

[6] Ralph L Rosnow. Inside rumor: A personal journey. American psychologist 46, 5 (1991), 484.

MOTIVATION

How dose dual emotion appear in fake news?



(a) Emotion resonance in a fake news piece: the *publisher emotion* and *social emotion* are both *angry*.

(b) Emotion dissonance in a fake news piece: the *publisher emotion* is *happy* while the *social emotion* is *angry*.

MOTIVATION

How dose dual emotion appear in fake news?

CONTENT **Publisher emotion: *Angry***

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 **Social emotion: *Angry***

Another **f**king** house demolition!

...

The **killers** will not confess their crimes easily!

...

The **disgusting** government!

(a) Emotion resonance in a fake news piece: the *publisher emotion* and *social emotion* are both *angry*.

CONTENT **Publisher emotion: *Happy***

Exciting! On the 70th Anniversary of the Victory of the Anti-Japanese War, Prime Minister of Japan resigns officially. It is worth **celebrating** for every Chinese!

COMMENTS **Social emotion: *Angry***

Don't **readily believe** in rumors!

...

So stupid of you. **Don't** spread it **any more!**

...

Too naive... Think twice before talking.

(b) Emotion dissonance in a fake news piece: the *publisher emotion* is *happy* while the *social emotion* is *angry*.

The relationship in dual emotion can be indicative of the news veracity!

PRELIMINARY ANALYSIS

Is dual emotion distinctive between fake and real news?

Chi-squared statistical significance test

- The two categorical variables: (1) News Veracity; (2) Dual Emotion Category
- Conclusion: dual emotion signals are statistically dependent on news veracity.

PRELIMINARY ANALYSIS

Is dual emotion distinctive between fake and real news?

Chi-squared statistical significance test

- The two categorical variables: (1) News Veracity; (2) Dual Emotion Category
- Conclusion: dual emotion signals are statistically dependent on news veracity.

Visualization of Dual Emotion Category

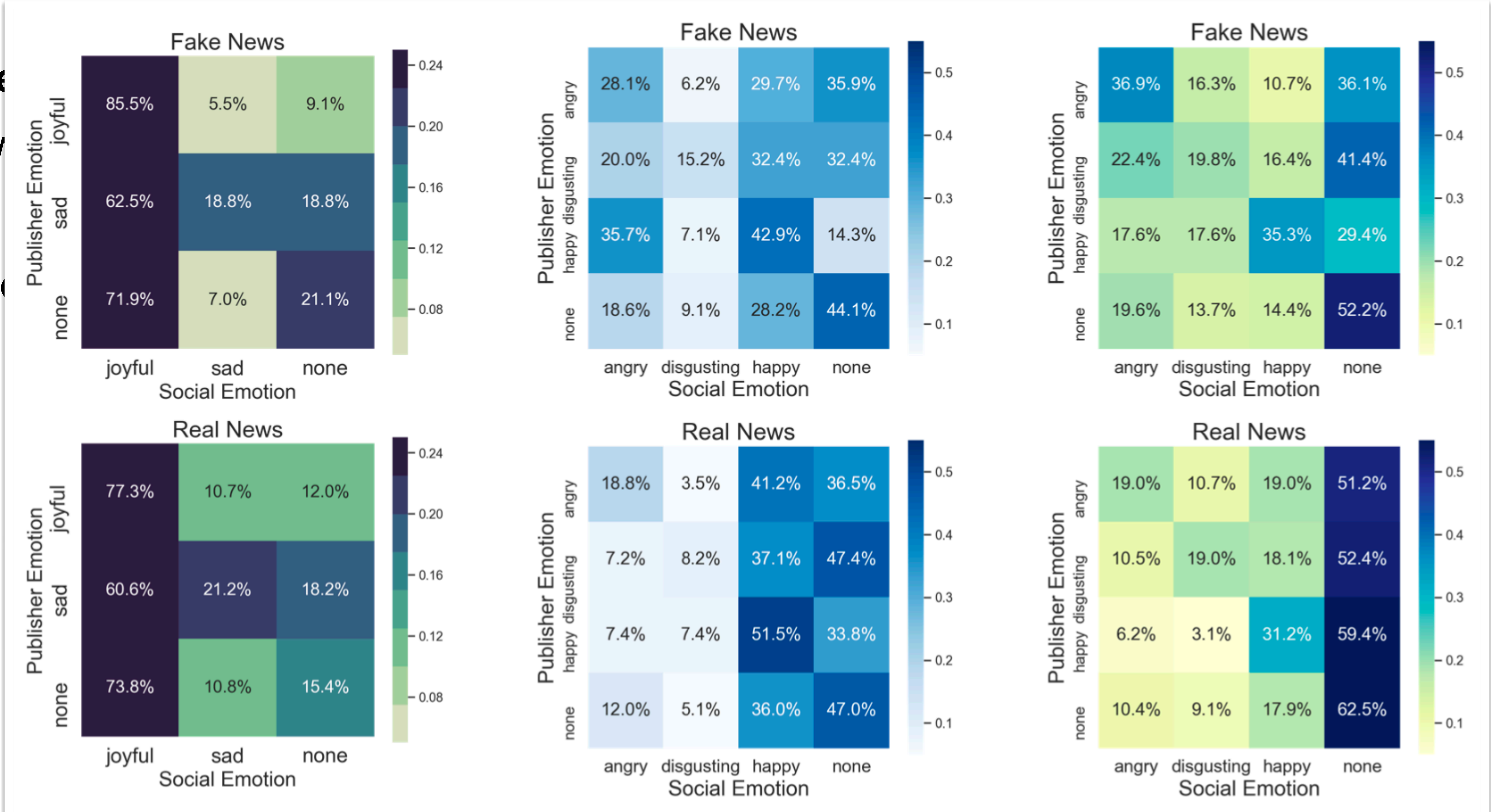
PRELIMINARY ANALYSIS

Is dual emotion distinctive between fake and real news?

Chi-square

- The tw
- Concl

Visualizati



(a) On *RumourEval-19*

(b) On *Weibo-16*

(c) On *Weibo-20*

PRELIMINARY ANALYSIS

Is dual emotion distinctive between fake and real news?

Chi-square

- The tw
- Concl

Visualizati

(a) On *RumourEval-19* (b) On *Weibo-16* (c) On *Weibo-20*

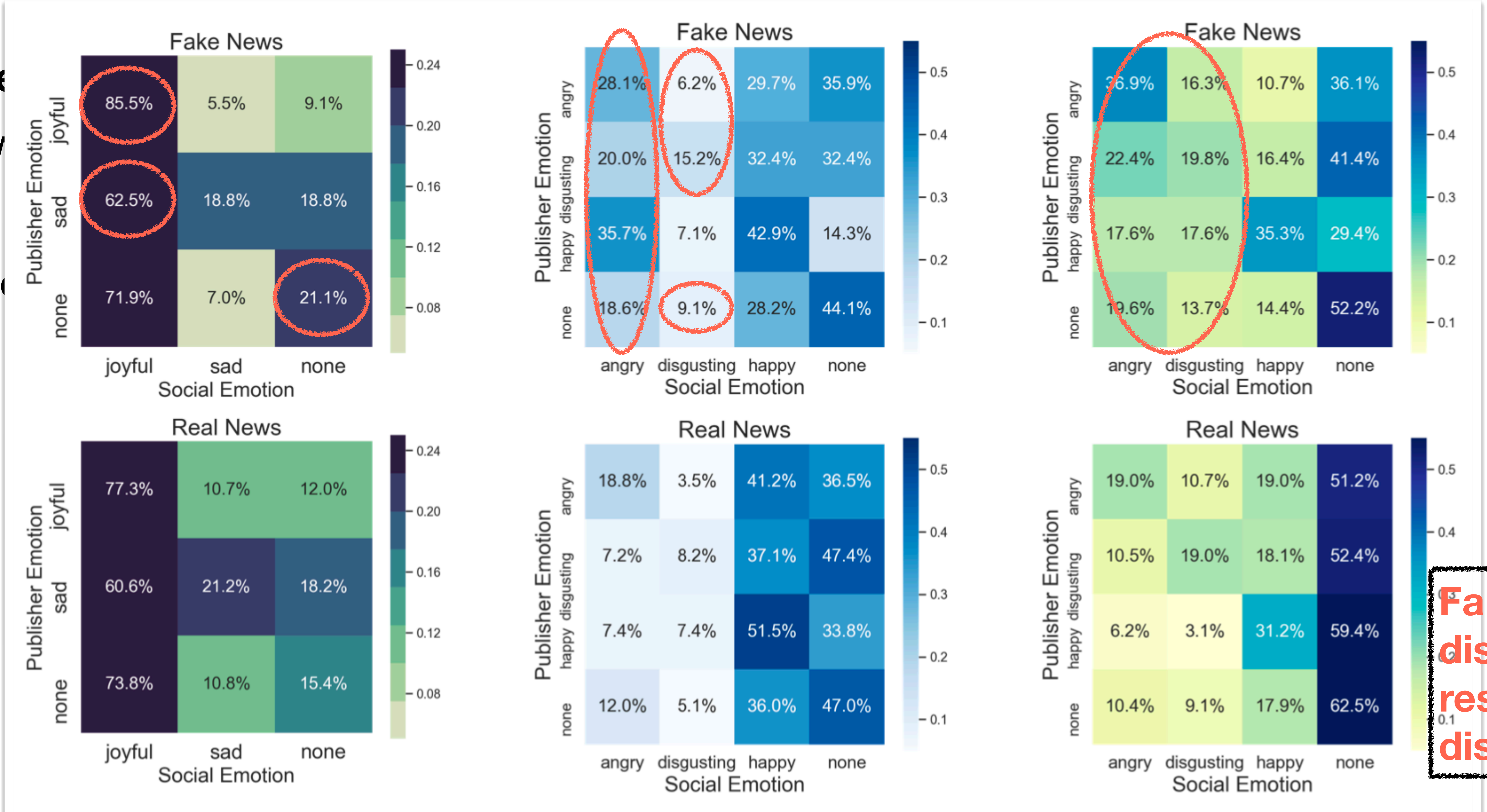
PRELIMINARY ANALYSIS

Is dual emotion distinctive between fake and real news?

Chi-square

- The tv
- Concl

Visualizati



(a) On *RumourEval-19*

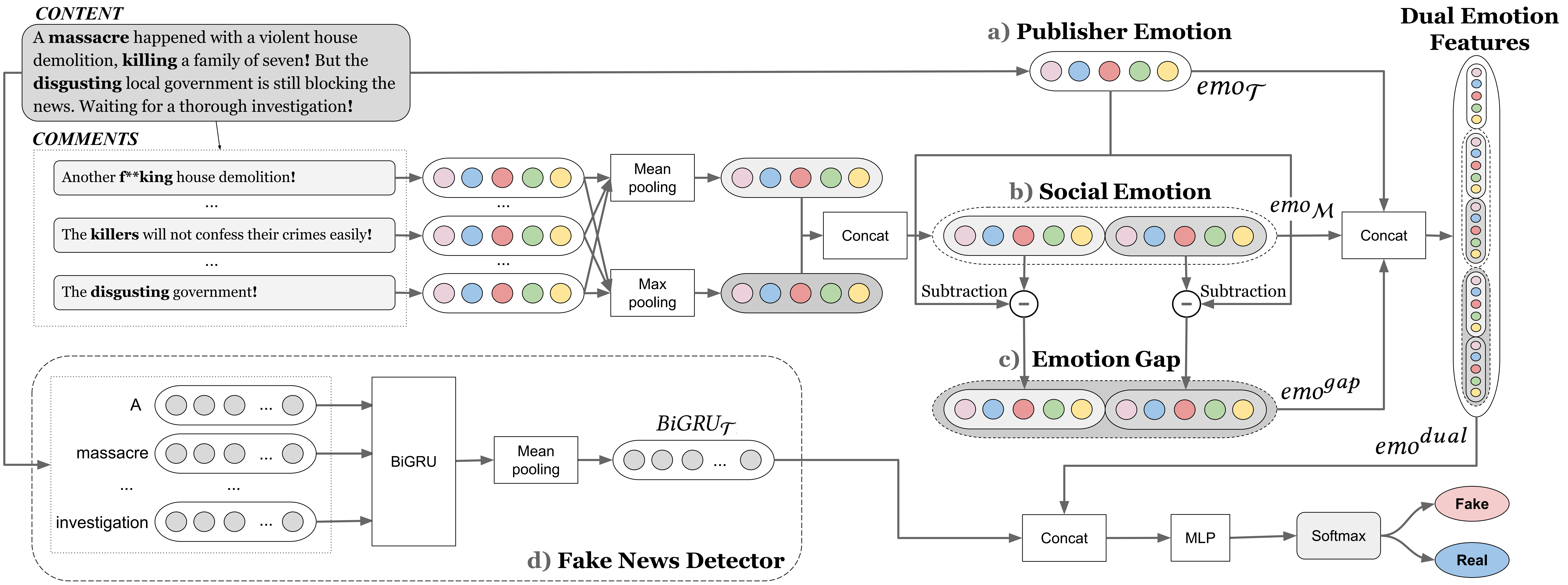
(b) On *Weibo-16*

(c) On *Weibo-20*

Fake news owns distinct emotion resonances and dissonances.

MODEL

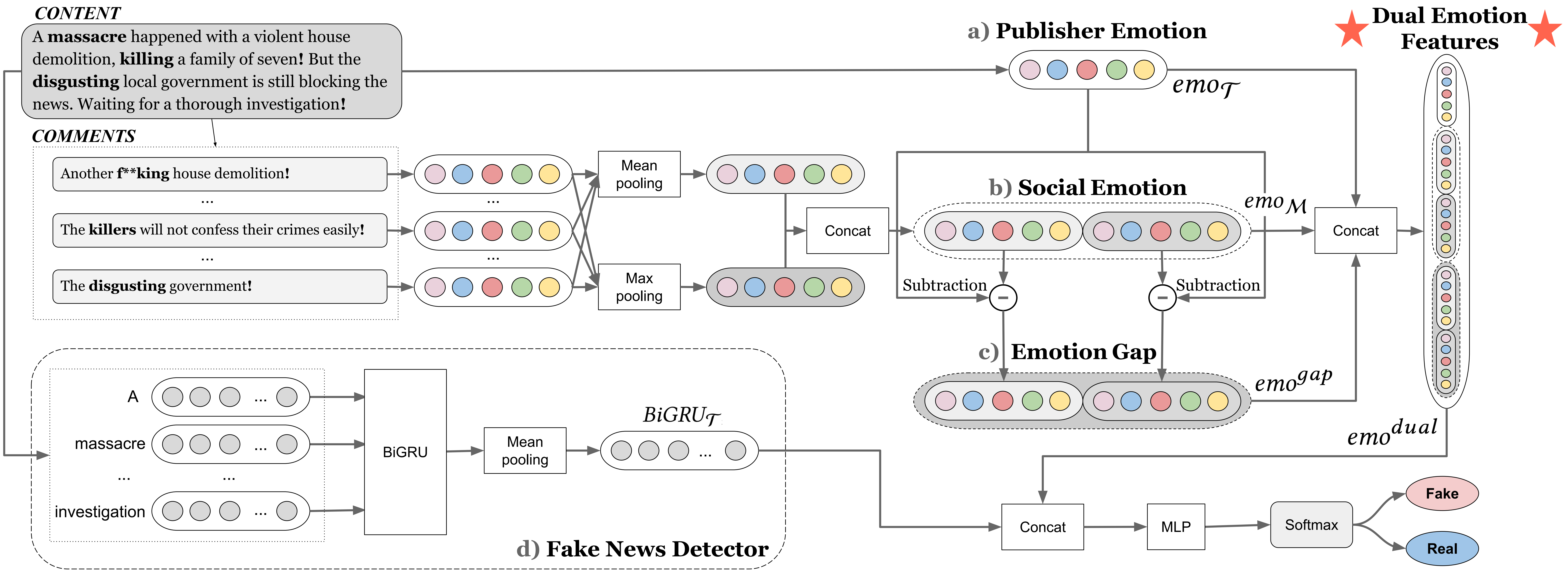
How to model the dual emotion jointly?



An overall framework of using Dual Emotion Features for fake news detection.
(Here we take BiGRU as an example of Fake News Detector.)

MODEL

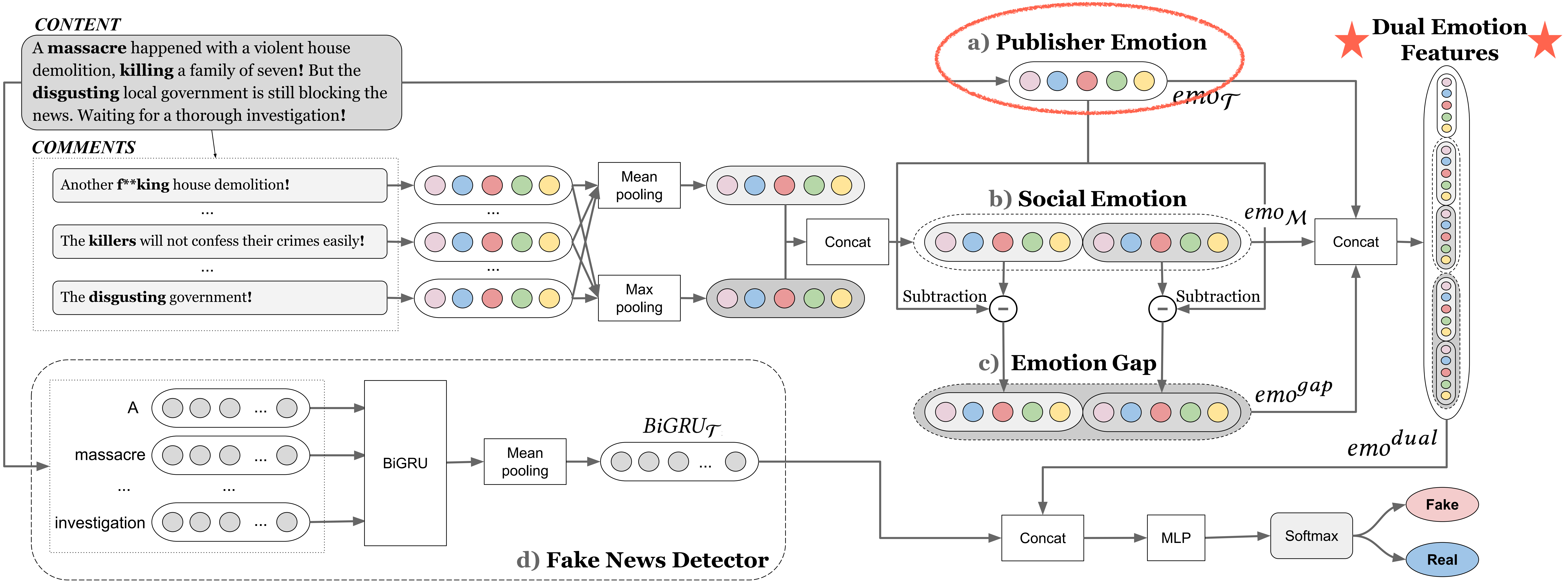
How to model the dual emotion jointly?



An overall framework of using Dual Emotion Features for fake news detection.
(Here we take BiGRU as an example of Fake News Detector.)

MODEL

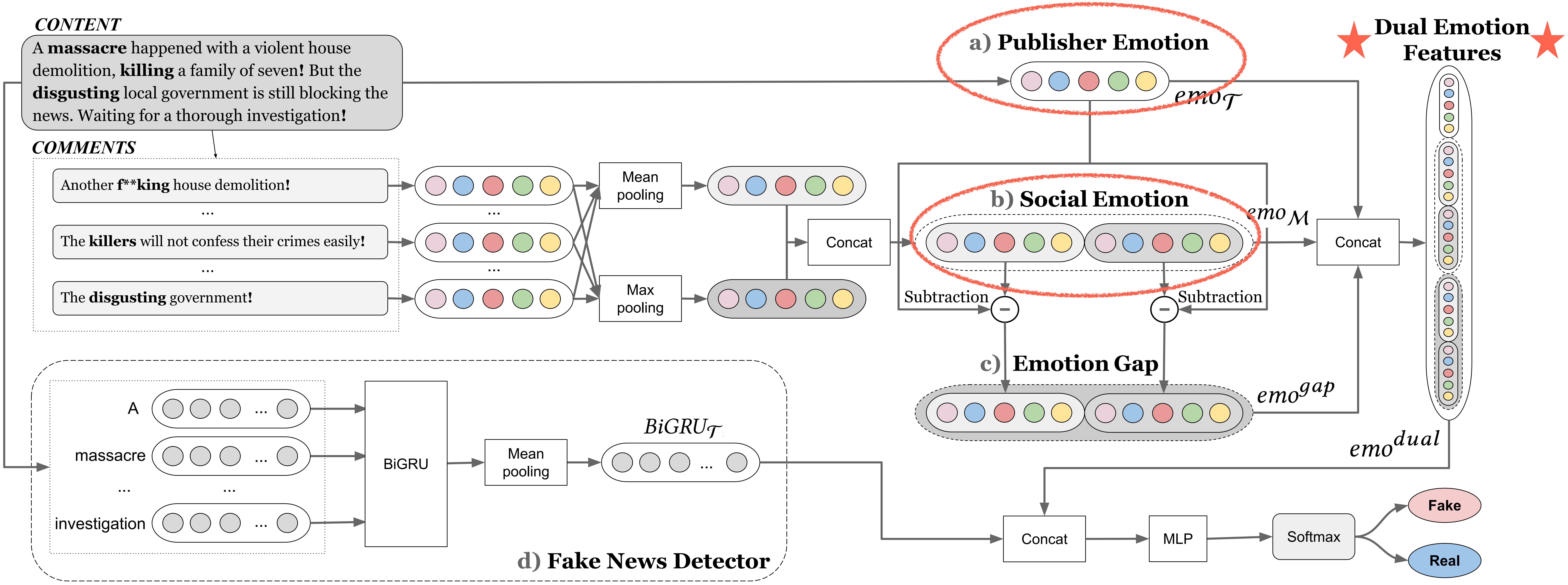
How to model the dual emotion jointly?



An overall framework of using Dual Emotion Features for fake news detection.
(Here we take BiGRU as an example of Fake News Detector.)

MODEL

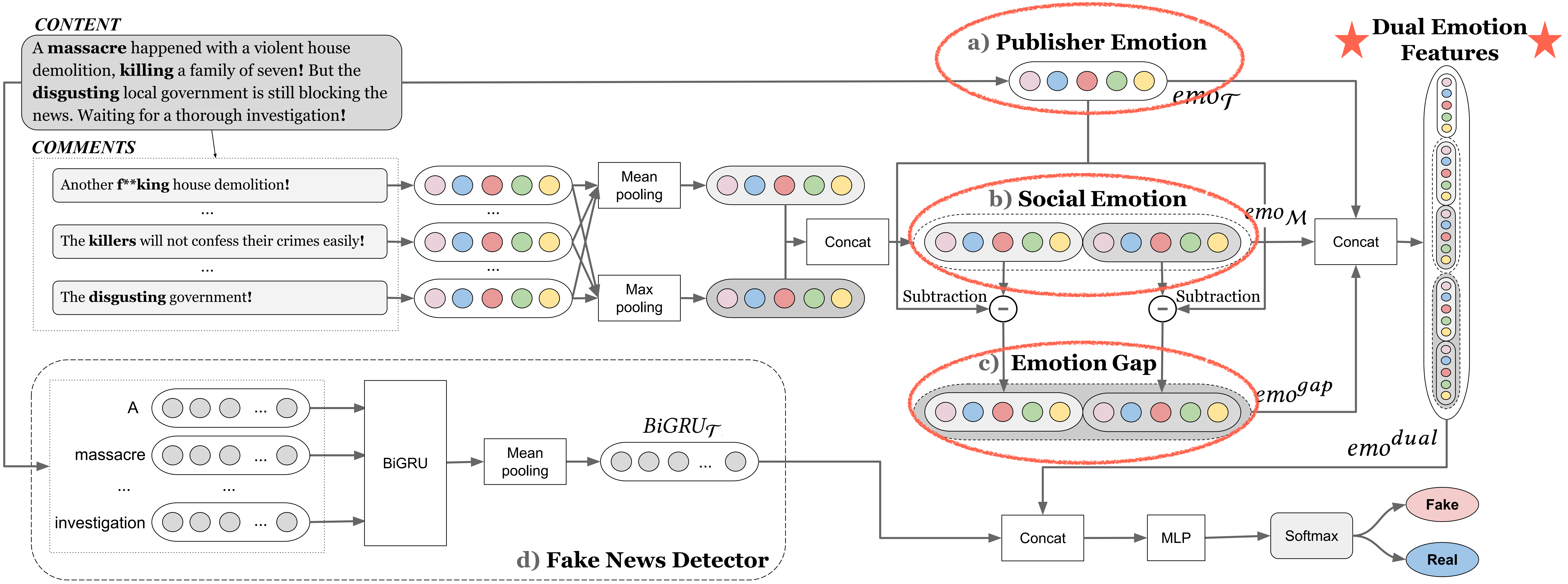
How to model the dual emotion jointly?



An overall framework of using Dual Emotion Features for fake news detection.
(Here we take BiGRU as an example of Fake News Detector.)

MODEL

How to model the dual emotion jointly?



An overall framework of using Dual Emotion Features for fake news detection.
(Here we take BiGRU as an example of Fake News Detector.)

MODEL

How to represent emotions?

Five types of emotion features

Type	Resources	Description	The level of emotional signals
Emotion Category	NVIDIA, Baidu AI	depending on pretrained emotion classifiers	overall level
Emotion Lexicon	NRC Emotion lexicon, Affective Lexicon Ontology	depending on handcrafted emotional dictionaries	word-level
Emotional Intensity			word-level
Sentiment Score	Vader, HowNet		overall level
Other Auxiliary Features	Wikipedia, HowNet, and others		word- and symbol-lvel

EXPERIMENTS

Dataset

- RumourEval-19 [7]
- Weibo-16 [8]
- Weibo-20

	Veracity	RumourEval-19		Weibo-16		Weibo-20	
		#pcs	#com	#pcs	# com	#pcs	#com
Training	Fake	79	1,135	801	649,673	1,896	749,141
	Real	144	1,905	1,410	482,226	1,920	516,795
	Unverified	104	1,838	-	-	-	-
	Total	327	4,878	2,211	1,131,899	3,816	1,265,936
Validating	Fake	19	824	268	222,149	632	137,941
	Real	10	404	470	146,948	640	185,087
	Unverified	9	212	-	-	-	-
	Total	38	1,440	738	369,097	1,272	323,028
Testing	Fake	40	689	286	193,740	633	245,216
	Real	31	805	471	179,942	641	149,260
	Unverified	10	181	-	-	-	-
	Total	81	1,675	757	373,682	1,274	394,476
Total	Fake	138	2,648	1,355	1,065,562	3,161	1,132,298
	Real	185	3,114	2,351	809,116	3,201	851,142
	Unverified	123	2,231	-	-	-	-
	Total	446	7,993	3,706	1,874,678	6,362	1,983,440

Table 2: Statistics of the three datasets. #pcs: number of news pieces; #com: number of comments.

[7] Genevieve Gorrell, Ahmet Aker, Kalina Bontcheva, Leon Derczynski, Elena Kochkina, Maria Liakata, and Arkaitz Zubiaga. SemEval-2019 Task 7: RumourEval, Determining Rumour Veracity and Support for Rumours. In SemEval@NAACL-HLT 2019. 845–854.
[8] Jing Ma, Wei Gao, Prasenjit Mitra, Sejeong Kwon, Bernard J Jansen, Kam-Fai Wong, and Meeyoung Cha. Detecting rumors from microblogs with recurrent neural networks. In IJCAI 2016. 3818–3824.

EXPERIMENTS

Baselines and Fake News Detectors

Baseline emotion features:

- Emoratio [4]
- EmoCred [5]

Fake News Detectors:

- BiGRU
- BERT [9]
- NileTMRG [10]
- HSA-BLSTM [11]

[9] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In NAACL-HLT (1) 2019: 4171-4186.
[10] Omar Enayet and Samhaa R. El-Beltagy. NileTMRG at SemEval-2017 Task 8: Determining Rumour and Veracity Support for Rumours on Twitter. In SemEval@ACL 2017. 470–474.
[11] Han Guo, Juan Cao, Yazi Zhang, Junbo Guo, and Jintao Li. RumorDetection with Hierarchical Social Attention Network. In CIKM 2018. 943–951.

EXPERIMENTS

Using Dual Emotion Features alone to detect fake news

Source	Emotion Features	R-19	W-16	W-20
Content	Emoratio	0.185	0.553	0.524
	EmoCred	0.253	0.564	0.542
	Publisher Emotion	0.290	0.571	0.573
Comments	**Social Emotion**	0.296	0.692	0.754
Content, Comments	**Emotion Gap**	0.332	0.716	0.746
	Dual Emotion Features	**0.337**	**0.728**	**0.759**

Table 3: Macro F1 scores when only using emotion features on the MLP model. R-19: RumourEval-19, W-16: Weibo-16, W-20: Weibo-20.

Removed type	R-19	W-16	W-20
Emotion Category	0.193	0.679	0.686
Emotion Lexicon	0.239	0.715	0.745
Emotional Intensity	0.216	0.725	0.750
Sentiment Score	0.245	0.723	0.743
Other Auxiliary Features	0.307	0.653	0.722

Table 4: Macro F1 scores of *Dual Emotion Features* when removing one specific type of emotion features on the MLP model. R-19: RumourEval-19, W-16: Weibo-16, W-20: Weibo-20.

EXPERIMENTS

Plug Dual Emotion Features into Fake News Detectors

Models	Macro F1 score	RMSE	F1 score		
			Fake News	Real News	Unverified News
BiGRU	0.269	0.804	0.500	0.222	0.083
+ Emoratio	0.275	0.823	0.463	0.160	**0.200**
+ EmoCred	0.311	0.797	0.456	0.295	0.182
+ **Dual Emotion Features**	**0.340**	**0.752**	**0.580**	**0.337**	0.104
BERT	0.272	0.808	0.533	0.105	0.176
+ Emoratio	0.271	0.857	0.406	0.240	0.167
+ EmoCred	0.308	0.833	0.367	**0.367**	0.189
+ **Dual Emotion Features**	**0.346**	**0.778**	**0.557**	0.244	**0.238**
NileTMRG	0.309	0.770	0.557	0.245	0.125
+ Emoratio	0.331	**0.754**	**0.571**	0.280	**0.143**
+ EmoCred	0.307	0.786	0.296	0.500	0.125
+ **Dual Emotion Features**	**0.342**	**0.754**	0.565	**0.565**	0.100

Table 5: Results on *RumourEval-19*.

EXPERIMENTS

Plug Dual Emotion Features into Fake News Detectors

Models	Weibo-16				Weibo-20			
	Macro F1 score	Accuracy	F1 score		Macro F1 score	Accuracy	F1 score	
			Fake	Real			Fake	Real
BiGRU	0.807	0.822	0.754	0.860	0.839	0.839	0.839	0.839
+ Emoratio	0.794	0.810	0.738	0.851	0.850	0.850	0.854	0.846
+ EmoCred	0.766	0.778	0.711	0.820	0.829	0.829	0.836	0.821
+ **Dual Emotion Features**	**0.826**	**0.838**	**0.781**	**0.871**	**0.855**	**0.855**	**0.857**	**0.852**
BERT	0.824	0.845	0.762	0.886	0.900	0.900	0.900	0.900
+ Emoratio	0.837	0.857	0.780	0.894	0.901	0.901	0.900	0.902
+ EmoCred	0.849	0.867	0.797	**0.901**	0.902	0.902	0.901	0.903
+ **Dual Emotion Features**	**0.867**	**0.873**	**0.837**	0.896	**0.915**	**0.915**	**0.913**	**0.918**
HSA-BLSTM	0.849	0.855	0.819	0.879	0.913	0.913	0.912	0.914
+ Emoratio	0.863	0.872	0.829	0.898	0.920	0.920	0.920	0.920
+ EmoCred	0.854	0.861	0.822	0.886	0.903	0.903	0.902	0.905
+ **Dual Emotion Features**	**0.908**	**0.913**	**0.885**	**0.930**	**0.932**	**0.932**	**0.932**	**0.933**

Table 6: Results on *Weibo-16* and *Weibo-20*.

EXPERIMENTS

Plug Dual Emotion Features into Fake News Detectors

- In the fields of fake news detection, when splitting datasets, most works just **shuffle** the datasets and split them into train / val. / test sets
- However, in the real-world scenarios, when a check-worthy news piece emerges, we only own the data **previously-emerging** to train the detector, which cannot be guaranteed when adopting the above data split.
- Solution: temporal data-split strategy

Models	Macro F1	Acc.	F1 score	
			Fake	Real
BiGRU	0.680	0.681	0.694	0.666
+ Emoratio	0.628	0.632	0.665	0.592
+ EmoCred	0.659	0.666	0.709	0.609
+ **Dual Emotion Features**	**0.701**	**0.702**	**0.714**	**0.689**
BERT	0.722	0.728	0.762	0.682
+ Emoratio	0.719	0.724	0.757	0.681
+ EmoCred	0.725	0.728	0.752	**0.699**
+ **Dual Emotion Features**	**0.734**	**0.734**	**0.773**	0.692
HSA-BLSTM	0.776	0.778	0.796	0.686
+ Emoratio	0.771	0.774	0.796	0.663
+ EmoCred	0.777	0.781	0.806	0.646
+ **Dual Emotion Features**	**0.805**	**0.808**	**0.827**	**0.694**

Table 7: Results on *Weibo-20* (temporal data split). Acc. is short for Accuracy.

EXPERIMENTS

Ablation study of Dual Emotion Features

Models		R-19	W-16	W-20
BiGRU+	Publisher Emotion	0.310	0.809	0.842
	Social Emotion	0.322	0.818	0.847
	Emotion Gap	0.336	0.811	0.849
	Dual Emotion Features	**0.340**	**0.826**	**0.855**
BERT+	Publisher Emotion	0.312	0.850	0.889
	Social Emotion	0.339	0.856	0.911
	Emotion Gap	0.338	0.858	0.906
	Dual Emotion Features	**0.346**	**0.867**	**0.915**
Nile TMRG+	Publisher Emotion	0.311	-	-
	Social Emotion	0.325	-	-
	Emotion Gap	0.337	-	-
	Dual Emotion Features	**0.342**	-	-
HSA-BLSTM+	Publisher Emotion	-	0.876	0.915
	Social Emotion	-	0.892	0.922
	Emotion Gap	-	0.901	0.926
	Dual Emotion Features	-	**0.908**	**0.932**

Table 8: Ablation study of the three components of *Dual Emotion Features*. The evaluation metric is macro F1 scores. R-19: RumourEval-19, W-16: Weibo-16, W-20: Weibo-20

CASE STUDY

Mining dual emotion as a remedy

Content — **Publisher emotion:** ***Angry***

Black Lives Matter **THUGS** Blocking Emergency Crews From Reaching Hurricane Victims.

Comments — **Social emotion:** ***Angry***

- **Run over** their **asses**.
- ...
- When thugs block emergency vehicles, vehicles should **run over** the **thugs**
- ...
- And why don't we have military support there to enforce the law?

	Fake	Real	Unverified
BiGRU	0.33	**0.61**	0.06
BiGRU + Emoratio	0.35	**0.57**	0.08
BiGRU + EmoCred	0.27	**0.64**	0.09
BiGRU + **Dual Emotion Features**	**0.65**	0.21	0.14

Content — **Publisher emotion:** ***Joyful***

Believe it or not, this is a shark on the freeway in #Houston, #Texas #HurricaneHarvy...

Comments — **Social emotion:** ***Joyful***

- Fresh water? **Nice** try.
- ...
- **Oh my!**
- ...
- It's trying to ask you for directions...must be a female shark**!!** 😆

	Fake	Real	Unverified
BiGRU	0.31	**0.50**	0.19
BiGRU + Emoratio	0.36	**0.56**	0.08
BiGRU + EmoCred	0.40	**0.54**	0.06
BiGRU + **Dual Emotion Features**	**0.65**	0.22	0.13

Content — **Publisher emotion:** ***None***

Category 6? If Hurricane Irma Becomes The Strongest Hurricane In History, It Could Wipe Entire Cities Off The Map.

Comments — **Social emotion:** ***Sad***

- Watch what the power of **prayer** does
- ...
- Get out of there people. Florida, South Carolina.
- ...
- **Pray to god** the only chance you have...

	Fake	Real	Unverified
BiGRU	0.47	**0.52**	0.01
BiGRU + Emoratio	0.40	**0.59**	0.01
BiGRU + EmoCred	0.28	**0.58**	0.14
BiGRU + **Dual Emotion Features**	**0.63**	0.17	0.20

Figure 4: Three fake news pieces on *RumourEval-19*, which are missed by original BiGRU but detected after using *Dual Emotion Features*. The prediction results of the four models are shown at the bottom, where the numbers represent confidence scores (a float value from 0 to 1). The scores that identify prediction labels are shown in bold.

REFERENCES

[1] Chuan Guo, Juan Cao, Xueyao Zhang, Kai Shu, Miao Yu. Exploiting Emotions for Fake News Detection on Social Media[J]. arXiv:1903.01728, 2019.
[2] Soroush Vosoughi, Deb Roy, Sinan Aral. The spread of true and false news online[J]. Science, 2018, 359(6380): 1146-1151.
[3] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. Information credibility on twitter. In WWW 2011. 675–684.
[4] Oluwaseun Ajao, Deepayan Bhowmik, and Shahrzad Zargari. Sentiment Aware Fake News Detection on Online Social Networks. In IEEE ICASSP 2019. 2507–2511.
[5] Anastasia Giachanou, Paolo Rosso, and Fabio Crestani. 2019. Leveraging Emotional Signals for Credibility Detection. In ACM SIGIR 2019. 877–880.
[6] Ralph L Rosnow. Inside rumor: A personal journey. American psychologist 46, 5 (1991), 484.
[7] Genevieve Gorrell, Ahmet Aker, Kalina Bontcheva, Leon Derczynski, Elena Kochkina, Maria Liakata, and Arkaitz Zubiaga. SemEval-2019 Task 7: RumourEval, Determining Rumour Veracity and Support for Rumours. In SemEval@NAACL-HLT 2019. 845–854.
[8] Jing Ma, Wei Gao, Prasenjit Mitra, Sejeong Kwon, Bernard J Jansen, Kam-Fai Wong, and Meeyoung Cha. Detecting rumors from microblogs with recurrent neural networks. In IJCAI 2016. 3818–3824.
[9] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In NAACL-HLT (1) 2019: 4171-4186.
[10] Omar Enayet and Samhaa R. El-Beltagy. NileTMRG at SemEval-2017 Task 8: Determining Rumour and Veracity Support for Rumours on Twitter. In SemEval@ACL 2017. 470–474.
[11] Han Guo, Juan Cao, Yazi Zhang, Junbo Guo, and Jintao Li. RumorDetection with Hierarchical Social Attention Network. In CIKM 2018. 943–951.

THANKS

Feel free to contact zhangxueyao19s@ict.ac.cn

The source code and datasets are released at https://github.com/RMSnow/WWW2021