Mining Dual Emotion for Fake News Detection

Xueyao Zhang[†]
Institute of Computing Technology,
Chinese Academy of Sciences
University of Chinese Academy of
Sciences
zhangxueyao19s@ict.ac.cn

Juan Cao[*][†]
Institute of Computing Technology,
Chinese Academy of Sciences
University of Chinese Academy of
Sciences
caojuan@ict.ac.cn

Xirong Li[*]
Key Lab of Data Engineering and
Knowledge Engineering, Renmin
University of China
Beijing, China
xirong@ruc.edu.cn

Qiang Sheng[†]
Institute of Computing Technology,
Chinese Academy of Sciences
University of Chinese Academy of
Sciences
shengqiang18z@ict.ac.cn

Lei Zhong[†]
Institute of Computing Technology,
Chinese Academy of Sciences
University of Chinese Academy of
Sciences
zhonglei18s@ict.ac.cn

Kai Shu
Illinois Institute of Technology
Chicago, Illinois, USA
kshu@iit.edu

ABSTRACT

Emotion plays an important role in detecting fake news online. When leveraging emotional signals, the existing methods focus on exploiting the emotions of news contents that conveyed by the publishers (i.e., *publisher emotion*). However, fake news often evokes high-arousal or activating emotions of people, so the emotions of news comments aroused in the crowd (i.e., *social emotion*) should not be ignored. Furthermore, it remains to be explored whether there exists a relationship between *publisher emotion* and *social emotion* (i.e., *dual emotion*), and how the *dual emotion* appears in fake news. In this paper, we verify that *dual emotion* is distinctive between fake and real news and propose *Dual Emotion Features* to represent *dual emotion* and the relationship between them for fake news detection. Further, we exhibit that our proposed features can be easily plugged into existing fake news detectors as an enhancement. Extensive experiments on three real-world datasets (one in English and the others in Chinese) show that our proposed feature set: 1) outperforms the state-of-the-art task-related emotional features; 2) can be well compatible with existing fake news detectors and effectively improve the performance of detecting fake news.[1] [2]

[*]Corresponding authors

[†]At Key Laboratory of Intelligent Information Processing of Chinese Academy of Sciences. Also at State Key Laboratory of Communication Content Cognition, People's Daily Online.

[1]Please kindly note that the examples in this paper contain offensive and swear words.
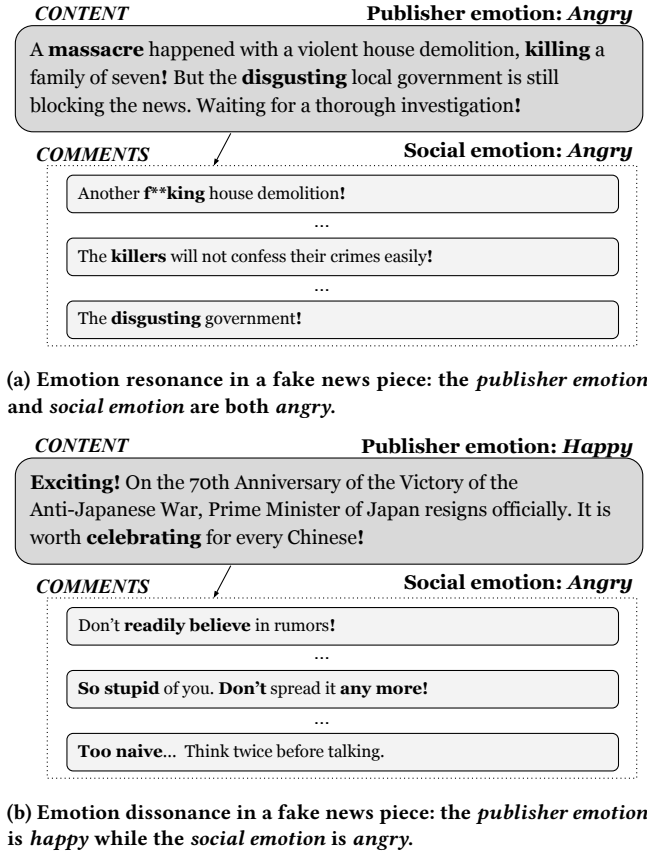[2]The code and datasets are released at https://github.com/RMSnow/WWW2021.

1 INTRODUCTION

In recent years, fake news on social media has threatened not only cyberspace security, but also the real-world order in politics [14], economy [12], society [2], etc. The most recent example is the concomitant *infodemic* during the COVID-19 pandemic across the world [41]. Thousands of news pieces with misleading content have been spreading through social media [44] and led to socio-economic disorder [6] and weakened the effect of pandemic prevention [4]. To tackle this issue, researchers have been devoted to developing automatic methods to detect fake news (i.e., designing a classifier to judge a given news piece as real or fake) by leveraging signals from text [5, 32, 34], images [20, 35], or social contexts [17, 24, 26–28, 39, 40]. [3]

In existing text-based works [1, 5, 15], the role of sentimental or emotional signals has been considered for fake news detection. Ajao et al. [1] point out that there exists a relationship between news veracity and the sentiments of the posted text, and append a sentimental feature (the ratio of the number of negative and positive words) to help text-only fake news detectors. Instead of appending a sole feature, Giachanou et al. [15] extract richer emotional features from the news contents based on emotional lexicons for fake news detection. To the best of our knowledge, most existing works leverage the emotional signals of fake news content conveyed by the publishers but rarely focus on the emotions of fake news comments aroused in the crowd. However, for spreading in the crowd virally, fake news often evokes high-arousal or activating emotions of the crowd [37]. Therefore, in addition to emotions of news contents, it is necessary to explore whether emotions of news comments and the relationship between the two emotions are helpful for fake news detection.

To describe the two emotions clearly, we define them respectively as 1) **publisher emotion**: the emotions conveyed by publishers of the news pieces; and 2) **social emotion**: the emotions aroused in the crowd facing to the news pieces. And we adopt **dual emotion** as a general term of these two emotions. For a news piece, *dual*

[3]In this paper, we use *news pieces* to refer to social media news posts. A news piece generally contains content and its attached comments.

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*.

Figure 1: Two fake news pieces on Chinese microblog platform Weibo, with different *Dual Emotion*. The texts are translated from Chinese to English manually.

emotion has two appearances: emotion resonances (i.e., the *publisher emotion* is same or similar to the *social emotion*) and emotion dissonances (i.e., the *publisher emotion* is different from the *social emotion*). We analyze the data and find that the two appearances have a statistically significant distinction between fake and real news (see details in Section 4.2). For example, as to the emotion resonance, there are more fake news pieces whose *dual emotion* are both *angry* than real news, while as to the emotion dissonances, more fake news pieces whose *publisher emotion* is *happy* while *social emotion* is *angry*. Figure 1 shows two representative examples selected from fake news pieces on Weibo [4]. In Figure 1a, the fake news publisher conveys its rage with expressions like "massacre", "killing", "disgusting". As a result, the great indignation of the crowd is evoked, shown by "f**king", "killers", and "disgusting". In Figure 1b, the fake news publisher expresses happiness with "Exciting!" and "celebrating". While the crowd considers it as a ridiculous news piece, and use "readily believe", "So stupid" and "Too naive" to express their disgust and contempt to the publisher. The data observation statistical findings highlight that the

---

[4]https://www.weibo.com

relationship in *dual emotion* can be indicative of the news veracity and should be considered when modeling.

To model the *dual emotion* and emotion resonances and dissonances for fake news detection, we propose *Dual Emotion Features* to represent *publisher emotion*, *social emotion* and the similarity and difference of the *dual emotion* jointly. Besides, it is convenient to implement and plug the features into existing fake news detectors as an enhancement.

In this paper, our contributions are summarized as follows:

- We propose and verify that the *dual emotion* (i.e., *publisher emotion* and *social emotion*) signal is distinctive between fake and real news.
- We firstly propose the feature set, *Dual Emotion Features*, to comprehensively represent dual emotion and the relationship between the two kinds of emotions, and exhibit how to plug it into the fake news detectors as a complement and enhancement.
- We conduct experiments on the real-world datasets, including a newly-constructed Chinese dataset. The results demonstrate that: 1) *Dual Emotion Features* outperforms the existing emotional features for fake news detection. 2) It can be compatible with existing fake news detectors and effectively improve the performance of the detectors.

2 RELATED WORK

Fake news detection is also known as false news detection, rumor detection, misinformation detection, etc. [33] and is closely connected to the field of information credibility evaluation. In the earliest study on information credibility evaluation, Castillo et al. [5] manually extract content features, publisher features, topic features, and propagation features from news pieces. And the work finds that sentiment-based features like the fraction of sentimental words and exclamation marks are effective for evaluating information credibility. In recent years, researchers begin to utilize deep learning models such as GRU-based and CNN-based models for fake news detection [26, 47]. Beyond news content, social contexts such as texts of comments and reposts [17, 26–28, 38], viewpoints and stances of the crowd [19, 22], and user credibility[24, 40] are emphasized as well.

There are also existing works focusing on discovering the distinctive emotional signals between fake and real news. Ajao et al. [1] verify that there exists a relationship between news veracity (real or fake) and the usage of sentimental words, and design an emotion feature (the ratio of the count of negative and positive words) to help detect fake news. Besides, Giachanou et al. [15] extract emotion features based on emotional lexicons from news contents for fake news detection. However, these works only leverage the emotional signals of fake news contents but ignore the emotions of fake news comments and the relationship between the two emotions. Recently, Wu and Rao [45] propose an adaptive fusion network for fake news detection, modeling emotion embeddings from the contents and the comments. However, this work focuses on adaptively fusing various features by advanced deep learning models, and do not explore the specific distinction of dual emotion signals between fake and real news. So far, the work that pays attention to mining dual emotion signals from publishers and crowds remains vacant.
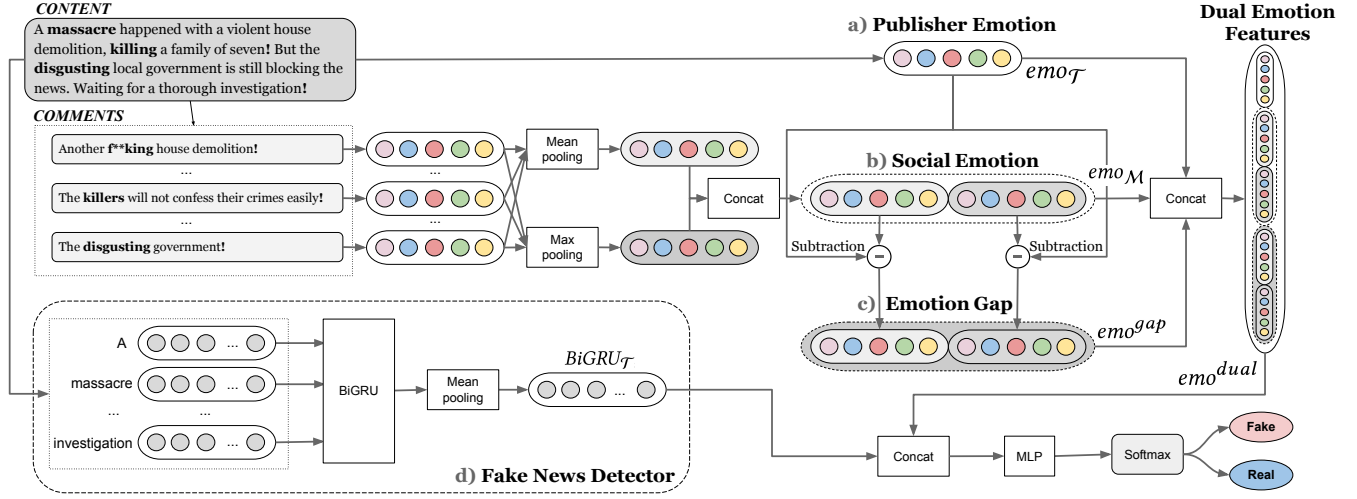
Figure 2: An overall framework of using *Dual Emotion Features* for fake news detection. *Dual Emotion Features* consist of three components: a) *Publisher Emotion* extracted from the content; b) *Social Emotion* extracted from the comments; c) *Emotion Gap* representing the similarity and difference between publisher emotion and social emotion. *Dual Emotion Features* are concatenated with the features from d) Fake News Detector (here, BiGRU as an example) for the final prediction of veracity.

3 MODELING DUAL EMOTION FOR FAKE NEWS DETECTION

To model dual emotion signals for fake news detection, we propose *Dual Emotion Features*, which can leverage publisher emotion, social emotion, and the similarity and difference of the dual emotion. Figure 2 exhibits the process of obtaining *Dual Emotion Features* and integrating them into an existing fake news detector as an enhancement to classify a given piece of news. In this section, we detail the feature extraction of publisher emotion and social emotion, and the modeling of emotion gap. Then, we describe the process to plug *Dual Emotion Features* into the existing fake news detectors.

3.1 Publisher Emotion

To comprehensively represent the *Publisher Emotion*, we use a variety of features extracted from news contents, including the emotion category, emotional lexicon, emotional intensity, sentiment score, and other auxiliary features. In the five kinds of features, emotion category, emotional intensity and sentiment score provide the overall information and the other two provide word- and symbol-level information.

Given the input sequence of the textual content with length L, $\mathcal{T} = [t_1, t_2, \ldots, t_i, \ldots, t_L]$, where t_i is the ith word in the text, the goal is to extract emotion features $emo_{\mathcal{T}}$ from the text $\mathcal{T}$.

3.1.1 Emotion Category. We use public emotion classifiers (which will be introduced in Section 4.2) to get emotion category features. Usually, the output of an emotion classifier is the probabilities that the given text contains certain emotions.

Given the emotion classifier f and the text $\mathcal{T}$, we assume the dimension of the output is d_f and thus the prediction of the text is

$f(\mathcal{T})$. So we can obtain the emotion category features $emo_{\mathcal{T}}^{cate} = f(\mathcal{T})$, where $emo_{\mathcal{T}}^{cate} \in \mathbb{R}^{d_f}$.

3.1.2 Emotional Lexicon. Usually, a piece of text conveys specific emotions by using several specific words (which are generally included in emotional lexicons). Thus, we next extract the features based on the emotional lexicon. The approach is dependent on the existing emotion dictionaries annotated by experts. In the emotion dictionary, we assume that there are d_e kinds of emotions, denoted as $E = \{e_1, e_2, \ldots, e_{d_e}\}$. For the emotion $e \in E$, the dictionary provides a list of emotional words $\mathcal{E}_e = \{w_{e,1}, w_{e,2}, \ldots, w_{e,L_e}\}$, where L_e is the length of the emotion lexicon of e in the dictionary.

Given the text $\mathcal{T}$, we gradually aggregate the scores of each word and the whole text across all the emotions for rich representation. For one of the emotions e, we firstly calculate the word-level score $s(t_i, e)$, where t_i is i^{th} word in the text $\mathcal{T}$. If the word t_i is in the dictionary $\mathcal{E}_e$, we consider not only its occurrence frequency, but also its contextual words (specifically, degree words and negation words). For example, in the sentence "I am not very joyful today" (the length of the sentence is 6), "joyful" belongs to the emotion *happy* and its occurrence frequency is 1/6. Assume that we only consider the left context and the window size is 2 (i.e., the context words are "not" and "very"). When we set the negation value of "not" as -1 and the degree value of "very" as 2, the final $s(joyful, e_{happy}) = -1 * 2 * (1/6) = -1/3$. In practice, we use the existing emotion dictionary to match and calculate the values of negation and degree words. As described above, $s(t_i, e)$ is defined in Equation 1:

$$s(t_i, e) = \frac{\mathbb{1}_{\mathcal{E}_e}(t_i) * neg(t_i, w) * deg(t_i, w)}{L} \qquad (1)$$

$$\mathbb{1}_{\mathscr{E}_e}(t_i) = \begin{cases} 1, & if \ t_i \in \mathscr{E}_e \\ 0, & otherwise \end{cases} \quad (2)$$

where w is the window size of the left context. And $neg(t_j)$ (Equation 3) and $deg(t_j)$ (Equation 4) are respectively the negation value and degree value of t_j, which can be looked up according to the emotion dictionary.

$$neg(t_i, w) = \prod_{j=i-w}^{i-1} neg(t_j) \quad (3)$$

$$deg(t_i, w) = \prod_{j=i-w}^{i-1} deg(t_j) \quad (4)$$

We then calculate text-level score on the specific emotion e, denoted as $s(\mathcal{T}, e)$, by summing the scores of each word in the text, as Equation 5 shows:

$$s(\mathcal{T}, e) = \sum_{i=1}^{L} s(t_i, e), \quad \forall e \in E \quad (5)$$

Finally, the emotional lexicon features $emo_{\mathcal{T}}^{lex}$ are obtained by concatenating all the scores of the d_e emotions (Equation 6), where $\oplus$ is the concatenation operator, and $emo_{\mathcal{T}}^{lex} \in \mathbb{R}^{d_e}$.

$$emo_{\mathcal{T}}^{lex} = s(\mathcal{T}, e_1) \oplus s(\mathcal{T}, e_2) \oplus \cdots \oplus s(\mathcal{T}, e_{d_e}) \quad (6)$$

3.1.3 Emotional Intensity. As for emotional lexicons, we also consider the emotional intensity of the lexicons. For example, when expressing the emotion *happy*, the word "ecstatic" owns a higher intensity than "joyful". The extracting process is similar to that of the emotional lexicon features, except for that we here include the intensity scores. Given the emotions E, the emotional word list $\mathscr{E}_e$ for every emotion e, and the text $\mathcal{T}$, we first calculate the intensity-aware text-level scores $s'(\mathcal{T}, e)$ by summing the intensity-weighted word-level scores, as shown in Equation 7:

$$s'(\mathcal{T}, e) = \sum_{i=1}^{L} s'(t_i, e) = \sum_{i=1}^{L} int(t_i) * s(t_i, e), \quad \forall e \in E \quad (7)$$

where $int(t_i)$ denotes the intensity score of the word t_i. If t_i is in the dictionary, $int(t_i)$ can be calculated according to the emotion dictionary, otherwise $int(t_i) = 0$.

The emotional intensity features $emo_{\mathcal{T}}^{int}$ can be obtained by concatenating all the intensity scores of d_e kinds of emotions, as shown in Equation 8:

$$emo_{\mathcal{T}}^{int} = s'(\mathcal{T}, e_1) \oplus s'(\mathcal{T}, e_2) \oplus \cdots \oplus s'(\mathcal{T}, e_{d_e}) \quad (8)$$

where $emo_{\mathcal{T}}^{int} \in \mathbb{R}^{d_e}$.

3.1.4 Sentiment Score. In addition to the emotion-level features described above, we also consider the coarse-grained sentiment score of the text. Usually, the sentiment score is a positive or negative value, which represents the degree of the positive or negative polarity of the whole text. And it can be calculated by using sentiment dictionaries or public toolkits. Assuming that the dimension of the sentiment score is d_s (usually, $d_s = 1$), we can get the sentiment score feature $emo_{\mathcal{T}}^{senti} \in \mathbb{R}^{d_s}$.

3.1.5 Other Auxiliary Features. Considering that the above features do not explicitly exploit the information beyond emotion dictionaries, we introduce a set of auxiliary features to capture the emotional signals behind the non-word elements, including emoticons, punctuations, and uppercase letters (only for English). Also, we add the frequency of sentimental words and personal pronouns to enhance the awareness of the users' word usages. Take emoticons as an example. The emoticons are universal for emotional expression across the world, such as ":)" for *happy*, ": (" for *sad*. Besides, punctuations like "!" and "?" can also convey people's moods and emotions. Table 1 summarizes the auxiliary features used in the *Dual Emotion Features*. Assume that there are d_a features, and we can extract the other auxiliary features $emo_{\mathcal{T}}^{aux} \in \mathbb{R}^{d_a}$.

Type	Features
Emoticons	The frequency of happy emoticons
	The frequency of angry emoticons
	The frequency of surprised emoticons
	The frequency of sad emoticons
	The frequency of neutral emoticons
Punctuations	The frequency of exclamation mark
	The frequency of question mark
	The frequency of ellipsis mark
Sentimental Words	The frequency of positive sentimental words
	The frequency of negative sentimental words
	The frequency of degree words
	The frequency of negation words
Personal Pronoun	The frequency of pronoun first
	The frequency of pronoun second
	The frequency of pronoun third
Others (For English corpus)	The frequency of uppercase letters

Table 1: Auxiliary Feature List

To get the *Publisher Emotion* of the text $\mathcal{T}$ from the content, we concatenate all five kinds of features described above and obtain $emo_{\mathcal{T}}$, as shown in Equation 9:

$$emo_{\mathcal{T}} = emo_{\mathcal{T}}^{cate} \oplus emo_{\mathcal{T}}^{lex} \oplus emo_{\mathcal{T}}^{int} \oplus emo_{\mathcal{T}}^{senti} \oplus emo_{\mathcal{T}}^{aux} \quad (9)$$

where $emo_{\mathcal{T}} \in \mathbb{R}^d$ (i.e., $d = d_f + 2d_e + d_s + d_a$).

3.2 Social Emotion

We first extract *Social Emotion* from the comments of a news piece and then aggregate them as the whole representation. The comments of a news piece are denoted as $\mathcal{M} = [\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_i, \ldots, \mathcal{M}_{L_{\mathcal{M}}}]$, where $\mathcal{M}_i$ is the ith comment of the news piece, and $L_{\mathcal{M}}$ is the length of comment list. As for $\mathcal{M}_i$, we can calculate its emotion vector $emo_{\mathcal{M}_i}$ by Equation 9, where $emo_{\mathcal{M}_i} \in \mathbb{R}^d$. Then we stack the transposed emotion vector (row vector) of every comment to obtain the whole emotion vector of comments $\widehat{emo_{\mathcal{M}}}$, as shown in Equation 10:

$$\widehat{emo_{\mathcal{M}}} = emo_{\mathcal{M}_1}^{\mathsf{T}} \oplus emo_{\mathcal{M}_2}^{\mathsf{T}} \oplus \cdots \oplus emo_{\mathcal{M}_{L_{\mathcal{M}}}}^{\mathsf{T}} \quad (10)$$

where $\widehat{emo_{\mathcal{M}}} \in \mathbb{R}^{L_{\mathcal{M}} \times d}$.

After getting $\widehat{emo_{\mathcal{M}}}$, we consider two aggregators to generate the *Social Emotion* of the whole comment list: 1) Mean pooling for representing the average emotional signals (Equation 11); and 2) max pooling for capturing the extreme emotional signals (Equation 12).

$$emo_{\mathcal{M}}^{mean} = mean(\widehat{emo_{\mathcal{M}}}) \quad (11)$$

$$emo_{\mathcal{M}}^{max} = max(\widehat{emo_{\mathcal{M}}}) \quad (12)$$

where $emo_{\mathcal{M}}^{mean}, emo_{\mathcal{M}}^{max} \in \mathbb{R}^d$.

Finally, we concatenate them as the *Social Emotion*:

$$emo_{\mathcal{M}} = emo_{\mathcal{M}}^{mean} \oplus emo_{\mathcal{M}}^{max} \quad (13)$$

where $emo_{\mathcal{M}} \in \mathbb{R}^{2d}$.

3.3 Emotion Gap

To model the resonances and dissonances of dual emotion, we propose *Emotion Gap* (denoted as emo^{gap}). It is designed as the subtraction between *Publisher Emotion* and *Social Emotion*. As shown in Equation 14, emo^{gap} is concatenated by the difference of $emo_{\mathcal{T}}$ and $emo_{\mathcal{M}}^{mean}$ and the difference of $emo_{\mathcal{T}}$ and $emo_{\mathcal{M}}^{max}$:

$$emo^{gap} = (emo_{\mathcal{T}} - emo_{\mathcal{M}}^{mean}) \oplus (emo_{\mathcal{T}} - emo_{\mathcal{M}}^{max}) \quad (14)$$

where $emo^{gap} \in \mathbb{R}^{2d}$. By this means, it can measure the differences (i.e., dissonances) between the dual emotion. For emotions resonances, the values in the *Emotion Gap* vector are tiny (nearly zero).

3.4 Dual Emotion Features

Finally, *Dual Emotion Features* are concatenated by the *Publisher Emotion*, the *Social Emotion* and the *Emotion Gap*. In Equation 15 we obtain the *Dual Emotion Features*, where $emo^{dual} \in \mathbb{R}^{5d}$.

$$emo^{dual} = emo_{\mathcal{T}} \oplus emo_{\mathcal{M}} \oplus emo^{gap} \quad (15)$$

After getting *Dual Emotion Features*, we can concatenate it with representations that extracted by the fake news detectors, which is exemplified by Figure 2. Assuming that the fake news detector is BiGRU and the output feature vector is denoted as $BiGRU_{\mathcal{T}}$, the concatenated vector $[BiGRU_{\mathcal{T}}, emo^{dual}]$ is fed into a multi-layer perceptron (MLP) layer and a softmax layer for the final prediction of news veracity $\hat{y}$, as shown in Equation 16:

$$\hat{y} = \text{Softmax}(\text{MLP}([BiGRU_{\mathcal{T}}, emo^{dual}])) \quad (16)$$

4 EXPERIMENTS AND EVALUATION

In this section, we conduct experiments to compare our proposed *Dual Emotion Features* and other baseline features and explore their roles in improving the performance of fake news detection. Specifically, we mainly answer the following evaluation questions:

- **EQ1:** Are *Dual Emotion Features* more effective than baseline features when used alone for fake news detection? How effective are the different types of features in *Dual Emotion Features*?
- **EQ2:** Can *Dual Emotion Features* help improve the performance of text-based fake news detectors?

- **EQ3:** How robust do the fake news detection models with *Dual Emotion Features* in real-world scenarios?
- **EQ4:** How effective are the components of *Dual Emotion Features*, including the publisher emotion, social emotion, and emotion gap?

4.1 Dataset

Although the emotions are believed universal, albeit affected by culture [11], how emotions are expressed and perceived varies across different socio-cultural backgrounds [36]. Thus, we conduct experiments on three real-world datasets in two languages (meanwhile, two countries with different cultures), one in English (*RumourEval-19*) and two in Chinese (*Weibo-16* and *Weibo-20*). The statistics of these datasets are shown in Table 2.

4.1.1 RumourEval-19. The dataset *RumourEval-19* is constructed for determining the veracity of the rumors on Twitter and Reddit. It is released in an academic evaluation[5] [16]. Each news piece is labeled as fake, real, or unverified. We keep the same dataset splits and evaluation criteria as what the organizers provide.

4.1.2 Weibo-16. The dataset *Weibo-16* is firstly proposed in [26] and has been a benchmark dataset of fake news detection in Chinese [17, 38, 47]. Each news piece is labeled as fake or real. It needs to be clarified that in the original dataset, the subset of fake news has many duplications. Concerned about the influence to learning and evaluation by duplications, we perform deduplication on the subset of fake news based on a clustering algorithm based on text similarity. As a result, the amount of clusters is only 59% of the original amount of fake pieces. We suppose that the duplication may increase the risk of data leakage when splitting training and testing sets and make models tend to learn some event-specific features[42] (as they may repeat multiple times in the training process), which limits the generalizability of models. Therefore, we filtered out the highly similar fake news pieces and produce a deduplication version of *Weibo-16* (Table 2). We also clustered real news pieces but found no duplications in *Weibo-16*. As an empirical supplement of our analysis, we conduct comparison experiments between the original and the deduplication version of *Weibo-16*, and verified the necessity of deduplication (see details in Appendix A). In our experiments in the main text, the deduplicated *Weibo-16* is divided into train / val. / test sets in the ratio of 3:1:1.

4.1.3 Weibo-20. As a benchmark Chinese dataset for fake news detection, *Weibo-16* contains fake news pieces ranging from Dec 2010 to April 2014, and is not extended until now. Besides, the scale of *Weibo-16* is smaller after deduplication (Section 4.1.2). Therefore, we constructed the dataset *Weibo-20* on the basis of *Weibo-16*.

We keep the two-class setting (i.e., fake or real for each news pieces). For fake news, we retain the 1,355 fake news pieces of *Weibo-16* and further collect news pieces judged as misinformation officially by Weibo Community Management Center[6] (the same source of fake news of *Weibo-16* [26]) ranging from April 2014 to Nov 2018. And we filter out the highly similar fake news pieces and guarantee there are no duplications. For real news, we retain the 2,351 real news pieces of *Weibo-16* and gather 850 unique real news

---

	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.

pieces in the same period as the fake news. The newly-collected real news pieces are real news verified by NewsVerify[7] which focuses on discovering and verifying suspicious news pieces on Weibo. Totally, *Weibo-20* contains 3,161 fake news pieces and 3,201 real news pieces. As for dataset splits, we split train / val. / test sets in the ratio of 3:1:1.

4.2 Preliminary Analysis of Dual Emotion Signals

To check whether it is statistically dependent or not between dual emotion signals and the veracity of news pieces, we construct two categorical variables to do a chi-squared statistical significance test. The one is *News Veracity*, whose value is *Fake* or *Real*. The other is *Dual Emotion Category*, whose value is combined publisher emotion category and social emotion category, such as *publisher emotion is none and social emotion is angry*. To calculate the value of *Dual Emotion Category*, we use the open-source emotion classification model released by NVIDIA[8] [21] for *RumourEval-19*, and use *Emotion Detection Service* on Baidu AI platform[9] for the two Chinese datasets. In the chi-squared statistical significance test, we firstly assume that the dual emotion signals are independent of the veracity of news pieces (i.e., the null hypothesis). Then we check whether the chi-squared statistic is over the critical value or not. Specifically, on the dataset *RumourEval-19*, the chi-squared statistic is 50.570, over the critical value of 48.602 for the probability of 95%, which means we can reject the null hypothesis. Similarly, on the dataset *Weibo-16*, the chi-squared statistic is 209.14, which is much more than the critical value of 50.892 for the probability of 99%. And on the dataset *Weibo 20*, the chi-squared statistic is 239.963, which is much more than the critical value of 46.963 for the probability of 99%. In conclusion, we can reject the null hypothesis on all three datasets, which indicates that dual emotion signals are statistically dependent on news veracity.

We visualize the variable *Dual Emotion Category* further. On *RumourEval 19*, we select three emotion categories to visualize, *joyful*, *sad* and *none* (over 98% of news pieces covered). And on Chinese datasets, we select four emotion categories, *angry*, *disgusting*, *happy* and *none* (over 97% of news pieces covered). We utilize the heatmap to exhibit the distribution of *Dual Emotion Category* in Figure 3. In the heatmap, each cell represents the percentage of news pieces whose *Dual Emotion Category* is the specific value. And we normalize the percentages for each row (i.e., each publisher emotion). For example, in the top sub-figure of Figure 3a, the upper-left cell indicates that among fake news pieces whose publisher emotion is *joyful*, the percentage of pieces whose social emotion is also *joyful* is 85.5%.

In Figure 3, we can see there are distinct emotion resonances and emotion dissonances in fake news from real news. For example, in Figure 3a, the percentage of dual emotion categories that are both *joyful* in fake news is 8.2% higher than that of real news. And the percentage of emotion dissonance with *sad* publisher emotion and *joyful* social emotion in fake news is 1.9% higher than real news. Evidence is stronger on the two Chinese datasets. Specifically, as for emotion resonances, there are more news pieces whose dual emotion categories are both *angry* and are both *disgusting* in fake news than real news. As for emotion dissonances, there are more news pieces emotion dissonances with are *happy*/*none* publisher emotion but *angry*/*disgusting* social emotion in fake news.

It needs to be recognized that the specific emotion resonances or dissonances may vary from English to Chinese datasets, since the expression styles of people using different languages may be also different. However, our analysis shows that on each dataset itself, no matter what its *dominant language* is, the fake news owns distinct emotion resonances and dissonances from real news, which can be helpful for distinguishing the fake and real news.
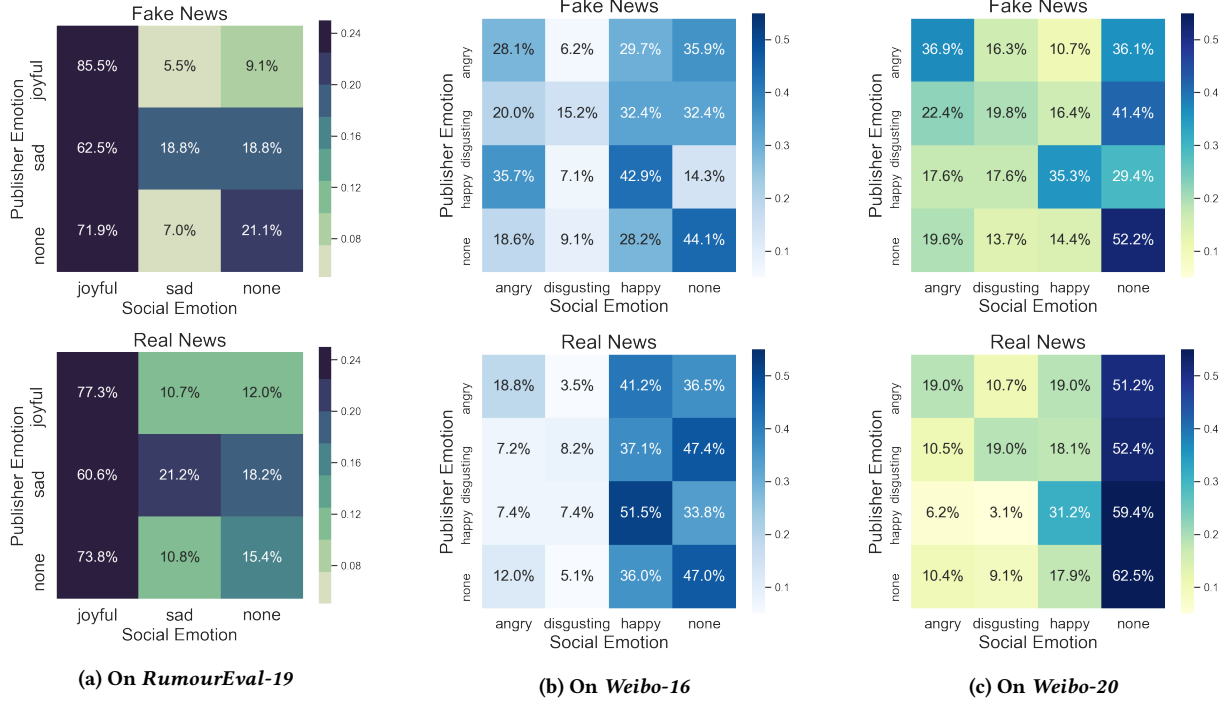
4.3 Experimental Setup

4.3.1 Emotion Resources. For emotion classifiers, as described in Section 4.2, we adopt the pretrained models of NVIDIA for English

---

[7]https://www.newsverify.com/
[8]https://github.com/NVIDIA/sentiment-discovery
[9]https://ai.baidu.com/tech/nlp/emotion_detection

Figure 3: The distribution of *Dual Emotion Category* on the three datasets. In fake news, there are distinct emotion resonances and emotion dissonances from real news.

and Baidu AI for Chinese. To ensure the robustness of the two models, per language we randomly sampled 100 instances and had their emotion categories manually and independently labeled by three annotators, resulting the accuracy of 87% for NVIDIA model and 83% for Baidu model. Therefore, the two classifiers are considered reliable for extracting emotions for fake news detection. As for other emotion resources, for English corpus, we adopt NRC Emotion lexicon[30] and NRC Emotion Intensity lexicon[29] to extract emotion lexicon and emotion intensity features, respectively. And we use the Vader package of NLTK[3] to calculate sentiment scores. For Chinese corpus, we adopt the Affective Lexicon Ontology[46] to extract emotion lexicon and emotion intensity features. And we utilize the dictionary HowNet[10] to calculate sentiment scores. As for auxiliary features in Table 1, for emoticons, we utilize the *List of emoticons* of Wikipedia[43] and divide emoticons into five emotions: *happy*, *angry*, *surprised*, *sad* and *neutral*. For sentimental words and degree words, we use the bilingual sentiment dictionary in HowNet[10]. For negation words, we compile the words list from Wikipedia, Oxford Dictionary, and Cambridge Dictionary.[10]

4.3.2   Fake News Detectors and Baselines. In the experiments, we select two baseline emotion features to evaluate the effectiveness of our *Dual Emotion Features*. These features are implemented with the same emotion dictionaries as *Dual Emotion Features*:

- **Emoratio**: Ajao et al. [1] propose an emotion feature that can be extracted from the content text of news pieces, named

emoratio. It is calculated by the ratio of *count of negative emotional words* and *count of positive emotional words*.

- **EmoCred**: Giachanou et al. [15] utilize the emotional lexicon and intensity features of the content texts. These features are calculated based on the lexicons' occurrence frequency.

For testing the ability of the emotional features to help the text-based fake news detectors (especially those that do not explicitly model the emotional signals), we select BiGRU (as Figure 2 shows), BERT, and other state-of-the-art fake news detectors as follows:

- **BiGRU**: Text-based models like GRU[8] and LSTM[18] are proven effective for fake news detection in [7, 26]. Here we use **BiGRU** to examine whether *Dual Emotion Features* can improve it or not. In practice, as for word embeddings, we use GloVe [31] for English and Chinese Word Vectors for Chinese [25]. The max sequence length of $BiGRU_{\mathcal{T}}$ is 100, and the dimensionality of hidden state of $BiGRU_{\mathcal{T}}$ is 32.
- **BERT** [9]: As a strong text classification model, **BERT** has been adopted to represent semantic signals when detecting fake news in [45]. In the experiments, we truncate the sequences to the maximum length of 512, and finetune the pretrained models[11] for our task.
- **NileTMRG** [13]: For *RumourEval-19* dataset, we use the model implemented by the competition organizers[12] [16], **NileTMRG**. The model is effective and outperforms other

---

[11]The pretrained models are downloaded from https://huggingface.co/models. We use *bert-base-uncased* for English and *bert-base-chinese* for Chinese.
[12]https://github.com/kochkinaelena/RumourEval2019

[10]The negation word lists are released together with our code and datasets.

contestants' models of the leaderboard except for the champion. The model is a linear SVM and uses text features, social features, and use comment stance features. In practice, we keep all the hyperparameters of the original model.

- **HSA-BLSTM** [17]: For the two Chinese datasets, we implement the **HSA-BLSTM**, which is widely used as a baseline on *Weibo-16* dataset. The authors propose a hierarchical attention neural network and utilize not only the contents of news pieces but also the comments. In experiments, we keep all the hyperparameters as those in the original model.

4.3.3 Model Parameters. The dimensionalities of sub features in *Dual Emotion Features*, i.e., d_f, d_e, d_s and d_a, are determined by the language-specific emotion resources. The value of d_f, as the output of pretrained emotion classifiers, is 16 for English and 8 for Chinese. The value of d_e is the size of emotion kinds of the English or Chinese emotion dictionaries, which is 8 or 21, respectively. For d_s, sentiment scores of English texts, produced by the Vader package of NLTK, correspond to four dimensions (positive, negative, neutral and compound), while sentiment scores of Chinese texts are calculated by HowNet, which have one dimension only. The value of d_a is the number of the heuristic features in Table 1, which is 16 for English and 15 for Chinese. The full dimension d is computed as Equation 9, which is 52 for English and 66 for Chinese. The window size is 2, which was determined by grid search that maximizes the performance on the validation set. As for the amount of comments, we set $L_{\mathcal{M}} = 100$, which means that only the earliest 100 comments (or less) of every news piece are considered. In Equation 16, the output dimensionality of MLP is 32.

4.3.4 Evaluation Metrics. On *RumourEval-19*, we adopt the official evaluation metrics, macro F1 score and RMSE (root mean squared error) [16]. Considering the imbalance of the dataset, we also consider the F1 scores of fake, real, and unverified news. On the two Weibo datasets, we use accuracy and macro F1 score as the evaluation metrics, the same as [17]. We also the F1 scores of fake and real news. The other experiments use the macro F1 score.

4.4 Results

4.4.1 Effectiveness of Dual Emotion Features. To answer **EQ1** under the circumstance that the confounding factor of fake news detectors is excluded, we utilize emotion features **alone** to detect fake news. We adopt a simple five-layer MLP and feed only emotion features into it. Table 3 displays the results on the three datasets.

Source	Emotion Features	R-19	W-16	W-20
	Emoratio	0.185	0.553	0.524
Content	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,	**Emotion Gap**	0.332	0.716	0.746
Comments	**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.

In Table 3, among the three emotion features that source from Content, *Publisher Emotion* is more effective than *EmoCred* and *Emoratio*, especially on *RumourEval*. It reveals the effectiveness of *Dual Emotion Features* in modeling emotional signals. What's more, we can see the more improvements of *Social Emotion* and *Emotion Gap*, which are first proposed to help detect fake news in this paper. Specifically, on *RumourEval-19*, using *Emotion Gap* owns 4.2% increase than *Publisher Emotion*. And on the two Chinese datasets, using *Social Emotion* or *Emotion Gap* can both improve the macro F1 score of more than 10%. Moreover, using *Dual Emotion Features* can further obtain enhancements on the three datasets. Especially on *RumourEval-19*, only using *Dual Emotion Features* for fake news detection owns a high macro F1 score of 0.337. And only using *Emotion Gap* is also effective, which is 0.332 for the macro F1 score. It is worth mentioning that such two emotion features even outperform the state-of-the-art model **NileTMRG** (0.309 for macro F1 score, shown in Table 5). That indicates the necessity of dual emotion signals and the importance of mining dual emotion and the relationship between them for fake news detection. Additionally, it needs to be clarified that comparing the three datasets to each other, the performances in *RumourEval-19* are rather worse than the two Chinese datasets. The reasons are discussed in [16, 23], that the amount of news pieces is small and there is a relatively low inter-annotator agreement for the dataset.

In Section 3.1, we adopt five types of emotion features when modeling emotional signals (Emotion Category, Emotion Lexicon, Emotional Intensity, Sentiment Score, and Other Auxiliary Features). To verify the effect of every type of emotion features, we remove one specific type of features from *Dual Emotion Features* every time, to observe the performance changes. As Table 4 shows, the macro F1 scores of *Dual Emotion Features* all decrease regardless of the removed type of emotion features. Thus, it reveals the necessity of using five types of emotion features jointly.

4.4.2 Performance Evaluation within Fake News Detectors. To answer **EQ2**, we exhibit the results of adding *Dual Emotion Features* into the existing fake news detectors on the three datasets.

Table 5 exhibits the results on *RumourEval-19* dataset. Overall, after using *Dual Emotion Features*, the three fake news detectors are both improved a lot. Specifically, on the text-based detectors, **BiGRU** and **BERT**, the use of *Dual Emotion Features* both improves the performance more than *EmoCred* and *Emoratio*. Especially, putting *Dual Emotion Features* into **BERT** owns 0.346 for macro F1 score, far more than the other two emotion features. On

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*.

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*.

the state-of-the-art model **NileTMRG**, using *Emoratio* and *Dual Emotion Features* both improves the macro F1 score further. And the improvement of *Dual Emotion Features* is 3.3%, which is 1.1% higher than *Emoratio*.

The experimental results on the two Weibo datasets are displayed in Table 6. Overall, we can see that our proposed *Dual Emotion Features* outperforms *Emoratio* and *EmoCred* on any models in both datasets. Specifically, on **BiGRU** and **BERT**, the improvements in macro F1 score of *Dual Emotion Features* are at least 1.5% higher on the two datasets. However, when using *Emoratio* or *EmoCred* on **BiGRU**, sometimes the metrics even decrease. It reveals that *Emoratio* and *EmoCred* are more likely to be overfitted, since both of them focus on the contents alone but ignore the comments. And learning dual emotion jointly can avoid this situation to some extent. On the state-of-the-art model **HSA-BLSTM**, after using *Dual Emotion Features* as an enhancement, all the metrics are improved further in both datasets. Especially in *Weibo-16*, the accuracy and macro F1 score both own about 6% improvement, far more than *Emoratio* and *EmoCred*.

4.4.3 Evaluation Under Real-World Scenario Simulation. In the fields of fake news detection, when splitting datasets, most works just **shuffle** the datasets and split them into train / val. / test sets [17, 26, 38, 47], including the datasets splits in Table 2. The kind of data split can somehow prove the effectiveness of proposed methods, but also has a shortcoming: 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. To answer **EQ3**, we simulate a real-world scenario by additionally performing a temporal data split, which means that instances in the train / val. / test sets are arranged in chronological order, to evaluate the ability of models to detect *future* news pieces.

In this section, we adopt the dataset *Weibo-20* and select the most recent 20% news pieces of them as the testing set. Among the remaining 80% news pieces, we next select the most recent 25% of them for validation and let the others be the training set. The results on temporally split *Weibo-20* are displayed in Table 7. Compared with Table 2, we can see that in Table 7 all the performances decrease a lot. It indicates that the temporal data-split strategy creates a more

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.

Models		R-19	W-16	W-20	W-20(t)
BiGRU+	Publisher Emotion	0.310	0.809	0.842	0.681
	Social Emotion	0.322	0.818	0.847	0.693
	Emotion Gap	0.336	0.811	0.849	0.693
	Dual Emotion Features	**0.340**	**0.826**	**0.855**	**0.701**
BERT+	Publisher Emotion	0.312	0.850	0.889	0.705
	Social Emotion	0.339	0.856	0.911	0.730
	Emotion Gap	0.338	0.858	0.906	0.731
	Dual Emotion Features	**0.346**	**0.867**	**0.915**	**0.734**
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	0.779
	Social Emotion	-	0.892	0.922	0.792
	Emotion Gap	-	0.901	0.926	0.800
	Dual Emotion Features	-	**0.908**	**0.932**	**0.805**

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, and W-20(t): temporally split Weibo-20.

challenging scenario, because the topics and writing styles of newly arrived instances are likely to change over time. Such a scenario can somehow expose the drawback of existing techniques and it requires a model of higher generalizability to cope with novel instances.

Under this hard setting, the models with our proposed *Dual Emotion Features* still outperform those with *Emoratio* and *EmoCred*. Sometimes the introduction of *Emoratio* or *EmoCred* even leads to a performance decrease. In contrast, using *Dual Emotion Features* still enhances both models and increases all the metrics, which reveals the effectiveness and generalization ability of *Dual Emotion Features* to some extent.

4.4.4 Ablation Study. To answer **EQ4**, we further conduct ablation experiments on *RumourEval-19*, *Weibo-16*, *Weibo-20* and *Weibo-20 (temporally)* (splitting datasets temporally, described in Section 4.4.3). The results are displayed in Table 8.

In Table 8, we can see that among the four datasets, adding *Dual Emotion Features* into the fake news detectors all obtain the highest macro F1 scores. Besides, compared with the original fake news detectors (Table 5 and Table 6), using any component of *Dual Emotion Features* all enhances the performances of them. During the three components of *Dual Emotion Features*, it exhibits that adopting *Social Emotion* or *Emotion Gap* improves the macro F1 scores more than *Publisher Emotion* on any models on all the datasets. So it concludes that *Social Emotion* and *Emotion Gap* matter more when detecting fake news.

4.5 Case Study

We provide a qualitative analysis of *Dual Emotion Features* in some cases. Take the detector **BiGRU** on *RumourEval-19* as an example, we select three fake news pieces that missed by the original **BiGRU** but detected after using *Dual Emotion Features* as an enhancement (Figure 4). In the figure, there are rich dual emotion signals in every case, such as emotion resonances of *angry* in the left case, of *joyful* in the middle case, and emotion dissonances with *none* publisher emotion and *sad* social emotion in the right case. However, it exhibits using *Emoratio* or *EmoCred* do not help **BiGRU** detect

rightly for the three cases. It reveals that mining dual emotion additionally sometimes is a remedy for the incompetence of only using semantics for detecting fake news.

5 CONCLUSION AND FUTURE WORK

In this paper, we bring a new concept of *dual emotion*, i.e., the publisher emotion and social emotion, into fake news research. We uncover the relationship between dual emotion signals (especially, the emotion gap) and the news veracity. Based on the data observation and analysis, we further propose a feature set, *Dual Emotion Features*, to expose the distinctive emotional signals for detecting fake news. Further, we exhibit that our proposed features can be easily plugged into existing fake news detectors as an enhancement. The extensive experiments conducted on three real-world datasets (including a newly-constructed Chinese dataset) have demonstrated that our proposed feature set outperforms the existing emotional features in fake news detection and essentially improves the performance of existing text-based methods. In future work, we plan to leverage multi-modal information (e.g., emotion in visual contents) to capture the emotions more precisely and use more sophisticated models for dual emotion representation.

Content   **Publisher emotion:** *Angry*   *Content*   **Publisher emotion:** *Joyful*   *Content*   **Publisher emotion:** *None*

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

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

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

Comments   **Social emotion:** *Angry*   *Comments*   **Social emotion:** *Joyful*   *Comments*   **Social emotion:** *Sad*

> **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?

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

> 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	0.33	**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

	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

	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] Oluwaseun Ajao, Deepayan Bhowmik, and Shahrzad Zargari. 2019. Sentiment Aware Fake News Detection on Online Social Networks. In *IEEE ICASSP 2019*. 2507–2511.

[2] BBC. 2020. Bangladesh lynchings: Eight killed by mobs over false child abduction rumours. Retrieved October 19, 2020 from https://www.bbc.com/news/world-asia-49102074

[3] Steven Bird, Ewan Klein, and Edward Loper. 2009. *Natural language processing with Python: analyzing text with the natural language toolkit.* " O'Reilly Media, Inc.".

[4] Leonardo Bursztyn, Aakaash Rao, Christopher P Roth, and David H Yanagizawa-Drott. 2020. *Misinformation During a Pandemic.* Working Paper 27417. National Bureau of Economic Research. https://doi.org/10.3386/w27417

[5] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. 2011. Information credibility on twitter. In *WWW 2011*. 675–684.

[6] Qingqing Chen. 2020. *Coronavirus rumors trigger irrational behaviors among Chinese netizens.* Retrieved October 19, 2020 from https://www.globaltimes.cn/content/1178157.shtml (in Chinese).

[7] Tong Chen, Xue Li, Hongzhi Yin, and Jun Zhang. 2018. Call Attention to Rumors: Deep Attention Based Recurrent Neural Networks for Early Rumor Detection. In *PAKDD 2018*. 40–52.

[8] Kyunghyun Cho, Bart van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. 2014. On the Properties of Neural Machine Translation: Encoder–Decoder Approaches. In *SSST@EMNLP 2014*. 103–111.

[9] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *NAACL-HLT*, Jill Burstein, Christy Doran, and Thamar Solorio (Eds.). Association for Computational Linguistics, 4171–4186. https://doi.org/10.18653/v1/n19-1423

[10] Zhendong Dong and Qiang Dong. 2003. HowNet: a hybrid language and knowledge resource. In *International Conference on Natural Language Processing and Knowledge Engineering, 2003*. IEEE, 820–824.

[11] Paul Eckman. 1972. Universal and Cultural Differences in Facial Expression of Emotion. In *Nebraska Symposium on Motivation*, Vol. 19. 207–284.

[12] Dina ElBoghdady. 2013. Market quavers after fake AP tweet says Obama was hurt in White House explosions. *The Washington Post* (2013).

[13] Omar Enayet and Samhaa R. El-Beltagy. 2017. NileTMRG at SemEval-2017 Task 8: Determining Rumour and Veracity Support for Rumours on Twitter. In *SemEval@ACL 2017*. 470–474.

[14] Marc Fisher, John Woodrow Cox, and Peter Hermann. 2016. Pizzagate: From rumor, to hashtag, to gunfire in DC. *Washington Post* 6 (2016).

[15] Anastasia Giachanou, Paolo Rosso, and Fabio Crestani. 2019. Leveraging Emotional Signals for Credibility Detection. In *ACM SIGIR 2019*. 877–880.

[16] Genevieve Gorrell, Ahmet Aker, Kalina Bontcheva, Leon Derczynski, Elena Kochkina, Maria Liakata, and Arkaitz Zubiaga. 2019. SemEval-2019 Task 7: RumourEval, Determining Rumour Veracity and Support for Rumours. In *SemEval@NAACL-HLT 2019*. 845–854.

[17] Han Guo, Juan Cao, Yazi Zhang, Junbo Guo, and Jintao Li. 2018. Rumor Detection with Hierarchical Social Attention Network. In *CIKM 2018*. 943–951.

[18] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.

[19] Zhiwei Jin, Juan Cao, Yongdong Zhang, and Jiebo Luo. 2016. News Verification by Exploiting Conflicting Social Viewpoints in Microblogs. In *AAAI 2016*. 2972–2978.

[20] Zhiwei Jin, Juan Cao, Yongdong Zhang, Jianshe Zhou, and Qi Tian. 2016. Novel visual and statistical image features for microblogs news verification. *IEEE Trans. Multim.* 19, 3 (2016), 598–608.

[21] Neel Kant, Raul Puri, Nikolai Yakovenko, and Bryan Catanzaro. 2018. Practical Text Classification With Large Pre-Trained Language Models. *arXiv:1812.01207* (2018).

[22] Elena Kochkina, Maria Liakata, and Arkaitz Zubiaga. 2018. All-in-one: Multi-task Learning for Rumour Verification. In *COLING 2018*. 3402–3413.

[23] Quanzhi Li, Qiong Zhang, and Luo Si. 2019. eventAI at SemEval-2019 Task 7: Rumor Detection on Social Media by Exploiting Content, User Credibility and Propagation Information. In *SemEval@NAACL-HLT 2019*. 855–859.

[24] Quanzhi Li, Qiong Zhang, and Luo Si. 2019. Rumor detection by exploiting user credibility information, attention and multi-task learning. In *ACL 2019*. 1173–1179.

[25] Shen Li, Zhe Zhao, Renfen Hu, Wensi Li, Tao Liu, and Xiaoyong Du. 2018. Analogical Reasoning on Chinese Morphological and Semantic Relations. In *ACL(2) 2018*. ACL, 138–143.

[26] Jing Ma, Wei Gao, Prasenjit Mitra, Sejeong Kwon, Bernard J Jansen, Kam-Fai Wong, and Meeyoung Cha. 2016. Detecting rumors from microblogs with recurrent neural networks. In *IJCAI 2016*. 3818–3824.

[27] Jing Ma, Wei Gao, and Kam-Fai Wong. 2017. Detect Rumors in Microblog Posts Using Propagation Structure via Kernel Learning. In *ACL 2017*. 708–717.

[28] Jing Ma, Wei Gao, and Kam-Fai Wong. 2018. Rumor Detection on Twitter with Tree-structured Recursive Neural Networks. In *ACL 2018*. 1980–1989.

[29] Saif Mohammad. 2018. Word Affect Intensities. In *LREC 2018*.

[30] Saif Mohammad and Peter D. Turney. 2013. Crowdsourcing a Word-Emotion Association Lexicon. *Comput. Intell.* 29, 3 (2013), 436–465.

[31] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. Glove: Global Vectors for Word Representation. In *EMNLP 2014*. 1532–1543.

[32] Verónica Pérez-Rosas, Bennett Kleinberg, Alexandra Lefevre, and Rada Mihalcea. 2018. Automatic Detection of Fake News. In *COLING 2018*. 3391–3401.

[33] Francesco Pierri and Stefano Ceri. 2019. False News On Social Media: A Data-Driven Survey. *SIGMOD Rec.* 48, 2 (2019), 18–27.

[34] Vahed Qazvinian, Emily Rosengren, Dragomir Radev, and Qiaozhu Mei. 2011. Rumor has it: Identifying misinformation in microblogs. In *EMNLP 2011*. 1589–1599.

[35] Peng Qi, Juan Cao, Tianyun Yang, Junbo Guo, and Jintao Li. 2019. Exploiting multi-domain visual information for fake news detection. In *ICDM 2019*. IEEE, 518–527.
[36] Peter J Richerson and Robert Boyd. 2008. *Not by genes alone: How culture transformed human evolution.* University of Chicago press.
[37] Ralph L Rosnow. 1991. Inside rumor: A personal journey. *American psychologist* 46, 5 (1991), 484.
[38] Natali Ruchansky, Sungyong Seo, and Yan Liu. 2017. CSI: A Hybrid Deep Model for Fake News Detection. In *CIKM 2017*. 797–806.
[39] Kai Shu, Suhang Wang, and Huan Liu. 2018. Understanding user profiles on social media for fake news detection. In *MIPR 2018*. IEEE, 430–435.
[40] Kai Shu, Suhang Wang, and Huan Liu. 2019. Beyond News Contents: The Role of Social Context for Fake News Detection. In *WSDM 2019*. 312–320.
[41] The Lancet Infectious Diseases. 2020. The COVID-19 infodemic. *The Lancet Infectious Diseases* 20, 8 (2020), 875. https://doi.org/10.1016/S1473-3099(20)30565-X
[42] Yaqing Wang, Fenglong Ma, Zhiwei Jin, Ye Yuan, Guangxu Xun, Kishlay Jha, Lu Su, and Jing Gao. 2018. EANN: Event Adversarial Neural Networks for Multi-Modal Fake News Detection. In *KDD 2018*. 849–857.
[43] Wikipedia. 2020. *List of emoticons.* Retrieved October 19, 2020 from https://en.wikipedia.org/wiki/List_of_emoticons
[44] Wikipedia. 2020. *Misinformation related to the COVID-19 pandemic.* Retrieved October 19, 2020 from https://en.wikipedia.org/wiki/Misinformation_related_to_the_COVID-19_pandemic
[45] Lianwei Wu and Yuan Rao. 2020. Adaptive Interaction Fusion Networks for Fake News Detection. In *ECAI 2020*. 2220–2227.
[46] Linhong Xu, Hongfei Lin, Yu Pan, Hui Ren, and Jianmei Chen. 2008. Constructing the Affective Lexicon Ontology. *Journal of the China Society for Scientific* 27, 2 (2008), 180–185. (in Chinese).
[47] Feng Yu, Qiang Liu, Shu Wu, Liang Wang, and Tieniu Tan. 2017. A Convolutional Approach for Misinformation Identification. In *IJCAI 2017*. 3901–3907.

APPENDIX A. THE REASONS WHY THE DATASET *WEIBO-16* NEEDS TO BE DEDUPLICATED

In Section 4.1.2, we mention that the original version of *Weibo-16* contains many duplications of fake news pieces. Table 9 shows the data statistics. Comparing to Table 2, the number of fake news pieces decrease from 2,312 to 1,355 after deduplication. And there are no duplications in real news pieces.

	Veracity	#pcs	#com
Training	Fake	1,386	789,841
	Real	1,410	482,226
	Unverified	-	-
	Total	2,796	1,272,067
Validation	Fake	463	255,833
	Real	470	146,948
	Unverified	-	-
	Total	933	402,781
Testing	Fake	463	224,795
	Real	471	179,942
	Unverified	-	-
	Total	934	404,737
Total	Fake	2,312	1,270,469
	Real	2,351	809,116
	Unverified	-	-
	Total	4,663	2,079,585

Table 9: Statistics of the original version of *Weibo-16*. #pcs: number of news pieces; #com: number of comments.

To further research the impact of duplications data on the ability of models, we conduct comparison experiments on the original and deduplicated versions of *Weibo-16* respectively. And the results are exhibited in Table 10. Here we choose **BiGRU** and **HSA-BLSTM** as fake news detectors. Considering the class imbalance of the

deduplicated version of the dataset, we train the models based on class weights on the deduplicated training set.

Models	Dataset Version		Macro F1	Acc.
	Train & Val	Test		
BiGRU	original	original	0.793	0.793
	deduplicated	original	**0.806**	**0.807**
		deduplicated	0.807	0.822
HSA-BLSTM	original	original	0.854	0.854
	deduplicated	original	**0.873**	**0.873**
		deduplicated	0.849	0.855

Table 10: Results of the comparison experiments on the original and deduplication versions of *Weibo-16*. Acc. is short for Accuracy.

In Table 10, we can see that if we train and validate the detectors on the deduplicated version of the dataset, the performances of the two detectors will increase on the original testing set (shown in bold in the table). Therefore, it verifies that training on the deduplicated datasets will enhance the generalization ability of the models to some extent. Moreover, if we fix the training and validation set deduplicated and just change the testing set from the original version to the deduplicated version, on **BiGRU** the macro F1 score and accuracy increase, while on **HSA-BLSTM** the metrics both decrease. We suppose the reasons are that on the original testing set, the detectors will predict the duplicated news pieces as highly similar results. So some clusters of duplicated pieces may be all predicted correctly, while others may be all predicted mistakenly, resulting in the unstable performance of the detectors. In a conclusion, deduplicating the dataset can help mitigate this issue.

APPENDIX B. THE METHOD TO CALCULATE THE DUAL EMOTION CATEGORY

It is mentioned in Section 4.2 that we use the pretrained emotion classifiers to calculate the value of *Dual Emotion Category*. The method to calculate the *Dual Emotion Category* are as follows:

For publisher emotion, we feed the text of the news content into the emotion classifier and take the emotion with the maximum probability as the publisher emotion category. For social emotion, we feed the news comments once a time. After getting the output vector of each comment, each dimension of which represents the probability of the given comment having a certain kind of emotion, we average the probability vector of all the comments in each dimension. Finally, we take the emotion with the maximum probability as the social emotion category (i.e., soft voting).

For example, assume that the the output of an emotion classifier is a probability vector on *angry*, *disgusting*, *happy* and *none* and the given news piece has two comments. The content probabilities are $[0.3, 0.1, 0, 0.6]$. So we can use the corresponding emotion of 0.6, *none*, as the publisher emotion category. The probability vector is $[0.8, 0.1, 0, 0.1]$ for the first comment, and $[0.6, 0.3, 0.1, 0]$ for the second comment. So we firstly average all the comment probability values and get $[0.7, 0.2, 0.05, 0.05]$. Then we use the corresponding emotion of 0.7, *angry*, as the news social emotion category. Thus, the categorical variable *Dual Emotion Category* is none for publisher emotion and angry for social emotion.