Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 6, No. 1, 2023

A Joint Fake News Detection Model based on Multi-Features

Shuxia Ren [1, *], Ning He [1] and Xuanzheng Zhang [2]

[1] School of Software Engineering, Tiangong University, Tianjin, 300387, China
[2] School of Computer Science and Technology, Tiangong University, Tianjin, 300387, China
*** Corresponding author:** Shuxia Ren (Email: t_rsx@126.com)

Abstract: Text analysis-based models have achieved outstanding results in fake news detection tasks in recent years, which is closely linked to the quantity and quality enhancement of feature information extracted from the text. Drawing upon the existing semantic detection frameworks, studies in this field concentrate on extracting various textual information through a solitary auxiliary feature, text stance feature or sentiment feature. However, it is challenging to depict the general attributes of the text using a single auxiliary feature, which frequently results in missing essential details and leaves problems with stance distortion and emotional resonance. To tackle the problem, this study proposes a joint model for identifying fake news, incorporating numerous textual characteristics. By extracting and blending various aspects of text features, i.e., semantic, stance and sentiment features, a more detailed and effective joint analysis of textual information is attained, resulting in improved performance in fake news detection. On the RumourEval-17 datasets, our model attains the Macro F1 Score of 0.891, surpassing current models for detecting rumors. Additionally, our model obtains a Macro F1 Score of 0.904 on the latest COVID-19 dataset, demonstrating strong competitiveness and promising prospects for fake news detection.

Keywords: Multi-Features; Joint Analysis; Fake News Detection; Deep Learning.

1. Introduction

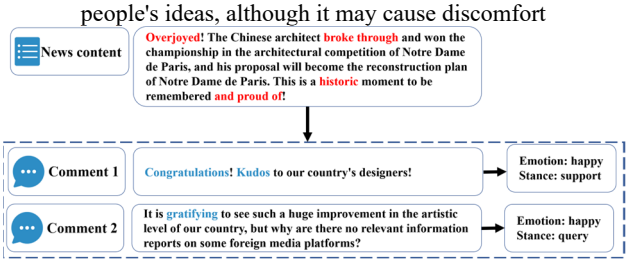
The past few years have witnessed a quick expansion of diverse social media, which has in turn led to the proliferation of false news in various fields owing to inadequate regulation. From daily news, such as the COVID-19 situation [1], to major topics like the 2016 presidential election campaign [2], all are impacted by the negative consequences of online false information. For example, the prevalence of pseudoscientific therapies and conspiracy theories on social media makes it challenging to encourage the COVID-19 vaccine [3]. Therefore, to minimize its influence on the general public, it is crucial to identify and prevent the dissemination of fake news on these platforms, and this work is referred as fake news detection [4].

In this field, existing methods mainly utilize four categories of features to detect fake information: text [2, 5–9], user and publisher profile [2, 10–13], social context [14–18], and images [19–23].

In text-based research, emotional, stance, and semantic information are valuable features for identifying news authenticity. In terms of emotional signals, Conroy et al claimed that negative emotional features could be excavated for fake news detection [24] in 2015. Based on the opinion, the rise of emotional lexicons by 2018, a dictionary containing sentimental polarity words [25, 26], achieved more comprehensive mining of the news context. Zhang et al stepped further in 2021, by introducing the emotions of publishers and commenters, which showed robust performance [27].As for the stance feature, this signal had been initially used in congressional floor debate to infer one's standpoint in 2006 [28]. Then, based on bidirectional RNN, Augenstein et al successfully detected the stance of text with high performance in 2016. From then on, with the transferring from parliaments to the Twitter platform [29] and the evaluation of the stance strength among different texts [30], the application of position detection has shown a more diverse development, Over the past few years, as claimed in 2019 by Genevieve Gorrell et al, stance detection has become the cornerstone of some stance-induced tasks, such as fake news detection [29].

However, in fake news detection, both the stance-semantic based model and the emotionsemantic based model, which are considered state-of-the-art, only utilize two of the aforementioned three features. This leaves room for potential issues. For the stance-semantic model [31], stance distortion issues exist wherein news publishers may attempt to manipulate the public's stance in a way that is similar to real news reporting. This is done in an objective tone to deliberately align stances with those of the publisher, which can evoke negative emotions from people. As a result, stance-semantic-based models may become confused by the falsified stance alignment present in fake news. As demonstrated in Fig. 1(a), fake news creators manipulate individuals' thoughts and perspectives by utilising respected organisations (such as China Daily) and specific locations (like Beijing), leading individuals to side with the false information despite any negative feelings towards it. This poses difficulties for stance-semantic-based models in identifying such instances of fake news. Emotional resonance problems also arise from the emotion-semantic based model [27, 32]. Imitating the writing style of genuine news, fake news employs sentimental language to intentionally elicit public sympathy, a sense of justice, or outrage. This poses a challenge to emotion-semantic based models. As depicted in Fig. 1(b), fake news publishers successfully use emotional words (e.g., Overjoyed) to elicit positive comments from readers, which can confuse the judgment of the models, even though some individuals may be skeptical of the news.

(a)A case of stance distortion problems: false news report distorts people's ideas, although it may cause discomfort



(b)A case of emotional resonance problems: false news post has aroused people's wrong feelings, although people have different views on it

Fig 1. Two pieces of fake news on social media that deliberately misleads people to have wrong feelings or position judgments

To tackle the problems, this paper designs a multi-feature-based joint detection model for fake news, which extracts and fuses the semantic feature, emotional feature, and stance feature in a post and its related comments. For the issue of stance distortion, the inclusion of an emotional feature provides a more accurate representation of the negative emotions that arise from fake news. This helps to eliminate the phenomenon of consistent stance created by rumor mongers. For the issue of emotional resonance, the inclusion of an emotional feature provides a more accurate representation of the views of commenters. This helps to avoid the effect of false emotions that arise from fake news. Eventually, the supplementary information for our model's numerous functionalities accomplishes a more appropriate collaborative analysis of false news.

In this study, our contributions are illustrated as follows:

We introduce a dual emotional information module in our fake news detection model. The module considers the sentiment of both the post and comments, which is highly effective in reducing stance distortion.

We add a stance information module to our fake news detection framework. Analyzing the distribution of stances in news comments, the module successfully avoids any emotional resonance.

We first realize the joint analysis of semantic, stance, and emotional information in fake news detection. On the RumourEval-17 dataset, our model obtains the highest Macro F1 Score of 0.891, respectively, among different rumor detection models. The model was also evaluated on a realtime COVID-19 dataset, achieving an impressive Macro F1 score of 0.904, demonstrating its excellent processing capability.

2. The Proposed Model

2.1. Model Structure

The core idea of the proposed multi-feature model is to extract the semantic and emotional features of the news as well as its related comments and combine them with the stance features of the commentators to form a joint rumor detection with multiple features. Fig. 2 shows the framework of the whole model. Specifically, to represent different features, two main representations are used in this paper: (1) semantic representation, and (2) dual emotion-stance representation. The former is used to represent the semantic features extracted from the news content and the news comment section, while the latter consists of two subcomponents, one is the dual emotional information, i.e., the sentiment information of the publisher and the commenter, and the other is commenters' stance information. In this section, the extraction of these two main features will be introduced in detail, as well as the workflow of the entire model.

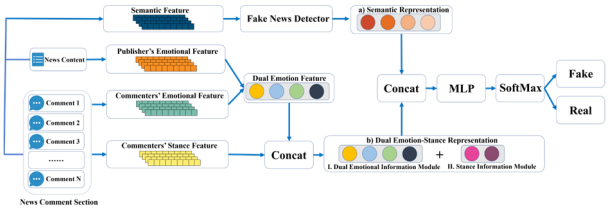


Fig 2. The framework of the proposed model

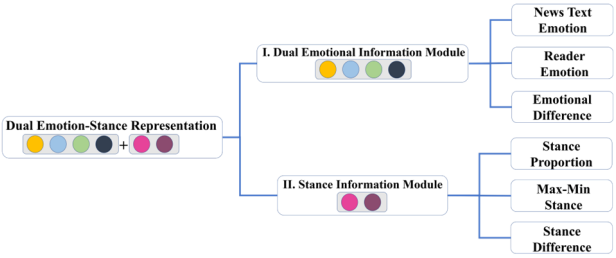


Fig 3. The component of dual emotion-stance representation

2.2. Dual Emotion-Stance Representation

Aiming at dealing with the problems of (1) stance distortion problems and (2) emotional resonance problems, the dual emotion-stance representation is designed. It is a vector contains two main parts of the comment set: dual emotional information module and stance information module. These two parts of the vector will be explained in detail in 3.3-3.5 and their overall structure is shown in Fig. 3.

2.3. Dual Emotional Information

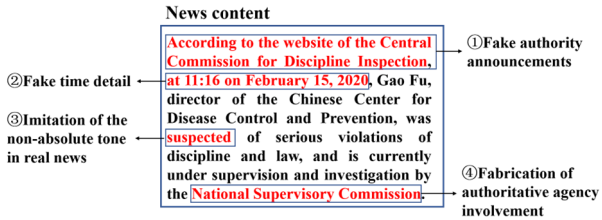


Fig 4. A case of online fake news using several tricks to influence public positions

The dual emotional information module is the representation of the features and differences between news content emotion and comment emotion, which is used to relieve the (1) stance distortion problems. In this problem, fake news publishers rarely use a large number of emotional words to guide people's emotions. Instead, they use a tone that imitates objective news reports and a large number of fake details to confuse people's judgments and positions. This includes making up something the authority didn't say and providing false but precise times, locations, names, or organizations in the news. Such a trick aligns people's stances with fake news, making it difficult for a stance-semantic-based model to verify fake news, since it's impossible to judge whether this alignment of commenter-publisher stances in news is deliberately created. An example of online fake

news that uses the above techniques and distorts people's positions is shown in Fig. 4.

But such false news, in order to distort people's stances, often violates people's original cognition by forging "facts". This practice is easy to cause a series of negative emotions of the public, such as shock, frustration, anger, etc., since people may be difficult to accept this "fact" that goes against the common sense. On the contrary, such negative emotions can hardly be aroused by real news, as the reports of these news are more consistent with the public's perception in most cases. Therefore, the weakness of stance-semantic based model in stance distortion problems could be made up by extracting and detecting the emotional features of both news texts and comments.

Based on this, we introduce the dual emotion information module emo^{dual} to describe news and comment sentiment. As shown in Fig. 3, it is made up of three various components: the news text emotion $\mathrm{emo_T}$, reader emotion $\mathrm{emo_C}$, and their emotional difference emo^{diff}. Among them, news text emotion $\mathrm{emo_T}$ is used to describe the writing sentiment of the news subject, reader emotion $\mathrm{emo_C}$ refers to the overall sentiment of all reader comments, emotional difference $\mathrm{emo^{diff}}$ is used to indicate the emotional difference between readers and news publishers and it is an indicator used to indicate whether a negative emotion is generated, because readers would likely have negative emotions after reading fake news, and their emotions may gradually lose resonance with the emotions of the news content. The extraction methods of the three various components in dual emotion information $\mathrm{emo^{dual}}$ will be introduced in this section.

2.3.1. News Text Emotion.

News text emotion $\mathrm{emo_T}$ refers to the emotion that the author reveals or wants to express when compiling the text. In this paper such emotion signal is composed of three different coarse-grained features: word-level features (lexical sentiment score $\mathrm{emo_T^{lex}}$, sentiment magnitude $\mathrm{emo_T^{int}}$), content-level features (general sentiment $\mathrm{sem_T^{cate}}$, general semantic value $\mathrm{emo_T^{senti}}$), and auxiliary features (latent sentiment information $\mathrm{emo_T^{latent}}$).

Word-level features include lexical sentiment score $\mathrm{emo_T^{lex}}$ and sentiment magnitude $\mathrm{emo_T^{int}}$. The former is used to represent the type of sentiment words in the text, and the latter reflects the intensity of related sentiment words. To capture such two features, the emotion lexicon is introduced. It is a dictionary containing different sentiment words, where the sentiment categories and intensities of these sentiment words are manually labeled by experts. By tagging emotional words present in news content, we can extract the word-level feature of news content. Here, we assume that the sentiment dictionary we use contains $\mathrm{d_e}$ different types of sentiments, and the full set of sentiments can be expressed as $\mathrm{E} = \{\mathrm{e_1}, \mathrm{e_2}, \dots, \mathrm{e_{d_e}}\}$. In addition, a specific emotion $e \in E$ corresponds to its emotion thesaurus $\mathcal{E}_e = \{w_{e,1}, w_{e,2}, \dots, w_{e,L_e}\}$, where $\mathrm{w_{e,i}}$ is the i^{th} emotional word in thesaurus $\mathcal{E}_e$ and $\mathrm{L_e}$ refers to the number of vocabularies that the thesaurus holds.

The lexical sentiment score $\mathrm{emo_T^{lex}}$ of a piece of text T can be given by splicing the scores of different sentiments, and one sentiment score can be given by summing the scores of each word $\mathrm{s(t_j, e)}$ in the corresponding emotional thesaurus. Here, $\mathrm{t_j}$ is the j^{th} word of text T. To give a more accurate word score, the context of the j^{th} word must take into account several factors, including degree words, negation words, and the frequency of word occurrences. Thus, given a window size w (suppose that only the left context is considered when scoring, that is, the w words to the left of the j^{th} word), the score of each score $\mathrm{s(t_j, e)}$ can be calculated from formula 1:

$$s(t_j, e) = \frac{\mathcal{M}_{\mathcal{E}_e}(t_j) * neg(t_j, w) * deg(t_j, w)}{L} \quad (1)$$

In formula 1, $\mathcal{M}_{\mathcal{E}_e}(t_j)$ is a matching value (formula 2), which represents whether the j^{th} word is in thesaurus $\mathcal{E}_e$.While $\mathrm{neg(t_j, w)}$ and $\mathrm{deg(t_j, w)}$ represent the negation amount (formula 3) and the degree amount (formula 4) in the window w, respectively.

$$\mathcal{M}_{\mathcal{E}_e}(t_j) = \begin{cases} 1, & if \ \ t_j \in \mathcal{E}_e \\ 0, & otherwise \end{cases} \quad (2)$$

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

$$\mathrm{deg(t_j, w)} = \prod_{k=j-w}^{j-1} \mathrm{deg}(t_k) \quad (4)$$

Therefore, for the whole text $\mathcal{T}$, the score of a certain emotion $s(\mathcal{T}, e)$ can be obtained by formula 5:

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

Finally, all sentiment scores are concatenated to obtain the lexical sentiment score $emo_{\mathcal{T}}^{lex}$, as shown in equation 6, where $\oplus$ refers to the concatenation operator and $emo_{\mathcal{T}}^{lex} \in \mathbb{R}^{d_e}$.

$$emo_{\mathcal{T}}^{lex} = s(\mathcal{T}, e_{d_1}) \oplus s(\mathcal{T}, e_{d_2}) \oplus \dots \oplus s(\mathcal{T}, e_{d_e}) \quad (6)$$

As for sentiment magnitude $emo_{\mathcal{T}}^{int}$, it is constructed by splicing magnitude score vectors of various emotions in news content. For one certain sentiment e of text $\mathcal{T}$, we use the same method as lexical sentiment score $emo_{\mathcal{T}}^{lex}$ to extract sentiment scores of words $s(t_j, e)$, but assign strength weights to each score according to the sentiment thesaurus, and then we obtain emotional intensity score $s'(\mathcal{T}, e)$ as shown in the following formula 7:

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

In formula 7, $\mathrm{int(t_j)}$ represents the emotional intensity of the t_j word in the emotional thesaurus $\mathcal{E}_e$. If the word is not included in the thesaurus $\mathcal{E}_e$, it is recorded as 0. And sentiment magnitude $emo_{\mathcal{T}}^{int}$ can be obtained through concatenating all the emotional intensity scores from formula 8:

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

Content-level features are made up of general semantic $sem_{\mathcal{T}}^{cate}$ and general sentiment value $emo_{\mathcal{T}}^{senti}$. The general semantic $sem_{\mathcal{T}}^{cate}$ describes the semantic expressed in general throughout the passage, while general sentiment value $emo_{\mathcal{T}}^{senti}$ represents the intensity of the overall sentiment of the whole passage.

As for general semantic $sem_{\mathcal{T}}^{cate}$ reflects the overall semantic of the news content $\mathcal{T}$, based on the semantic detector f_e proposed by NVIDIA [31], it can be obtained by formula 9:

$$sem_{\mathcal{T}}^{cate} = f_e(\mathcal{T}) \quad (9)$$

In formula 9, $sem_{\mathcal{T}}^{cate} \in \mathbb{R}^{d_{f_e}}$, where d_{f_e} refers to the dimension of the semantic detector f_e output.

Likewise, in order to measure the general sentiment value $emo_{\mathcal{T}}^{senti}$ of the content, such as strong, flat, etc., an emotion lexicon-based detection detector f_{lex} is used in this paper to extract the corresponding sentiment of text $\mathcal{T}$, as shown in

formula 10. For general sentiment value $emo_{\mathcal{T}}^{senti}$ with d_s dimension, the final sentiment value $emo_{\mathcal{T}}^{senti} \in \mathbb{R}^{d_s}$.

$$emo_{\mathcal{T}}^{senti} = f_{lex}(\mathcal{T}) \quad (10)$$

In addition to words and sentences, there is also latent sentiment information $emo_{\mathcal{T}}^{latent}$ in the text, including non-verbal signals and underlying habits. Non-verbal signals include capitalization, punctuation, and emoticons, while underlying habits include the usage frequency of personal pronouns and some emotional words. Both types of words have the function of conveying and expressing emotion. For example, the use of non-verbal all-caps "GET OUT OF MY HOUSE!" conveys stronger emotions than lowercase; and in underlying habits, use emoticon ":)" to express happiness, ":(" to express sadness. Assuming that the dimension of the latent feature information is d_a, then we can get the latent sentiment information $emo_{\mathcal{T}}^{latent} \in \mathbb{R}^{d_a}$.

Finally, the news text emotion of text $\mathcal{T}$, denoted as $emo_{\mathcal{T}}$, can be gained through concatenating all the subordinate ingredients of word-level features, content-level features and auxiliary features, as shown in formula 11:

$$emo_{\mathcal{T}} = sem_{\mathcal{T}}^{cate} \oplus emo_{\mathcal{T}}^{lex} \oplus emo_{\mathcal{T}}^{int} \oplus emo_{\mathcal{T}}^{senti} \oplus emo_{\mathcal{T}}^{latent} \quad (11)$$

where $emo_{\mathcal{T}} \in \mathbb{R}^d$, d refers to the dimension of $emo_{\mathcal{T}}$.

2.3.2. Reader Emotion.

The reader emotion $emo_{\mathcal{C}}$ is divided into two parts: reader specific sentiment and reader general sentiment. For the first part, the sentiment score for each comment can be calculated in the same way as for news text emotion $emo_{\mathcal{T}}$ in section 2.3.1, denoted as emo_r, and calculate the score of the entire comment set $\widehat{emo_{\mathcal{C}}}$ through Equation 12:

$$\widehat{emo_{\mathcal{C}}} = emo_{r_1}^{T} \oplus emo_{r_2}^{T} \oplus \ldots \oplus emo_{r_n}^{T} \quad (12)$$

where $\widehat{emo_{\mathcal{C}}} \in \mathbb{R}^{r_n \times \mathrm{d}}$.

The second part, reader general sentiment has two subordinate components: the average sentiment of the review set $\mathcal{C}$ and the extreme sentiment of the review set $\mathcal{C}$. The average sentiment can be obtained by the mean pooling of Equation 13, and the extreme sentiments can be obtained by the max pooling of Equation 14.

$$emo_{\mathcal{C}}^{mean} = mean(\widehat{emo_{\mathcal{C}}}) \quad (13)$$

$$emo_{\mathcal{C}}^{max} = max(\widehat{emo_{\mathcal{C}}}) \quad (14)$$

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

Finally, the reader emotion $emo_{\mathcal{C}}$ is derived from formula 15:

$$emo_{\mathcal{C}} = emo_{\mathcal{C}}^{mean} \oplus emo_{\mathcal{C}}^{max} \quad (15)$$

Where $emo_{\mathcal{C}} \in \mathbb{R}^{2\mathrm{d}}$.

2.3.3. Emotional Difference.

We design emo^{diff} to highlight the emotional differences between authors and reviewers to show whether reviews, highlighting public responses to news. As shown in formula 16, it can be obtained by the difference between news text emotion $emo_{\mathcal{T}}$ and two different parts, average sentiment $emo_{\mathcal{C}}^{mean}$ and the extreme sentiment $emo_{\mathcal{C}}^{max}$ in the review set $\mathcal{C}$:

$$emo^{diff} = (emo_{\mathcal{T}} - emo_{\mathcal{C}}^{mean}) \oplus (emo_{\mathcal{T}} - emo_{\mathcal{C}}^{max}) \quad (16)$$

Where $emo^{diff} \in \mathbb{R}^{2\mathrm{d}}$.

2.3.4. Dual Emotion Features Module.

After performing the steps 3.3.1-3.3.3, dual emotional information module emo^{dual} can be connected by its three sub-parts: the news text emotion $emo_{\mathcal{T}}$, reader emotion $emo_{\mathcal{C}}$, and their emotional difference emo^{diff}. It is generated as shown in formula 17:

$$emo^{dual} = emo_{\mathcal{T}} \oplus emo_{\mathcal{C}} \oplus emo^{diff} \quad (17)$$

2.4. Stance Information.

The stance information module is introduced to alleviate the (2) emotional resonance problem. In this problem, due to people's strong sense of group belonging, they tend to receive group cues and group emotional infection. Therefore, they are easily guided by specific textual emotions, and immersed in the whirlpool of emotions. In some cases, fake news publishers have exploited this feature by using specific, carefully crafted emotional sentences or words in fake news to evoke emotional resonance in people, even though people may have different positions on the news. This emotional resonance brings detection difficulties to emotion-semantic based models because it is difficult to confirm whether the consistency between news texts and readers' emotions is deliberately created. An example of deliberately using specific emotional words to guide public sentiment is shown in Fig. 5.

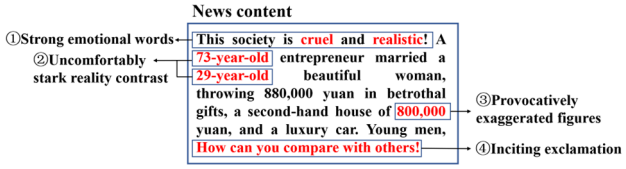


Fig 5. A case of online fake news trying to evoke emotional resonance

But for this type of fake news, it is difficult to convince the public since the author uses a high proportion of emotional words and lacks factual descriptions. Therefore, even if the emotional resonance of the public is evoked, people's stance on the news still won't change much. Considering that fake news lacks the support of facts, it is more likely to be opposed by the public than real news. Therefore, fake news can be detected by extracting stance information $I_{\mathcal{C}}$ in the comment area to avoid the negative impact of emotional resonance.

In this paper, the stance information module $I_{\mathcal{C}}$ is defined as the stance characteristic of a certain news in comment set $\mathcal{C}$. In order to comprehensively describe the distribution of different stance in the comment area, three features are taken into account: stance proportion SP, max-min type MMP, and stance difference SD. The stance proportion $SP = \{fs_1, fs_2, \ldots, fs_i\}$, where $\mathrm{fs_i}$ is the percentage of $i-th$ type of stances in comment set $\mathcal{C}$, describes the proportion of different stances in the entire comment area; the max-min stance $MMP = \{maxs, mins\}$, describes the most common stance types $maxs$ and the less common stance types $mins$ in the comment area; the stance difference SD describes the distribution difference of the stance categories, which is represented by the variance and standard deviation of the proportions of all stance classes, namely vs and sds, respectively. The detailed structure of this feature is shown in Table 1:

In addition, to obtain the stance of each comment, Bert stance classifier f_B is introduced, and Fig. 6 shows its structure in detail. Classifier receives comment sentences as its input, inside [CLS] and [SEP] tags. The result of stance classification obtains from [CLS], the final token includes a semantic feature vector representing the entire comment. Therefore, in a stance detection task with r types of stances, for one comment c_m in comment set, the stance category can obtain as formula 18.

$$s = f_B(c_m) \quad (18)$$

Table 1. The structure of stance features

Type	Features
Stance Proportion	Frequency of stance type 1 (fs_1)
	Frequency of stance type 2(fs_2)
	……
	Frequency of stance type $i(fs_i)$
Max-Min Stance	The most common stance types ($maxs$)
	The less common stance types ($mins$)
Stance Difference	Variance of all stance class proportions (vs)
	Standard deviation of all stance class proportions (sds)

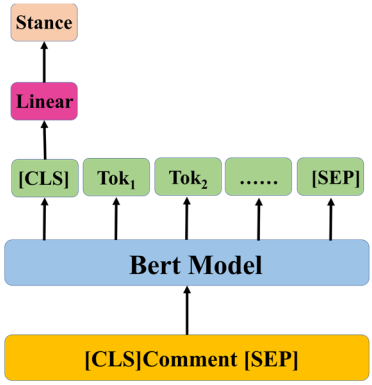


Fig 6. The structure of the Bert stance classifier

In the actual operation of the stance information module, the Bert classifier is firstly used to extract the stances of a specific comment set $\mathcal{C} = \{r_1, r_2, \ldots, r_n\}$ one by one to obtain the comment stance set $\mathcal{C}_s = \{s_1, s_2, \ldots, s_n\}$, as shown in Equation 19.

$$\mathcal{C}_s = f_B(\mathcal{C}) \tag{19}$$

Then, according to the comment stance set $\mathcal{C}_s$, a series of parameters related to the stance's distribution can be calculated, including the frequency of all stance type $fs_1, fs_2, \ldots, fs_i$ (formula 20), the most common stance types $maxs$ and the less common stance types $mins$(formula 21), the variance vs and standard deviation sds(formula 22).

$$\begin{cases} fs_1 = \frac{num_1}{n} \\ fs_2 = \frac{num_2}{n} \\ \ldots\ldots \\ fs_i = \frac{num_i}{n} \end{cases} \tag{20}$$

$$\begin{cases} maxs = maxtype(fs_1, fs_2, \ldots, fs_i) \\ mins = mintype(fs_1, fs_2, \ldots, fs_i) \end{cases} \tag{21}$$

$$\begin{cases} vs = var(fs_1, fs_2, \ldots, fs_i) \\ sds = dev(fs_1, fs_2, \ldots, fs_i) \end{cases} \tag{22}$$

In equation 20, num_i represents the number of the $i-th$ stance in the comment area. As for equation 21, $maxtype$ and $mintype$ are the maximum and minimum index functions. These two functions are used to extract the most frequent and least frequent stance categories in the comment area, respectively. In equation 22, var and dev are the variance and standard deviation functions.

After the process mentioned above, three sub-parts of stance information $I_{\mathcal{C}}$, namely stance proportion SP, max-min type MMP, and stance difference SD, can be calculated by formulas 23-25:

$$SP = fs_1 \oplus fs_2 \oplus \ldots \oplus fs_n \tag{23}$$

$$MMP = maxs \oplus mins \tag{24}$$

$$SD = vs \oplus sds \tag{25}$$

The stance information module $I_{\mathcal{C}}$ can be obtained through the splicing of its three sub-parts, as shown in Equation 26.

$$I_{\mathcal{C}} = SP \oplus MMP \oplus SD \tag{26}$$

2.5. Dual Emotion-Stance Representation

Dual emotion-stance representation $R_{stance,dual}$ is concatenated by dual emotional information module emo^{dual} and stance information module $I_{\mathcal{C}}$, as shown in formula 27:

$$R_{stance,dual} = emo^{dual} \oplus I_{\mathcal{C}} \tag{27}$$

The model involving this representation $R_{stance,dual}$ performs better than models based on sentiment or stance, both in detecting general fake news and well-crafted fake news: (1) When representing general fake news and comment text information, it can provide two-dimensional descriptions of stance and emotion, bringing finer and higher-grained input information to subsequent fake news analysis (2) When dealing with fake news that tries to arouse emotional resonance and when dealing with fake news that distorts public stances, the performance can be improved by analyzing the anomalies of public positions; Besides, when facing the fake news that distorts public stances, anomalies in news text and comments' sentiments can be captured to reduce errors in judgment only from stances. The Fig. 7 shows how different features synergize to mitigate the problem of stance distortion and emotional resonance.

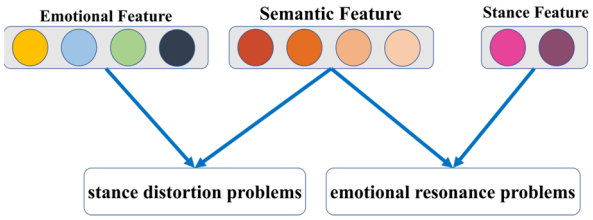


Fig 7. Schematic of joint collaboration of three features

In order to further extract the key information in the news text and the comment section, we connect our representation $R_{stance,dual}$ with existing semantic detection frameworks, as shown in Figure 2. This paper uses the BiGRU detector, whose corresponding output feature vector is $BiGRU_{\mathcal{T},\mathrm{C}}$. Thus, we can obtain splicing vector $[R_{stance,dual}, BiGRU_{\mathcal{T},\mathrm{C}}]$ that responds to the characteristics of each dimension of a piece of news. Finally, as shown in formula 28, this splicing vector $[R_{stance,dual}, BiGRU_{\mathcal{T},\mathrm{C}}]$, as the input, passes through the multi-layer perceptron (MLP) and a softmax layer to obtain the news authenticity prediction $\hat{y}$:

$$\hat{y} = Softmax(MLP([R_{stance,dual}, BiGRU_{\mathcal{T},\mathrm{C}}])) \tag{28}$$

In the proposed model, since the inputs contain feature vectors spliced with different dimensions, they jointly have an impact on the final prediction result $\hat{y}$. So, the model in this paper has a fuller understanding of the information of the text.

3. Experiments

In this section, we conduct different experiments to evaluate the plausibility and performance of our model. Specifically, we focus on the following four evaluation questions:

- EQ1: Do the features(emotional, stance, semantic features) used by the model and joint feature operations improve the results of rumor detection?
- EQ2: How well does our model perform in the real world?

3.1. Dataset Description.

3.1.1. RumourEval-17.

The RumourEval-17 dataset is constructed on the release of SemEval-2017 Task 8 [32] that is involved with a source dataset containing rumors on 9 distinct breaking news. The raw dataset comprises Twitter conversations that are initiated by a false tweet and 297 conversational threads. The conversations contain commenters' tweets replying to those false tweets. The tweets in comment area are annotated with the label of support, deny, query, or comment (SDQC). There are two subtasks in the SemEval-2017 Task 8 which are stance classification and veracity prediction. The subtask A of stance classification is set to label the stance of comments, while the purpose of the subtask B-veracity prediction is to determine the veracity of a given rumor. We make the division of our dataset and evaluation metrics the same as what the raw dataset shows.

3.1.2. COVID-19.

The data used in the dataset of COVID-19 are provided by a CodaLab competition1 with the title of " COVID19 Fake News Detection in English". The news posts are all related to coronavirus and carefully selected from the Twitter. The COVID-19 dataset contains three JSON files, which are named train, test and val, respectively. We split the quantity of post in the three files by the ratio of 4:1:1. In training set, there are 100 records of post including two types of data-Real and Fake. In testing set, there are 25 records of post also annotated with the label of Real and Fake. In validation set, there are merely 25 records of post with the label of Real and Fake.

3.2. Dataset Preprocessing.

The publicly available tool of robust sentiment discovery published by NVIDIA [31] is used to calculate the value of emotion type for RumourEval-17 and COVID-19.

3.3. Experimental Setup.

3.3.1. Sentiment Material

So as to verify the effectiveness of the English model, we stochastically selected 100 samples and annotated their emotion types manually and separately, which achieves the precision of 86% for NVIDIA model. Therefore, The NVIDIA model could be used to extract the latent emotional information for fake news detection. For the remaining emotion signals, the NRC Emotion dictionary [33] and NRC Emotion Intensity dictionary [34] are utilized to extract emotional words and emotional intensity features, separately. Besides, sentiment scores are obtained through the Vader library of NLTK [35]. With respect to supplementary features in table 1, in emoticons, the statistical list of emoticons from Wikipedia [36] is applied to classify emoticons into five categories: pleasure, rage, wonder, sadness and neutrality. In emotional words and degree words, the bilingual emotional lexicon in HowNet [37] is used to calculate the frequency of them. In negative words, the words list from Wikipedia, Oxford lexicon, and Cambridge lexicon are collected.

3.3.2. Parameter Scale.

The dimensions of sub features contained in dual sentiment features, like d_{fe}, d_e, d_s and d_a, are relevant to English sentiment material. The parameter d_{fe} taken as the dimension of the emotional detector output is assigned with 16. The parameter d_e with the value of 8 represents the scale of various sentiments. The parameter d_s as the dimension of emotional scores of English texts, generated via the Vader package of NLTK, has four dimensions that are positive, negative, neutral and compound, respectively. For d_a, the dimensionalities of auxiliary features in Table 1, has the value of 16. The value of parameter d is 52. For every tweet post, at most 100 comments in the outset are selected. In the Target Aware BERT for stance detection [45], the parameter seed is specified as 4214 for the current run. The cross_ validation_ num, as the facilitation of input for cross validation in RumourEval-17 and COVID-19 datasets, is 4. The dataset_name, restricting the type of datasets for the experiments involved, has the value of RumourEval-17 and COVID-19. The batch_size, used to determine the direction of descending, is set as 16. The lr, deciding whether target function converge to the local maximum and when to achieve it, is 1e-4.

3.4. Result and Analysis.

Table 2. Performance comparison on RumourEval-17

Models	Macro F1 Score	RMSE	F1 Score		
			Fake News	Real News	Unverified News
MLP	-	-	-	-	-
+ Stance Feature	0.166	0.812	0	0.500	0
+ Dual Emotion Features	0.161	0.807	0	0.484	0
+ Dual Emotion-Stance Features	0.351	0.783	0.15	0.530	0.363
BiGRU	0.321	0.781	0	0.625	0.307
+ Stance Feature	0.384	0.771	0.447	0.444	0.260
+ Dual Emotion Features	0.429	0.732	0.142	0.57	0.476
+ Dual Emotion-Stance Features	0.891	0.564	0.892	0.886	0.894

We conducted several sets of experiments to answer several of the questions raised in this section. For EQ 1, we performed ablation experiments on RumourEval-17 to observe the contribution of different features to the detector. The experimental results are shown in Table 2, where the data items in bold font represent the leading data. It is worth noting that in all experimental results, our semantic extractor used only news text information by default, and the tag "*" is used to mark the experimental group that used semantic information of news text and comments at the same time. From a macro point of view, the models with feature fusion are ahead in the two experimental groups. Specifically, the macro F1 scores of the experimental group with only the stance feature and that with only the dual emotion features are close to each other; At the same time, it is difficult to identify Fake News and Unverified News with a single feature, especially in the MLP experimental group that does not use any semantic feature extractor. Both the fake news and unverified news experimental groups obtain F1 scores of 0. The experimental group that achieved the fusion of two features obtained better results: F1 macro scores were higher than those of the rumor detection model based on single feature, and the model with the fusion of three sorts of features achieved the best results among all experimental groups. At the same time, the fusion of dual emotional-stance

features alleviates the problem of 0 macro F1 score in false news and unverified news, and scores in both categories. This indicates that the dual emotion-stance fusion feature has a certain effect on the performance of a single feature.

For EQ 2, we tested the performance of our model on the COVID-19 fake news dataset to simulate the rumor detection capabilities of a specific topic in real situations. The original dataset contains only fake and real labels, which is inconsistent with the output of model authenticity label. Therefore, we modify the range of authenticity label to {F, R}(Fake news, Real news) to accommodate the format of the COVID-19 dataset. As shown in Table 3, in the test set, we obtained an Macro F1 Score of 0.904, among which, the scores in Fake News and Real classes were relatively close and balanced, and our model could obtain better detection results for different types of posts in real conditions.

Table 3. The performance of fake news detection task on COVID-19 English dataset

Models	Macro F1 Score	RMSE	F1 Score	
			Fake News	Real News
MLP+Stance Feature	0.683	0.842	0.687	0.685
MLP+DualEmotionFeatures	0.756	0.697	0.759	0.752
MLP+DualEmoStaFeatures	0.891	0.591	0.894	0.887
BiGRU+Stance Feature	0.691	0.829	0.696	0.692
BiGRU+DualEmotionFeaturs	0.763	0.682	0.765	0.761
BiGRU+DualEmoStaFeatures	0.904	0.579	0.907	0.903

4. Conclusion

In this study, a new multi-feature rumor detection model is proposed to achieve better performance in the fake news detection task by integrating semantic, emotional and stance features. Then, the model is compared with other commonly used fake news detection frameworks on RumorEval-17 and COVID-19 datasets and achieves the best results on Marco F1 and RMSE. At the same time, the proposed model has similar model size and better performance as the Dual Emotion based model, which shows excellent potential in terms of low running cost and offline client rumor detection. Finally, we also simulated the actual operation of the proposed model on the COVID-19 dataset and obtained a Marco F1 value of 0.904. The next step of this research is to explore the characteristics of different dimensions of fake news detection (such as images, videos, voice, etc.), and combine multi-dimensional information in news to achieve a more comprehensive understanding of fake news and better detection ability.

Acknowledgments

This work was financially supported by Natural Science Foundation of Tianjin Municipality(19JCYBJC18700).

References

[1] Kar D, Bhardwaj M, Samanta S, et al. No rumours please! a multi-indiclingual approach for covid fake-tweet detection. In: 2021 Grace Hopper Celebration India (GHCI). IEEE, p. 1–5.

[2] Zhang X, Ghorbani, AA. An overview of online fake news: Characterization, detection, and discussion. Information Processing & Management 57(2):102025.

[3] Naeem SB, Bhatti R, Khan A (2021) An exploration of how fake news is taking over social media and putting public health at risk. Health Information & Libraries Journal 38(2):143–149.

[4] Zhang Q, Zhang S, Dong J, et al. Automatic detection of rumor on social network. 2021 Grace Hopper Celebration India. Springer, p. 113–122.

[5] Yang S, Shu K, Wang S, et al. Unsupervised fake news detection on social media: A generative approach. Proceedings of the AAAI conference on artificial intelligence, p. 5644–5651.

[6] Girgis S, Amer E, Gadallah M. Deep learning algorithms for detecting fake news in online text. 2018 13th international conference on computer engineering and systems (ICCES). IEEE, p. 93–97.

[7] Ahmed H, Traore I, Saad S. Detecting opinion spams and fake news using text classification. Security and Privacy 1(1):e9.

[8] Bharadwaj P, Shao Z. Fake news detection with semantic features and text mining. International Journal on Natural Language Computing (IJNLC). Vol. 8.

[9] Reddy H, Raj N, Gala M, et al. Text-mining-based fake news detection using ensemble methods. International Journal of Automation and Computing 17(2),p. 210–221.

[10] Shu K, Zhou X, Wang S, et al. The role of user profiles for fake news detection. Proceedings of the 2019 IEEE/ACM international conference on advances in social networks analysis and mining, p. 436–439.

[11] Shu K, Wang S, Liu H. Understanding user profiles on social media for fake news detection. 2018 IEEE conference on multimedia information processing and retrieval (MIPR). IEEE, p. 430–435.

[12] Hamdi T, Slimi H, Bounhas I, et al. A hybrid approach for fake news detection in twitter based on user features and graph embedding. Distributed Computing and Internet Technology: 16th International Conference, ICDCIT 2020, Bhubaneswar, India, January 9–12, 2020, Proceedings 16. Springer, p. 266–280.

[13] Chowdhury R, Srinivasan S, Getoor L. Joint estimation of user and publisher credibility for fake news detection. Proceedings of the 29th ACM International Conference on Information & Knowledge Management, p. 1993–1996.

[14] Shu K, Wang S, Liu H. Beyond news contents: The role of social context for fake news detection. Proceedings of the twelfth ACM international conference on web search and data mining, p. 312–320.

[15] Shu K, Mahudeswaran D, Wang S, et al. Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. Big data 8(3), p. 171–188.

[16] Nguyen VH, Sugiyama K, Nakov P, et al. Fang: Leveraging social context for fake news detection using graph representation. Proceedings of the 29th ACM international conference on information & knowledge management, p. 1165–1174.

[17] Raza S, Ding C. Fake news detection based on news content and social contexts: a transformer-based approach. International Journal of Data Science and Analytics, p. 335–362.

[18] Della Vedova ML, Tacchini E, Moret S, et al. Automatic online fake news detection combining content and social signals. 2018 22nd conference of open innovations association (FRUCT), p. 272–279.

[19] Huh M, Liu A, Owens A, et al. Fighting fake news: Image splice detection via learned self-consistency. Proceedings of the European conference on computer vision (ECCV), p. 101–117.

[20] Giachanou A, Zhang G, Rosso P. Multimodal multi-image fake news detection. 2020 IEEE 7th international conference on data science and advanced analytics (DSAA). IEEE, p. 647–654.

[21] Singh B, Sharma DK. Predicting image credibility in fake news over social media using multi-modal approach. Neural Computing and Applications 34(24), p. 21503– 21517.

[22] Mangal D, Sharma DK. Fake news detection with integration of embedded text cues and image features. 2020 8th international conference on reliability, infocom technologies and optimization (trends and future directions)(ICRITO). IEEE, p. 68–72.

[23] Singhal S, Shah RR, Chakraborty T, et al. Spotfake: A multi-modal framework for fake news detection. 2019 IEEE fifth international conference on multimedia big data (BigMM). IEEE, p. 39–47.

[24] Conroy NK, Rubin VL, Chen Y. Automatic deception detection: Methods for finding fake news. Proceedings of the association for information science and technology 52(1), p. 1–4.

[25] Klyuev V. Fake news filtering: Semantic approaches. 2018 7th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO). IEEE, p. 9–15.

[26] Guo C, Cao J, Zhang X, et al. Exploiting emotions for fake news detection on social media. arXiv preprint arXiv: 190301728.

[27] Zhang X, Cao J, Li X, et al. Mining dual emotion for fake news detection. Proceedings of the web conference 2021, p. 3465–3476.

[28] Thomas M, Pang B, Lee L. Get out the vote: Determining support or opposition from congressional floor-debate transcripts. arXiv preprint cs/0607062.

[29] Thomas M, Pang B, Lee L. Get out the vote: Determining support or opposition from congressional floor-debate transcripts. arXiv preprint cs/0607062.

[30] Gorrell G, Kochkina E, Liakata M, et al. Semeval-2019 task 7: Rumoureval 2019: Determining rumour veracity and support for rumours. Proceedings of the 13th International Workshop on Semantic Evaluation: NAACL HLT 2019. Association for Computational Linguistics, p. 845–854.

[31] Han X, Huang Z, Lu M, et al. Rumor verification on social media with stance aware recursive tree. Knowledge Science, Engineering and Management: 14th International Conference, Tokyo, Japan, August 14–16, 2021, p. 149–161.

[32] Wang G, Tan L, Shang Z, et al (2022) Multimodal dual emotion with fusion of visual sentiment for rumor detection. arXiv preprint arXiv:220411515.

[33] Kant N, Puri R, Yakovenko N, et al. Practical text classification with large pretrained language models. arXiv preprint arXiv:181201207.