*Research Article*

# A New Early Rumor Detection Model Based on BiGRU Neural Network

**Xiangning Chen,[1,2] Caiyun Wang,[1,2] Dong Li,[1,2] and Xuemei Sun [1,2]**

[1]*School of Computer Science and Technology, Tiangong University, Tianjin 300387, China*
[2]*Tianjin Key Laboratory of Autonomous Intelligence Technology and Systems, Tianjin, China*

Correspondence should be addressed to Xuemei Sun; sunxuemei@tiangong.edu.cn

With the progress of society and the rapid development of computer technology, rumors arise on social media, which seriously affects the social economy. How to detect rumors accurately and rapidly has become one hot research topic. In this paper, a new early rumor detection model is proposed. The aim of this model is to increase the efficiency and the accuracy of rumor detection simultaneously. Specifically, in this model, the input data is firstly refined through account filtering and data standardization, then the BiGRU is used to consider the context relationship, and a reinforcement learning algorithm is applied to detection. Experimental results show that compared with other early rumor detection models (e.g., checkpoints), the accuracy of the proposed model is improved by 0.5% with the same speed, which testifies the effectiveness of this model.

## 1. Introduction

Rumors refer to statements that have no corresponding factual basis but are fabricated and promoted through certain means. In today's highly developed situation of information dissemination media, it can spread quickly through social media, and malicious rumors may affect economy and society significantly. The negative impact of rumors may increase significantly when certain major events occur, such as the traceability of COVID-19 in 2019. This makes people realize that if malicious rumors are not discovered in time, they may continue to cause significant damage, so the timing of their detection is crucial. Figure 1 shows an example of a rumor propagating on Twitter which is named "German Wings Crash." The source message started a claim about the crash could be an Airbus A320 German wings. A German wing Airbus A320 with 150 people on board crashed in Barcelonite, southern France, with no one surviving. The message was retweeted by multiple users on Twitter, either by reposting, commenting, or questioning the original source message. We extracted several related tweets within 24 hours.

Currently, the most research on rumors uses the Twitter social platform and Weibo platform as the main research objects. Considering that the research of rumor detection technology on the Weibo dataset has been relatively mature, and the latest accuracy rate has reached about 95%, this paper uses public standard Twitter dataset as the main research object. Unlike funny videos and celebrity gossips that are popular on Weibo, hot social events are the most popular topics on Twitter.

Based on the platform of Twitter, many researchers have conducted research on rumor detection. When most researchers focus on improving the accuracy of rumor detection, Ma et al. [1] and Kwon et al. [2] have proposed the use of presettings in recent years. The method of fixed checkpoints can evaluate the timeliness of the discovery of rumors, but this method has the disadvantage of not being able to capture the changes in different rumors spreading modes. On this basis, Farajtabar et al. [3] proposed a combination of reinforcement learning and a point process network activity model to detect false news and achieved good results. This assessment of the timeliness of rumor detection is also one of the focuses of this paper.

ID: 580317556516483072
Favorite Count: 37
Followers Count: 57279
"We got reports the crash could be an Airbus A320 Germanwings between Barcelona and Dusseldorf."

2015-03-24

10:37:41
START

ID: 580321844089212928
Favorite Count: 94
followers Count: 891879

"Airbus A320 with 142 passengers and 6 crew members crashes in Digne, France 4U9525."

10:54 :44

ID: 580364656608428032
Favorite Count: 20
Followers Count: 1452091

"AFP: Aviation official says Germanwings crew did not send u distress signal."

13:44:51

ID: 580379486732066817
Favorite Count: 45
Followers Count: 3360512

"Distress call linked with Germanwings jet came from air traffic control, not plane--aviation official."

14:43:47

2015-03-25

ID: 580697361799876608
favorite Count: 69
followers Count: 4972128

"Reports of "moving body" amidst Germanwings wreckage A320."

11:46:54

ID: 580706841769820160
Favorite Count: 17
Followers Count: 328255

"At least 72 Germans and 35 Spanish killed in Germanwings crash: Airline."
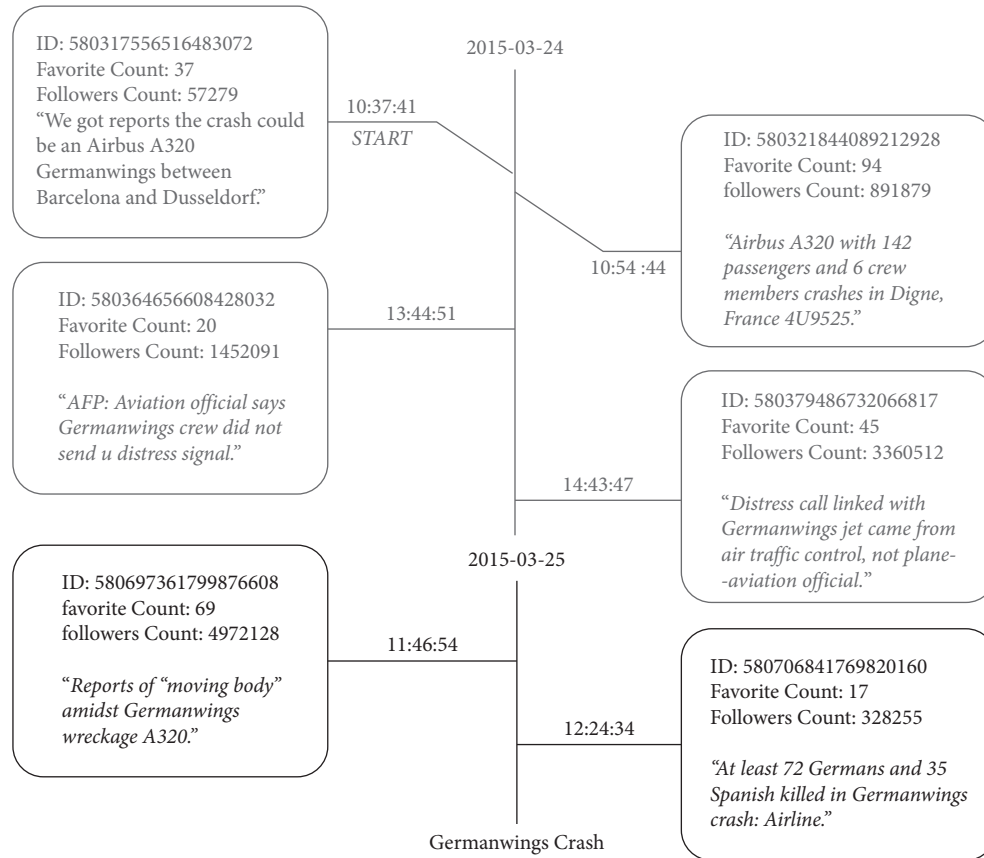
12:24:34

Germanwings Crash

Figure 1: Rumors spread about the schematic "German Wings Crash" events (excerpt).

At the same time, although data preprocessing has been widely used in NLP (Natural Language Processing) to improve performance, few researchers conduct preprocessing on target data for rumor detection. Moreover, this paper finds that most studies focus on the extraction of various features the Twitter standard dataset to capture rumor indicators but ignore the data complexity on Twitter as a large social platform. Firstly, Twitter has a large number of spam accounts. The tweets sent by these accounts usually have a certain commercial purpose, and the research on rumor detection for these tweets is usually weakly relevant or invalid. They are intended to be included in the tweets for this paper. The extracted value information has caused serious interference, so it is necessary to filter it out. Secondly, the short text and randomness of tweets in Twitter makes it difficult for people and computers to accurately understand the information of tweets, so it is necessary to preprocess the tweets to standardize the tweets on the Twitter standard dataset.

The key contributions of our work are summarized as follows:

(1) In rumor detection, for the early rumor detection with checkpoints, this paper proposes to apply BiGRU to the early rumor detection model. Using BiGRU to consider the characteristics of context relations, combined with the sequence before and after, the two posts input before and after the event

are included in the detection, so as to improve the effect of rumor detection.

(2) In rumor detection, aiming at the early rumor detection with checkpoints, this paper proposes a data preprocessing method based on account filtering and text standardization for the first time. The account filtering method is used to remove the junk account. Text standardization is used to standardize tweets in Twitter standard dataset so that the data can better express the meaning of the text so as to improve the accuracy of rumor detection.

(3) In this paper, Q-learning, a reinforcement learning algorithm, is applied for rumor detection to dynamically determine checkpoints, thereby improving the timeliness of rumor detection.

## 2. Related Work

In the contemporary era of the emergence of various social media, rumor detection has attracted the attention of all parties, and the current research on rumor detection has achieved initial results. The current research on rumor detection is mainly divided into two categories: one is based on traditional machine learning for rumor detection, and the other is based on feature learning to extract main features for rumor detection.

Rumor detection methods based on traditional machine learning mainly use decision trees, SVM, and other classifiers to classify events. Liang et al. [4] no longer use a single classifier, such as KNN and SVM but propose a BP neural model and an improved excitation function and add an impulse term to make it possible to detect rumors in the propagation process. Lu et al. [5] noticed that there is an imbalance in the data, which has affected the implementation of rumor detection and proposed a Co-Forest algorithm to improve the imbalanced data and balance the data distribution. Mao et al. [6] used an integrated classifier to detect rumors based on characteristics such as emotional orientation and communication process. In recent years, the rise of artificial intelligence has made the application of deep learning increasingly widespread. Similarly, in the field of rumor detection, deep learning plays an important role. Takahashi and Igata [7] developed a system to detect rumors by studying the spread of rumors and conducted experiments on Twitter and found that they can effectively detect rumors, which opened up a new chapter in rumor detection. Karamchandani and Franceschetti [8] proposed a method of detecting the source of rumors to control the rumors, which extended the best estimator of rules and irregularities to achieve the purpose of detecting the source of rumors. Similarly, Wu et al. [9] also considered propagation in the rumor detection technology and proposed a hybrid SVM classifier based on the graph kernel, which not only captures semantic features such as topics and emotions but also captures high. The first-order propagation mode improves the classification accuracy.

With the rise of artificial intelligence, rumor detection has entered a new stage. The method of rumor detection based on feature learning mainly uses advanced artificial intelligence ideas. Li et al. [10] combine the convolutional layer of the convolutional neural network to extract text features, use the GRU network to process the features, and then judge whether it is a rumor. Ren et al. [11] considered that Weibo text is a graph structure, and information such as the attitude of users' comments will affect the spread of Weibo text and proposed a rumor detection model based on time series. Liao et al. [12] considered the potential information of some Weibo texts and partial user information and proposed a social media rumor detection method based on a hierarchical attention network. Srinivasan and Dhinesh Babu [13] proposed a double convolutional neural network method with a new activation function for the sparse data with little available information that can be used to distinguish rumors at the beginning, because this method has faster generalization speed and more high precision and has a very good effect in rumor detection. After summarizing many studies, Zhou et al. [14] found that their research seldom considered the timeliness of rumor detection and proposed an early rumor detection model to process rumors through two modules, a rumor detection module and a checkpoint module. The rumor detection module is used to extract features, and the checkpoint module is used to solve the problem of timeliness, which is used to trigger the rumor detection module to ensure the timeliness of rumor detection while ensuring accurate identification of rumors. Lin

et al. [15] raised the issue of word independence and found that some common words appear in rumors. Once these words appear, they can be judged as rumors. They proposed a deep sequence model to consider the two aspects of rumors: falsehood sex and influence, using long- and short-term memory units to learn falsehood, and combining deep sequences and social characteristics to learn influence. Asghar et al. [16] proposed a bidirectional long-term short-term memory model based on convolutional neural network, which uses convolutional neural network to extract post features and uses bidirectional long-term short-term memory method to store points and consider contextual information, effectively detecting rumors on Weibo, focusing on the research of rumor detection in Arabic, extracting information from users and content, and proposing the use of semisupervised expectation maximization (EM) to train newsworthy tweets topics to achieve the purpose of rumor detection [17].

In recent years, the research on rumor detection has mainly focused on extracting features and analyzing features. According to the text content of the given data, the main features expressed by the data are extracted and analyzed. Since the data comes from Twitter accounts, there are some spam accounts that guide public opinion, which interferes with the process of rumor detection and affects the results of the detection. At the same time, the longer the rumors spread, the more harmful it will be to the society, and timely detection of rumors is an important aspect of rumors detection research. Aiming at two aspects, this paper starts from the two directions of preprocessing and timeliness. On the one hand, the accuracy of rumor detection is improved through account screening and tweet standardization. On the other hand, reinforcement learning algorithms are used to save the time of rumor detection as much as possible.

## 3. DDR Model Architecture

The model mainly consists of three submodels: a data preprocessing model based on account filtering and standardization (DP model), a rumor detection model based on deep learning (DL model), and a checkpoint model based on reinforcement learning (RL model). Therefore, the proposed model is called DDR (data preprocessing and rumor detection based on deep learning and reinforcement learning) model for short as shown in Figure 2. In the DP model, this paper proposes basic information of users to analyze users, filter spam accounts in Twitter data, refine the data, and use standardized Twitter text to enhance the data to achieve accurate detection of rumors purpose. In the DL model, data characteristics are mainly analyzed, and BiGRU is used to consider the context, analyzes Twitter text data, and detects rumors. The RL model mainly solves the problem of the timeliness of rumors detection. The Q-learning algorithm is used to judge the detection results, and the reward and punishment mechanism is set to make the model trade-off between timeliness and accuracy to improve the timeliness of detection. At the same time, the accuracy of detection is improved.
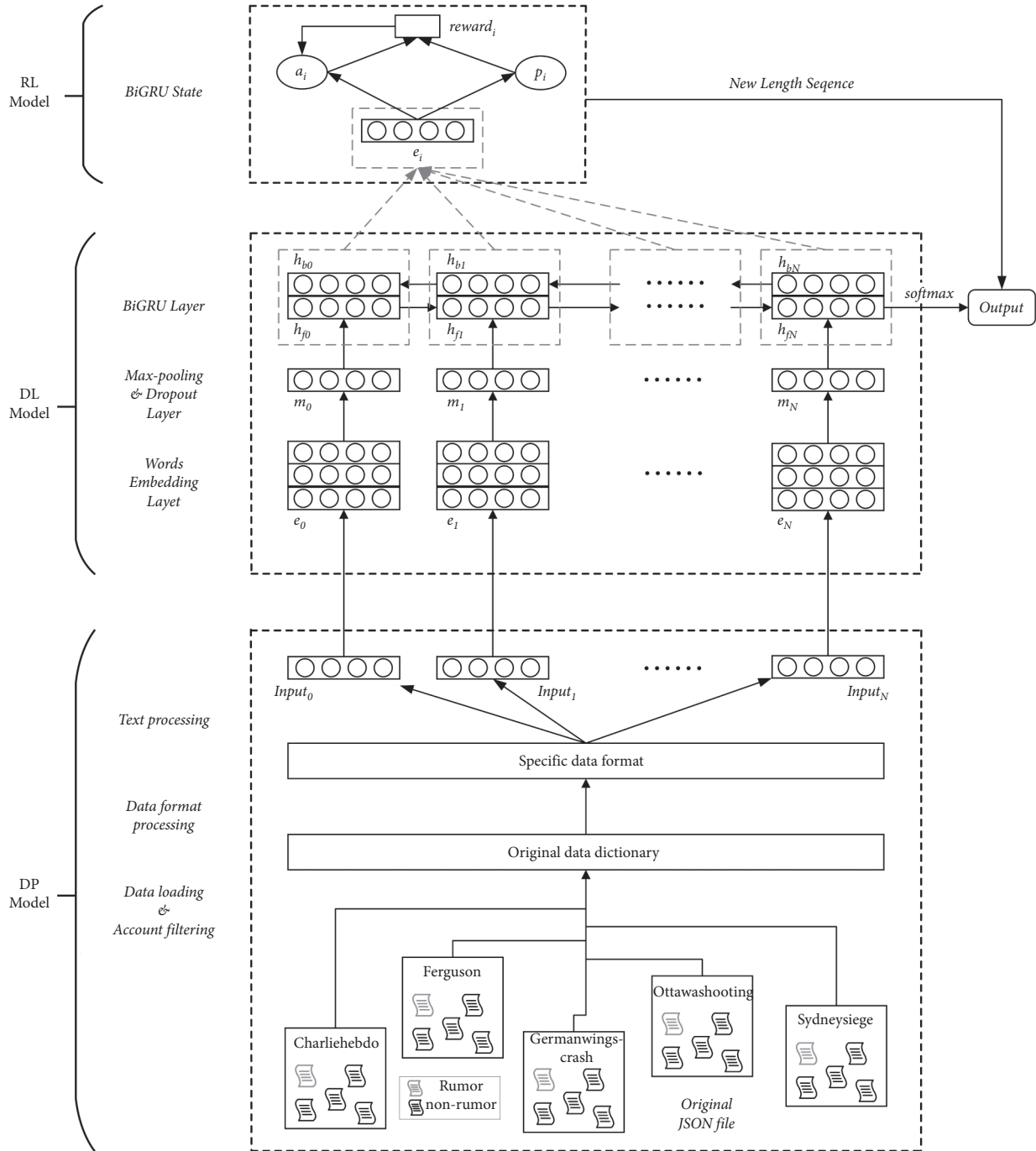
FIGURE 2: Architecture of DDR (data preprocessing and rumor detection based on deep learning and reinforcement learning).

## 3.1. DP Model

### 3.1.1. Account Filtering.
For a Twitter account, its main features are generally divided into features based on user portraits and features based on tweet. User portrait features are a series of features that can be directly extracted from user information, such as the number of people that users follow and number of fans. User tweet features are statistical features extracted from user tweets, such as the proportion of tweets containing URLs and the proportion of liked tweets to the total tweets. Since the tweets of a single Twitter account in the dataset used for rumor detection usually is not larger enough to extract the user-based tweet text features, it is impossible to obtain complete user tweet features. At the same time, if a tweet of a single Twitter account has a high number of likes, the tweet can be considered valuable, thereby reducing the probability of the account being filtered.

Based on the above description, it can be concluded that when a certain type of Twitter account meets certain conditions, the tweets belonging to this account are usually weakly relevant or invalid for rumor detection. In this paper,

this type of account is defined as "filtered account," or FA, and using the user's portrait features and the number of likes on user tweets to determine whether an account is FA. The definition of FA is as follows:

(1) An FA will follow many people, while few users will follow the FA. Based on this feature, the authenticity definition of user accounts is proposed, as shown in the following:

$$\text{authenticity} = \frac{\text{FollowersCount}}{\text{FollowersCount} + \text{FriendsCount}}, \quad (1)$$

where FollowersCount denotes user followers and FriendsCount refers to the number of users following.

(2) Generally, the personal information of spam accounts in Twitter is not complete enough, and it is rare to fill in user description and user location information. Therefore, HasDesc and HasLoc are defined to indicate whether the user has description information and location information. Thus, HasDesc or HasLoc is 1; otherwise, it is 0. Based on the above analysis, the definition of user authority is proposed, as shown in the following:

$$\text{authority} = \text{authenticity} + 0.5 \cdot (\text{HasDesc} + \text{HasLoc})$$

$$+ \frac{\text{TweetsLike}}{\text{AvgLike}}, \quad (2)$$

where TweetsLike denotes the total number of likes corresponding to the current user's tweet and AvgLike refers to the average of the total number of likes of all users.

(3) The authority of all users is sorted in a nonincreasing manner. It is concluded that the bottom 5% of users in the ranking have lower authority, and these users are defined as FA.

We filter out the FAs in the Twitter standard dataset and filter the tweets belonging to these FAs.

### 3.1.2. Tweet Standardization.

The standardization of text is a part of text preprocessing, which mainly refers to the correction of some irregularities or errors in the text, thus transferring it to a text that people can understand correctly. Based on the characteristics of tweets in the Twitter standard dataset, a standardization method for tweet text is proposed in this section.

Tweets in Twitter are usually random and short. On the one hand, tweets are generally limited to 140 words. On the other hand, compared with traditional standard texts, tweets contain many irregularities or errors in terms of wording, grammar, format, and so on, such as spoken language, colloquialisms, acronyms, Internet terms, or emoji expressions, which greatly increase the difficulty of computer understanding of the text, disturbing factors in the difficulty of understanding the text. At the same time, the tweets also contain some symbols and network links that have no actual meaning, and other factors that have no relation with the semantics of the text.

In view of the above characteristics of tweets, in order to strengthen the computer's understanding of tweets, this paper carries out the following standardized processing on tweets:

(1) Unit replacement is as follows: replacing the unit in the text with a unified format, such as replacing "4 kgs" and "4 kg" with "4 kg"

(2) Acronym replacement is as follows: replacing the acronyms in the text with complete words, such as replacing "can't" with "can not"

(3) Spelling proofreading is as follows: replacing some network terms or punctuation of words with irregular spelling, such as replacing "rep" with "reply"

(4) Punctuation is as follows: adding spaces on both sides of all punctuation

(5) Symbol replacement is as follows: replacing all logical symbols with words, such as "and" with the word "and"

(6) Redundant information processing is as follows: removing extra spaces, "@" and "#" symbols in hashtags and removing all hyperlink information

(7) Delete stopwords is as follows: deleting a series of stop words such as "if" and "to"

(8) Part-of-speech restoration is as follows: restoring an English word of any form to its general form, such as "does," "did," and "done" unified reduction to "do"

### 3.2. DL Model.

Rumor detection model based on deep learning processes the tweets after tweet standardization, dividing into words embedding layer, max-pooling and dropout layer, and BiGRU layer. It is used to transform a piece of text into the final state and to judge whether the text is a rumor through the softmax function.

### 3.2.1. Words Embedding Layer.

In the words embedding layer, this paper first performs word segmentation on the text $\text{Input}_i$ that has been standardized. Considering that simple splitting will destroy the semantics of compound words such as "eleven-years-old," this paper uses a word segmentation method based on phrase dictionary matching for text segmentation.

After word segmentation, we map the words to word vectors $w_i^n$ according to word frequency. This paper sets $E_i = \{\text{Input}_1, \text{Input}_2, \ldots, \text{Input}_n\}$ to indicate that there are $n$ tweets at a time, where $\text{Input}_i = \{w_i^1, w_i^2, \ldots, w_i^n\}$ means that the tweet has word vectors, and these word vectors are combined together and obtain the vector matrix $e_i$ of the tweet formed after $\text{Input}_i$ is processed by the words embedding layer.

*3.2.2. Max-Pooling and Dropout Layer.* In order to get the most prominent features of posts, the maximum pooling method is used for pooling, so that keywords or sentence features are reduced, and parameters are reduced. Finally, a fixed-size vector $m_i$ can be generated. At the same time, in order to slow down overfitting and enhance the model generalization ability, the dropout layer is added.

*3.2.3. BiGRU Layer.* In order to strengthen the model's understanding of contextual semantics, this paper uses BiGRU to simultaneously combine the before and after sequences to make predictions. BiGRU is composed of two GRU stacked on top of each other, and its main structure is a combination of two unidirectional GRU. For each time $t$, the input will be provided to the two GRU in opposite directions at the same time, and the output will be jointly determined by the two unidirectional GRU.

As shown in Figure 3, $x_t$ is the input data, $h_t$ is the output of the GRU unit, $z_t$ is the update gate, $z_t$ and $r_t$ jointly control the calculation from the hidden state of $h_{t-1}$ to the hidden state of $h_t$, and the update gate also controls the current input data and previously memorized information $h_{t-1}$, output a value $z_t$ between 0 and 1, and $z_t$ determines how much $h_{t-1}$ is transferred to the next state; the specific unit is calculated as the following formulae show:

$$z_t = \sigma \left( W_z \cdot [h_{t-1}, x_t] \right), \qquad (3)$$

$$r_t = \sigma \left( W_r \cdot [h_{t-1}, x_t] \right), \qquad (4)$$

$$\widetilde{h}_t = \tanh \left( W \cdot [r_t \times h_{t-1}, x_t] \right), \qquad (5)$$

$$h_t = \left( 1 - z_t \right) \times h_{t-1} + z_t \times \widetilde{h}_t, \qquad (6)$$

where $\sigma$ is the Sigmoid function and $W_z, W_r, W$ are the weight matrix of update gate, reset gate, and candidate hidden state, respectively. The reset gate controls the importance of $h_{t-1}$ to the result $h_t$. When the previous memory $h_{t-1}$ is completely related to the new memory, the reset gate can be used to increase the impact of the previous memory. According to the calculation results of reset gate, update gate, and hidden state, the output $h_t$ at the current moment can be obtained by formula (6), thereby obtaining the relationship between BiGRU and a large number of posts. Then, we use the final state $h_N$ ($N$ = the number of posts received so far) to judge the rumors through the softmax function:

$$p = \text{softmax} \left( \omega^T \cdot \tanh \left( h_N \right) \right). \qquad (7)$$

*3.3. RL Model.* In addition to the accuracy of detection, this paper also considers the timeliness of detection. This paper uses the Q-learning algorithm to dynamically determine the best checkpoint to improve the timeliness of rumor detection. The Q-learning algorithm has a calculation action value and a reward mechanism. The action value function calculates the Q value according to the obtained state expression. As shown in formula (8), combining the Q value and the post state expression, the action value function is used to determine whether to terminate or continue, and as shown in formula (9), the characteristic representations $h_{fi}$ and $h_{bi}$ are used as inputs to calculate the action value:

$$Q_{i+1} (s, a) = E \left[ r + \gamma \max Q_i (s', a') \mid s, a \right], \qquad (8)$$

$$a_i = W_a \left( \text{ReLu} \left( W_h \left( h_{fi} + h_{bi} \right) + b_h \right) \right) + b_a, \qquad (9)$$

where $\gamma$ is the discount rate, $r$ is the reward value, $a^0$ is rumor, and $a^1$ is nonrumor.

According to the action value and state value, we get the max reward value, used to optimize the action value.

The main features of the posts are maintained in the words embedding layer unchanged. Through the BiGRU layer, the new state value is obtained by combining the main features of the current input post and the previously obtained state value as the input of reinforcement learning. Then, the action value obtained by reinforcement learning decides whether to terminate or continue. If it terminates and the prediction is correct, the model will give a reward. If it terminates but the prediction is wrong, this model will give a big penalty. If it continues, this model will also give a small penalty. The calculation formula is shown in formula (10). In this way, the model will make a consideration between whether to continue or terminate in the end and, at the same time, make a trade-off between accuracy and timeliness:

$$\begin{cases} \log N & \text{terminate with correct prediction} \\ -P & \text{terminate with incorrect prediction} , \quad (10) \\ -\varepsilon & \text{continue} \end{cases}$$

where $N$ is the number of correct predictions accumulated thus far, $P$ is a large value to penalize an incorrect prediction, and $\varepsilon$ is a small penalty value for delaying the detection.

## 4. Experiments and Analyses

*4.1. Dataset.* This paper uses the public standard Twitter dataset proposed by Ma et al. [18]. This dataset was proposed in 2016 and has been recognized by the academic community. It has since been widely used in the field of rumor detection and is a classic dataset on the problem of rumor detection. It contains 5802 events, each of which contains several Twitter-related JSON files. Each JSON file represents a tweet, including the information of the tweet and the basic information of the user who posted the tweet. The specific data information of the dataset is shown in Table 1. This paper uses part of the information to perform account filtering and then standardizes tweets, uses the creation time of the tweets as the test basis for early detection, encapsulates the data in a certain format, and uses the ratio of 8 : 2 to divide the preprocessed dataset into a training set and a test set for the application of the subsequent sequence in the training and testing of the model.

*4.2. Evaluation Indicators.* For the evaluation criteria of the model, this paper uses four indicators consistent with the literature [19, 20], namely, Accuracy, Precision, Recall, and *F*1. The calculation formula is shown as follows:
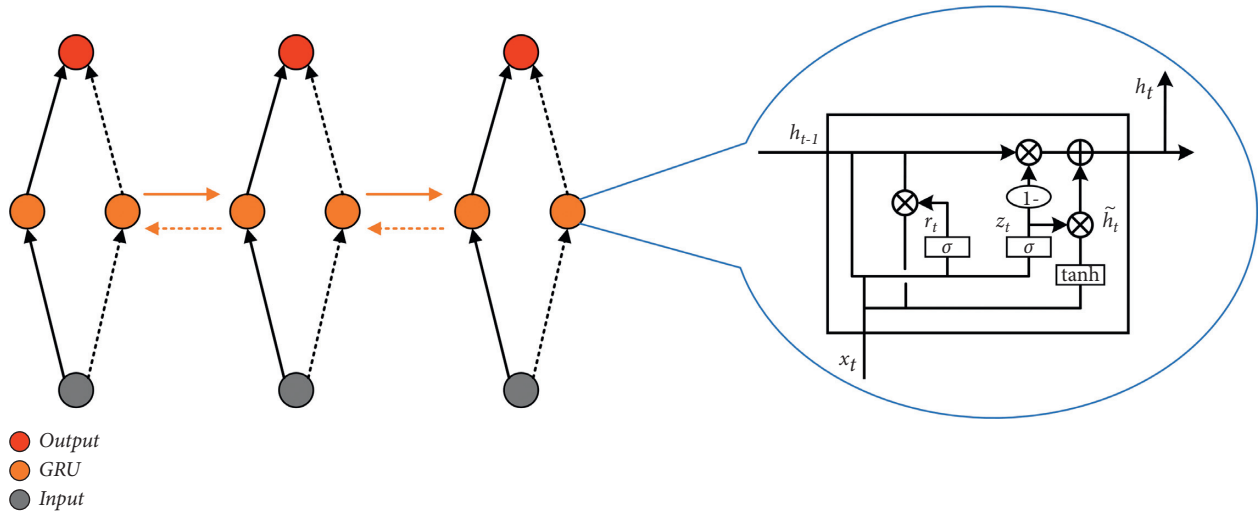
FIGURE 3: BiGRU (left) and GRU update process (right).

TABLE 1: Twitter standard dataset information.

| Statistics | Twitter |
| --- | --- |
| User# | 49,345 |
| Posts# | 103,212 |
| Events# | 5,802 |
| Rumors# | 1,972 |
| Nonrumors | 3,830 |
| Avg. hours per event | 33.4 |
| Avg. # of posts per event | 17 |
| Max # of posts per event | 346 |
| Min # of posts per event | 1 |

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}}, \quad (11)$$

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (12)$$

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (13)$$

$$F1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}, \quad (14)$$

where TP, FN, FP, TN are shown in Table 2 [21].

*4.3. Environment and Hyperparameters.* In this paper, the model is implemented under the Linux system, and the model is trained under the GPU environment of Python 3.6 and TensorFlow 1.13.1. In the training process, this paper uses the per-trained GloVe as the initialization of the word vector to input, which contains 840 billion words. Also, the dimensionality of word embedding is set to 300, while the dropout rate to embedding layer is 0.5. The Adam optimiser (Kingma and Ba) [22] with a learning rate of 0.01 for DL model and 0.001 for RL model are used as the optimization method. Set the size of each batch to 50 and the number of DL model and RL model alternate training rounds to 20.

TABLE 2: Confusion matrix.

| | | Actual | |
| --- | --- | --- | --- |
| | | 1 | 0 |
| Predicted | 1 | True positive (TP) | False positive (FP) |
| | 0 | False negative (FN) | True negative (TN) |

*4.4. Results*

*4.4.1. Training Loss and Reward.* The training loss and reward values over iterations are presented in Figures 4 and 5. In this paper, the DL model and RL model are trained in an alternating manner. It can be seen that the training loss of the DL model tends to be dynamically balanced after about 5000 iterations, and the loss value drops below 0.2, achieving the optimal value. In addition, the reward curve fluctuates more as the reward was calculated based on the accuracy of DL model. When switching between training DL and RL model, the reward value tends to change abruptly. But with the improvement of accuracy over time, a consistent improvement of reward value can be seen.

*4.4.2. Detection Model and Comparison.* In order to effectively testify the accuracy and effectiveness of the model, this paper compares the proposed model with the following models. The RNN model mainly uses recurrent neural networks to analyze data and detect whether a text is a rumor. The LSTM model is an improvement of the recurrent neural network model, which considers the contextual relationship. The GRU-2 model [18] is an improvement of the LSTM model, which reduces the parameters and improves the efficiency. Specifically, this model first divides the event into time periods, then uses the tf-idf method to calculate the text representation of each time period, uses a two-layer GRU network to learn the hidden layer representation of each event, and finally realizes the classification of the event. The HMM model [23] uses the group's point of view for analysis and finally achieves the classification effect. The GAN-GRU model [24] uses a generative adversarial network
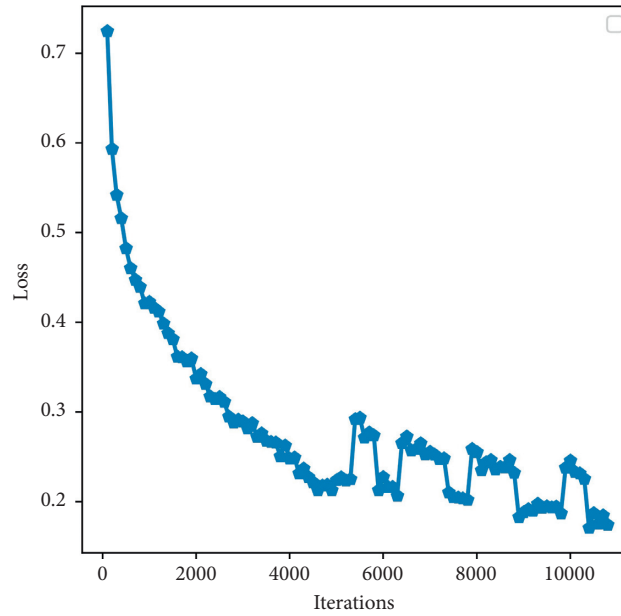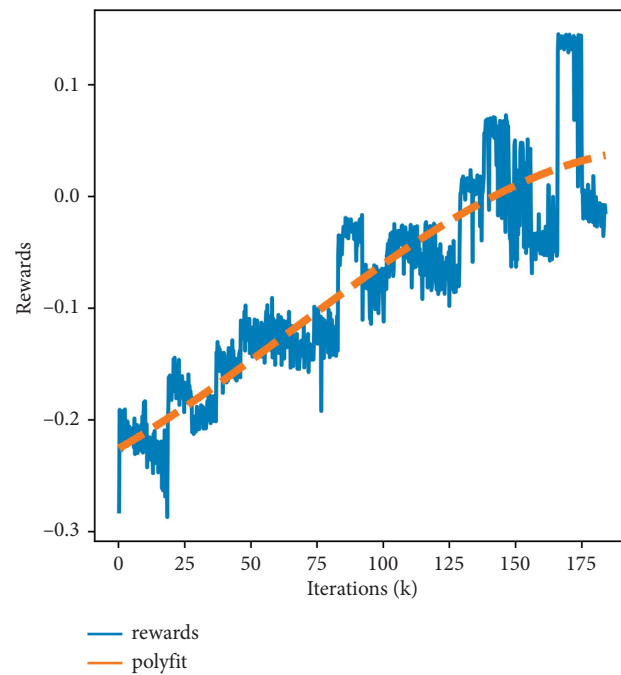
FIGURE 4: Changes in loss.



FIGURE 5: Reward during training.

to train the discriminator for rumor detection. The generator is mainly designed to produce uncertain or conflicting sounds, which complicates the original dialogue thread to make the different supercharged to learn stronger indicative representations of rumors from the enhanced and more challenging examples. The RDM model is a part of the ERD model. It mainly uses a single-layer GRU network to extract main information and uses reinforcement learning to conduct rumor detection in time.

In this section, this paper evaluates DDR based on four evaluation indicators on the test set, as shown in Table 3. This paper is mainly based on the improvement of the ERD model proposed in 2019 and proposes a preprocessing module based on account filtering and standardization. By processing the input of the model, the data is refined, and the accuracy of rumor detection is improved. It can be seen from the chart that the indicators of DDR are basically better than traditional models, such as RNN model, LSTM model,

TABLE 3: Detection accuracy on Twitter.

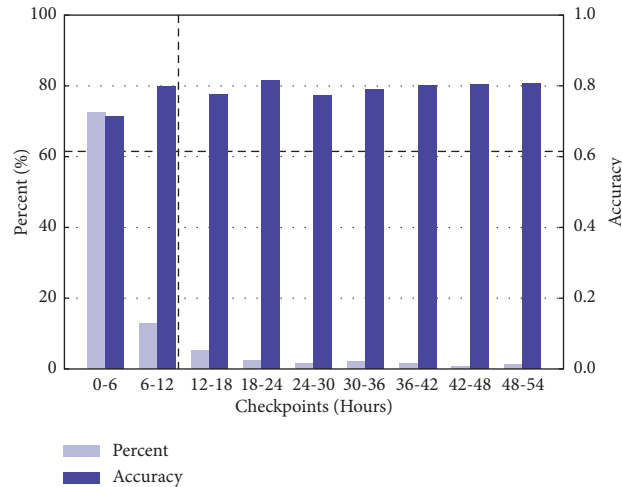| Method | Accuracy | Precision | Recall | F1 |
|---|---|---|---|---|
| Baseline | 0.612 | 0.355 | 0.465 | 0.398 |
| RNN | 0.785 | 0.707 | 0.659 | 0.682 |
| LSTM | 0.796 | 0.719 | 0.683 | 0.701 |
| GRU-2 | 0.808 | 0.741 | 0.694 | 0.717 |
| HMM | — | — | — | 0.524 |
| GAN-GRU | 0.781 | 0.773 | 0.796 | 0.784 |
| RDM | **0.873** | 0.817 | 0.823 | 0.820 |
| ERD | 0.858 | **0.843** | 0.735 | 0.785 |
| DDR (ours) | **0.863** | 0.816 | **0.858** | **0.836** |



FIGURE 6: Proportion of Twitter events classified and detection accuracy by DDR over time. Vertical dashed line indicates the optimal checkpoint for GRU-2 and horizontal dotted line indicates GRU-2's accuracies.

HMM model, and GAN-GRU model. At the same time, the model in this paper is compared with the ERD model that also considers the timeliness of rumor detection. The ERD model has an accuracy of 0.858 in the accuracy of detection. After improving on the basis of the ERD, the accuracy of the model has reached 0.863, an increase of 0.005.

Comparing the three indicators of Precision, Recall, and F1, the model proposed in this paper is 0.027 lower than the ERD model on the Precision indicator, and the other two indicators have increased by 0.123 and 0.051, respectively. It can be seen that the model proposed in this paper has a good effect in the process of rumor detection. However, when reinforcement learning is added, reinforcement learning must consider the timeliness of detection, and you will want to get results faster when making trade-offs. Therefore, reinforcement learning has an impact on the accuracy of the model, compared to not considering timeliness. For the RDM model, the accuracy of the model proposed in this paper is slightly lower than 0.01.

In summary, the model proposed in this paper improves the accuracy of rumor detection while considering the timeliness of model detection. It can be seen from the graph that the proposed model is effective in detecting rumors.

*4.4.3. Detection Timeliness.* Then, in order to evaluate the timeliness of detection, based on the Twitter standard dataset, this paper focuses on comparing the DDR model

with the GRU-2 model. The biggest difference between the two models is that GRU-2 uses fixed checkpoints. The DDR dynamically determines checkpoints through the DL model.

This paper presents the proportion of events that are classified by DDR and the classification accuracy over time (6-hour interval) in Figure 6. Firstly, it can be seen that 70% of rumors are discovered within 6 hours, and the best checkpoint of GRU-2 (vertical dashed line) is 12 hours, so DDR can detect most rumors earlier than GRU-2. Secondly, it can be clearly seen that the classification accuracy of DDR is better than GRU-2 (horizontal dotted line) at all checkpoints.

In summary, it can be seen that DDR improves the timeliness of rumor detection compared with the GRU-2 model.

## 5. Conclusions

This paper presents a new early rumor detection model. The model is divided into three submodels: DP model, DL model, and RL model. In rumor detection, according to the early rumor detection model, this paper proposes to filter the account according to the user's portrait characteristics and the user tweet's praise number. The text standardization for tweets is defined. The original data is processed more precisely and more precisely to improve the accuracy of the detection. At the same time, BiGRU is proposed to enhance

the understanding of context semantics, fully consider the relationship between the post text, and improve the effect of model training. The model is trained by the method of intensive learning to ensure the timeliness of rumor detection. The results are compared with the early rumor detection model. The results show that the accuracy of DDR model is 86.3% in the test set, 0.5% higher than ERD model, and 5.5% higher than that of GRU-2 model. It is proved that the model proposed in this paper has achieved good results in rumor detection. 70% of them are found in 6 hours, and the detection of GRU-2 model takes 12 hours. Therefore, this paper improves the accuracy of rumor discovery on the premise of ensuring the timeliness of rumor discovery. With the development of science and technology, there is an increasing trend about the investigation of the real world problems such as the development of Faults Detection [25], Vortex Theory [26], Periodic Orbit [27], Dynamics [28] and other research fields. It is hoped that the model proposed in this paper can be applied to these problems in the near future.

## Data Availability

No data were used to support this study.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

## References

[1] J. Ma, W. Gao, Z. Wei et al., "Detect rumors using time series of social context information on microblogging websites," in *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, vol. 3, pp. 1751–1754, Melbourne Australia, October 2015.

[2] S. Kwon, M. Cha, and K. Jung, "Rumor detection over varying time windows," *PLoS One*, vol. 12, no. 1, pp. e0168344–19, 2017.

[3] M. Farajtabar, J. Yang, X. Ye et al., "Fake news mitigation via point process based intervention," in *Proceedings of the 34th International Conference on Machine Learning*, pp. 1097–1106, Sydney, Australia, August 2017.

[4] C. Liang, Y. Qiu, and S. Lu, "Research on microblog rumor detection method," *Computer applications and software*, vol. 30, no. 2, pp. 226–228, 2013.

[5] T. Lu, B. Shi, Z. Yan et al., "A semi supervised learning algorithm for microblog rumor detection," *Computer application research*, vol. 33, no. 3, pp. 744–748, 2016.

[6] E. Mao, G. Chen, X. Liu et al., "Research on microblog rumor detection based on deep feature and ensemble classifier," *Computer application research*, vol. 33, no. 11, pp. 3369–3373, 2016.

[7] T. Takahashi and N. Igata, "Rumor detection on twitter," in *Proceedings of the 6th International Conference On Soft Computing and Intelligent Systems*, pp. 452–457, Kobe, Japan, November 2012.

[8] N. Karamchandani and M. Franceschetti, "Rumor source detection under probabilistic sampling," in *Proceedings of the IEEE International Symposium On Information Theory Proceedings (ISIT)*, Istanbul, Turkey, July 2013.

[9] K. Wu, S. Yang, and K. Q. Zhu, "False rumors detection on sina Weibo by propagation structures," in *Proceedings of the IEEE 31st International Conference On Data Engineering (ICDE)*, pp. 651–662, Seoul, South Korea, April 2015.

[10] L. Li, G. Cai, and J. pan, "Microblog rumor detection method based on C-GRU," *Journal of Shandong University*, vol. 49, no. 2, pp. 102–106, 2019.

[11] W. Ren, B. Qin, and T. Liu, "Research on rumor detection based on time series network," *Intelligent computer and application*, vol. 9, no. 3, pp. 300–303, 2019.

[12] X. Liao, Z. Huang, D. Yang et al., "Social media rumor detection based on hierarchical attention network," *Chinese Science: Information Science*, vol. 48, no. 11, pp. 1558–1574, 2018.

[13] S. Srinivasan and L. D. Dhinesh Babu, "A parallel neural network approach for faster rumor identification in online social networks," *International Journal on Semantic Web and Information Systems*, vol. 15, no. 4, pp. 69–89, 2019.

[14] K. Zhou, S. Chang, B. Li et al., "Early rumour detection," *Association for Computational Linguistics*, vol. 1, pp. 1614–1623, 2019.

[15] D. Z. Lin, B. Ma, D. L. Cao, and J. H. Lau, "Chinese microblog rumor detection based on deep sequence context," *Concurrency and Computation: Practice and Experience*, vol. 31, no. 23, 2019.

[16] M. Z. Asghar, A. Habib, A. Habib et al., "Exploring deep neural networks for rumor detection," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, 2019.

[17] S. M. Alzanin and A. M. Azmi, "Rumor detection in arabic tweets using semi-supervised and unsupervised expectation-maximization," *Knowledge-Based Systems*, vol. 185, 2019.

[18] J. Ma, W. Gao, P. Mitra et al., "Detecting rumors from microblogs with recurrent neural networks," in *Proceedings of the 25th International Joint Conference on Artificial Intelligence*, pp. 3818–3824, New York, NY, USA, July 2016.

[19] F. Yu, Q. Liu, S. Wu et al., "A convolutional approach for misinformation identification," in *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pp. 3901–3907, Melbourne, Australia, August 2017.

[20] N. Ruchansky, S. Seo, and Y. Liu, "CSI: a hybrid deep model for fake news detection," in *Proceedings of the ACM on Conference on Information and Knowledge Management*, pp. 797–806, ACM, Singapore, Asia, November 2017.

[21] M. Yuan and H. Tang, "Research on rumor identification based on graph convolution network," *Computer Engineering and Application*, vol. 57, no. 13, 2020.

[22] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," in *Proceedings of the International Conference on Learning Representations (ICLR)*, San Diego, CA, USA, May 2015.

[23] S. Dungs, A. Aker, N. Fuhr et al., "Can rumour stance alone predict veracity," in *Proceedings of the 27th Inter-national Conference on Computational Linguistics*, pp. 3360–3370, Santa Fe, NM, USA, August 2018.

[24] J. Ma, W. Gao, and K. F. Wong, "Detect rumors on twitter by promoting information campaigns with generative adversarial learning," in *Proceedings of the World Wide Web Conference*, pp. 3049–3055, San Francisco, CA, USA, May 2019.

[25] M. Miguel, J. A. Antonino-Daviu, F. D. C. Castellá Pedro, and C. Alberto, "Higher-order spectral analysis of stray flux signals for faults detection in induction motors," *Applied Mathematics and Nonlinear Sciences*, vol. 5, no. 2, pp. 1–14, 2020.

[26] M. Sharifi and B. Raesi, "Vortex theory for two dimensional Boussinesq equations," *Applied Mathematics and Nonlinear Sciences*, vol. 5, no. 2, pp. 67–84, 2020.

[27] A. A. Abozaid, H. H. Selim, K. A. K. Gadallah, I. A. Hassan, and E. I. Abouelmagd, "Periodic orbit in the frame work of restricted three bodies under the asteroids belt effect," *Applied Mathematics and Nonlinear Sciences*, vol. 5, no. 2, pp. 157–176, 2020.

[28] S. S. Hassan, M. P. Reddy, and R. K. Rout, "Dynamics of the modified n-degree Lorenz system," *Applied Mathematics and Nonlinear Sciences*, vol. 4, no. 2, pp. 315–330, 2019.