Quest Journals
Journal of Software Engineering and Simulation
Volume 11 ~ Issue 2 (2025) pp: 09-15
ISSN(Online) :2321-3795 ISSN (Print):2321-3809
www.questjournals.org

Research Paper

Research on Fact Verification Based on GRU Integrated with Community Q&A Information

Jiankang Han[1]
[1]*School of Control and Computer Engineering, North China Electric Power University,*
Baoding, 071051, China
Corresponding Author:Jiankang Han

ABSTRACT: *This paper aims to explore methods for fact verification of user answers in the community question-and-answer domain. A novel Gated Recurrent Unit (OE-GRU) model based on global evidence information is proposed. The model improves performance by introducing global evidence information and a scaling factor, and designs a general framework for fact verification in community Q&A. Experiments conducted using the AnswerFact dataset compare the performance of existing CNN, RNN, LSTM, GRU, and the improved GRU models. Results show that the OE-GRU model achieves the best accuracy and F1 scores. Additionally, ablation experiments and comparisons with different scale values and pooling methods further validate the effectiveness of the proposed model's components. Overall, this study provides new methods and technical support for fact verification in the community Q&A domain.*

KEYWORDS: *Community Q&A; Fact Verification; GRU*

***R**eceived 08 Feb., 2024; **R**evised 16 Feb., 2025; **A**ccepted 18 Feb., 2025 © The author(s) 2025.*
P**ublished with open access at **www.questjournas.org

I. INTRODUCTION

Nowadays, with the development of social media, traditional e-commerce websites have begun to introduce community Q&A features to assist customers in selecting the products they desire. According to a Statista 2023 report, 83% of global e-commerce platforms have integrated community Q&A features, but reliability issues in user-generated content (UGC) have led to approximately 32% of consumers making incorrect purchasing decisions. However, the responses provided by users are not always accurate. Due to differences in users' age, education level, and geographical location, these responses may contain erroneous information. For example, in electronics Q&A, variations in regional voltage standards often result in incorrect advice, while in healthcare discussions, disparities in educational backgrounds may lead to misunderstandings of technical terms. Furthermore, some users might provide misleading comments for personal gain or malicious competition. Recent Amazon platform monitoring revealed that around 15% of reviews contain profit-driven misinformation, particularly severe in 3C (computer, communication, consumer electronics) product categories. Therefore, fact-checking user responses in the community Q&A domain is essential[1]. However, only a few researchers have focused on this area. Researchers have developed various machine learning and deep learning models to automatically assess the quality of answers provided by the community. These models typically consider the completeness, accuracy, reliability, and readability of the answers. However, there has been relatively less research on the factuality of the answers themselves. Thus, this paper aims to explore methods for fact-checking user responses in the community Q&A domain to improve users' ability to access accurate information. Through data annotation, model training, and experimental comparison, this research hopes to provide new research ideas and technical support for this field.

II. RELATED WORK

With the development of internet technology, an increasing amount of data, especially text data, is emerging on social networks. Effectively identifying the authenticity of this textual information to facilitate its utilization has become an important research direction. In the field of community Q&A, textual authenticity generally refers to the correctness of user responses. Due to the diversity of text forms and the uncertainty of information sources, this task is challenging. In recent years, with the advancement of artificial intelligence, particularly in natural language processing (NLP) techniques, researchers have begun to apply NLP technologies to fact-checking in community Q&A.

Fact-checking is a relatively novel and significant research direction in the field of NLP. Researchers such as A Tchechmedjiev[2] and N Vedula[3] have attempted to adopt machine learning and graph neural networks for fact-checking structured data. However, these methods are insufficiently generalizable and only applicable to specific domains.

Pre-trained language models are one of the significant advances in the field of NLP, demonstrating outstanding performance in various tasks in recent years. These models are pre-trained on large-scale corpora, learning rich linguistic knowledge and enabling transfer learning in multiple downstream tasks. Through pre-trained language models, this research encodes human textual information into machine-recognizable vector representations.

In the field of fact-checking, pre-trained language models can be combined with deep learning models to predict the authenticity of text. Common deep learning models include Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU).

2.1 CNN

CNNs are models that extract local features and patterns from input data through convolution and pooling operations[4]. While CNNs excel at capturing local information in text, they cannot effectively comprehend the overall semantics of sequential text, which makes CNNs less effective for long-text sequences.

2.2 RNN

RNNs are a class of deep learning models adept at handling sequential data[5]. Unlike traditional feedforward neural networks, RNNs have recurrent connections, enabling them to process and remember contextual information in sequential data. However, RNNs can suffer from vanishing and exploding gradient problems when dealing with long sequences, affecting the model's performance.

2.3 LSTM & GRU

To address these issues, LSTM and GRU were developed. LSTM introduces memory cells and gating mechanisms to effectively maintain gradients when processing long sequences, thereby capturing long-term dependencies in the text[6]. GRU, with its streamlined gating mechanism, achieves similar performance to LSTM but with higher computational efficiency.

Question: Does this case fit the S4 with the inductive charging back? It is slightly thicker than the original back.

Answer: No, it will not is only for the S2.

Verdict: FALSE

s1: Love these cases...they fit the Galaxy S4 so well, they even accommodate the wireless charger back plate.

s2: It fits the s4 perfect, the cut outs are perfect and its not bulky.

s3: I had a very similar case for my Galaxy S2, so I bought this one hoping it would hold up as well as the first.

s4: I wish it was available in more colors for the Galaxy S4.

s5: The case didn’t work with extended battery and cover.

Table 1:Example Data from the AnswerFact Dataset

III. RESEARCH CONTENT

The main research content of this paper includes the following three points:

Proposing a Novel GRU Model: Based on the field of fact-checking in community Q&A, a novel GRU model is designed. This model introduces new global evidence information input and a scaling factor into the GRU to enhance model performance.

Proposing a General Framework for Fact-Checking in Community Q&A: Through the analysis of the fact-checking field in community Q&A, a general framework for fact-checking in community Q&A is designed. This framework improves the model's generalizability and provides a reference for future fact-checking in community Q&A.

Model Performance Analysis: By comparing the performance of existing CNN, RNN, LSTM, GRU, and the improved GRU, the strengths and weaknesses of each model are analyzed. Experimental results demonstrate the effectiveness of the improved GRU.

IV. RESEARCH PROCESS

For fact-checking in the field of community Q&A, this study is defined as follows: Given an answer aa and its corresponding question q, the goal of this research is to predict the authenticity of the answer by utilizing k relevant evidence sentences $s1, s2, \ldots, sk$. The authenticity of the answer belongs to one of the predefined types of authenticity. The authenticity labels are

$$y = \{true\ ,\ false\ ,\ unsure\}$$

This study employs the AnswerFact dataset[7]. This dataset includes five domains with the most Q&A pairs: electronics, home and kitchen items, sports and outdoor gear, health and personal care products, and mobile phones and accessories, totaling approximately 2.7 million Q&A pairs. In this section, we present a general framework for fact-checking in the community Q&A domain. For each word in the given text sequence, whether it is a question, an answer, or an evidence sentence, we map it to a vector representation using an embedding matrix. To capture temporal interactions between words, we utilize a bidirectional GRU to transform word embeddings w_t into context-aware representations h_t. We note that there is an interrelationship between evidence and Q&A pairs. Evidence may help better support or refute a given answer. Therefore, we incorporate the overall information of the evidence into the GRU gating mechanism, designing a novel GRU that helps better capture the logical relationship between evidence and Q&A pairs.

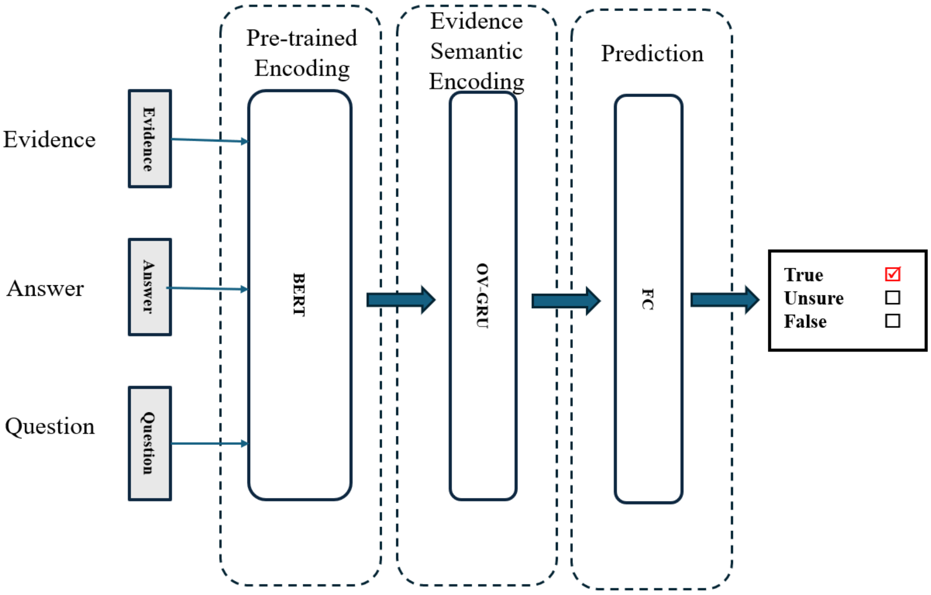

Figure 1 Overall Architecture Diagram

Specifically, for the set of evidence $s = s1, s2, \ldots, sk$ encoded by the pre-trained language model, we concatenate them end-to-end to form a sentence S, and then perform max pooling on S to obtain the overall vector representation of the evidence.

$$s = \max(S)$$

We believe that this s can effectively represent the overall information in the evidence set.

Additionally, we designed a novel OE-GRU based on global evidence information. This model improves the input gate and update gate of the original GRU to capture the potential logical connections in fact-checking within the community Q&A domain. Specifically, given a question q, an answer a, and the overall vector representation s of the evidence, we calculate the input gate r_t, update gate z_t, and new memory state n_t as follows:

$$r_t = \sigma\left(W_{ir} \cdot x_t + W_{hr} \cdot h_{t-1} + \text{scale} \cdot W_{yr} \cdot s + b_r\right)$$
$$z_t = \sigma\left(W_{iz} \cdot x_t + W_{hz} \cdot h_{t-1} + \text{scale} \cdot W_{yz} \cdot s + b_z\right)$$
$$n_t = \tanh\left(W_{in} \cdot x_t + r_t \cdot \left(W_{hn} \cdot h_{t-1} + \text{scale} \cdot W_{yn} \cdot s\right) + b_n\right)$$

Here, x can be q or a, t ranges from $[0, l-1]$, and l is the length of x. Wir,Whr,Wyr,Win,Whn,Wyn,br,bz,bn are learnable parameters, and $scale$ is an introduced scaling factor to control the amount of evidence information in the GRU gates.

Finally, combining the results of the gates, we have:

$$h_t = (1 - z_t) \cdot n_t + z_t \cdot h_{t-1}$$

Thus, we obtain the global evidence encoding of q, $Q = q1, q2, q3, \dots, ql$, and the global evidence encoding of a, $A = a1, a2, a3, \dots, al$.

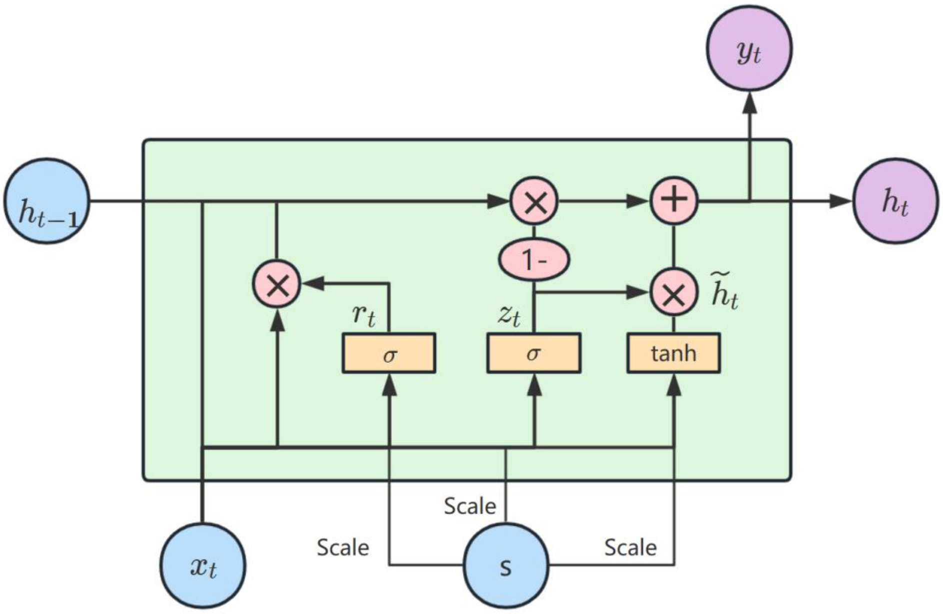

Figure 2 OE-GRU Structure Diagram

To integrate the obtained Q and A, this study pools them into a single vector for final computations. There are two pooling methods: Max Pooling and Mean Pooling.

Max Pooling: For the global evidence encoding Q and A, the vectors after max pooling are denoted as $Q_{\max}$ and $A_{\max}$, respectively. The formula for max pooling is as follows:

$$Q_{\max} = \max(q_1, q_2, \dots, q_l)$$
$$A_{\max} = \max(a_1, a_2, \dots, a_l)$$

Mean Pooling: For the global evidence encoding Q and A, the vectors after mean pooling are denoted as $\boldsymbol{Q}_{\text{mean}}$ and $\boldsymbol{A}_{\text{mean}}$, respectively. The formula for mean pooling is as follows:

$$Q_{\text{mean}} = \frac{1}{l}\sum_{i=1}^{l} q_i$$
$$A_{\text{mean}} = \frac{1}{l}\sum_{i=1}^{l} a_i$$

By using max pooling or mean pooling, this study can transform Q and A into fixed-size vectors Q_{pooled} and A_{pooled}, making them suitable for subsequent computations and model processing.

Finally, this study employs a fully connected layer to predict the results. Using the globally pooled evidence vectors Q_{pooled}, A_{pooled}, and S_{pooled}, the study conducts the final classification prediction.

Initially, these pooled vectors are concatenated to form a comprehensive representation vector:

$$H = [Q_{\text{pooled}}; A_{\text{pooled}}; S_{\text{pooled}}]$$

where$[\cdot;\cdot;\cdot]$ denotes the concatenation operation.

Subsequently, the comprehensive representation vector is input into the fully connected layer for classification prediction:

$$y = \text{softmax}(W \cdot H + b)$$

where: W is the weight matrix of the fully connected layer. b is the bias vector of the fully connected layer. softmax is the activation function for multi-class classification.

V. EXPERIMENTAL RESULTS

In the experiment, this study utilized the AnswerFact dataset. AnswerFact is a representative dataset for community question-answer factual verification tasks, containing a total of 60,864 pairs of questions and answers. The original dataset categorizes the results of factual verification into five classes. For better evaluation of the model, this study merges similar classes into three categories, transforming the task into a three-class classification problem. This study adopts accuracy and F1-score to evaluate the model's performance. To assess the model's effectiveness, common deep learning models were selected for performance comparison, as shown in the following table:

Model	Accuracy	F1 Score
FC	0.646	0.483
CNN	0.648	0.442
RNN	0.658	0.488
LSTM	0.649	0.492
GRU	0.668	0.517
OE-GRU	0.672	0.533

Table 2 Experimental Results of Different Models

The experimental results show that different models perform differently on the AnswerFact dataset. By comparing accuracy and F1 scores, it can be observed that the fully connected (FC) model has relatively low performance, indicating that a simple fully connected layer cannot adequately capture the complex relationships in question-answer data. While convolutional neural networks (CNN) have an advantage in capturing local features, their performance is limited when dealing with time-series data, resulting in lower F1 scores. Recurrent neural networks (RNN), long short-term memory networks (LSTM), and gated recurrent units (GRU) excel in capturing time-series information and perform better than the previous two models, though long-term dependency issues remain. The optimized gated recurrent unit (OE-GRU) model, which is based on global evidence information, further improves performance on the GRU and achieves the highest accuracy and F1 scores, indicating that the optimized GRU model is most effective in handling this dataset. In summary, the experimental results demonstrate that the OE-GRU model performs best in the community question-answer factual verification task, showcasing its superiority in this task. Additionally, ablation experiments were conducted on the GRU modules used in this study, and the results of the ablation experiments are as follows:

Model	Accuracy	F1 Score
OE-GRU	0.672	0.533
-scale	0.666	0.492
-pooling	0.661	0.451
-scale and pooling	0.646	0.483

Table 3 Module Ablation

Regarding scale, this study replaced the value of scale with 0 in the gating computation formula. For pooling, this study achieved it by replacing the pooling module with the hidden layer of the GRU. From the ablation experiments, it can be seen that the designed model's contributions in various parts are significant. Removing the scale part or the pooling part led to a certain degree of performance degradation, proving the effectiveness of these two parts. Removing pooling resulted in less comprehensive extraction of global information from the question-answer pairs, affecting subsequent judgment results. Removing scale led to less accurate weight allocation for different features, decreasing the model's ability to distinguish features.

This study also tested the effect of different scale values on the experimental results, as shown in the following figure:

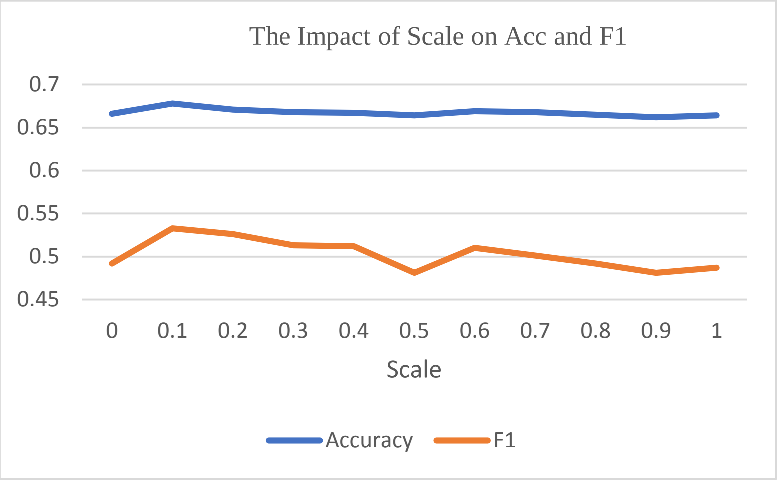

Figure 3 Effect of Scale on Accuracy and F1 Score

Based on the experimental results, it can be seen that different scale values have a significant impact on the performance of the model. Specifically, when the scale value is 0.1, the model performs best in terms of accuracy and F1 score, achieving 0.672 and 0.533, respectively. This indicates that an appropriate scale value can effectively improve the model's performance. As the scale value increases, the model's performance shows a downward trend. When the scale value exceeds 0.2, both accuracy and F1 score decline. This suggests that an excessively high scale value negatively affects model performance, possibly due to uneven weight distribution among features, thus affecting the model's classification effectiveness. Notably, when the scale values are 0.6 and 0.7, the model's performance rebounds but does not reach the optimal level. This implies that adjusting the scale value within a specific range can have an optimizing effect on the model's performance.

Additionally, this study also tested the effects of two pooling methods on the experimental results, as shown below.

Pooling method	Accuracy	F1 Score
max	0.672	0.533
mean	0.674	0.524

Table 4 Comparison of Different Pooling Methods

Experimental results indicate that different pooling methods have significant impacts on model performance. When using max pooling, the model's accuracy and F1 score are 0.672 and 0.533, respectively. In contrast, using average pooling slightly increases the model's accuracy to 0.674, but the F1 score drops to 0.524. This suggests that, although average pooling may improve model accuracy in certain cases, it negatively affects the model's consistency and classification performance, leading to a decrease in the F1 score. Conversely, max pooling demonstrates more stability in maintaining model consistency and classification performance.

VI. CONCLUSION

This paper proposes a novel model based on the Gated Recurrent Unit (OE-GRU) for fact verification in community question answering, leveraging global evidence information, and designs a general framework for fact verification in this domain. Experimental results indicate that the OE-GRU model performs best in the task of fact verification in community question answering, demonstrating high accuracy and F1 scores, thereby proving its superiority in handling question-answer data.

Additionally, this paper conducts ablation experiments and comparative analyses of different scale values and pooling methods, further validating the effectiveness of each component of the proposed model. The experimental results show that appropriate scale values and the use of max pooling methods can significantly enhance model performance, thus improving the effectiveness of fact verification in community question answering.

Overall, this study not only provides new methods and technical support for fact verification in the field of community question answering but also lays the foundation for future research and applications. The study aims to continually enhance the model's performance and practicality through further research and optimization, thereby providing users with more accurate and reliable information services.

REFERENCES

[1]. Mihaylova T, Nakov P, Màrquez L, et al. Fact checking in community forums[C]//Proceedings of the AAAI conference on artificial intelligence. 2018, 32(1).

[2]. Tchechmedjiev A, Fafalios P, Boland K, et al. ClaimsKG: A knowledge graph of fact-checked claims[C]//The Semantic Web–ISWC 2019: 18th International Semantic Web Conference, Auckland, New Zealand, October 26–30, 2019, Proceedings, Part II 18. Springer International Publishing, 2019: 309-324.

[3]. Vedula N, Parthasarathy S. Face-keg: Fact checking explained using knowledge graphs[C]//Proceedings of the 14th ACM International Conference on Web Search and Data Mining. 2021: 526-534.

[4]. Yang Y, Zheng L, Zhang J, et al. TI-CNN: Convolutional neural networks for fake news detection[J]. arXiv preprint arXiv:1806.00749, 2018.

[5]. Atanasova P, Nakov P, Màrquez L, et al. Automatic fact-checking using context and discourse information[J]. Journal of Data and Information Quality (JDIQ), 2019, 11(3): 1-27.

[6]. Rai N, Kumar D, Kaushik N, et al. Fake News Classification using transformer based enhanced LSTM and BERT[J]. International Journal of Cognitive Computing in Engineering, 2022, 3: 98-105.

[7]. Zhang W, Deng Y, Ma J, et al. AnswerFact: Fact checking in product question answering[C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2020: 2407-2417.