Financial Fact-Check Via Multi-modal Embedded Representation and Attention-Fused Network

Padmapriya Mohankumar, Vishal Kumar Singh, and Ashraf Kamal

Abstract Nowadays, fact-check has become an important problem to address. Although there has been immense exploration and research in this direction, considering it from financial data perspective is still unexplored. This paper presents a new deep learning-based multi-modal approach for fact-checking of financial claims. `MuDal-FinFaCk`, a new model is considered which is responsible for classifying financial claims as either supported or refuted based on textual and image-related evidences. Our proposed model is evaluated on two benchmark datasets, and it shows impressive results in terms of *F-score* and *Accuracy*. Also, it performs significantly better as compared to the relevant methods and previous works.

Keywords Misinformation · Fact check · Claim verification · Financial fact-check · Information retrieval

1 Introduction

In the last two decades, online information has been broadly span on the Web [1]. It has been used in the detection of several problems, such as satire [2], hate speech [3], structure generation networks [4], fake news [5], sentiment analysis [6], and non-literal text [7]. It also causes the spread of online misinformation propagation. Recently, it has become a serious issue worldwide. Considering these facts, computational fact-check verification has been seen as a need of the hour to ensure the trustworthiness of any received information. It helps in preventing the spread of misinformation and ensures that decisions and opinions are based on verified and

P. Mohankumar · V. K. Singh (✉) · A. Kamal
PayPal, Chennai, India
e-mail: vishalksingh@paypal.com

P. Mohankumar
e-mail: pamohankumar@paypal.com

A. Kamal
e-mail: askamal@paypal.com

A. Swaroop et al. (eds.), *Proceedings of Data Analytics and Management*, Lecture Notes in Networks and Systems 1301, https://doi.org/10.1007/978-981-96-3372-2_3

credible information. Also, multi-modal-related misinformation, such as the combination of text and images, text and audio, or images and audio, has been conveyed more often nowadays. Thus, multi-modal-related fact-checking is also important by verifying information using multiple data sources, such as text, images, and videos. It helps to detect any potential manipulation or context distortion that may occur when information is presented in numerous ways [8].

The rise of misinformation in the financial domain has become a pressing concern, with potential impacts on public trust, investor decisions, and overall market stability. Although, fact-check in the financial domain is yet less explored. Financial fact-check involves verifying information related to financial matters, such as statements about companies, markets, investments, economic indicators, and financial policies [9]. It is crucial because inaccurate or misleading financial information can have significant consequences for investors, businesses, and economy. Further, multi-modal financial fact-check involves verifying financial information using several data formats. Admitting these facts, the financial fact-check verification in a multi-modal setting is the need of the hour, wherein evidences are taken from different aforementioned data sources and formats [10].

1.1 Our Contributions

This study presents a new multi-modal approach for financial fact-check. A new model, namely, `MuDal-FinFaCk`, a new model based on deep learning approach is established. It takes financial claim and their corresponding textual and image-related evidences as input and further it is passed to the corresponding embedding layer. Thereafter, the outcome of the embedding layer is passed to the corresponding convolutions neural network (CNN) which is responsible to extract syntactic and semantic factual information, bi-directional long short-term memory (BiLSTM) which captures latent fact-check-related contextual sequences, and attention layer which considers important information related to financial fact-check verification with respect to claims and evidences. Thereafter, a fusion attention layer is employed which concatenates context vectors generated from the attention layers of the textual and image evidences to retrieve important factual insights from both multi-modal sources. Further, the outcome of the fusion attention layer and claim attention layer is used to measure the contrastive loss-related latent representations between input claims and evidences. The output of the contrastive loss is passed to the dense and output layers for financial fact-check verification. As a result, it classifies/verifies the particular input claim either *supported* or *refuted*.

The key contributions of this study can be summarized as follows:

- Introduced a multi-modal approach for financial fact-check problem.
- Implemented `MuDal-FinFaCk`, a new model to financial fact-check verification.
- Performed empirical evaluation of our proposed `MuDal-FinFaCk` model on two benchmark datasets.

- Compared our proposed MuDal-FinFaCk model with relevant baselines methods.

Section 2 presents the relevant studies. Section 3 presents the proposed approach and briefly discusses our new MuDal-FinFaCk model. Section 4 presents the experimental setup, results, and comparison. Finally, Sect. 5 presents conclusion of this study and emphasizes on relevant important future works.

2 Related Work

The section presents the relevant works for financial fact-check claim verification. In [9], authors considered check-worthy claims to know the truth among general public. Authors highlighted that each claim refers to binary decision or ranking of claims. In [10], authors considered truth discovery for a particular problem based on true facts and conflicting data source. Authors excluded factual claims and assumed the input of contradicting tuples which highlight property values of objects. In [11], authors considered knowledge graphs and highlighted the shortest path between nodes to fact-checking.

In [12], authors considered multi-modal and multi-lingual content-related fact verification task and evaluated their performance on three benchmark datasets. In [13], authors performed fact-check using via external resources like Web. In [14], authors published a financial fact-check dataset in multi-modal settings. Also, they highlighted explanation generation in the same study. In [15], authors proposed a large-scale dataset namely, Mocheg for fact-check verification consisting of text and images. They claimed that this is the first dataset using multi-modal setting.

The above discussion shows that the computational fact-check is a well-known problem. However, there has been extensive study in this direction of research, especially in the uni-modal setting. On the other hand, considering this problem towards financial domain is yet less explored. Also, financial fact-check verification for financial claims based on their corresponding evidences via multi-modal settings is a worth research investigation task and need of the hour. To the best of our knowledge, this is the first study towards financial fact-check-related claim verification/classification via multi-modal settings.

3 Proposed Approach

In this section, we present the proposed approach. It includes the problem description, dataset collection followed by pre-processing, and finally, a brief discussion of the proposed MuDal-FinFaCk model.

3.1 Problem Description

This study introduces a new problem for financial fact-check for finance-based claims based on their multi-modal evidences. It is considered as two-class problem, wherein a particular financial claim is verified or classified as either *supported* or *refuted*.

3.2 Dataset Collection and Pre-processing

This study considers two publicly available benchmark datasets. A short description of datasets is given below:

- `Fin-Fact` [14]: This dataset is based on multi-modal financial fact-check. It contains 3562 claims, out of which 1807 are *True/Supported*, 1315 are *False/Refuted* claims. Rest of the 440 claims have not enough information. So, in this study, we consider only *true* and *false* claims.
- `Mocheg` [15]: This dataset is based on multi-modal fact-checking. It contains 15,601 claims and 33,880 textual evidences and 12,112 image-based evidences. We have taken financial keywords from [1] work to filter only finance-related claims and evidences.

Thereafter, we apply several pre-processing steps over all collected claims and evidences from these two benchmark datasets to clean the collected data. Hence, we have eliminated comma, punctuation, unwanted dots, question marks, alpha-numeric characters, URLs, mentions, hashtags, exclamation marks, and in the end convert collected data into lower-case form.

3.3 Proposed `MuDal-FinFaCk` Model

This section presents the newly proposed `MuDal-FinFaCk` model. Figure 1 presents the architectural workflow of the proposed `MuDal-FinFaCk` model. The complete description of all layers of the proposed model is demonstrated in the following sub-sections.

Input Layer: The input layer receives inputs such as textual claim, textual evidence, and pictorial/image evidence. Words available in the input textual claims and input textual evidence are tokenized. Further, it assigns a number as index value to generate a dictionary for it and convert as an textual claim vector, c_i, and textual evidence vector, e_i. A fixed-padding is applied and generate padded claim vector, c_p, and padded evidence vector, e_p accordingly to maintain the same padding length and passed to their corresponding embeddings in the next layer. Likewise, for padded pictorial/image evidence is converted into a visual padded vector, v_i.

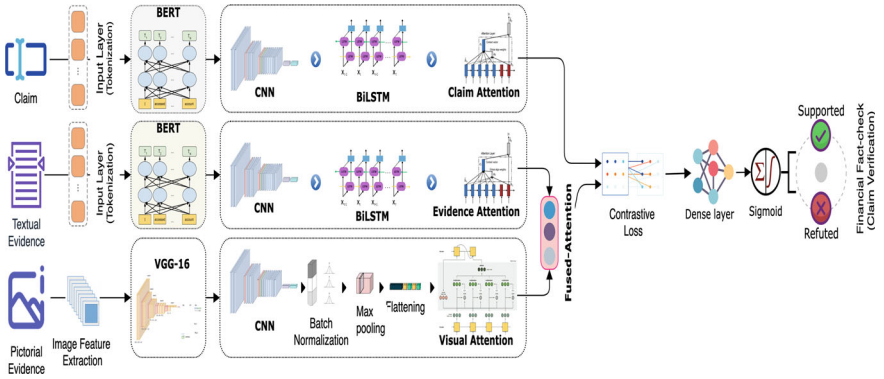

Fig. 1 Architectural workflow of our proposed `MuDal-FinFaCk` model

Embedding Layer: In this study, three embeddings are used, wherein `BERT` [16], one of the most popular pre-trained language models is used for input padded claim vector and input padded evidence vector separately. Both embeddings generate relevant and contextual word vector representation based on the fact-claim and fact-evidences through 768-dimensional vector using `Transformer`. On the other hand, `VGG16` [17], a popular embedding in computer vision is used to generate padded pictorial/image evidence-related embedded vector. In the end, the encoded representation of three embeddings is passed to the common layers.

Common layers: In our proposed model, the common layer is consisted of CNN, BiLSTM, and attention layers. They receive the embedding layer outcome as input. The attention layer is specific to the related input claim and evidences. A brief information about these layers is given below:

- **CNN layer**: In this study, CNN receives the outcome of their corresponding embedded vectors of input claims and textual and image-related evidences. It is employed to extract relevant semantic and syntactic features from the claims and evidences accordingly for financial fact-check purpose via applying different convolutional and pooling operations. Our proposed model considers one-dimensional convolution layer for input claims and input evidence, while two-dimensional convolution layer is applied for image-based evidences with different filters and window size settings. The role of all filters is to perform convolutional operation and the pooling layer retrieves the maximum relevant financial fact-based features. As a result, features of different size sequences are extracted and CNN generates feature sequences for all input financial claims and evidences filters accordingly.
- **BiLSTM layer**: In this study, BiLSTM, a popular recurrent neural network which activates in both directions via three gates receives the outcome of the CNN layers generated from the input claim and evidences. The main role of BiLSTM in our proposed model is to obtain contextual and latent semantic-related financial fact-oriented feature representation in terms of forward and backward sequences in forward direction and backward directions, respectively. Equation 1 shows a fused

representation of BiLSTM functioning in parallel and opposite directions.

$$lstm_i = [\overrightarrow{lstm_f}, \overleftarrow{lstm_b}] \tag{1}$$

- **Attention layer**: In this study, three attention layers are used. They play a crucial by assigning different weights to different tokens according to the factual information present in the input claim and evidence sequences. Firstly, it obtains BiLSTM output from the corresponding input layers. It supervises the model to receive important relevant contextual information efficiently. As a result, it gives three context vectors as c_a, e_a, and v_a for claim attention, evidence attention, and visual attention, respectively.

Fused-attention layer: This layer takes the context vectors generated from the evidence attention, e_a and visual attention, v_a and fused it together with an aim to generate relevant comprehensive contextual finance-based factual information. As a result, this layer gives a fused-context vector, f_a.

Contrastive loss: The contrastive role in our proposed model takes two context vectors. One which is generated from claim attention c_a, and another one is fused-context vector, f_a. It captures the representations of input claim and both input evidences such that it measures the factual similarity and dissimilarity between them. It generates a contrastive vector, c_v as output and passes it to the next layer.

Dense layers and Classification: A common dense layer is used in our proposed model which receives the contrastive vector, c_v. Thereafter, `Sigmoid`, a non-activation function is used which is responsible for final classification/verification of the input claim as either *supported* or *refuted*.

4 Experimental Setup and Results

In this study, both benchmark datasets are divided as 70% for training, 10% for validation, and 20% of testing for empirical evaluation of our proposed model. We use `Intel` processing machine, `Ubuntu` OS, 64-GB RAM, and `NVIDIA` GPU. Our model is developed via `PyTorch`,[1] a machine learning framework in `Python`. We use $1e-5$ learning rate, binary cross-entropy, 100 batch-size, 64 padding-size, `adam` optimizer, 25 epochs, 128 neurons in BiLSTM, and 128 filters in image evidence-based CNN, and 64 filters in textual input-based claim and evidence-related CNNs.

[1] https://pytorch.org/.

Table 1 Performance evaluation results on two benchmark datasets in terms of *F-score* and *Accuracy*

Datasets →	Fin-Fact [14]		Mocheg [15]	
Methods ↓	F-Score	Accuracy	F-Score	Accuracy
Our Proposed model	**0.90**	**0.92**	**0.88**	**0.90**
Padma et al. [5]	0.85	0.84	0.83	0.81
Karadzhov et al. [13]	0.80	0.78	0.81	0.80
CNN	0.75	0.72	0.70	0.65
BiLSTM	0.72	0.71	0.74	0.70
RoBERTa	0.84	0.82	0.80	0.81
RoBERTa+CNN	0.78	0.76	0.72	0.80
RoBERTa+BiLSTM	0.77	0.75	0.71	0.80

4.1 Results and Comparative Analysis

In this section, we discuss the received results of our proposed model and compare it with existing studies and five comparable baseline methods. Table 1 shows the performance evaluation results on both benchmark datasets. Observe that, our proposed MuDal-FinFaCk model shows remarkable *F-score* and *Accuracy* on both datasets. The proposed model receives *F-score* of 0.90 and *Accuracy* of 0.92 on Fin-Fact [14] dataset.

Further, observe that, the proposed model outperforms existing studies and baseline methods. In existing studies, Padma et al. [5] show better results on both datasets. The proposed model shows 0.05% and 0.095% better *F-score* and *Accuracy*, respectively, than Padma et al. [5] work on Fin-Fact dataset. In baseline models, RoBERTa performs better on both datasets. The proposed model shows 0.071% and 0.12% better *F-score* and *Accuracy*, respectively than RoBERTa on Fin-Fact dataset.

These results give a remarkable indication that our proposed model is quite efficient to perform significantly on multi-modal data. It shows that consideration of the context-aware relevant embeddings and fusion of evidence-based attention layers contribute an important role in enhancing the model performance for financial fact verification.

5 Conclusion and Future Works

This paper has presented a multi-modal approach for financial fact-check verification of claims. To this end, a new model called MuDal-FinFaCk is introduced. It takes textual and image evidences for a particular piece of input financial textual claim which is also a kind of limitation of this study. The empirical evaluation of our proposed model is conducted on two benchmark datasets. It has received magnificent

results and also outperforms the comparable methods. The consideration of this study on multi-lingual data, especially in low-resource language setting could be a worth research exploration from future perspectives. Also, other multi-modal data sources can be considered in future works.

References

1. Kamal A, Mohankumar P, Singh VK (2022) IMFinE: an integrated BERT-CNN-BiGRU model for mental health detection in financial context on textual data. In: Proceedings of the 19th international conference on natural language processing (ICON). ACL, IIIT Delhi, India, pp 139–148
2. Kamal A, Abulaish M, Jahiruddin, (2024) Contextualized satire detection in short texts using deep learning techniques. J Web Eng 23(1):27–52
3. Kamal A, Anwar T, Sejwal VK, Fazil M (2023) BiCapsHate: attention to the linguistic context of hate via bidirectional capsules and hatebase. IEEE Trans Comput Soc Syst (TCSS) 11(2):1781–1792
4. Faizi SAA, Singh NK, Kamal A, Raza K (2024) Generative adversarial networks in protein and ligand structure generation: a case study. Deep learning applications in translational bioinformatics. Academic Press, Elsevier, pp 231–248
5. Mohankumar P, Kamal A, Singh VK, Satish A (2023) Financial fake news detection via context-aware embedding and sequential representation using cross-joint networks. In: Proceedings of the 15th international conference on communication systems and networks (COMSNETS). IEEE, Bengaluru, India, pp 780–784
6. Singh VK, Mohankumar P, Kamal A (2023) Fin-STance: a novel deep learning-based multi-task model for detecting financial stance and sentiment. 2023 14th international conference on computing communication and networking technologies (ICCCNT). IEEE, IIT Delhi, India, pp 1–6
7. Kamal A (2021) A unified data mining approach for detecting figurative language in Twitter. https://shodhganga.inflibnet.ac.in/handle/10603/441185
8. Kamal A, Mohankumar P, Singh VK (2023) Financial misinformation detection via RoBERTa and multi-channel networks. International conference on pattern recognition and machine intelligence (PReMI). Springer Nature Switzerland, ISI Kolkata, Cham, pp 646–653
9. Hassan N, Li C, Tremayne M (2015) Detecting check-worthy factual claims in presidential debates. In: Proceedings of the CIKM, Melbourne, Australia, pp 1835–1838
10. Li Y, Gao J, Meng C, Li Q, Su L, Zhao B, Fan W, Han J (2016) A survey on truth discovery. ACM SIGKDD Explorat Newsletter 17(2):1–16
11. Ciampaglia GL, Shiralkar P, Rocha LM, Bollen J, Menczer F, Flammini A (2015) Computational fact checking from knowledge networks. PloS one 10(6):e0128193
12. Roy A, Ekbal A (2021) Mulcob-mulfav: multimodal content based multilingual fact verification. In: Proceedings of the IJCNN, IEEE, pp 1–8
13. Karadzhov G, Nakov P, Màrquez L, Barrón-Cedeño A, Koychev I (2017) Fully automated fact checking using external sources. In: Proceedings of the RANLP, Varna, Bulgaria, pp 344–353
14. Rangapur A, Wang H, Shu K (2023) Fin-fact: a benchmark dataset for multimodal financial fact checking and explanation generation. ArXiv e-prints, pp 1–8
15. Yao BM, Shah A, Sun L, Cho JH, Huang L (2023) End-to-end multimodal fact-checking and explanation generation: a challenging dataset and models. In: Proceedings of the 46th international ACM SIGIR conference on research and development in information retrieval, Washington D.C, USA, pp 2733–2743

16. Devlin J, Chang MW, Lee K, Toutanova K (218) BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the NAACL-HLT, New Orleans, Louisiana, pp 4171–4186
17. Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. In: Proceedings of the ICLR, San Diego, CA, USA, pp 1–14