CryptoOpinionMining: A Comparative Analysis of Hierarchical and Flat Classification Models for Social Media Content

Rudra Roy[1], Pritam Pal[2], Rishika Jha[3] and Dipankar Das[4]

[1]*Indian Institute of Engineering Science and Technology, Shibpur, Howrah, West Bengal, 711103*
[2]*RCC Institute of Information Technology, Kolkata, West Bengal, 700015*
[3]*Institute of Engineering and Management, Kolkata, West Bengal, 700091*
[4]*Jadavpur University, Kolkata, West Bengal, 700032*

Abstract

With the recent popularity of cryptocurrency, social media such as Twitter and Reddit have become the main epicentre for cryptocurrency enthusiasts to share opinions and discussions about cryptocurrency. This paper presents a three-level hierarchical and single-level flat classification framework that can classify cryptocurrency-related unstructured social media opinions by leveraging the power of pre-trained transformer-based models. Along with opinion classification, we also proposed a question-answering framework utilizing the BiLSTM model, that can identify the most relevant social media comment regarding a cryptocurrency-related question from social media. By training the models with training data provided by FIRE 2024 CryptoQA shared task organizers and evaluating the models with test data, our best-performing opinion classification framework achieved a macro F1 score of 0.778 and 0.542 for Twitter and Reddit test data respectively which is also the best-performing framework in 'FIRE 2024 CryptoQA Task 1: Opinion Classification from CryptoCurrency related Social Media Posts'.

Keywords

Opinion Classification, Hierarchical Classification, Question Answering, Social Media Analysis

1. Introduction

In the past few years, social networks have become one of the most important sources of information within society and are also helpful in sharing opinions, and news in various fields. An example of a field that has recently been actively discussed on social networks is cryptocurrency. The recently developed digital currencies that include Bitcoin, Ethereum, and several other altcoins have sparked significant discussions, controversies, and expectations on social media platforms. Cryptocurrencies are also highly volatile and their value can change in the blink of an eye with reference to public opinion. This has therefore created a very interesting factor that has not been easy to handle when doing natural language processing (NLP) and sentiment analysis. The discussions related to cryptocurrencies on social media contain technical terms, followed by self-referencing factors and dynamic narratives, unlike traditional financial markets where sentiment analysis has already been done.

The specific problem we address in this research is the accurate understanding of opinions expressed in cryptocurrency-related social media posts. This task is particularly challenging due to the diverse nature of these posts, which can range from objective price discussions to subjective predictions, from technical analyses to promotional content. We noticed that traditional language-sentiment analysis techniques do not suffice when it comes to analyzing complex cryptocurrency language.

Therefore, the main objective of the proposed study is to develop an accurate and efficient opinion classification model for social media posts regarding cryptocurrencies. We should be able to create a system that's capable of categorizing information that falls under noise, fact, opinion, and fine-graded

sentiment within the fact and opinion type and a system that can give efficient answers related to cryptocurrency.

Along with the classification task, we focused on the different questions and doubts that arise among the potential crypto investors and to check whether the comment with respect to the query is relevant or not. The main contributions in this paper can be summarized as follows:

- We developed a novel three-level hierarchical classification framework for cryptocurrency-related social media posts utilizing a series of fine-tuned RoBERTa-base[1] models to progressively classify posts with increasing specificity.
- Next, we developed a single-level classification framework by employing one fine-tuned XLM-RoBERTa-base[2] model.
- Furthermore, we also developed a cryptocurrency question-answer framework by using BiLSTM[3].

In the following sections, we divide the whole task into two parts: one is 'Opinion classification on cryptocurrency-related social media posts' (Task 1) and another is 'Question answering from cryptocurrency-related social media posts' (Task 2).

2. Dataset

2.1. Task 1

This task utilizes two different datasets, consisting of cryptocurrency related Twitter and Reddit posts, both provided by FIRE (Forum for Information Retrieval Evaluation) 2024 CryptoQA[4] shared-task organizers. The datasets are created to be highly diverse in terms of opinions and types of content regarding cryptocurrencies and thus provide a sound basis for this multi-level classification task. There are 5000 opinions in the Reddit dataset and 4987 opinions in the Twitter dataset. The distribution of opinions in both datasets is given in Table 1.

2.1.1. Hierarchical Labeling Structure

Each of the entries in the two datasets was labeled based on a three-tiered hierarchical classification scheme, designed to capture the complexity of opinions regarding cryptocurrencies. The labelling structure is described in Figure 1.
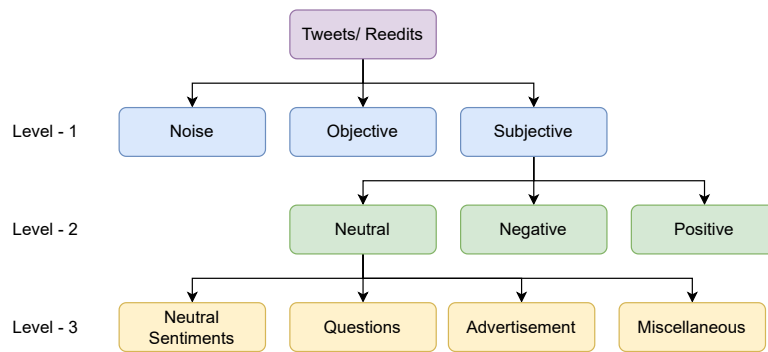

Figure 1: Hierarchical Labeling of Opinions

2.2. Task 2

The dataset for Task 2 was also provided by FIRE 2024 CryptoQA shared task organizers, which contains a total of 7932 questions with corresponding comments. The dataset annotated the relevancy of a

Classification Level	Category	Twitter Count	Reddit Count
Level 1	Noise	1553	645
	Objective	1689	503
	Subjective	1745	3852
Level 2	Neutral	1286	3183
	Negative	74	410
	Positive	385	259
Level 3	Neutral Sentiments	272	476
	Questions	279	2390
	Advertisements	676	105
	Miscellaneous	59	212

Table 1
Distribution of Opinions in Training Dataset.

comment for a specific question with 0 or 1, where 0 represents non-relevant and 1 represents relevant. The training dataset contains five labels-title, MAIN, selftext, comment_body and relevance. It has a total of 6787 non-relevant comments and 1145 relevant comments. The distribution of data is provided in Figure 2.
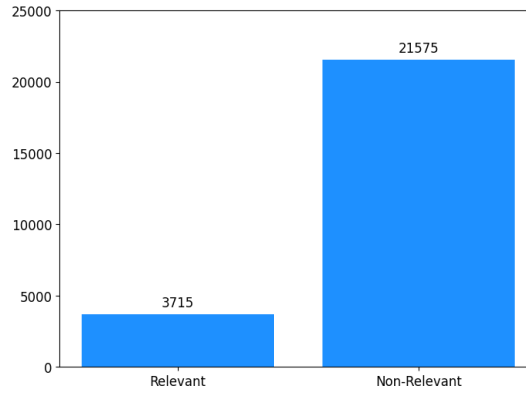

Figure 2: Distribution of relevancy label for question-answering task

3. Methodology

This section presents a comprehensive description of our approach to opinion classification in social media posts and question-answering related to cryptocurrencies such as text preprocessing, framework development, training etc.

3.1. Task 1

For Task 1, we present a multi-level classifier and a single-level classifier that is capable of categorizing these nuanced opinions about cryptocurrencies. We propose a methodology to tackle fine-grained opinion classification, particularly featuring novel method that cater the peculiarities of cryptocurrency discussions in social media platforms.

3.1.1. Text Preprocessing

In our approach to classifying cryptocurrency-related social media posts, we purposefully chose not to use conventional text processing methods. This decision was based on our experimental observations and the unique characteristics of the dataset. Our experiments revealed that any form of preprocessing, including common techniques such as lowercasing, punctuation removal, or stop word elimination,

negatively impacted the model's ability to identify the 'noise' class accurately. Consequently, we opted to use the raw text as input for our models. This approach preserves all original features of the posts, including capitalization, punctuation, and special characters.

3.1.2. Tokenization

Our research employed two classification frameworks: a hierarchical classification model and a single-level classification model. For both approaches, we utilized pre-trained models and their corresponding tokenizers from the Hugging Face Transformers library.

For the three-level hierarchical model, we used the RoBERTa base model and its tokenizer and for the single-level approach, we employed the XLM-RoBERTa (XML-R) model and its corresponding tokenizer.

For both frameworks, the tokenized sequences and their attention masks were used as input for the respective classification models. This dual approach allowed us to compare the effectiveness of hierarchical versus single-level classification for cryptocurrency-related social media posts.

3.1.3. Proposed Framework

We developed two frameworks to accomplish the hierarchical classification framework, leveraging the power of the transformer-based RoBERTa base model and XLM-RoBERTa base.

The RoBERTa-base model is a transformer-based architecture with 12 encoder layers, 768 hidden units, and 12 attention heads, totalling 125 million parameters. It uses a vocabulary of 50,265 tokens and supports a maximum sequence length of 512 tokens. Unlike BERT, RoBERTa was trained on a larger and more diverse dataset (160GB of text from Common Crawl), with dynamic masking and without the Next Sentence Prediction (NSP) objective. It was optimized with large batch sizes and a higher learning rate, resulting in improved performance across various NLP tasks such as text classification, named entity recognition, and question answering.

On the other hand, the XLM-RoBERTa is an improved version of XLM that builds upon the RoBERTa model and is trained in 100 languages. The overall model framework is depicted in Figure 3.

3-level classification framework: The flow diagram for the 3-level hierarchical classification framework is provided in Figure 3(a) where we employed separate 'RoBERTa-base' models for each level of classification, leveraging its pre-trained knowledge for effective transfer learning in the domain of cryptocurrency-related social media content.

1. **Level 1 (Model 1)**: Classifies posts into Noise, Objective, and Subjective categories.
2. **Level 2 (Model 2)**: Classifies posts into Neutral, Negative, and Positive sentiments.
3. **Level 3 (Model 3)**: Classifies posts into Neutral-sentiment, Questions, Advertisements, and Miscellaneous categories.

Single-level classification framework: In this scheme, we first convert the hierarchical annotation to a single-level annotation: Noise, Objective, Negative, Positive, Neutral Sentiments, Questions, Advertisements, and Miscellaneous. Then, we employed one 'XLM-RoBERTa-base' model to classify the posts into one of the 8 labels. Figure 3(b) provides the overall flow diagram for the single-level classification framework.

3.1.4. Training

For 3-level classification, the training process involves training of three independent models, each trained on specific subsets of the data:

1. **Model 1 Training**:
    - Input: The entire training dataset.
    - Output: Classification into Noise, Objective, and Subjective categories.
2. **Model 2 Training**:
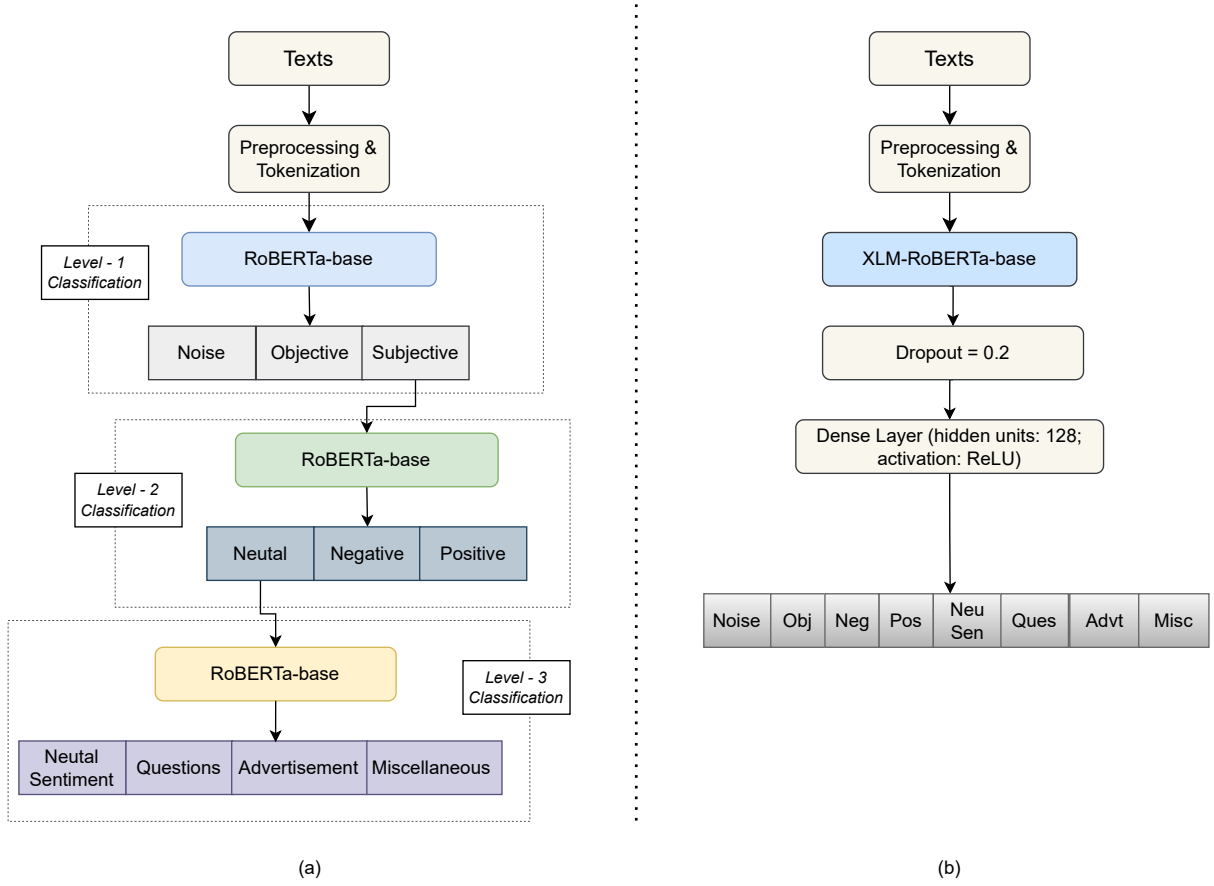
Figure 3: Proposed frameworks. (a) 3-level classification with three RoBERTa models; (b) single-level classification with one XLM-RoBERTa model

- Input: Posts labelled as "subjective" in the training dataset.
- Output: Classification into Neutral, Negative, and Positive sentiments.

3. **Model 3 Training**:

- Input: Posts labelled as "neutral" in the training dataset.
- Output: Fine-grained classification into Neutral-sentiment, Questions, Advertisements, and Miscellaneous categories.

In the case of single-level classification, we provided the entire training dataset as an input and classified the output into Noise, Objective, Negative, Positive, Neutral Sentiments, Questions, Advertisements, and Miscellaneous categories. For both 3-level and single-level frameworks the output layers used softmax as their activation function.

As no validation split was provided in the 'CryptoQA' shared task data, we randomly split the train data into a 4:1 ratio, where 80% of data was taken as training split and 20% of data was considered as validation split to examine the performances at the time of training.

3.1.5. Tuning Parameter

At the time of the training process, we fine-tune the parameters of the RoBERTa and XLM-RoBERTa models to achieve the best result possible in the validation dataset. For the 3-level hierarchical framework, the models were trained up to 3 epochs with a batch size of 16. The learning rate was taken as 2e-5 with AdamW[5] optimizer and weight decay of 0.01.

In the case of the single-level classification framework, the proposed framework was trained up to 5 epochs with a batch size of 16 and the learning rate was taken as 2e-5 with Adam[6] optimizer.

3.2. Task 2

Regarding Task 2, we developed a framework to accomplish the question-answering task using the BiLSTM model. As mentioned in Section 2.2, the comments for questions were annotated with relevant and non-relevant in the original dataset, thus we treat this task as a relevancy classification task where we aim to classify a comment as relevant or non-relevant.

The BiLSTM has long-term memory which allows memorizing important information from long sequences of text. Also, the bidirectional capability enables the model to capture both past and future contexts within a sequence or text, enabling the model to better language understanding than unidirectional LSTMs.
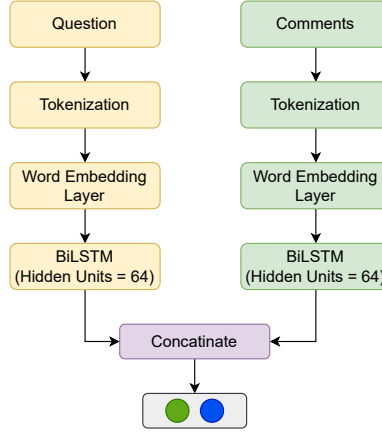

Figure 4: Proposed framework for question answering task in cryptocurrency-related social media posts. (Task 2)

The overall model flow diagram is depicted in Figure 4 where we passed the word embedding representation of the tokenized sequence of questions and the corresponding comments into two separate BiLSTM layers with 64 hidden units as a key input. Next, the output of two BiLSTM layers was concatenated and passed to the final output layer to identify whether the comment was relevant to the question or not. The output layer used two hidden units with softmax as an activation function.

To train the proposed framework, we split the training data as same as discussed in 3.1.4. We selected CategoricalCrossentropy as a loss function with a learning rate of 0.001. The optimizer was taken as Adam, and the model was trained up to 5 epochs with 128 batch size.

4. Evaluation and Result

This section presents the evaluation and results of our opinion classification task (Task 1) and question-answering task (Task 2) for cryptocurrency-related social media posts. We conducted extensive testing on a validation dataset to assess the framework's effectiveness. We provide a detailed account of the prediction process and discuss the outcomes observed on the test datasets provided by FIRE 2024 CryptoQA task organizers.

4.1. Prediction

Three-level hierarchical framework: The prediction process followed a hierarchical structure, mirroring our three-level training architecture:

1. **Level 1 Classification:** All posts in the test dataset were initially processed by our first-level model, which categorized them into three classes: Noise, Objective, and Subjective.
2. **Level 2 Classification:** Posts classified as "Subjective" by the first model were then fed into the second-level model. This model further categorized these posts into Neutral, Negative, or Positive sentiments.

3. **Level 3 Classification:** Posts classified as "Neutral" by the second model underwent a final classification by the third-level model. This model provided fine-grained categorization into Neutral-sentiment, Questions, Advertisements, and Miscellaneous.

Single-level classification framework: In the case of single-level classification, the prediction strategy was relatively simple where all posts in the test dataset were passed as input to the framework and categorized into eight classes: Noise, Objective, Negative, Positive, Neutral Sentiments, Questions, Advertisements, and Miscellaneous. The results for these two approaches are provided in Table 2.

Run file Name	Summary	Twitter Data (Macro-f1)	Reddit Data (Macro-f1)
run-1.csv	Hierarchical Framework	0.725	0.518
run-2.csv	Single-level Framework	0.778	0.542

Table 2
Results for proposed two schemes of the opinion classification framework for cryptocurrency-related social media posts (All the reported scores are macro F1 score and the performances were evaluated by shared task organizers)

4.2. Result

From Table 2 it is observed that the single-level classification framework achieved the best result with F1 scores of 0.778 and 0.542 for Twitter and Reddit test data respectively. However, the performance of the hierarchical classification framework slightly downgraded to 6.94% and 2.4% for Twitter and Reddit data respectively compared to the single-level framework. In addition, our proposed single-level framework achieved the best score than the performances of other shared task participants in 'FIRE 2024 CryptoQA Task 1: Opinion Classification from CryptoCurrency related Social Media Posts'.

Regarding Task 2, our proposed framework didn't perform well and provided a result of 0.1465[1](Macro-f1).

5. Conclusion and Future Work

In this paper, we presented a novel hierarchical approach and single-level approach to classifying cryptocurrency-related social media posts. Our proposed three-level hierarchical approach and single-level approach performed well and among them, the single-level approach provided the best result in classifying cryptocurrency-related social media posts.

While our model shows promising result, there are areas for further improvement and research. Future Work could focus firstly on incorporating more advanced natural language processing techniques to better capture context and nuance. Secondly on addressing potential error propagation through the hierarchical levels. These advancements may improve our categorization accuracy even further.

Declaration on Generative AI

The authors do not use any Generative AI tool.

References

[1] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, Roberta: A robustly optimized BERT pretraining approach, CoRR abs/1907.11692 (2019). URL: http://arxiv.org/abs/1907.11692. arXiv:1907.11692.

---

[1]Evaluated by shared task organizers.

[2] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, V. Stoyanov, Unsupervised cross-lingual representation learning at scale, CoRR abs/1911.02116 (2019). URL: http://arxiv.org/abs/1911.02116. arXiv:1911.02116.

[3] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (1997) 1735–1780. URL: https://doi.org/10.1162/neco.1997.9.8.1735. doi:10.1162/neco.1997.9.8.1735.

[4] K. R. K. G. Sougata Sarkar, Gourav Sen, Understanding cryptocurrency related opinions and questions from social media posts (cryptoqa 2024), in: Forum of information retrieval (FIRE), 2024. URL: https://sites.google.com/view/cryptoqa-2024.

[5] I. Loshchilov, F. Hutter, Decoupled weight decay regularization, 2019. URL: https://arxiv.org/abs/1711.05101. arXiv:1711.05101.

[6] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2017. URL: https://arxiv.org/abs/1412.6980. arXiv:1412.6980.