Sentiment Analysis for Cryptocurrency in Social Media

Marcos Geraldo
Thomas Claiborne

April 17, 2025

Introduction

- We are interested in detecting the **"financial sentiment"** of social media with respect to **Cryptocurrencies**.

- Previous work has solved the problem of sentiment in social media, but the crypto financial angle is under-investigated

- We use **LoRA** to find an **effective** and **efficient** way to fine-tune a pre-trained model for this purpose.

Methodology
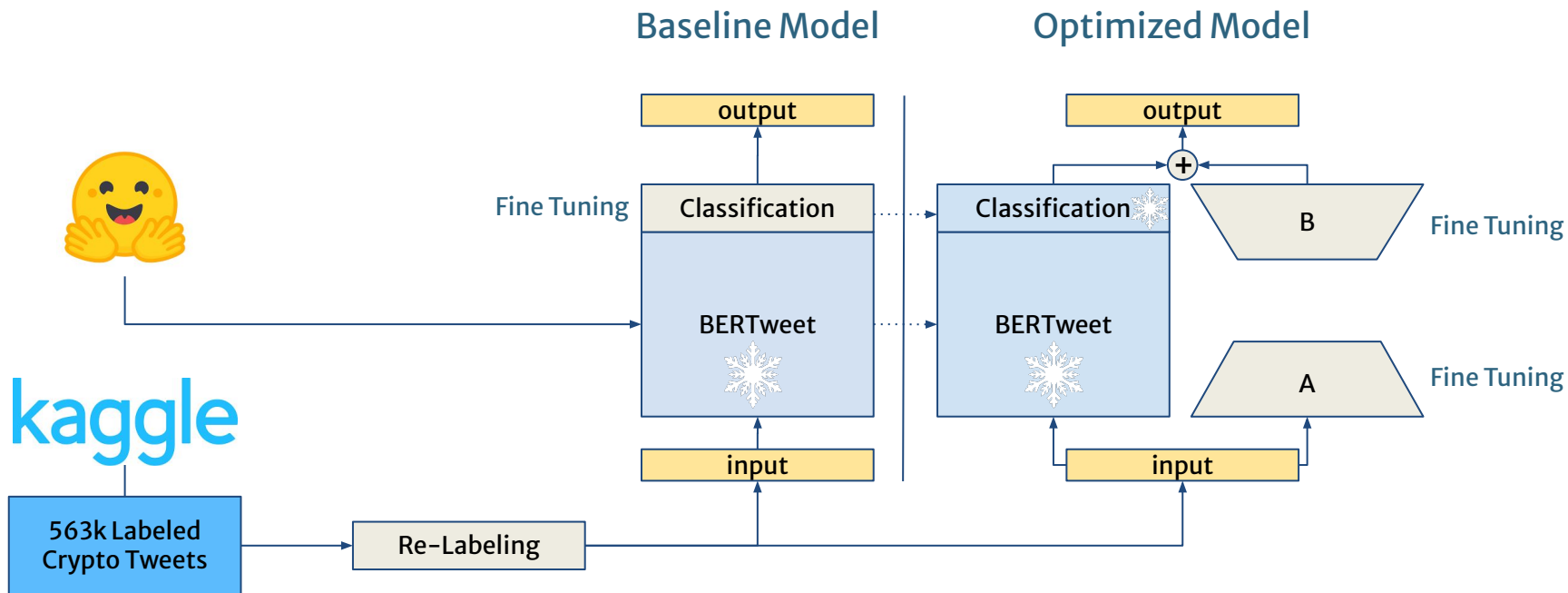
Label Generation

Lorem ipsum dolor

- Used high absolute "Vader" composite score

- Key Words

- Excessive use of emojis

- Joke Markers

- "Nonsense" Markers

- 3 level sentiment indicator

- The contingent at . Special thanks to the hackers for all the tooling feedback we'll be taking back to Brooklyn! 🙏 🙏 🙏

- Wow! I'm only 100 subscribers away from reaching 2,000 on YouTube. Please help me reach that milestone. â□¤ï□ I make videos about bitcoin, politics and self-empowerment. Subscribe here:

- this is why fooo is biggity BAWLS deep into ethereum it is the infrastructure upon which is built so much of this space it also is like dry tinder to the flames of idealism VERY sexy narratives surround eth it is a near-perfect story, and one that lends to tribalism well

- Someday this box will be full of !!! I love it so much! 🚀 🌕 📈

- A kindhearted light enlightened me to Brave Rewards for content creators. So if you install Brave (the web-browser); use the link below, that way I get some crypto-currency offerings from the Machine-God! I use Brave instead of Chrome.

Architecture

Baseline Model

Layer (type: depth – idx)	Param #
RobertaForSequenceClassification	—
└─Roberta Embeddings: 2-1	—
└─Embedding: 3-1	51,471,360
└─Embedding: 3-2	526,336
└─Embedding: 3-2	1,024
└─Embedding: 3-4	2,048
└─Dropout: 3-5	—
└─RoBERTa encoder: 2-2	—
└─ModuleList: 3-6	302,309,376
Roberta Classification Head: 1-2	
└─Linear: 2-3	1,049,600
└─Dropout: 2-4	—
└─Linear: 2-5	3,075
Total params:	355,362,819
Trainable params:	**1,052,675**
Non-trainable params:	354,310,144

LoRA Model

Layer (type: depth – idx)	Param #
PeftModelForSequenceClassification	—
└─LoraModel:1-1	—
└─Roberta For Sequence Classification: 2-1	--
└─RobertaModel: 3-1	354,703,360
└─ModulesToSaveWrapper: 3-2	2,105,350
Total params: 356,808,710	
Trainable params: 1,445,891	
Non-trainable params: 355,362,819	

Fine Tuning

BERTweet Data Download
- 563,799 rows
- Sentiment score

Relabel
- 563,799 rows
- Imbalanced classes
- Financial sentiment label

Sampling
- 15,000 rows
- Balanced classes
- Financial sentiment label

Fine tuning

Base Model Fine Tuning
- Freeze all RoBERTa layers
- Train the classification layer
- 1,052,674 parameters trained
- 5 epochs

LoRA Config
- Attach Lora to base model
- Freeze base model

LoRA training
- Train LoRA parameters
- 1,445,891 parameters trained
- 5 epochs

Optimization
- LoRA Hyperparameter optimization

Hyperparameter Optimization

id	Rank	Alpha	Accuracy
1	8	16	0.678
2	8	32	0.696
3	8	64	0.727
4	4	32	0.696
5	4	128	0.756
6	4	64	0.732
7	16	32	0.694
8	16	16	0.671
9	16	64	0.727
10	8	256	0.785
11	**4**	**512**	**0.807***
Baseline model			0.444

Epochs: 5
Dropout: 0.1
Learning Rate: 1e-5
Batch Size: 16
Decay: 0.01

Accuracy vs LoRA Scaling Factor
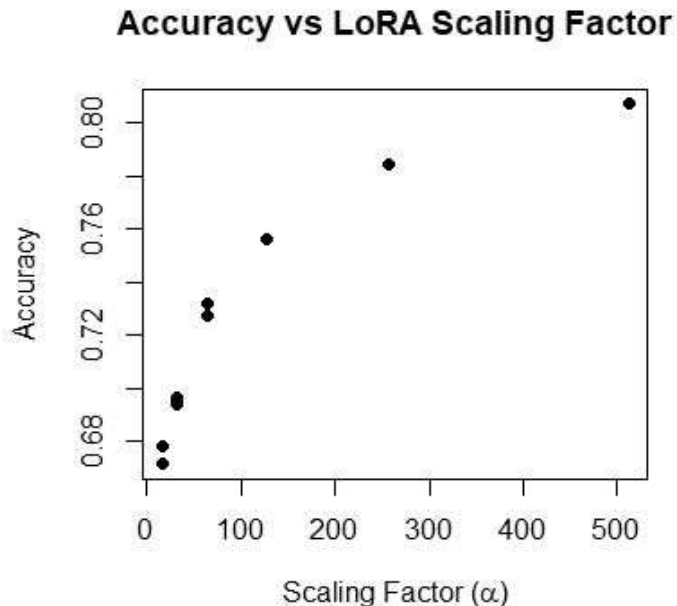
Results

- Evaluation with 100,000 tweets, label ratio the same as overall dataset
- The LoRA model achieves higher performance across all metrics.
- The best performance was achieved with scaling factor of 512 and rank 4.
- The LoRA model shows False positives in the labels Negative and Positive affecting its precision, but strong recall.

Baseline Model

Label	Precision	Recall	F1	Support
Neutral	0.72	0.56	0.63	62,437
Negative	0.18	0.40	0.25	9,581
Positive	0.38	0.42	0.40	27,982
Accuracy			**0.51**	100,000

Optimized Model

Label	Precision	Recall	F1	Support
Neutral	0.96	0.69	0.81	62,437
Negative	0.43	0.88	0.58	9,581
Positive	0.66	0.84	0.74	27,982
Accuracy			**0.75**	100,000

Conclusion

- Significant improvement over baseline model

  - Benefits of Lora

- Potential implications for crypto financing

  - Sentiment insights

- Limitations

  - Relying on previously labelled data

  - Internet full of biased nonsense

- Future Directions

  - Beyond Twitter

  - Financial history

Thank You