Cyclical Pre-Training for Cryptocurrency Price Predictions

Stanford CS224N Custom Project

Andrew Simpson
Department of Computer Science
Stanford University
asimpso8@stanford.edu

Abstract

Cryptocurrency markets are known for their extreme volatility. News on different cryptocurrencies may send prices soaring or falling in very short period of time. Knowing how to use this information is a form of temporal sentiment analysis with wider applications. In this paper multiple approaches were attempted in order to predict future prices. Using BERT with bidirectional LSTMs with and without attention was found to be ineffective providing results no better than a coin flip. Using a cyclical pre-training approach with GPT by pre-training on the news of the day followed by fine-tuning on past questions, asking future questions, and then fine-tuning on both past and future questions did much better than chance. The overall accuracy of this approach was found to be 55%, or 5% above guessing if the price will increase or decrease during random times in the next 24 hour period.

1 Key Information to include

- Mentor: Anna Yang
- External Collaborators (if you have any): None
- Sharing project: No

2 Introduction

Cryptocurrency such as Bitcoin [1] use a decentralized network for transaction authentication, which means that no one party controls a cryptocurrency. Over the last few years other cryptocurrency such as Ethereum have offered additional features to allow for smart contracts and decentralized applications [2]. The underlying technology of the Block chain has many uses outside a repository of value such as healthcare records, supply chain management etc. [3].

Over the past few years, there has been enormous interest in cryptocurrencies. Since cryptocurrencies such as Bitcoin have no intrinsic value, their value is only what people believe it is. Since cryptocurrencies are a new technology many people are unsure of their value and applications. This uncertainty creates an extremely volatile market. Prices for certain cryptocurrencies may swing double digit percentages in a very short time. An example of this volatility can be seen in Figure 1.

News articles, Tweets, etc. may send a certain coin rapidly in one direction. Interpreting this data is a form of sentiment analysis where a temporal aspect in introduced. Natural language processing techniques (NLP) have been researched for the general case of sentiment analysis for many years. In this paper different approaches using different the different yet related tasks of sentiment analysis and question answering will be addressed to answer the fundamental question of is it possible to use news headlines to predict future cryptocurrency prices.

Figure 1: Example of Bitcoin prices in a 24 hour period.

3 Related Work

Cai et al. [4] used a method of taking the final output layer of BERT (Bidirectional Encoder Representations from Transformers see [5]) to pass into a hidden layer to be used by an input to an LSTM network to predict sentiment of investors and consumers for an energy market. This was done using a database of approximately 100,000 Chinese investor and consumer comments. The authors showed that using BERT with bidirectional LSTMs substantially outperformed BERT and bidirectional LSTMs on their own in terms of recall, accuracy, precision, and the F1 score.

One of the issues that BERT has is that ability to perform on tasks involving numeracy. Wallace et al. explored this issue further [6]. It was shown that word vector representations such as GLoVe [7] or word2vec [8] showed a reasonable degree of numeracy on certain numeric data sets. Character level embeddings such as in ELMo [9] had the best overall numeracy abilities.

There are huge bodies of text available, but only a few have labels useful for machine learning tasks. One of the most important breakthroughs in NLP over the last few years has been the use of pre-training on unlabeled data. This pre-training has been found to greatly improve downstream tasks such as question answering with a common approach being masking parts of a sentence and predicting the missing parts [10] [11]. After pre-training, a fine tuning task such as question answering is performed that shows great potential in the quality of the results [12].

The quality of results shown from pre-training shows that transformer models are being used as a knowledge base. Roberts et al. showed that fine-tuning of pre-trained models can answer questions without access to any external context or knowledge [13]. This means that transformer models can answer questions they have never seen in the fine-tuning process and are using the parameters of the network to store information and make inference from pre-training corpus. Knowing this and the previous information it seems possible that a transformer model could be used to incorporate current information about cryptocurrencies and make an informed prediction about future prices.

4 Approach

The first baseline approach was to use an approach similar to Cai et al. [4] by using a pre-trained version of BERT and using the last hidden state, in this case just the [CLS] hidden state. Each news headline and price sentence (data set described in next section) were passed into BERT to get vectors to pass to an bidirectional LSTM encoder whose final hidden state was used for the LSTM decoder as see in Figure 2. The decoder bidirectional LSTM network takes as input a question about the price of cryptocurrency such as "Did the price of Bitcoin increase or decrease 0 days 1 hour and 7 minutes ago?" The final output of the decoder network was then passed to a stack of fully connected layers to predict an increase or a decrease.[1]

[1]There was an additional experiment using BERT trained on reviews specifically for sentiment analysis. The quality of the results produced on news headlines were not satisfactory with most, even obviously positive, headlines producing low scores. This approach was dropped in favor of cyclical GPT pre-training.

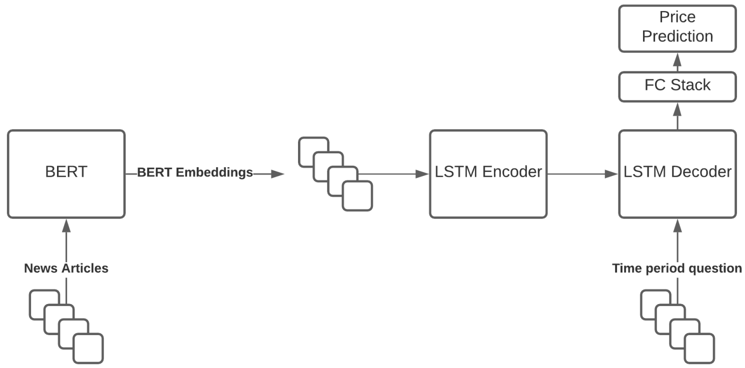


Figure 2: BERT LSTM encoder-decoder architecture

The baseline approach was further enhanced by adding an attention mechanism between the LSTM encoder outputs and the LSTM decoder. This approach used a multiplicative attention mechanism. This approach is similar to that used by Luong et al. [14]. This mechanism helps to focus on which news article headlines and price sentences are the most relevant.

Since the previously mentioned approaches required multiple passes through BERT, which is a large model, only a very small batch size could be used and the number of news headlines and prices sentences needed to be truncated to only the last 25 sentences. Another approach was to use the Longformer, a transformer that can take as sequences of up to length 4096, to use a substantially larger and in many cases the entire 24 hour period of information. The last hidden state of the first [CLS] token was used, and it was passed to a hidden layer and a leaky ReLU (set to 0.2) that transformed it to the size of the hidden state of the LSTM decoder as can be seen in Equation 1.

$$a_i = LReLU(W_a L_i + b_a) \tag{1}$$

Since character level embeddings were found to have the best numeracy and that transformers can be used as a knowledge base to answer never seen questions, the main novel approach is to use a cyclical form of pre-training where a transformer model is pre-trained on a 24 hour period of news and price information using span corruption, then fine-tuned on only past questions, calculate the accuracy of future questions, fine-tune on both past and future questions, pick a new random 24 hour period, and repeat as shown in Figure 3.

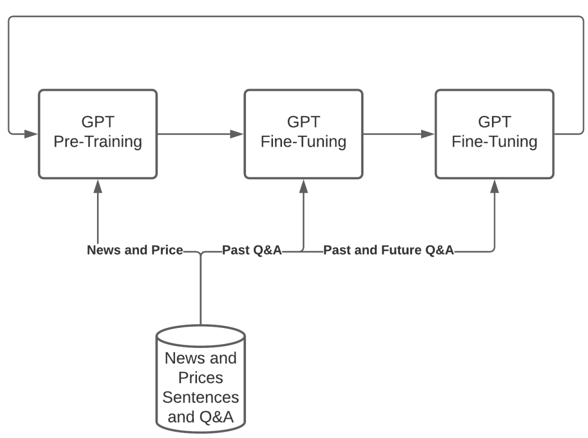


Figure 3: Cyclical Pre-training with GPT

In order to attempt this approach a transformer model that can be pre-trained rapidly was required. The total length of news articles and price sentences per day is limited and often only a few thousand total words. Large models such as BERT take large computational resources and time to pre-train effectively. A smaller model based around GPT-1 was used that can be pre-trained on a days worth of information in a much shorter period of time [10].

The data sets include multiple types of questions that require understanding of sentiment and numeracy such as asking how much a coin change between two points in time. GPT is a relatively small model compared to others such as BERT. In order to mitigate issues with complicated tasks a simplified version of the task was created asking simple questions of whether a coin increased or decreased in a positive (future) or negative (past) amount of time from the publish time of the last article in a 24 hour period (described in more detail in the next section).

Knowing what information is relevant to the model was also explored. The main data set used news information followed by price sentences intercalated between the next news information. Using combined data, news only, and prices only was examined by running the model with the same parameters and 24 hour period fixed random seed selection.

What can we learn from the past? The final approach explores using pre-training on similar days as well as the current 24 hour period. This approach first used a subset of the total data (1000 random 24 hour periods) to calculate a distance matrix between each 24 hour period. The similarity was calculated using a pre-trained version of BERT for sentence Embeddings [15]. Since BERT has a limited input size and each 24 hour period has multiple sentences, the approach was to calculate an embedding for each sentence in the 24 hour period and then calculate the centroid for each 24 hour period. A distance matrix was then constructed. The top 5 most similar 24 hour periods were stored for each 24 hour period and one was randomly selected at training time.

The training approach for similar news periods was as follows. First, select a random closest 24 hour period from the top 5 closest 24 hour periods. Second, combine the closest period with the current period for pre-training. Third, pre-train with only the current period. Fourth, fine-tune future questions about the closest 24 hour period. Fifth, fine-tune on past questions for the current day. Sixth, find the accuracy on the future questions of the current 24 hour period. Seventh, fine-tune on future questions for the current 24 hour period. Finally, pick a new 24 hour period and repeat.

5 Experiments

5.1 Data

The base data set consists of 134,828 news articles (examples can be seen in Table 1) between November 2018 and January 2021 related three popular cryptocurrencies Bitcoin, Ethereum, and Litecoin with influential terms such as regulation restriction, fear, attack, etc. or record explosive acceptance, etc. The data set also contains prices for each of the coins in five minute increments.

The base data set was used to generate 500,000 randomly selected periods of time in the data set. News articles from up to a day before were included along with prices of cryptocurrencies from up to a day before and a day after. From this day length data point, temporal questions about price changes and a specific cryptocurrency were generated. The a non exhaustive list of questions can be seen in Table 2 (complete version found in the appendix Table 6). This data set can be used for multiple tasks. The data set is designed be used for question answering, but can also be easily extended to be used for temporal sentiment analysis (increase or decrease in a period of time to label articles), temporal classification (increase, decrease, near zero, etc.) as used in the baseline as related experiments, or even temporal regression (L2 loss of price change at a point in time).

2020-01-16T16:35:26
Cryptocurrency Industry Debates Digital Dollar as Former Regulators Form Think Tank
Yesterday rumors circulated the internet that the US Federal Reserve will announce the creation of a cryptocurrency-like digital dollar.

Table 1: Example of news article data

In [TIME PERIOD] how much will [COIN] change?	[COIN] will change [PERCENT CHANGE] IN [TIME PERIOD]
How much will [COIN] change from [FUTURE TIME 1] to [FUTURE TIME 2]?	[COIN] will change [PERCENT CHANGE] between [FUTURE TIME 1] and [FUTURE TIME 2]
How much did [COIN] change from [PAST TIME 1] to [PAST TIME 2]?	[COIN] changed [PERCENT CHANGE] between [PAST TIME 1] to [PAST TIME 2]

Table 2: Example list of cryptocurrency questions.

Each news article started with a timestamp sentence such as "Article time Friday at 01 50 PM ." In between each news article price information was intercalated where the change of a coin's price between the time two news articles were published was also added added. A complete example can be seen in Figure 4.

```
Article time Friday at 03 11 PM .
ADA, LINK, SKY, XTZ, TRX: Crypto Price Analysis and Overview May 29
One similar feature among the cryptocurrencies is that at the overbought region, the coins are repelled and a sideways trend will follow.
The price of Litecoin changed negative zero point two six percent in 0 days 0 hours and 30 minutes between Friday at 03 10 PM and Friday at 03 40 PM .
Article time Friday at 03 38 PM .
SEC Establishes $25 Million Compensation Fund for BitClave ICO
The US Securities and Exchange Commission (SEC) recently settled with BitClave. The regulator has established a $25 million compensation fund to return
```

Figure 4: Cyclical Pre-training with GPT

Additionally, a simplified version of the questions was constructed using only one past and one future question type that differed in only one character as can be seen in Table 3. In this simplified case the decision to use only one type of past and future questions differing by only one character was used for two reasons. The first reason is that fine-tuning on the past question type will hopefully generate similar attention in the model. The second is that both the '-' and especially the '+' characters are not frequently found in the news and price data, so this improves the chances that pre-training will not completely wipe out any parameters directly or indirectly related to these two characters.

What will [COIN] do in - [TIME PERIOD] ?	[COIN] [increase/decrease] .
What will [COIN] do in + [TIME PERIOD] ?	[COIN] [increase/decrease] .

Table 3: Simplified past and future questions.

For pre-training a character level span corruption was used

5.2 Evaluation method

For BERT-BiLSTM, the metric used was accuracy on the output of the fully connected stack. For using GPT with question answering, the main metric used for future increase or decrease was a relaxed accuracy. For a human reader, having the exact match of "The price of Bitcoin will increase" conveys the same meaning as "Coin Positive." To create a relaxed version of accuracy, the generated string would need to match "inc" or "pos" for a price increase or "dec" or "neg" for a price decrease. In experimentation many times it was common to see the word "positive" or "negative" in the answer after pre-training as the corpus data contains those words. The relaxed accuracy metric over many different 24 hour periods is used to determine the overall success of the model and not the loss function. Since the data set is so large, this relaxed accuracy is also a stand in for a validation set as no 24 hour period was repeated (post pre-training and past question fine-tuning). Note that there is not a default value. If the model does not match the requirements it counts against the accuracy.[2]

5.3 Experimental details

For the BERT BiLSTM baselime, bidirectional LSTMs were used with a hidden size of 1024. The fully connected stack consisted of a power of two reduction from the hidden size until 32 with a final output of size two. The optimizer used was Adam with default parameters and a learning rate of 0.00005. Gradients were also clipped with clip gradient norm with a max norm of 0.01 in order to prevent exploding gradients. The baseline model was trained for approximately one day, and no

[2]For all GPT models, the accuracy is calculated off of 16 randomly generated future questions.

gradients were updated for the BERT model [3]. This same procedure was repeated with the baseline with attention.

The longformer approach used the same approach as the baseline although the AdamW optimizer was used. The maximum number of words for the news and price data for a 24 hour period was limited to 2500 words. The longformer is a large model and the batch size was set to two. This model was also trained for approximately one day.

Cyclical pre-training for the GPT model was first done by pre-training for an extended number, 600, of epochs on epochs on a randomly selected 24 hour period. The learning rate was set to 6×10^{-3}. The cyclical part of the training starts with selecting one 24 hour period and pre-training with a learning rate of 6×10^{-3} for 50 epochs. The next step is fine-tuning for 10 epochs [4] with a learning rate of 6×10^{-4}. After this step, accuracy on the future question of increase or decrease is calculated as previously mentioned. Finally fine-tuning on past and future questions is done with a learning rate of 6×10^{-4} for 10 epochs.

For the simplified GPT version the same parameters and number of epochs as before were used. The only difference is that the questions are now only two simplified types previously mentioned. For the news and prices experiment, the same setup was used, but the data set was modified to only have news headlines or only prices. The order of 24 hour period selection was fixed in all GPT methods.

The similarity experiment modifies the preceding GPT parameters. The approach was to first find a randomly selected top 5 closest 24 hour period to the current 24 hour period as previously described. The model then pre-trains on the combined news and price information for both 24 hour periods for half as many epochs. This was followed by only pre-training on the current 24 hour period for half the number epochs as well. Then the model was fine-tuned with the future question for the closest 24 hour period followed by only the past question for the current 24 hour period. Then accuracy on the future question was calculated, followed by finishing on the future question for the current 24 hour period.

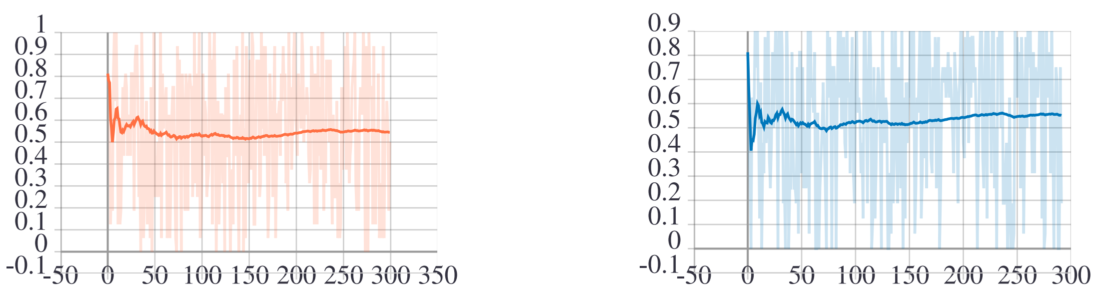
5.4 Results

BERT BiLSTM	**BERT BiLSTM + Attention**	**Cyclical GPT**	**Cyclical GPT (simple)**
50.00	50.05	49.08	55.38

Table 4: Percent accuracy comparison between models.

GPT	**GPT (news)**	**GPT (prices)**	**GPT (similarity)**
55.38	54.06	53.79	55.60

Table 5: Percent accuracy comparison GPT (simple) model experiments.



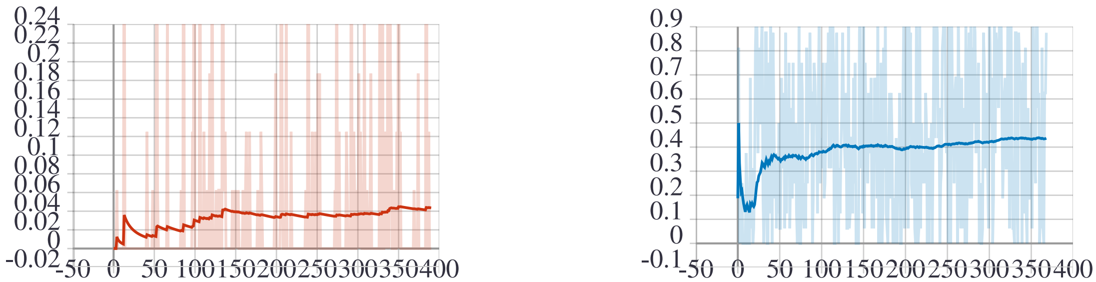
(a) Future question accuracy for simplified GPT.

(b) Future question accuracy for simplified GPT with simmilarity.

Figure 5: Future question answering accuracy for simplified GPT and simplified GPT with similarity.

[3] A small example was tested with gradients into BERT a batch size of two and a maximum of 25 sentences. This did not produce results better than a coin flip as well, but it does overfit on a small data set.

[4] For GPT, the questions are dynamically generated and set to a fixed epoch size of 2000.



(a) Future question accuracy after pre-training for new only. only.

(b) Future question accuracy after pre-training for prices only. only.

Figure 6: Future question answering accuracy directly following pre-training (no past question fine-tuning) on news only and prices only

6 Analysis

As can be seen in Table 4 both BERT with bidirectional LSTMs and BERT BiLSTM with attention did not perform well. The results are no better than a coin flip. This is not necessarily unexpected. BERT was pre-trained on Wikipedia and BookCorpus. This underlying knowledge, while useful for many areas of NLP, was not useful in helping predict future prices based on news headlines and price information of a 24 hour period. These models do over-fit a small data set indicating that they do learn; however, they aren't able to use news data in batches to predict future prices.

With the full set of future and past questions, the model was not able to break 50% accuracy (see Table 4). This may have been due to the fact that past and future questions did not have much overlap and used past and future tenses that the model may not have been able to infer that the past and future questions were asking the same thing just at different temporal points.

Cyclical pre-training with simplified questions proved to be the most powerful concept. The results in Table 4 show a very substantial increase in accuracy above chance using this approach. The goal of cyclical pre-training was to have the model retain the most pertinent data for a 24 hour period related to question answering and using the model as a knowledge base for past and future inference updated between 24 hour periods. Overall, using news and price information together was found to produce the highest accuracy. It appears using the information of both helps the model to more accurately predict future prices. Most humans would likely also agree that seeing the trends for the day and headlines would be important information to consider when buying.

One of the more interesting properties to emerge from multiple rounds of cyclical pre-training was how the network answers questions after the pre-training step. For the case of both news and prices, the network does not output sensible responses; however, there is evidence that the network still has concepts of these questions after a pre-training step as can be seen in Figure 7. Although the responses are not close to the correct response, when viewed together, it's clear that the network still understands the core concept of increase or decrease in the future based on its response. This likely means that the parameters of the are able to hold onto the questions between rounds especially with the '-' and '+' characters as this behavior was more rare in the non-simplified version.

Prices only had interesting emergent results that can be seen in Figure 6. In all cases except the prices only experiment, the future question accuracy after pre-training and before past question fine-tuning produces nonsense results as mentioned previously; however, in the prices only experiment, after a period of time, the model starts to produce high quality answers to future price questions. This surprising ability to regularly answer future questions with out fine-tuning might be caused by multiple reasons. One reason could be the limited and repetitive vocabulary used. Except for amounts and times, predictions are trivial as they are the same from price sentence to price sentence. This could have allowed the model to use fewer parameters for prediction while leaving others used for question answering largely untouched in pre-training.

Using similarity did not offer any substantial improvement. This may have been for many reasons, but it is possible that the centroid approach using BERT was not most optimal way to encode the similarities between different 24 hour periods. Using a whole document approach may have produced a higher overall accuracy. However, the similarity experiment did produce interesting results seen in Figure 5. The overall accuracy is about the same, but there is substantial reduction in times where the model was very wrong (picking almost the opposite of what will happen). It may be the case that using similarity helps to "hedge the bets" of the model based on what happened in the related day instead of going "all in" on what it believes.

```
What will Bitcoin do + 0 days 4 hours and 50 minutes ?□ Bitcoin decrease .  in the markets in the the biggg
What will Bitcoin do + 0 days 7 hours and 40 minutes ?□ Bitcoin decrease .  in the markets in the the biggg
What will Bitcoin do + 0 days 17 hours and 0 minutes ?□ Bitcoin increase .  the large for the worlds of the
What will Bitcoin do + 0 days 10 hours and 45 minutes ?□ Bitcoin decrease .  and the ar
What will Bitcoin do + 0 days 14 hours and 40 minutes ?□ Bitcoin increase .  and the arge for the worlds of
What will Bitcoin do + 0 days 4 hours and 20 minutes ?□ Bitcoin decrease .  the large for the worlds of the
What will Bitcoin do + 0 days 14 hours and 45 minutes ?□ Bitcoin increase .  and the large for the worlds of
What will Bitcoin do + 0 days 19 hours and 20 minutes ?□ Bitcoin increase .  and the argets to the biggggest
What will Bitcoin do + 0 days 3 hours and 50 minutes ?□ Bitcoin decrease .  in the markets in the the biggg
What will Bitcoin do + 0 days 10 hours and 40 minutes ?□ Bitcoin decrease .  and the ar
What will Bitcoin do + 0 days 5 hours and 5 minutes ?□ Bitcoin decrease .  the large for the worlds of the
What will Bitcoin do + 0 days 18 hours and 10 minutes ?□ Bitcoin increase .  and the argets to the biggggest
```

Figure 7: Question and answer after multiple rounds of cyclical pre-traning.

7 Conclusion

Overall, it was found that using cyclical pre-training with simplified questions produced the best results with over 55% accuracy in predicting future increases or decreases for a particular cryptocurrency. The baseline model, baseline with attention, and the longformer were not able to infer future prices given news and price information about a 24 hour period giving an accuracy of a coin flip. The cyclical pre-training GPT model with more advanced questions was also not able to infer future prices better than chance. Similarity was not found to have any substantial improvements in accuracy, but it did appear to mitigate some of the cases where the model was very wrong. A document level approach instead of centroid based approach for similarity could have produced better results as well. Some sort of attention similarity could be attempted as well where a certain number of existing news periods are always passed into a modified version of the model.

The main limitation of this work is the requirement for simplified questions. The ideal case would be to have the model also understand the basics of English grammar and understand the relation between past and future tenses. Future work in this area work includes training on larger and more sophisticated models. GPT is a relatively small model. Pre-training on a larger model such as ELMo over a longer period of time may have provided even better results. Additionally, the hyper-parameters used have additional optimization that can be attempted. One area of future work is to shoehorn in an L2 loss to improve numeracy. For the case of predicting percent change, an additional loss could be added by using the position of digits and passing them through an additional network to calculate the L2 loss directly. Using this approach for trading cryptocurrencies, the reward of making and losing money could also be introduced to incorporate a reinforcement learning based approach.

The key finding was that cyclical pre-training allows for a much deeper understanding numeracy and temporal information. This is an area that should be explored in other areas of NLP and may be important to helping models understand not only time but potentially cause and effect. Outside of predicting cryptocurrency prices this research has applications in many other fields. The most obvious is use in stock markets, but there are also uses in predicting future customer sentiment based on current social media posts. This approach could also be used to help predict election outcomes and most importantly could be used to predict disease outbreaks time and locations based on current behavior on social media.

References

[1] Craig S Wright. Bitcoin: A peer-to-peer electronic cash system. *Available at SSRN 3440802*, 2008.

[2] Vitalik Buterin et al. A next-generation smart contract and decentralized application platform. *white paper*, 3(37), 2014.

[3] Shiroq Al-Megren, Shada Alsalamah, Lina Altoaimy, Hessah Alsalamah, Leili Soltanisehat, Emad Almutairi, et al. Blockchain use cases in digital sectors: A review of the literature. In *2018*

IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), pages 1417–1424. IEEE, 2018.

[4] R. Cai, B. Qin, Y. Chen, L. Zhang, R. Yang, S. Chen, and W. Wang. Sentiment analysis about investors and consumers in energy market based on bert-bilstm. *IEEE Access*, 8:171408–171415, 2020.

[5] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2019.

[6] Eric Wallace, Yizhong Wang, Sujian Li, Sameer Singh, and Matt Gardner. Do nlp models know numbers? probing numeracy in embeddings. *arXiv preprint arXiv:1909.07940*, 2019.

[7] Jeffrey Pennington, Richard Socher, and Christopher D Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.

[8] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space, 2013.

[9] Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. Deep contextualized word representations, 2018.

[10] Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. Improving language understanding by generative pre-training. 2018.

[11] Mandar Joshi, Danqi Chen, Yinhan Liu, Daniel S Weld, Luke Zettlemoyer, and Omer Levy. Spanbert: Improving pre-training by representing and predicting spans. *Transactions of the Association for Computational Linguistics*, 8:64–77, 2020.

[12] Michael Glass, Alfio Gliozzo, Rishav Chakravarti, Anthony Ferritto, Lin Pan, GP Bhargav, Dinesh Garg, and Avirup Sil. Span selection pre-training for question answering. *arXiv preprint arXiv:1909.04120*, 2019.

[13] Adam Roberts, Colin Raffel, and Noam Shazeer. How much knowledge can you pack into the parameters of a language model? *arXiv preprint arXiv:2002.08910*, 2020.

[14] Minh-Thang Luong, Hieu Pham, and Christopher D Manning. Effective approaches to attention-based neural machine translation. *arXiv preprint arXiv:1508.04025*, 2015.

[15] Nils Reimers and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. *CoRR*, abs/1908.10084, 2019.

A Appendix

In [TIME PERIOD] how much will [COIN] change?	[COIN] will change [PERCENT CHANGE] IN [TIME PERIOD]
How much will [COIN] change from [FUTURE TIME 1] to [FUTURE TIME 2]?	[COIN] will change [PERCENT CHANGE] between [FUTURE TIME 1] and [FUTURE TIME 2]
How much did [COIN] change from [PAST TIME 1] to [PAST TIME 2]?	[COIN] changed [PERCENT CHANGE] between [PAST TIME 1] to [PAST TIME 2]
How much did [COIN] change from [PAST TIME] until now?	[COIN] changed [PERCENT CHANGE] percent from [DAYS] days [HOURS] hours and [Minutes] minutes ago until now.
How much did [COIN] change between [PAST TIME 1 (hours and minutes)] to [PAST TIME 2 (hours and minutes)] hours ago?	[COIN changed [PERCENT CHANGE] percent
Did [COIN] increase or decrease between [PAST TIME 1] and [PAST TIME 2]?	The price of [COIN] [INCREASE/DECREASE]
Will [COIN] increase or decrease in [TIME PERIOD]?Will [COIN] increase or decrease in [TIME PERIOD]?	The price of [COIN] will [INCREASE/DECREASE]

Table 6: Complete list of cryptocurrency questions.