Predicting Taiwan Stock Marketing Using Social Moods

English name

Introduction

Mining social media data to forecast the future has been a popular research in recent years. With the development of social media, large amount of information can be derived from the social media such as Facebook and PTT. Stock market prediction has attracted the attention from academia in recent years. The trend of stock market has been considered to be related to the emotion of crowd. Hence, how to analyze the feelings of crowd is the issue worth discussion. The main goal of this study is to forecast the stock market behavior in Taiwan and discuss the correlation between online emotions and stock market.
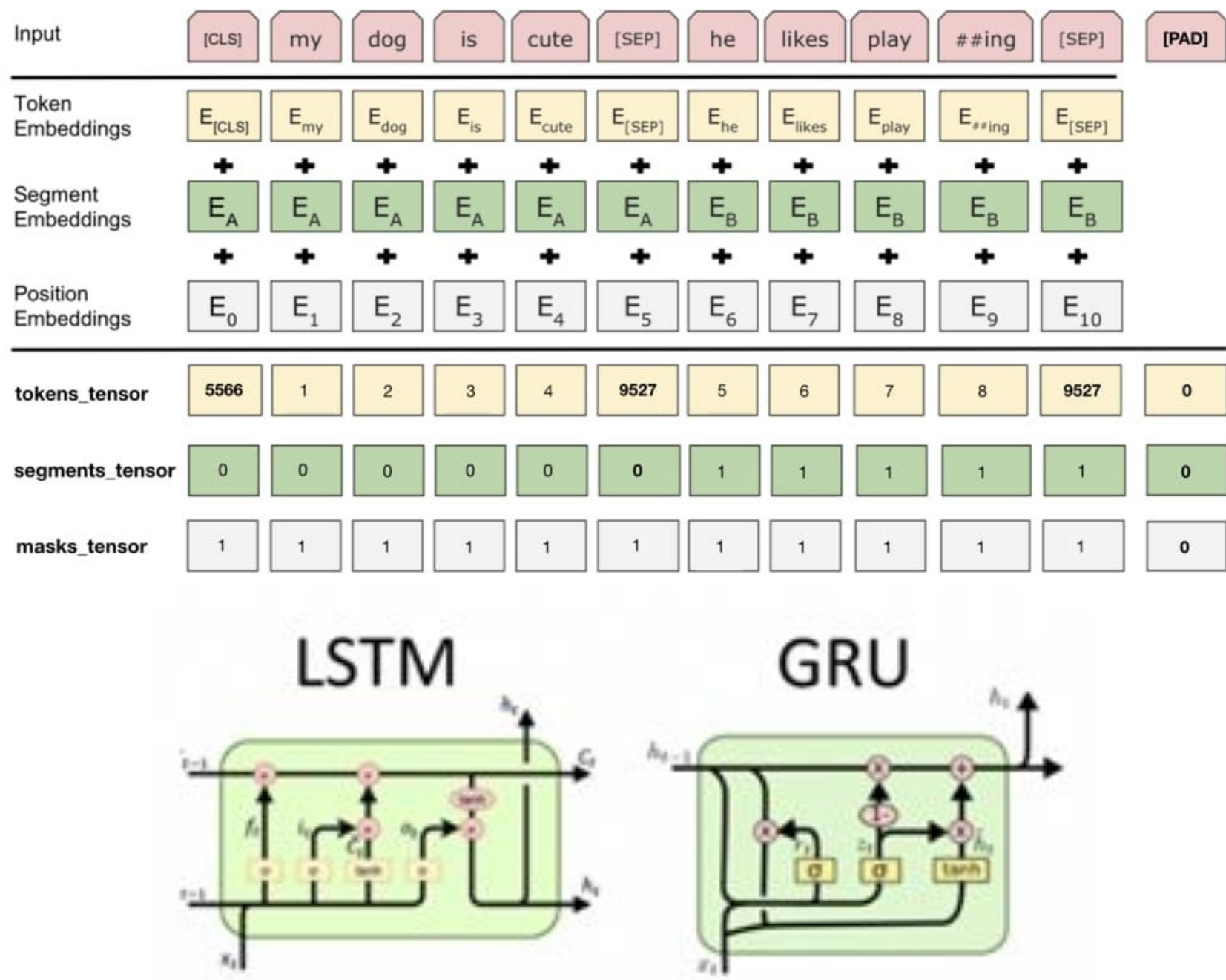
Method

SnowNLP: A library for traditional Chinese natural language processing. Generating the sentiments score from each article's topic by native bayes classifier.

$$P(c \mid x) = \frac{P(x \mid c)P(c)}{P(x)}$$

$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

Bert: A Transformer-based machine learning technique for natural language processing (NLP) pre-training developed by Google.

Model: We propose to apply an LSTM model because it is powerful with sequence prediction problems, and the LSTM model could store past information. Furthermore,w we determine to apply GRU, a LSTM with a forget gate, but has fewer parameters than LSTM. GRUs have been shown to exhibit better performance on certain smaller and less frequent datasets.

Data

Our training data could be divided into two parts, 'Technical index' and 'Sentiment Analysis'.

Technical index: We gathered data of TSMC from 2018 to 2020, which contains news and technical analysis factors. There are several technical analysis factors that we use, such as moving average(MA), KD, RSI, MACD, standard deviation of stocks. The data is normalized for training.

Sentiment Analysis: we obtain the online message in TSMC News on PPT. We deal the message with the Natural Language Processing (NLP) by using the 'BERT' and 'SnowNLP' models, respectively. Then we define the Social Mood with the index output from model mentioned above, therefore we can evaluate the social mood and make it as a feature for training.
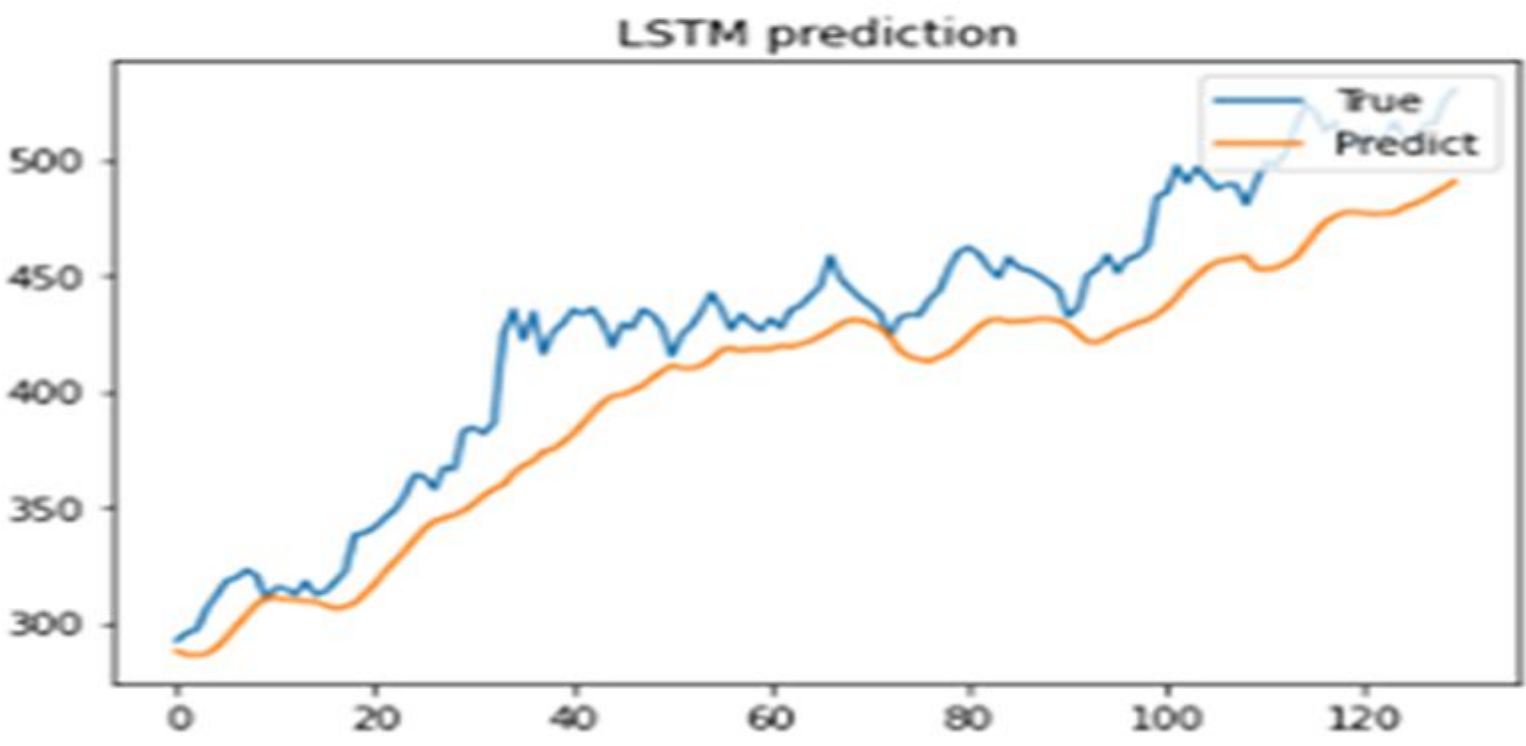
Result

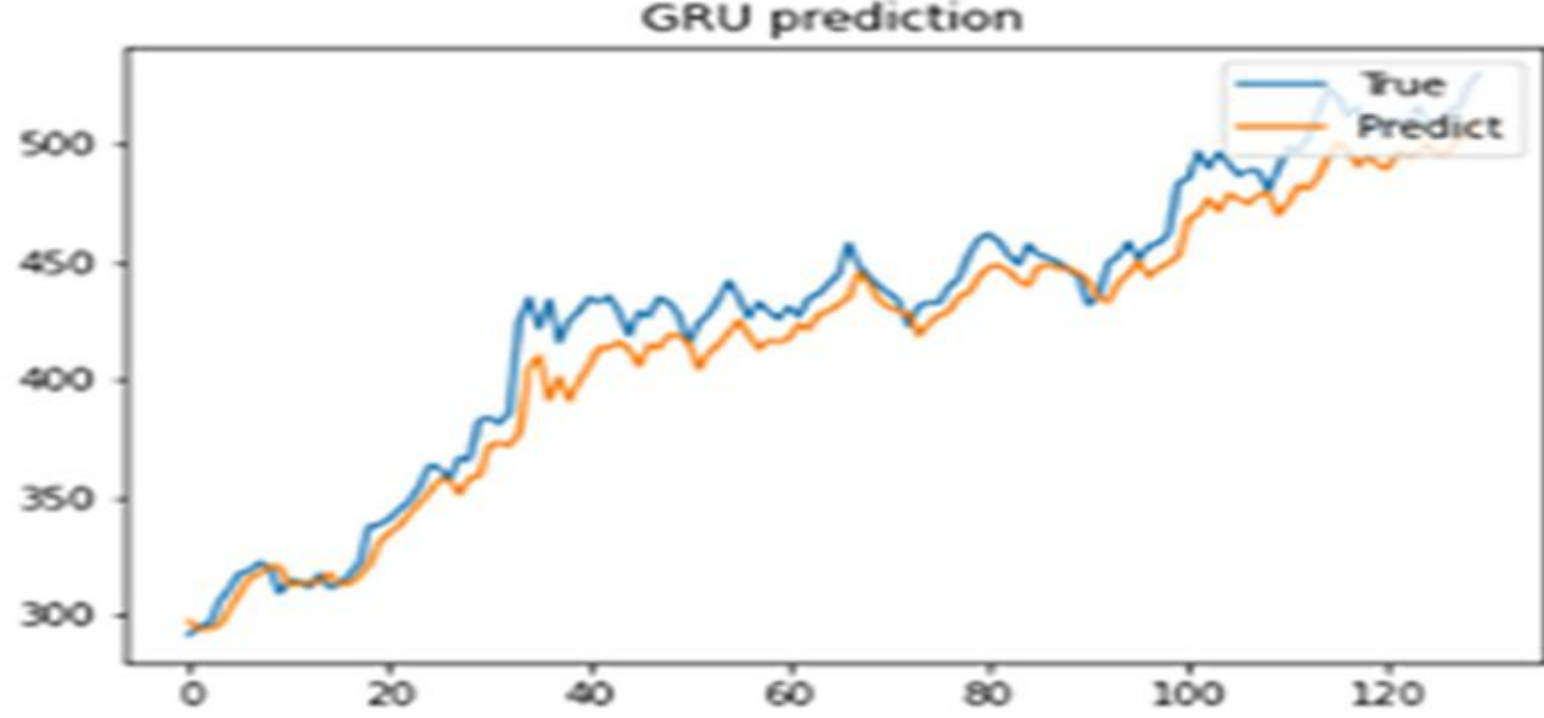Fig1 : Use technical index and sentiments to predict close via LSTM

Fig2 : Use technical index and sentiments to predict close via GRU
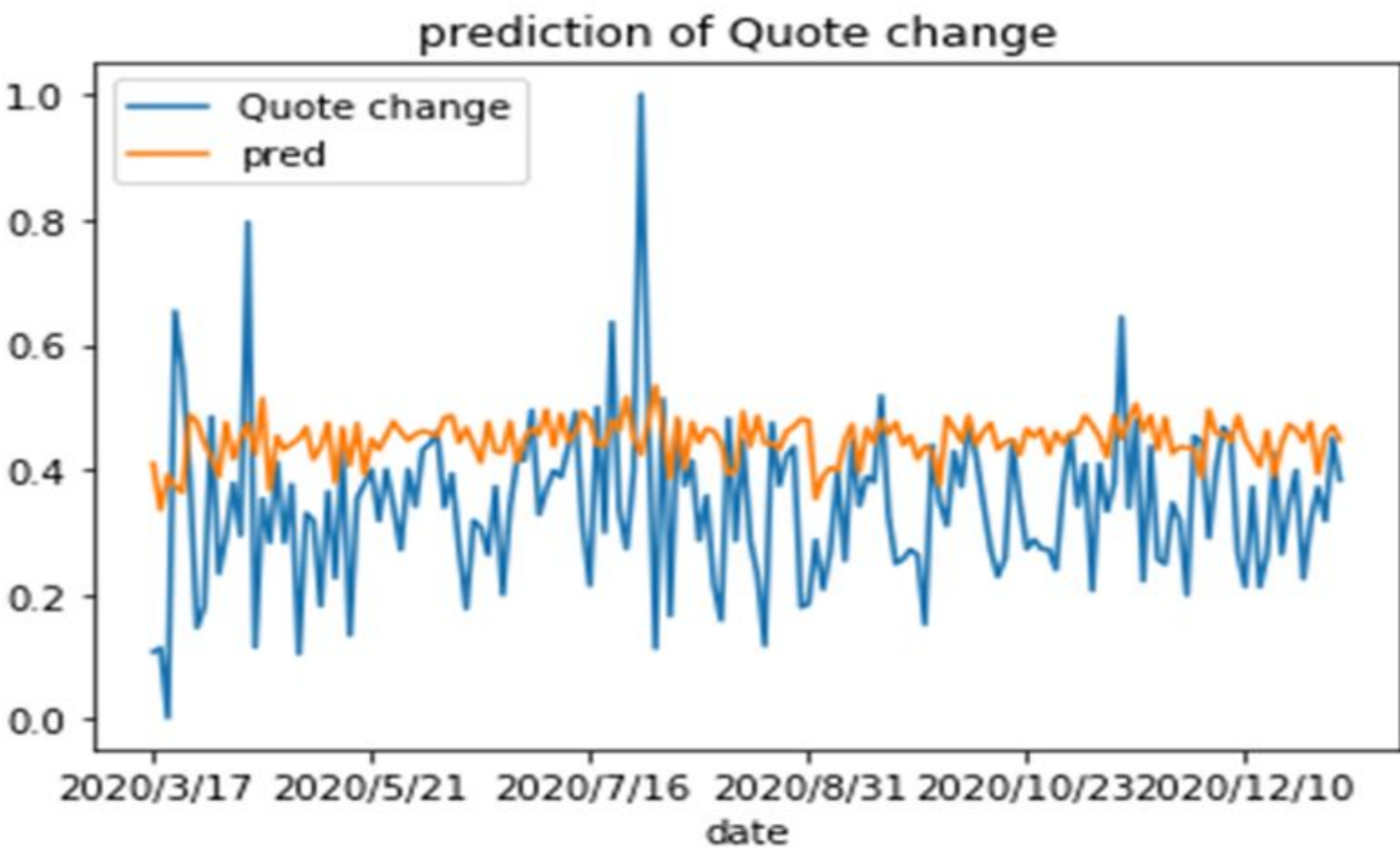

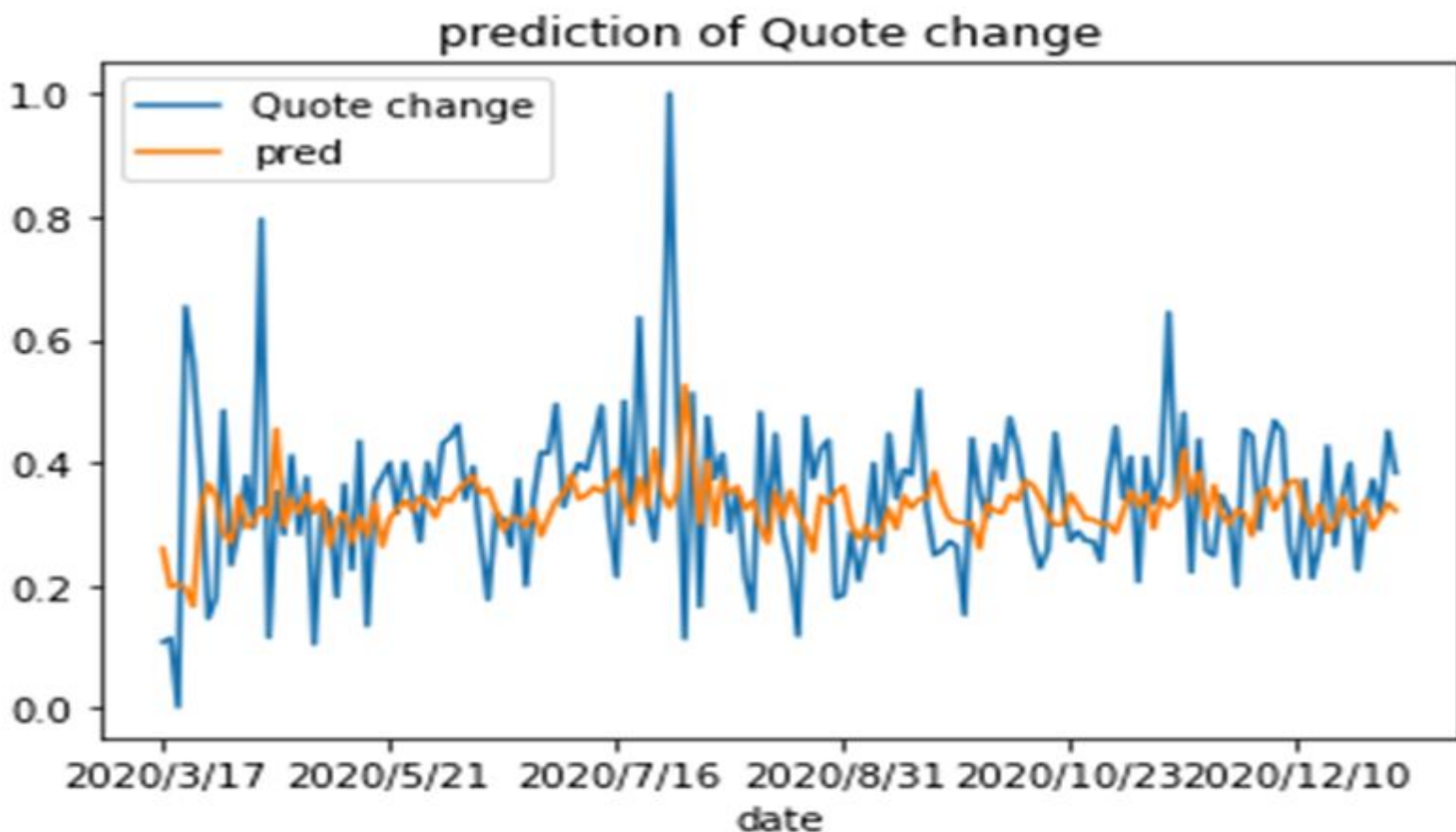Fig3 : Use snownlp to predict quote change via GRU

Fig4 : Use bert to predict quote change via GRU

Conclusion

In this topic, we use sentiment analysis and technical indexes to predict the closing price. After experiments, we found that sentiment analysis is indeed positive for predicting the closing price. This is also reflected in our use of Bert and Snownlp predicts the ups and downs of TSMC stocks, and from the results of the two sentiment analysis, Bert's results seem to be better. It may be speculated that Bert's training corpus uses twitter, which is closer to the word of ptt that can fully express the emotional content. Through this topic, we also confirmed that the sentiment of PTT can indeed predict the Taiwan stock market to some extent.

Contact

Fintech group 1
Email:

References

1.
2.
3.
4.
5.
6.
7.
8.
9.
10.