

# Forecasting Currency Exchange Rate Variations using Sentiment Analysis (Project Proposal)

Nianmin Guo

Wanhui Han

{guo.ni,han.wa}@husky.neu.edu

Northeastern University

## 1 Introduction

Using Natural Language Processing (NLP) methods to predict financial market movements has been an interesting topic for the past decade. Many researchers have proposed sentiment analysis models for time series predictions. In this paper, we want to use sentiment analysis to predict the movement of currency. This is an appealing problem because exchange rates of currencies reflects the economic variables and influences trading strategies of international corporations. In this paper, we want to evaluate a model based on historical news and currency exchanges data to show the correlation between sentiments factor and currency exchanges.

## 2 Related Work

Various researches has been done on analyzing sentiments of natural languages, with or without the context of a market. One group utilizes sentiment analysis and neural networks to predict the variations of the market, while the sentiment used is of general moods (*happy*, *alert*, etc.) Bollen et al. in 2008 [1] analyzed the moods of tweets using mood tracking modules for sentiment analysis and attempted to correlate the mood of the twtiverse of that day to the variations on the stock market (DJIA) with 86% accuracy.

One article or tweet can have different effects on different currency pairs. To avoid creating separate classifiers, embedding the currency pair information in the neural network is needed. In 2016, Wang et al. proposed an Attention-based LSTM with Aspect Embedding Neural Network [4] that can produce sentiment for each aspect defined. With a three-way (positive, negative, neutral) test, the proposed neural network achieved 77% accuracy analyzing restaurant reviews. The result was produced with a dataset of around 4500 entries, while we can observe whether more data can improve the three-way accuracy of said neural network.

With the advent of Deep Learning, Attention-based LSTM has achieved state-of-the-art results in sentiment analysis. Meanwhile, there have been previous successes on predicting market variations with sentiment analysis through other machine learning mechanisms. We thus believe that evaluating the currency exchanges with the Attention-based LSTM will produce novel and original results.

## 3 Datasets

For news data, we plan to use news published by the New York Times, CNN, Reuters, and 13 other media outlets between 2016 and 2017. There is a Public dataset available (*All the news*) with the criteria above, and it contains some 120,000 articles matching the conditions. The dataset can be found at <https://components.one/datasets/>

For social media data, we plan to use Twitter data during the above time period with specific profiles (e.g., Significant figures in international politics or market) and hashtags as filters. As of this time, only tweets in English is considered.

For currency exchange data, we plan to use a dataset of daily currency exchange rates covering the time period above, focusing on major currency pairs. There is also a Kaggle dataset available at <https://www.kaggle.com/thebass/currency-exchange-rates/home>.

The data above will be divided into parts for development, training, and testing. We are also considering using web spiders and public APIs for obtaining real-time data for testing purposes.

## 4 Methodology and Evaluation

We propose to implement an attention-based LSTM neural network [3], [4] to produce a sentiment classes that specifically expresses the sentiment of the text input regarding to the potential currency exchange market variations (*variation sentiment*), per aspect (currency pairs). The day-before, week-before, and month-before exchange rate variations will also be fed into the neural network.

A bag-of-words approach for the sentiment class generation can also be used to prove the effectiveness of the LSTM neural network.

As for the evaluation of this neural network, the correlation between the predicted sentiment class of a news article and the variations of the exchange market will be considered as the main metric, as described by Jiang and Li ([2]).

A stretch goal of this project will be the classification of *effecting terms* (i.e. when can the effect of a specific news article be observed) in addition to the *degree of variation*.

## References

- [1] Johan Bollen, Huina Mao, and Xiaojun Zeng. 2011. Twitter mood predicts the stock market. *Journal of computational science* 2, 1 (2011), 1–8.

- [2] Junfeng Jiang and Jiahao Li. 2018. Constructing Financial Sentimental Factors in Chinese Market Using Natural Language Processing. (2018).
- [3] Huicheng Liu. 2018. Leveraging Financial News for Stock Trend Prediction with Attention-Based Recurrent Neural Network. (2018).
- [4] Yequan Wang, Minlie Huang, and Li Zhao. [n. d.]. Attention-based lstm for aspect-level sentiment classification. In *Proceedings of the 2016 conference on empirical methods in natural language processing*. 606–615.