

# High-Frequency News Sentiment and Its Application to Forex Market Prediction

Frank Z. Xing  
Nanyang Technological University  
zhutian.xing@ntu.edu.sg

Duc-Hong Hoang  
Lund University  
duc-hong.hoang@nek.lu.se

Dinh-Vinh Vo  
Lund University  
dinh-vinh.vo@nek.lu.se

## Abstract

*Financial news has been identified as an important alternative information source for modeling market dynamics in recent years. While most of the attention goes to stock markets, the foreign exchange (Forex) market, in contrast, is much less studied. Most of the existing text mining research for the Forex market combine news sentiment with other text features, making the contribution of each factor unclear. To this end, we want to study the role of news sentiment exclusively. In particular, we propose a FinBERT-based model to extract high-frequency news sentiment as a 4-dimensional time series. We examine the efficacy of this news sentiment for Forex market prediction without involving any other semantic feature. Experiments show that our model outperforms alternative sentiment analysis approaches and confirm that news sentiment alone may have predictive power for Forex price movements. The sentiment analysis method seems to have a big potential to improve despite that the current predictive power is still weak. The results deepen our understanding of financial text processing systems.*

## 1. Introduction

Analyzing financial news was traditionally the job of professional analysts. However, the amount of information has exploded and the flow of information accelerated in the past decades. This transition poses challenges to the capacity of manual financial information processing. Fortunately, the intelligibility of machines has improved a lot with the recent advances in natural language processing (NLP) and deep learning. The progress makes it possible to automatically analyze text and subsequently apply the insights to financial forecasting systems. The financial text may be either coming from company releases [1], earning calls [2, 3], news [4], or social media data [5, 6]. The application tasks include stock market prediction [7], Forex rates forecasting [8], portfolio optimization and asset

allocation [9], decision-making and decision-support systems [2], and many more [10].

Compared to stock markets, the problem of Forex market prediction using financial news is significantly less investigated. The first reason may be that the global Forex market is over-the-counter (OTC), not exchange-traded, and stratified. Consequently, no participant could access the exhaustive trading information, and different market participants are quoted with different price spreads. This makes performance evaluations sensitive and less reliable compared to the back-testing practices for stock markets. The second reason may be the early pessimistic conclusion from Meese and Rogoff that Forex rates prediction models rarely beat the random walk baseline [11, 12, 13]. Although many researches were attempted later to dispute this result, the majority of them [14, 15, 16] studied only low-frequency (daily or monthly) predictions. It is difficult to incorporate financial news to such researches because financial news arrives much more frequently than the updates of fundamentals, whereby its predictive power is believed to work at shortened horizons. It can be concluded from various researches [8, 17, 18] that both the useful news information and predictive horizon are within a 5-minute to 2-hour's time. Therefore, a large amount of news data has to be sampled and processed at high frequencies.

Another challenge lies in choosing the appropriate NLP methods to process the financial news. Earlier studies have applied sentiment analysis [18], document term matrix (DTM) [17], term frequency-inverse document frequency (TF-IDF) [19] to represent the news. Recently, representing financial text with deep learning and document embedding models has become popular for their strong expressive power. However, this trend compromises interpretability. Most financial forecasting tasks are not designed for a single pursuit of the best performance, but to understand what is going on, because in practice, no one would apply such a model with unknown or unbounded risk.

Moreover, Forex market prediction may require a set of different, well-adapted NLP techniques because common practices are prone to errors. Take financial sentiment analysis [20] as an example. As Semiromi et al. [17] pointed out, in the stock market, news gives a clear indication of whether to take long or short positions. Whereas Forex rates concern currency pairs (base currency/quote currency, such as EUR/USD). A piece of positive news concerns the EU economy and a piece of positive news concerns the US economy drive the Forex rate to opposite directions, such that more sophisticated analysis, i.e., aspect-based sentiment analysis may be necessary.

Finally, news-based financial forecasting is by nature a two-phase task: in the first phase, the sentiment or content of news is analyzed, and in the second phase, the result is applied to describe financial markets. It is possible that the sentiment of the first phase is correct and accurate, but not useful for the second phase. Therefore, the evaluation of the two phases should be almost independent.

To tackle these challenges, we collect financial news with accurate (millisecond level) timestamps. Unlike most of the previous researches which only consider a mixed usage of both semantic and sentiment information, we separate sentiment apart from other information and examine whether sentiment alone has predictive power for the Forex market. This way, we mitigate the “black-box” problem of machine learning-based NLP methods and move one step closer to interpretable financial forecasting. Specifically, we propose to use FinBERT, one of the state-of-the-art NLP model for financial sentiment analysis, for sentiment time series construction and test it against strong baselines, including a Naive Bayes classifier, lexicon-based classifiers, and a recurrent neural network (RNN)-based classifier. We discovered that this model outperforms others under the condition that no price information from the Forex market has been accessed.

The remainder of this paper is organized as follows. Section 2 briefly discusses economic theories on Forex rate models; Section 3 provides further details on related works; Section 4 elaborates the architecture of our

proposed model and the structure of sentiment time series; Experiment settings and results are discussed in Section 5; Finally, Section 6 concludes this work with limitations and future work.

## 2. Theoretical Framework

Previous economics studies, e.g., [12, 22], conclude that the exchange rates are determined by the present discounted value of the current and expected future macroeconomic fundamentals. According to this theoretical framework, a general exchange rate model can be roughly formulated as follows [23]:

$$\log p_t = (1 - \theta) \sum_{i=0}^{\infty} \theta^i \mathbb{E}_t(f_{t+i} | \Omega_t), \quad (1)$$

where  $p_t$  is the exchange rate (as a price ratio) at time  $t$ ,  $\theta$  is the discount factor,  $f_{t+i}$  are relevant future economic fundamentals, and  $\Omega_t$  is the information available to the public at time  $t$ . Engel and West [22] documented that investors’ expectations of future fundamental variables are more important determinants of exchange rates compared to the current fundamentals. Therefore, the model needs an effective way to estimate investors’ belief. Possible ways include asking market participants in a survey or mining private information from the order flow [24]. In the digital age, though, another approach is to measure the market sentiment based on news from financial media, which is elaborated in this article.

## 3. Related Work

In this section, we discuss related work from two perspectives, i.e., the Forex predictive models and techniques for financial sentiment analysis.

### 3.1. Forex predictive models

There exists a spectrum of Forex predictive models, from the traditional econometric models to machine learning ones. In terms of the range of information sources exploited, econometric models typically focus on the historical prices and related

**Table 1. Literature on Forex market prediction using financial news and corresponding model configurations.**

Year	Literature	News Feature	Input	Predictive Model	Supervision	Span/Horizon
2015	Nassirtoussi [19]	Sentiment+TF-IDF	T	SVM	SentiWordNet	2hour/1hour
2019	Seifollahi [18]	WSD-Sentiment+TF-IDF	T	SVM	SentiWordNet	2hour/1hour
2019	Chen [8]	BHAM	T+I	Softmax	Price	40min/5min
2020	Semiromi [17]	Sentiment+TF	T+I	XGBoost	FXDIC	–/30min
2020	Rognone [21]	RavenPack	T+I	VAR-X	–	15min/15min

economic fundamentals. For instance, Cai and Zhang [25] proposed the Autoregressive Conditional Multinomial-Autoregressive Conditional Duration (ACM-ACD) model to allow more flexibility to time indexing. The ACM-ACD model utilizes only historical data to do one-step forward forecasting. In such models, although the patterns of high-frequency order flows may transmit macroeconomic announcements [26], the introduction of external information is very limited. Machine learning models, on the other hand, generally consider more factors from the digital age, such as news sentiment. The only econometric model we surveyed that considers news sentiment is the exogenous autoregressive model (VAR-X) [21]. However, the sentiment data is obtained from RavenPack, a news analytics service, hence the model does not directly analyze financial text.

Table 1 summarizes the predictive models from the recent literature. We observe that statistical learning models, such as the support vector machine (SVM), Softmax, and boosting methods are preferred, probably because of their efficient training process. In particular, Semiromi et al. [17] consider spontaneous influential news from a calendar, hence the information span is not applicable. Together with technical analysis, their winning model achieves around 60% directional accuracy; Other researches gather financial news from a time interval of 15 minutes to 2 hours, then use this information to predict the price movement after 5 minutes to 1 hour. The input of predictive models usually contains both financial text and technical indicators (T+I). For example, Semiromi et al. [17] included MACD, Bollinger bands, trend line, and many others. Even when text is exclusively used, the text feature usually contains semantic information, in order to achieve a higher accuracy measure [19, 18]. This fact casts a shadow on our understanding of the role of news sentiment, *as a basic spirit of science is to hold control variables constant*.

### 3.2. News sentiment in the Forex market

In terms of how to involve news sentiment, previous researches have different ideas. Nassirtoussi et al. [19] looked up the sentiment scores from SentiWordNet [27] for words from the news. The sentiment scores are later used to calculate the polarity of news and to weight its TF-IDF features. Similar general-purpose sentiment dictionaries include MPQA [28] and SenticNet [29]. Seifollahi and Shajari [18] improved the sentiment score retrieval results of [19] by adding a word sense disambiguation module. For example, [*fall*, *noun*] is a neutral word whereas [*fall*, *verb*] carries

negative sentiment in financial news. With the correct part-of-speech tag, the retrieved word polarity will be more accurate. Their experiments showed that without sentiment weighting, the accuracy of using only TF-IDF to predict drops greatly.

Chen et al. [8] used the sentiment information together with semantic information by learning a BERT-based Hierarchical Aggregation Model (BHAM). However, their method is problematic because the pre-trained BERT model was fine-tuned using price as a “distant supervision”. With such an end-to-end setting, it is impossible to examine any meaningful factor individually. It seems that a large portion of the predictive power comes from “technical analysis”, despite that there may be a nonlinear boosting effect between news sentiment and other economic variables. Their results are also not stable with a slightly changed training-test split. Semiromi et al. [17] provided a more reliable supervision (FXDIC) that is modified from the Loughran-McDonald Dictionary [30]. This research, in line with [19, 18], also use sentiment to weigh the TF and TF-IDF features.

Our approach discards all the semantic information and constructs sentiment time series at high-frequency. We later investigate the relation between the sentiment time series and the Forex rates.

## 4. The Proposed Method

With two principles in mind: (1) our model does not use price as distant supervision for sentiment analysis and, (2) the sentiment information has to be distilled and separated from other factors, we propose an architecture as displayed in Figure 1.

We formulate the Forex market prediction problem as a binary classification task, i.e., to classify directional movements as upward, if price  $p_t > p_{t-1}$ , and downward, if  $p_t < p_{t-1}$ . Formally, the tick return for a currency pair  $B/Q$  is:

$$R_t(B/Q) = (p_t - p_{t-1})/p_{t-1}. \quad (2)$$

Each financial news has a timestamp  $t_i$  and the associated representation  $N_i, i = 1, 2, \dots, n$ .

Because the news is unlabeled, we have to resort to external resources to build the financial sentiment analysis module. A recent transformer-based language model “BERT” [31] is proved to be a promising contextualized sentence representation. BERT is pre-trained on a vast amount of natural language data, which includes Wikipedia among others. According to Devlin et al. [31], the BERT-base version has 12 layers of encoder and 110 million parameters in total. The model is trained with two tasks, i.e., prediction of

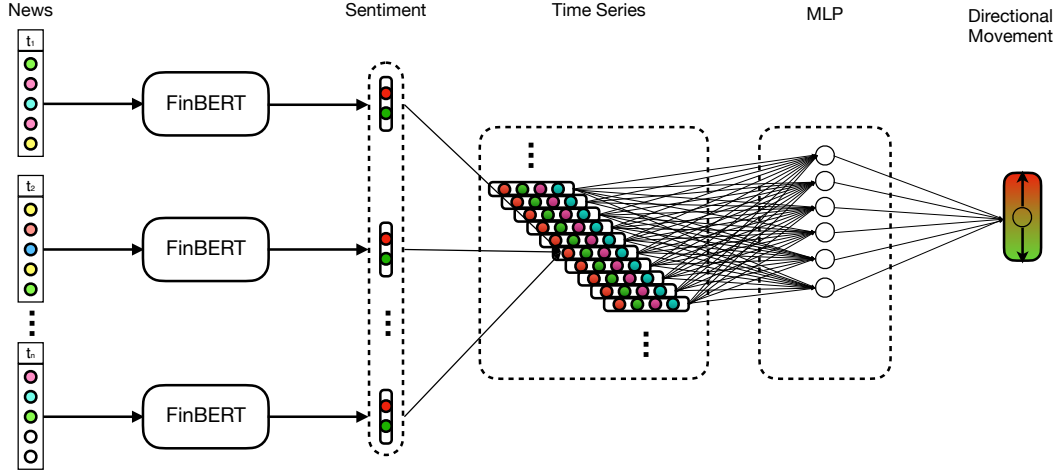


Figure 1. Overview of the model architecture.

randomly masked words in a sentence and prediction of the next sentence (whether or not there is a following relation for two given sentences). The input of BERT includes token embeddings, segment embeddings, and positional embeddings. For token embeddings, two special symbols, that are [CLS] and [SEP], are added to the beginning and end of each sentence.

To further facilitate BERT’s adaptation to the financial news domain, our proposed system uses FinBERT<sup>1</sup> [34], a recent innovation specifically developed for financial sentiment analysis. The FinBERT model first fine-tunes the original BERT on TRC2 data in order to obtain a language model in the finance domain. Subsequently, the model is further trained on Financial PhraseBank [35], learning to predict the sentiment as positive or negative given a piece of text. Following the suggestion of [34], we use the [CLS] token embeddings as the input for classification tasks. In this context, the sentiment score of each news  $N_i$  is calculated from its token embedding:

$$s(N_i) = \sigma(W * N_i^{[CLS]} + b), \quad (3)$$

where  $\sigma(x) = 1/(1 + e^{-x})$  is the sigmoid function;  $W$  and  $b$  are trained FinBERT parameters.

#### 4.1. Sentiment time series

Our goal being to aggregate sentiment signals from each news to form a time series, we preserve more statistical information instead of simply averaging them. Specifically, we define the (aggregated) news sentiment

<sup>1</sup>We note that this name has been competitively used by studies like [32, 33], though the variant we used here is the only available version when we conducted our research.

#### Algorithm 1. Build sentiment time series.

---

**Data:** News stream  $\{t_i, N_i\}$   
**Result:** Sentiment time series  $s_t(B/Q)$

```

1 for  $i = 1, 2, \dots$  do
2   if  $t - 1 < t_i < t$  then
3      $s(N_i) \leftarrow \sigma(W * N_i^{[CLS]} + b)$ ;
4     if  $s(N_i) > 0$  then
5        $s_t^I(+) \leftarrow$ 
6          $(k - 1/k)s_t^I(+) + (1/k)s(N_i)$ ;
7        $s_t^V(+) \leftarrow s_t^V(+) + 1$ ;
8     else if  $s(N_i) < 0$  then
9        $s_t^I(-) \leftarrow$ 
10         $(k - 1/k)s_t^I(-) + (1/k)s(N_i)$ ;
11        $s_t^V(-) \leftarrow s_t^V(-) + 1$ ;
12      $k \leftarrow k + 1$ ;
13   else
14      $t \leftarrow t + 1$ ;  $[s_t(B/Q), k] \leftarrow \mathbf{0}$ ;
15   end
16 end

```

---

at time  $t$  as a quadruple, that is, to calculate sentiment polarity and intensity for each relevant message and aggregate them in a discrete-time axis.

The quadruple inherits the same structure as in [5] and records the intensity and volume information for positive news and negative news respectively:

$$s_t(B/Q) = (s_t^I(+), s_t^I(-), s_t^V(+), s_t^V(-)), \quad (4)$$

In Equation 4,  $s_t^I(+)$ , e.g., is the average intensity for all the positive news, and  $s_t^V(-)$  the count of negative news regarding  $B/Q$ . Algorithm 1 elaborates on how the 4 dimensions of  $s_t(B/Q)$  are updated.

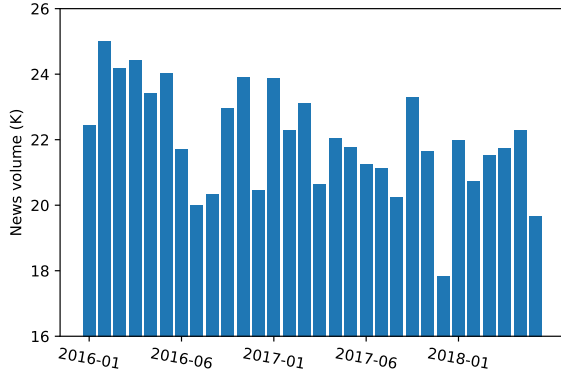


Figure 2. Monthly news volume on EUR/USD.

## 4.2. Using news sentiment to predict

Since  $s_t(B/Q)$  only samples news with  $t_i < t$ , the whole sequence  $s_t(B/Q), s_{t-1}(B/Q), s_{t-2}(B/Q)...$  can be used to predict Forex rate movements without information leakage. Although a longer sequence contains richer historical information in theory, empirical evidence suggests that the sequence does not have to be too long because recent sentiment is of higher importance. We denote the sequence with  $d$ -step lag as  $s_t^d(B/Q)$ , then the learning task is to train a binary classifier  $f : s_t^d(B/Q) \mapsto \text{sgn}(R_t(B/Q))$ .

We use a multi-layer perceptron (MLP) model with single hidden layer of  $q$  neurons to learn this mapping, that is:

$$f(x) = \text{Softmax}(W_2 * \text{ReLU}(W_1 * x + b_1) + b_2), \quad (5)$$

where ReLU is the activation function;  $W_1, W_2, b_1, b_2$  are MLP parameters and  $\|W_1\| = q$ .

The training objective is to minimize the regularized cross-entropy loss function:

$$\ell(\mathbf{W}, \mathbf{b}) = -y \ln \hat{y} - (1 - y) \ln(1 - \hat{y}) + \alpha \|\mathbf{W}\|^2, \quad (6)$$

where  $y = \text{sgn}(R_t(B/Q))$ ;  $\hat{y}$  is 0 or 1, depending on the value of  $f(x)$ ; the second term  $\alpha \|\mathbf{W}\|^2$  punishes complex models to prevent over-fitting.

## 5. Experiments and Results

### 5.1. Datasets

Two major datasets, i.e. the Forex dataset and the financial news dataset are used in our experiments. First, we collect the well-vetted high-frequency (5-minute level) exchange rates for EUR/USD from 2016-01-01 to 2018-06-30. The Forex data is provided by Olsen Financial Technologies, where frivolous or bad price quotes have been filtered out. The data comprises the

“bid” and “ask” prices. We compute the mid price by averaging them. If in some 5-minute interval there is no transaction made, the price, as well as the return for the next 5 minutes, will be null. Such data samples will be excluded from evaluation. Table 2 provides a snippet of the Forex dataset.

Table 2. High-frequency Forex data (EUR/USD).

Time	Bid	Ask	Mid
24.06.2018 21:35	1.16580	1.16630	1.16605
24.06.2018 21:40	null	null	null
24.06.2018 21:45	null	null	null
24.06.2018 21:50	1.16594	1.16614	1.16604
24.06.2018 21:55	1.16560	1.16590	1.16575

Table 3. News headline examples on 2018-05-01.

Time	News
05:01:47.115	ireland apr manufacturing pmi 55.3 vs 54.1 in mar
06:00:00.576	sentix euro break-up index falls to all-time low
06:21:54.055	australia says granted permanent exemption from u.s. metal tariffs

We further access financial news from Dow Jones Newswire<sup>2</sup> for the same time period. Figure 2 shows the news volume, where we observe approximately 20K news per month and periodic fluctuations: less news in mid-year and December. This number roughly agrees with the news frequency of around 300 per hour by Reuters reported in [8]. Each news has a millisecond level timestamp and a headline, a body, as well as other attached information such as industry, subject, region of the news, etc. We filtered the news such that only relevant news to the Forex movements, such as monetary policy, central bank announcements, macroeconomics, tax information is kept.

We use headlines only as the surrogate of content  $N_i$  for the sake of both noise smoothing and computational efficiency, and we align the timezone of Forex data to the same of the news data. Table 3 offers three examples of news headlines. We remove stop-words and apply basic preprocessing, so that the contractions are expanded and name entities mapped to the standard form.

Unlike previous researches, which either used very small test datasets [18] or a single training-test split [8], we use a sliding window to sample multiple training-test splits so that the results can cross-validate with each other. Specifically, we use 6 months’ data to train

<sup>2</sup><https://professional.dowjones.com/newswires>

**Table 4. Classification performances of baselines and the proposed model using only news sentiment.**

Quarter	NB		LMDIC		FXDIC		bi-LSTM		FinBERT	
	F-1	accuracy	F-1	accuracy	F-1	accuracy	F-1	accuracy	F-1	accuracy
2016-Q3	46.4	50.0	46.7	50.5	45.9	49.9	58.9	50.3	53.2	50.2
2016-Q4	54.6	50.2	54.2	50.3	58.0	50.1	57.1	50.3	53.7	49.9
2017-Q1	53.1	49.9	53.6	50.0	53.6	49.9	56.2	49.9	54.7	50.2
2017-Q2	51.3	50.4	51.2	49.9	48.5	49.8	54.6	49.8	52.8	50.0
2017-Q3	50.9	50.4	46.3	49.7	55.5	49.3	56.1	50.0	54.9	50.1
2017-Q4	57.6	50.2	53.6	50.4	57.7	50.0	57.8	50.6	54.9	50.4
2018-Q1	47.6	49.9	52.8	50.2	47.9	49.9	59.2	50.7	54.4	50.7
2018-Q2	55.0	50.3	45.1	49.8	54.0	49.9	58.4	50.6	51.5	50.5
Average	52.1	<b>50.2</b>	50.4	50.1	52.6	49.9	<b>57.3</b>	<b>50.3</b>	<b>53.8</b>	<b>50.3</b>
$\pm$ S.D.	( $\pm 3.5$ )	( $\pm 0.2$ )	( $\pm 3.5$ )	( $\pm 0.3$ )	( $\pm 4.3$ )	( $\pm 0.2$ )	( $\pm 1.5$ )	( $\pm 0.3$ )	( $\pm 1.1$ )	( $\pm 0.2$ )

the classifier and predict the directional movements for the next 3 months. For example, 2016-01 to 2016-06 are training samples to predict the range of 2016-07 to 2016-09; then 2016-04 to 2016-09 are training samples to predict the range of 2016-10 to 2016-12... In this way, we run 8 rounds of experiments and the results are reported in Table 4.

## 5.2. Comparison with baselines

We compare our model with other possible architectures by substituting the financial sentiment analysis module. To have good coverage of different types of method, we include both lexicon-based (LMDIC, FXDIC) and machine learning (NB, bi-LSTM) baselines as follows.

- **Naive Bayes (NB):** This method trains a Naive Bayes classifier using the bag-of-words features with the Movie Review dataset [36]. For this reason, it can be expected that the language domain does not completely overlap with the news data.
- **Loughran-McDonald Dictionary (LMDIC):** This method leverages a dictionary that is manually-crafted specifically for the finance domain from 10K documents by Loughran and McDonald [30]. The dictionary contains 354 positive words and 2349 negative words [37]; sentiment score is calculated from positive/negative word counts:  $s(N_i) = (Pos - Neg)/(Pos + Neg)$ .
- **Forex Dictionary (FXDIC):** The same method as LMDIC except that the LM dictionary is revised and augmented with keywords that are frequent and characteristic for the Forex market [17].

- **Bidirectional LSTM (bi-LSTM):** This method, like NB, is trained with the Movie Review dataset. However, it uses the output of a bi-LSTM as news features instead of bag-of-words. The LSTM reads the 840B.300d GloVe embedding<sup>3</sup> [38] for each word and outputs the final state.

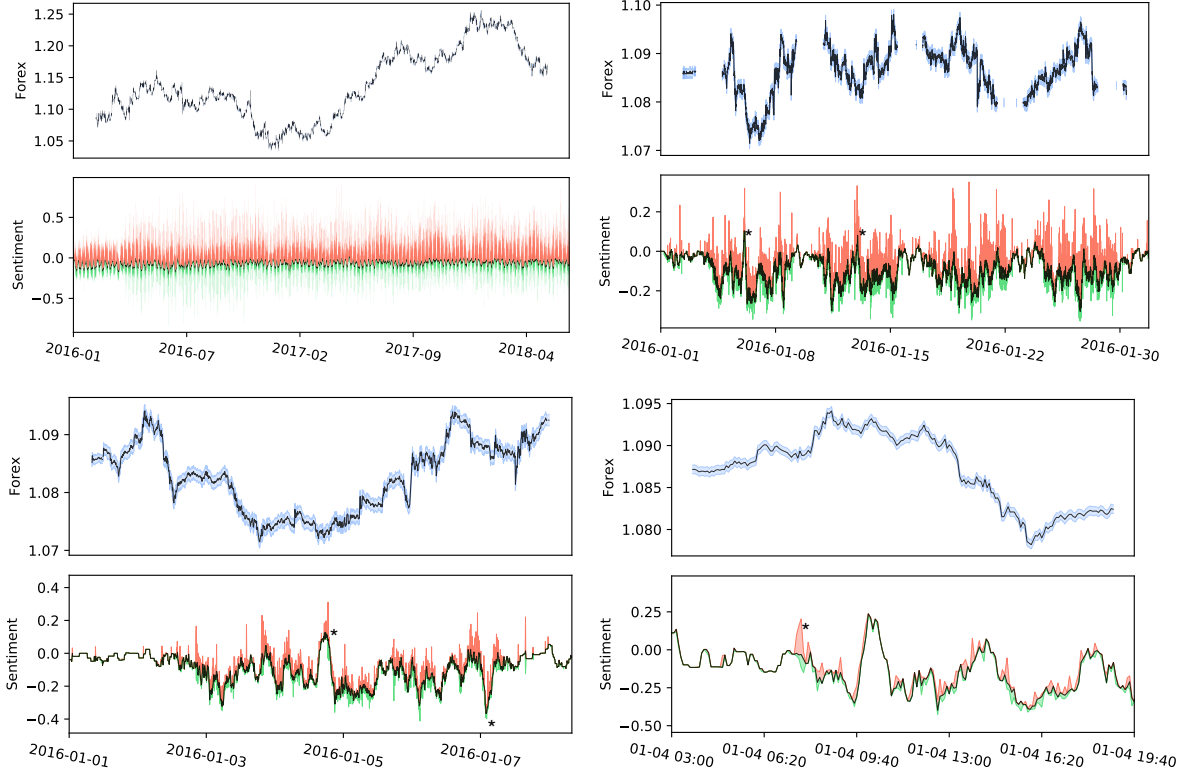
The financial sentiment analysis modules are used along with Algorithm 1. The result sentiment time series along with Forex rates are illustrated in Figure 3. The blue price band refers to the bid and ask price spread; The net sentiment (black line) is moving-averaged for better visual effects. The red and green deviations denote the strength of positive and negative sentiment respectively (strength is defined as intensity times volume).

## 5.3. Discussion on results

The high-frequency Forex market is proven to be very efficient. Early researches using low-frequency data have shown that most econometric models do not predict the Forex rates. By training an MLP with no news sentiment, we confirm that the same phenomenon also holds for high-frequency settings, and the past price series has no predictive power on Forex movements. Furthermore, most of the baseline sentiment analysis methods in Table 4 can not produce meaningful sentiment time series. Although there are some peaks in Figure 3 that seemingly signal the Forex movements (pointed by stars), they are multi-steps forward and hard to generalize. Therefore, we resort to more rigorous metrics.

Table 4 lists the F-1 and accuracy scores of using different sentiment analysis methods across 8 rounds

<sup>3</sup><https://nlp.stanford.edu/projects/glove>



**Figure 3. Visualization of EUR/USD rates and the aligned, moving-averaged news sentiment time series at different time scales: year (upper left), month (upper right), week (lower left), and intraday (lower right).**

of independent experiments. The two metrics are mathematically defined as follows:

$$\text{F-1 score} = \frac{2 \times tp}{2 \times tp + fn + fp}, \quad (7)$$

$$\text{Accuracy} = \frac{tp + tn}{tp + fp + tn + fn}, \quad (8)$$

where the number of true positive, false positive, true negative, and false negative classified samples are denoted by  $tp$ ,  $fp$ ,  $tn$ ,  $fn$ , respectively. We do not report the Matthews Correlation Coefficient (MCC) because our data is balanced in terms of both sentiment and directional movements.

We can observe that the accuracies are much more stable than F-1: standard deviations (S.D.) usually differ by one order of magnitude. Since type I and type II errors are assumed to cost similar, accuracy is more straightforward and important for the movement classification task. For these reasons, to omit accuracy (as in [8]) is problematic. The two metrics also not necessarily agree with each other. For example, FXDIC has lower accuracy, but higher F-1 than LMDIC. The two lexicon-based methods do not perform significantly

better than random guesses. We suspect that this is mainly due to the coverage issue because NB and bi-LSTM, even trained on another language domain, perform better. Noticing that none of these methods used the movements to train for sentiment analysis, the systematical out-performance of bi-LSTM and FinBERT (bold in Table 4) can be seen as evidence that they model the movements better than a random walk.

#### 5.4. Parameter sensitivity and significance

We perform grid-search for the best combination of two hyperparameters, i.e., the hidden layer size of the MLP ( $q$ ), and the length of the history that sentiment information is considered ( $d$ ).

We illustrate the model performance with two heatmaps (Figure 4), regarding our results of accuracies and F-1s. Moreover, to better understand the performance stability, we show two cross sections with error bars in Figure 5 and Figure 6. The cross sections are determined by simultaneously considering the two metrics with more emphasis on accuracy. Error bars are calculated and re-scaled from the 8 individual experiments. Combined with Table 4, we conclude



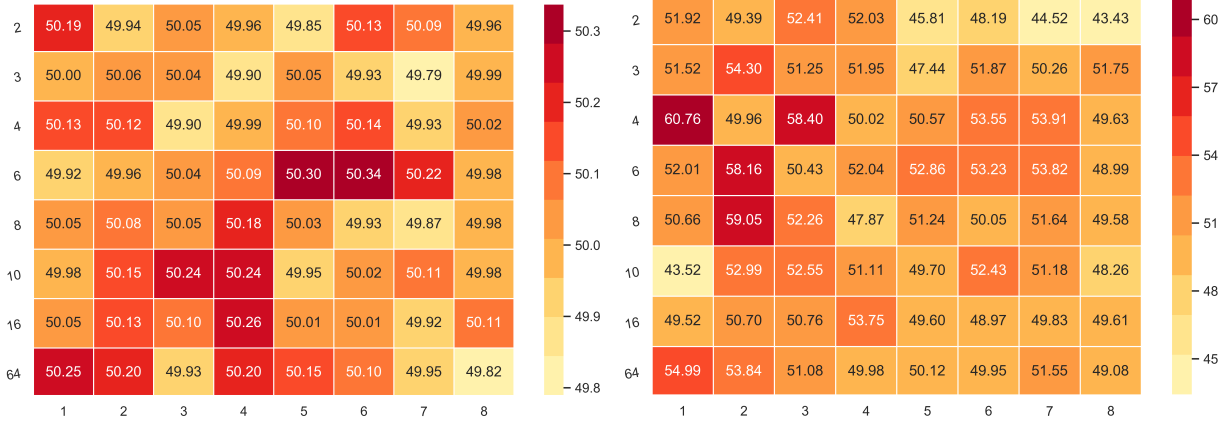


Figure 4. Heatmaps for average performance metrics for  $[q, d]$  combinations: accuracies (left) and F-1 (right).

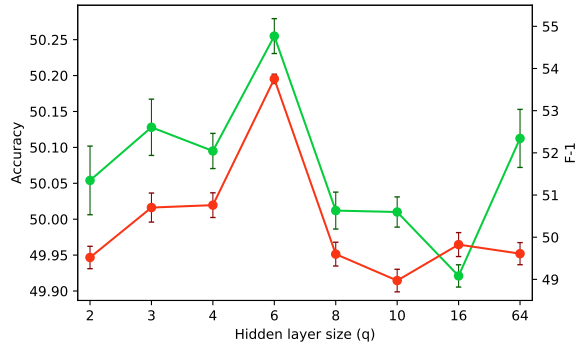


Figure 5. Sensitivity analysis for different values of  $q$ : accuracies (red) and F-1 (green).

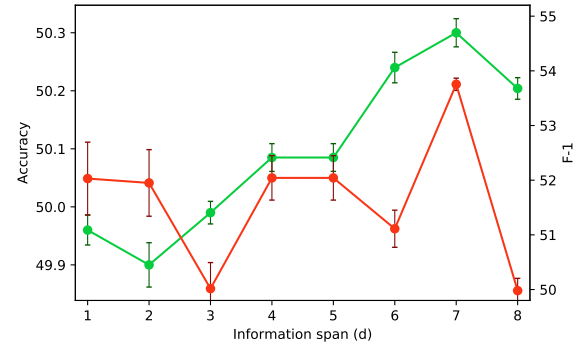


Figure 6. Sensitivity analysis for different values of  $d$ : accuracies (red) and F-1 (green).

that the model significantly outperforms random guess, though not by a large margin, when  $q$  and  $d$  are optimized.

Intuitively, increasing the information span will provide richer information, hence the model should have a larger number of parameters to fully exploit the information. We observed this corresponding relation in our experiments. Finally, we set the parameters to  $q = 6$  and  $d = 7$ .

**Hidden layer size ( $q$ )** is proportional to the number of trainable parameters in an MLP model. However, there are few guidelines and the hidden layer size is often searched trial-by-error. A too-large  $q$  will slow down the training process and make it difficult to converge, while a too-small  $q$  will limit the expressive power of the MLP. We search in the range of 2 to 64 and train the model with the Adam optimizer [39], since we have roughly 30k training samples in 6 months' time.

**Sentiment information span ( $d$ )** also has a big impact on model performances. A too-short span may not cover enough news and their propagation life-cycle on the market, while a too long span brings in noise. Previous researches [8, 18] suggest that  $d \approx 8 \sim 12$ . Our results suggest a slightly smaller  $d = 7$ , which means to predict the price movement after 5 minutes, the model considers news in the last 35 minutes. This may be due to the different years we experimented and that our data (2016-2018) is newer as the Forex market adapts faster to the information technologies. We learned from this fact that to build a Forex prediction system in practice, one should regularly adjust this hyperparameter or use a  $d$  learned from data.

Interestingly, accuracy and F-1 also have the best stability when their numbers are optimized with regard to  $q$  and  $d$ . Our hypothesis to this phenomenon is that when the model is well-fitted, it also generalizes better to different test data periods, so the performance metrics vary less.



## 6. Conclusion and Future Work

In this article, we propose an application scenario of sentiment time series, i.e., to extract high-frequency news sentiment and to use it for Forex movement prediction. Specifically, we experimented with the state-of-the-art (SOTA) financial sentiment analysis technique, i.e., FinBERT, as well as other common methods, namely NB, LMDIC, FXDIC, and bi-LSTM, for document-level sentiment detection. Unlike most of the previous researches on Forex market prediction using news, we separately consider sentiment and exclude other text features. To the best of our knowledge, *this is the first research that confirms the predictive power of high-frequency news sentiment even without any other feature or technical analysis*. We have discovered a very weak, yet detectable predictive power of the sentiment time series using bi-LSTM and FinBERT. Sensitivity analyses confirm that the model is effective only when the right model complexity and information span are selected.

There are several limitations to this preliminary research. The first limitation is that we only experimented with 2.5 years of Forex rates of a single currency pair (i.e., EUR/USD) due to limited computational power, though the data since euro's introduction in 1999 is accessible; The second limitation is on the regressive nature of the current sentiment analysis method, i.e., it only gives a sentiment score with no further information specified. However, it is clear that more sophisticated techniques, such as aspect-based sentiment analysis, concept-level sentiment analysis, and logical reasoning, are indispensable for conquering this task (of financial sentiment analysis and then Forex market prediction); A last major limitation is the lack of ground truth for the evaluation of news sentiment from a psycho-linguistic perspective. For this reason, we only evaluate the sentiment analysis results indirectly and empirically, via directional classification of Forex rates.

In future work, we plan to include a longer time span for training and test as many other researches do. The purpose is to test whether a larger training set could further improve the model performances. Although with the current settings, we have roughly 30 thousand training samples for each experiment, which seems enough. We also plan to include more currency pairs, though EUR/USD is the most traded currency pair and has the most media coverage. Finally, we plan to manually label a set of news to enable supervised sentiment learning or fine-tuning and, to enable direct evaluation of financial sentiment analysis methods.

## Acknowledgments

We are grateful for the support from the Vinnova grant: sentiment in the Forex markets, and the PPF grant: innovative approaches to portfolio optimization. We would also thank the anonymous reviewers for their suggestions and comments.

## References

- [1] B. Lutz, N. Prollochs, and D. Neumann, "Sentence-level sentiment analysis of financial news using distributed text representations and multi-instance learning," in *Proceedings of the Hawaii International Conference on System Sciences (HICSS)*, pp. 1116–1125, 2019.
- [2] K. A. Keith and A. Stent, "Modeling financial analysts' decision making via the pragmatics and semantics of earnings calls," in *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 493–503, 2019.
- [3] J. Li, L. Yang, B. Smyth, and R. Dong, "MAEC: A multimodal aligned earnings conference call dataset for financial risk prediction," in *Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM)*, pp. 1–8, 2020.
- [4] X. Ding, Y. Zhang, T. Liu, and J. Duan, "Deep learning for event-driven stock prediction," in *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 2327–2333, 2015.
- [5] F. Z. Xing, E. Cambria, and Y. Zhang, "Sentiment-aware volatility forecasting," *Knowledge-Based Systems*, vol. 176, pp. 68 – 76, 2019.
- [6] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *Journal of Computational Science*, vol. 2, no. 1, pp. 1–8, 2011.
- [7] J. Tan, J. Wang, D. Rinprasertmeechai, R. Xing, and Q. Li, "A tensor-based eLSTM model to predict stock price using financial news," in *Proceedings of the Hawaii International Conference on System Sciences (HICSS)*, pp. 1658–1667, 2019.
- [8] D. Chen, S. Ma, K. Harimoto, R. Bao, Q. Su, and X. Sun, "Group, extract and aggregate: Summarizing a large amount of finance news for forex movement prediction," in *Proceedings of the Second Workshop on Economics and Natural Language Processing*, pp. 41–50, 2019.
- [9] L. Malandri, F. Z. Xing, C. Orsenigo, C. Vercellis, and E. Cambria, "Public mood-driven asset allocation: the importance of financial sentiment in portfolio management," *Cognitive Computation*, vol. 10, no. 6, pp. 1167–1176, 2018.
- [10] F. Z. Xing, E. Cambria, and R. E. Welsch, "Natural language based financial forecasting: A survey," *Artificial Intelligence Review*, vol. 50, pp. 49–73, 2018.
- [11] P. Zhu, C. Dong, Y. Liu, and H. Liao, "The meese and rogoft puzzle in exchange rate forecasting: Answers from nonparametric methods," *Systems Engineering – Theory & Practice*, vol. 40, no. 6, pp. 1495–1508, 2020.
- [12] R. A. Meese and K. Rogoff, "Empirical exchange rate models of the seventies: Do they fit out of sample?," *Journal of International Economics*, vol. 14, no. 1-2, pp. 3–24, 1983.

- [13] Y.-W. Cheung, M. D. Chinn, and A. G. Pascual, "Empirical exchange rate models of the nineties: Are any fit to survive?," *Journal of international money and finance*, vol. 24, no. 7, pp. 1150–1175, 2005.
- [14] C. Liu, W. Hou, and D. Liu, "Foreign exchange rates forecasting with convolutional neural network," *Neural Processing Letters*, vol. 46, pp. 1095–1119, 2017.
- [15] M. Ca'Zorzi and M. Rubaszek, "Exchange rate forecasting on a napkin," *Journal of International Money and Finance*, vol. 104, p. 102168, 2020.
- [16] E. Colombo and M. Pelagatti, "Statistical learning and exchange rate forecasting," *International Journal of Forecasting*, vol. 36, no. 4, pp. 1260–1289, 2020.
- [17] H. N. Semiromi, S. Lessmann, and W. Peters, "News will tell: Forecasting foreign exchange rates based on news story events in the economy calendar," *North American Journal of Economics and Finance*, vol. 52, p. 101181, 2020.
- [18] S. Seifollahi and M. Shajari, "Word sense disambiguation application in sentiment analysis of news headlines: an applied approach to forex market prediction," *Journal of Intelligent Information Systems*, vol. 52, pp. 57–83, 2019.
- [19] A. K. Nassirtoussi, S. Aghabozorgi, T. Y. Wah, and D. C. L. Ngo, "Text mining of news-headlines for forex market prediction: A multi-layer dimension reduction algorithm with semantics and sentiment," *Expert Systems with Applications*, vol. 42, pp. 306–324, 2015.
- [20] F. Xing, L. Malandri, Y. Zhang, and E. Cambria, "Financial sentiment analysis: An investigation into common mistakes and silver bullets," in *Proceedings of the 28th International Conference on Computational Linguistics (COLING)*, pp. 1–10, 2020.
- [21] L. Rognone, S. Hyde, and S. S. Zhang, "News sentiment in the cryptocurrency market: An empirical comparison with forex," *International Review of Financial Analysis*, vol. 69, p. 101462, 2020.
- [22] C. Engel and K. D. West, "Exchange rates and fundamentals," *Journal of Political Economy*, vol. 113, no. 3, pp. 485–517, 2005.
- [23] M. Fratzscher, "Communication and exchange rate policy," *Journal of Macroeconomics*, vol. 30, no. 4, pp. 1651–1672, 2008.
- [24] M. Fratzscher, D. Rime, L. Sarno, and G. Zinna, "The scapegoat theory of exchange rates: the first tests," *Journal of Monetary Economics*, vol. 70, pp. 1–21, 2015.
- [25] C. X. Cai and Q. Zhang, "High-frequency exchange rate forecasting," *European Financial Management*, vol. 22, no. 1, pp. 120–141, 2016.
- [26] M. Evans and R. Lyons, "Order flow and exchange rate dynamics," *Journal of Political Economy*, vol. 110, no. 1, pp. 170–180, 2002.
- [27] S. Baccianella, A. Esuli, and F. Sebastiani, "Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining," in *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*, pp. 2200–2204, 2010.
- [28] L. Deng and J. Wiebe, "MPQA 3.0: An entity/event-level sentiment corpus," in *Proceedings of NAACL-HLT*, pp. 1323–1328, 2015.
- [29] E. Cambria, Y. Li, F. Z. Xing, S. Poria, and K. Kwok, "SenticNet 6: Ensemble application of symbolic and subsymbolic AI for sentiment analysis," in *Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM)*, pp. 105–114, 2020.
- [30] T. Loughran and B. McDonald, "When is a liability not a liability? textual analysis, dictionaries, and 10-ks," *Journal of Finance*, vol. 66, no. 1, pp. 35–65, 2011.
- [31] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of NAACL-HLT*, pp. 4171–4186, 2019.
- [32] Z. Liu, D. Huang, K. Huang, Z. Li, and J. Zhao, "Finbert: A pre-trained financial language representation model for financial text mining," in *International Joint Conference on Artificial Intelligence (IJCAI): Special Track on AI in FinTech*, pp. 4513–4519, 2020.
- [33] Y. Yang, M. C. S. UY, and A. Huang, "Finbert: A pretrained language model for financial communications," in *Arxiv:2006.08097[v2]*, pp. 1–15, 2020.
- [34] D. Araci, "Finbert: Financial sentiment analysis with pre-trained language models," in *Arxiv:1908.10063[v1]*, pp. 1–11, 2019.
- [35] P. Malo, A. Sinha, P. Korhonen, J. Wallenius, and P. Takala, "Good debt or bad debt: Detecting semantic orientations in economic texts," *Journal of the American Society for Information Science and Technology*, vol. 65, no. 4, pp. 782–796, 2014.
- [36] A. L. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts, "Learning word vectors for sentiment analysis," in *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 142–150, 2011.
- [37] F. Z. Xing, F. Pallucchini, and E. Cambria, "Cognitive-inspired domain adaptation of sentiment lexicons," *Information Processing & Management*, vol. 54, no. 3, pp. 554–564, 2019.
- [38] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1532–1543, 2014.
- [39] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *International Conference on Learning Representations (ICLR)*, pp. 1–13, 2015.