2018 IEEE International Conference on Data Mining Workshops (ICDMW)

Investigating Timing and Impact of News on the Stock Market

1st Simone Merello, 2nd Andrea Picasso Ratto
4th Oneto Luca
DIBRIS - University of Genova
Via Opera Pia 11A, I-16145 Genova, Italy
{simone.merello, andrea.picasso}@smartlab.ws
luca.oneto@unige.it

3rd Yukun Ma
Rolls-Royce@NTU Corporate Lab
Nanyang Technological University
50 Nanyang Ave, Singapore
mayu0010@e.ntu.edu.sg

5th Erik Cambria
SenticNet Lab
Nanyang Technological University
50 Nanyang Ave, Singapore
cambria@ntu.edu.sg

***Abstract*—Predicting stock market movements is an interesting and challenging problem: researchers and traders have approached this task with different techniques, from time series prediction to technical and fundamental analysis. Nowadays, a huge amount of textual data can be used to lead a new source of information on this task, well known to be highly stochastic and temporal dependent. In this paper, we investigate the problem of analyzing the timing and the impact that news have on the stock market. We focus on two different important aspects: the possible contribution that new information can have on the stock price and its possible relation with the aggregate news recently published. Our approach involves using the information available now to predict different prices movements, from the recent past to the far future. Results on US market show that the information contained in news can be used as an accurate predictor for past price movements. When the future is considered, however, the correlation between news and stock market becomes less clear.**

***Index Terms*—Financial forecasting, Time series analysis, Stock market prediction, Sentiment analysis**

I. Introduction

Traditionally, there are two schools of thought regarding what information to resort to for predicting future prices of stocks. Technical analysts [1] believe there exist patterns or motifs that would repeat in the future. Consequently, many data mining techniques are applied to historical data to find these patterns. While for fundamental analysis, what information to look at is of more significance, knowledge outside the historical data such as geopolitics, financial environment and business principles become more relevant. Since many macroeconomic factors are unstructured and scattered from different sources, it is a field where natural language processing (NLP) techniques [2] are frequently employed.

Different text sources have been exploited as input: tweets, microblogs and news articles. Twitter, in particular, result a perfect source for collecting public sentiment and opinions about current events [3]. Recent work revealed that the public aggregate mood extracted from twitter may well be correlated with Dow Jones Industrial Average Index (DJIA) [4] and with popular socio-political events [5]. Tweets together with historical prices have been used to predict the stock market with different focus: some paper propose specific new models for the task of stock market prediction [6], [7], while others approached the problem as more related to NLP, looking for different representations of sentiment or text [8]–[10].

A different approach is connected to the efficient market hypothesis (EMH) [11]. There are three forms of EMH: weak, semi-strong and strong, here we are interested in describing two of them. In his weak form, EHM states that financial market movements depend on news, current events and product releases. Future prices cannot be predicted by analyzing the past because they are determined entirely by information not contained in the price series. According to semi-strong form, the trading price of a stock is adjusted to publicly available new information very rapidly and in an unbiased fashion. Even if recent work has proved that EMH is not always true [12], news are still commonly utilized it accordance with the EMH principle: event-driven approaches to stock market prediction have been proposed [13], daily news from Wall Street Journal are correlated with trading volumes [14], features extracted from news together with historical prices are used to predict the stock market [9], events [15], or other kinds of structured information [16] are extracted from news and used for price prediction.

In recent works, deep learning has been largely used for the task from feed-forward neural networks [17]–[20] to sequence modeling through recurrent neural networks (RNNs) [21], [22]. Degrees of accuracy at 56% hit rate in the predictions are often reported as satisfying results for stock predictions [7], [23] especially because minor improvement can potentially lead to large profits.

Stock market prediction is a problem related to time: traders exploit some information available in the present moment to make educated guesses about the future. According to EHM, new information released is rapidly used by the market participants so that the price is adapted to it. New information is constantly consumed by traders which exploit their new knowledge to buy or sell stocks until the price reflects that knowledge and is supplied by news, coming events and releases. For instance, when a company suffers from a major scandal at time t_1 the market participants are going to exploit that new information until time $t_2 > t_1$. t_2 defines the instant in which the new knowledge is totally consumed and the price reflects the information of the scandal.

2375-9259/18/$31.00 ©2018 IEEE
DOI 10.1109/ICDMW.2018.00191

In this paper, two possible causes of the stock price movements are evaluated. Firstly, we relate the stock price trend with the single news. A collection of recently published news is considered as input. The proposed model is supposed to have the ability to select which are the most important items in the collection and to track the flow of the extracted information until it is totally consumed. Secondly, we investigate the aggregation of multiple news as input. A unique vector is used to represent a set of news recently published and a collection of them, from the most recent one up to a given point, is fed to the model. As before, the model is required to select which are the most important representations and to track their evolution through time.

In this work, the information at the current moment is used to predict the price movements in different periods to obtain a measured evaluation of which are the moments affected by the information available today. Our initial experiment regards predicting the past movements as a measure of the correlation between the information present in the current news and the past. Later, the forecast of future price movements is considered with the purpose to check if news articles contain useful information for stock market prediction. To the best of our knowledge, this kind of analysis has never been carried out before.

Results show that researchers should take care in predicting the future price fluctuations starting from news articles. According to our experiments, the latter are more related with the past than with the future and some attempts [24], [25] to address the issue have already been made.

The rest of the paper is organized as follows: Section II formalizes the problem; Section III defines our approach; Section IV provides an overview of the collected data; Section V explains in detail our experiments and results; finally, Section VI points out our conclusions and future work.

II. Problem Formalization

Our problem is time dependent. The predictions are made at fixed and discretized time steps t relying on text published on dates identified as $\bar{t}$.

The labels y_t are defined according to the price difference:

$$m_t = p_t - p_{t-w_m}$$

$$y_t = \mathbb{1}(m_t)$$

where $p_t \in \mathbb{R}^+$ denotes the close price at time step t, $\mathbb{1} : \mathbb{R} \rightarrow \{0,1\}$ represents the unit step function and $w_m \in \mathbb{R}^+$ defines the trend window. The momentum indicator m_t is commonly used in technical analysis as an index of strength or weakness in the stock's price able to signal developing up or down trends. When in t the future movements starting from y_{t+w_m} are considered, a crossing up through zero of our label is used as a signal to buy since in w_m time steps the price will be higher than the current moment. Similarly a crossing down through zero is used as a signal to sell.

Because our goal is to check if the market can be predicted using textual information, only text is used as input.

The input can be defined as:

$$x_t \in \mathbb{R}^{n \text{ x } e}$$

where n is the number of news considered as related to the price movements and e is the dimension of our representation of the news.

For each time step t all the news published within a time window are considered as possible influencer of the market movements. The representation of the news published at time $\bar{t}$ is defined as $n_{\bar{t}} \in \mathbb{R}^e$. Therefore our window is composed of:

$$[n_{\overline{t-k}}, n_{\overline{t-k+1}}, ..., n_{\bar{t}}]$$

where the publication date $\overline{t-k}$ is such that $\overline{t-k} \geq t{-}4$ days and $\bar{t} < t$. According to EHM, if the market is considered efficient 4 days represent an upper bound to the possible reflection of new information over the price. The task of choosing which are the most informative news in the window is left to the model.

A representation for the news is extracted counting the number of words in the news that are present in the different categories of LoughranMcDonald dictionary [26] ($e = 7$). In their work, the authors underline how financial writings can be different by usual text starting from the syntax level. Commonly used words such as "bull" or "bear" are not related with "animal" as pretrained word embeddings could have learn in different contexts. Furthermore, some commonly conceived negative words such as "tax" or "cost" in the financial domain may not have that negative meaning given in other fields. LoughranMcDonald dictionary includes positive, negative, litigious, interesting, uncertainty and other categories of words specific for finance and recently updated. It was constructed on annual reports of US enterprises considering also companies quoted in NASDAQ and because our focus is on the same market we believe is highly related to our task.

As an example, the aggregate information at time step t can be modelled as a simple moving average of the past news published within w time steps before (N: number of news found in the considered period):

$$\hat{x}_t = \frac{\sum_{\bar{t} \in [t-w,t)} n_{\bar{t}}}{N}$$

Fig. 1 shows the price p_t, our labels y_t and two of the features extracted, respectively counting of the number of words in the categories 'POSITIVE' and 'NEGATIVE' $\in \hat{x}_t$. It's possible to observe that in certain moments these two categories seem linked with the present stock market trend.

The target of our work is to use x_t as a representation of the textual information given in the time step t to predict

$$[y_t, y_{t+7}, y_{t+14}, y_{t+21}, y_{t+28}, y_{t+35}, y_{t+70}, y_{t+105}, y_{t+210}]$$

that in our case means using the news of the current moment to predict the price movements ending in $[0, 1, 2, 3, 4, 5, 10, 15, 30]$ days in the future. Fig 2 represents the experiments where $t = 0$ is considered as present. The experiments start considering the relation between the

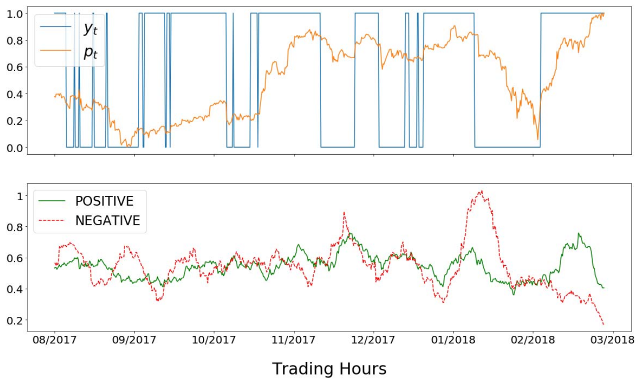

Fig. 1. AAPL stock, $w_m = 28$, $w = 40$ hours. The plotted values are normalized on the same scale.

information extracted from the textual sources and the past prices movements and then the window is slowly shifted to consider future fluctuations. According to the window w_m chosen as $w_m = 28$, y_{t+28} is the first prediction of price movements completely in the future.

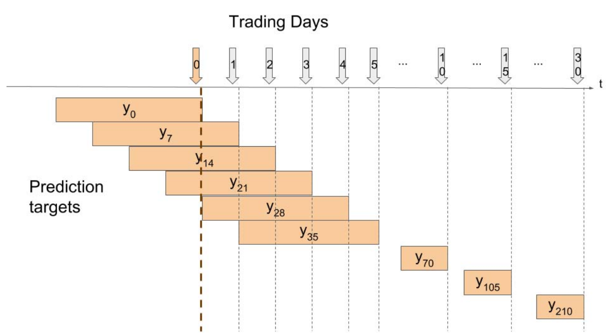

Fig. 2. Time line of the experiments. The present moment is considered as $t = 0$.

III. Proposed Approach

The main target of the experiments is to predict the stock price movements $y_t \in \{0, 1\}$ as a binary classification task. Each sample x_t summarizes the new information that can potentially have an impact on y_t. To do so a window of news published in the previous 4 days is taken in account. The proposed model can be divided in three stages: the first stage is needed to extract the most informative element in the window while the second stage model the evolution of the information over the different time steps t. Finally the third stage is needed to make the prediction $\hat{y}_t$. The model is depicted in Fig. 3.

The first approach tried to extract the information from a set of news is the moving average used to define $\hat{x}_t$. Moving average is an operation able to summarize the news in the window in only one representation of a single article. It can be useful to extract the average news from the initial set so that potential noisy information is smoothed but all the news that take part to the average are weighted equally, hence the sequence information between them is deleted. A better method is to define the extracted information c_t as:

$$c_t = \sum_{\bar{t} \in [t-w, t)} n_{\bar{t}} \cdot \alpha_{t,\bar{t}}$$

where the weights $\alpha_{t,\bar{t}} \in [0, 1]$ are learned so that $\sum \alpha_{t,\bar{t}} = 1$. This procedure is called Attention model [27]. $\alpha_{t,\bar{t}}$ represent the ability to select independently for each t which element of the window is more useful to lead the stock market price.

Given c_t, the next step is to model its evolution through time. RNNs are commonly used to process sequential data without considering its length. With RNNs, parameters are shared across the time steps so that the relation between each sample x_t and the state doesn't depend on t and what matters for generating the output is only the previous history until time step t: $x_v \forall v \in [0, t]$.

Applying this principle to our problem means that the publication date of the news is not important and given all the history of news until now, the impact that the last news has on the market is always the same. RNN allows to exploit the relations between samples x_t to model the evolution of the state h_t that acts as a summarization of the past:

$$h_t = f(h_{t-1}, x_t)$$

$$o_t = g(h_t)$$

where $f(\cdot)$ is the function that transform the state through the time steps. $g(\cdot)$ is the activation function for computing the output of the RNN represented by o_t. In this equation, h_t is forced to change every time step according to $f(\cdot)$, h_{t-1} and x_t, hence learn dependencies in the input sequence across a long series of sample can result difficult. Long short-term memory (LSTM) networks [28] solve this problem adding a new state s_t which evolves according to different factors:

$$s_t = \Gamma_f s_{t-1} + \Gamma_u \tilde{c}_t$$

$$h_t = \Gamma_o g(s_t)$$

s_t takes information directly from the past s_{t-1} through a forget gate Γ_f and is updated according to the update gate Γ_u and $\tilde{c}_t$. $\tilde{c}_t$ is a function of h_{t-1} and the input x_t . Γ_o is the output gate used to compute h_t, which is defined as the output of the network: $o_t = h_t$.

In our problem o_t summarizes the evolution of textual information until the current moment and is used to predict the stock market movements in accordance with our target.

At the third stage a dense layer is used to compute the prediction:

$$\hat{y}_t = \sigma(o_t w_\sigma + b_\sigma)$$

where $\sigma(\cdot)$ represents the sigmoid function with weights w_σ, b_σ. As usual to convert $\hat{y}_t \in [0, 1]$ to the class $\{up, down\}$ a threshold is applied

$$\begin{cases} up \text{ if } \hat{y}_t > 0.5 \\ down \text{ otherwise} \end{cases}$$

The above described model is applied with two different input settings:

- x_t defined as the window of recent news (4 days) ordered by time is fed as an input. The attention model is used to learn which are the most relevant news considering the temporal sequence in the window. As reported in Section I, the recent news should be more valuable because can potentially contain information not already consumed by traders. We will refer to this configuration as single news window (SNW) model.
- According to the recent work on aggregate mood, in this setting the collection of multiple news is considered more relevant. The proposed model is fed with the aggregated representation of a news defined as $\hat{x}_t$ in Section II where the dimension of the window w of the moving average is considered an hyperparameter. Since our features are numerical counts of the words spotted in a news, the moving average operation means defining the aggregate mood as the mean of the counting on that period.
  The input to our model x_t remains defined as the same window of the previous setting but this time $\hat{x}_t$ is considered as member of the window instead of $n_{\bar{t}}$. The attention model is fed with the aggregations of news detected in the recent time steps and as before has the task to choose which is the most relevant. At the second stage instead the LSTM is used to model the evolution of the aggregate representation through the different time steps. We will refer to this setting as aggregate news window (ANW) model.

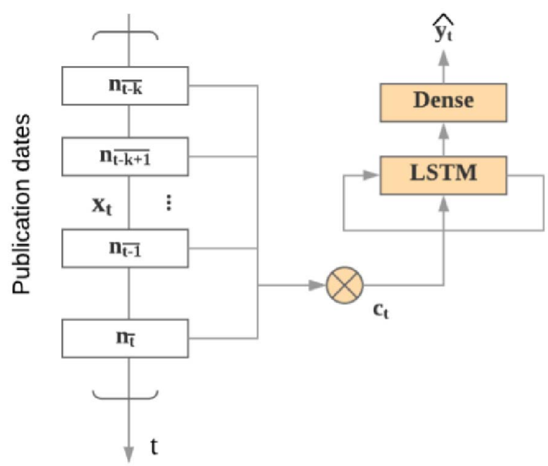

Fig. 3. SNW model

IV. Available Data

In this section, the two kinds of data used for the experiments are described: stock data and textual data. The problem of stock market forecasting is mapped to the prediction of the price movements of the top 10 stocks in capital size of NASDAQ. Previous work has already focus on similar experiments [6]. Some of the stocks traded there are in fact well known for their popularity and consequently a huge amount of textual information regarding them is constantly published.

The stock prices are collected from public sources[1] and only the periods in which the market was open are considered as time steps t. Stock price data are discretized hourly. One hour is a lower bound from the point of view of the available news published for each time step and can be also considered ad upper bound respect to the "20-minute" Theory according to which there exists an optimum time window to foresee the impact of new information released and the market correction to equilibrium [29].

In technical analysis, commonly used windows w_m for the momentum indicator m_t are 10 or 15 days, but since our focus is on hourly prediction $w_m = 28$ hours (4 days) was empirically chosen as a reasonable window that reflect the short-term trend of the market. y_t is hence labelled as positive if 4 days before t the price was lower than t.

Textual data are collected from aggregated news services that gather the information from various professional periodicals[2]. We distinguish between news articles and social media posts such as tweets because according to the EMH news which contain newly received or noteworthy information should have an impact on the stock price, instead tweets have been used as an indicator of the aggregate public feelings but in general cannot be related with new information released to the public hence are not related to our study.

TABLE I
Training and test data.

Stock	Train start	Train end - Test start	Test end
AAPL	2016-12-19	2017-10-17	2018-03-07
AMZN	2017-07-10	2017-12-23	2018-03-07
GOOGL	2017-08-03	2017-12-12	2018-03-07
MSFT	2017-04-06	2017-11-24	2018-03-07
FB	2017-07-05	2017-12-21	2018-03-07
INTC	2017-03-17	2017-11-16	2018-03-07
CSCO	2017-03-17	2017-11-16	2018-03-07
CMCSA	2017-03-17	2017-11-28	2018-03-07
NVDA	2017-04-04	2017-11-22	2018-03-07
NFLX	2017-03-17	2017-11-16	2018-03-07

V. Experiment

In this section, we detail our experiment and results.

A. Experiment Setup

Our data are considered as a time series of samples not independent ($\exists(i,j) : p(x_j, y_j | x_i, y_i) \neq p(x_j, y_j)$). The temporal dependencies between the samples are supposed to contain useful information that be exploited to improve our predictions, accordingly every kind of shuffling is avoided. In particular, given our definition of y_t there exist a temporal dependency so that (x_t, y_t) is deterministically correlated with (x_{t-v}, y_{t-v}), $v \in [0, w_m]$.

[1] http://finance.google.com/finance/
[2] https://finance.yahoo.com/, https://www.nasdaq.com/news/

Model selection is performed using cross validation as depicted in Fig. 4. The folds are selected so that the last fold takes into account all the training points, but leaves out the validation points to ensure the temporal independence between Train and Test. During the model selection phase, the hyperparameters of the the model are chosen to optimize the mean validation accuracy of the folds, where we select the validation accuracy as the max accuracy achieved by the model above all the epochs of training. The early stopping technique is applied to optimize the number of epochs on the validation set and avoid overfitting. The model is trained with Adam optimizer [30] with initial learning rate of 0.001.

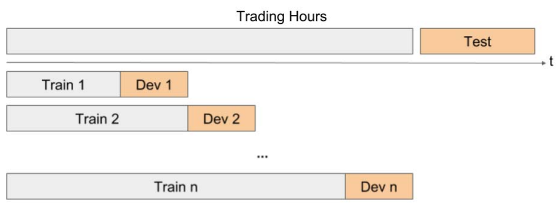

Fig. 4. Dataset division.

In the examined period, the general trend of market was positive and as a result the collected samples are unbalanced. If during the training phase a model is fed with classes largely unbalanced it will likely learn that predict the most popular class gives a good feedback from the cost function. Previous works [6], [22] have solved the issue selecting two special thresholds t_u, t_l so that the price movements are divided in classes of equal size (three balanced classes if the problem is related to predict {'*buy*', '*hold*', '*sell*'}, or if the focus is on binary classification, the middle class is disregarded). We believe that each data point can bring a contribution to the task and since the data are considered a sequence, omitting sample points from it would affect the temporal relations. Instead, in order to give the correct feedback to the proposed model, a training cost ℓ able to take in account the degree of skewness of the classes is used.

$$\ell = \sum_t -c_s \cdot y_t \cdot log(\sigma(\hat{y_t})) - (1 - y_t) \cdot log(1 - \sigma(\hat{y_t}))$$

$c_s = \frac{n_n}{n_p} \in \mathbb{R}^+$ is a constant related to the number of positive n_p and negative n_n samples.

B. Evaluation Metrics

Two assessment metrics are used. First, a standard and intuitive approach to measure the performance of classifiers is accuracy. Unfortunately, accuracy can be misleading when classes are unbalanced because a classifier that predicts always the most popular class can achieve great accuracies without understanding the real problem. In line with previous works [6], [15], [16], we choose our second evaluation metric as the Matthews correlation coefficient (MCC).

$$MCC = \frac{tp \cdot tn - fp \cdot fn}{\sqrt{(tp + fp)(tp + fn)(tn + fp)(tn + fn)}}$$

where tp: True positive, fp: False positive, tn: True negative, fn: False negative extracted from the Confusion matrix. $MCC \in [-1, 1]$, -1 means a classifier that predicts the exact opposite, while 1 implies perfect predictions.

The achieved performances are evaluated on average above all the stocks of our experiment. Averaging on different stocks from the finance point of view means consider the overall performance of the predictions over a portfolio of different financial assets as if we had bet on their performance and so obtaining earnings or losses by all of them. Since this work focus on analyzing the time span in which the new information is still usable to predict the stock price, considering different stocks means trying to generalize our result to a subset of the companies of NASDAQ and avoid specific conclusions for the single stock.

Different models are compared on the same problem formulated in Section II:

- ARMA (Auto Regressive Moving Average) is a benchmark based on time series prediction: given the past $[..., y_{t-2}, y_{t-1}]$ it tries to predict the next value y_t.
- Considering Fig. 1 a rule classifier (RC) can be defined as:
  $$\begin{cases} \hat{y_t} = 1 \text{ if } \hat{x}_t[POSITIVE] + b > \hat{x}_t[NEGATIVE] \\ \hat{y_t} = 0 \text{ otherwise} \end{cases}$$
  where $b \in \mathbb{R}$ is defined as the difference between the mean of $\hat{x}_t[NEGATIVE]$ and the mean of $\hat{x}_t[POSITIVE]$ of the training set.
- SNW model (described in Section III) .
- ANW model (described in Section III).

C. Results

Table II shows the MCCs obtained using the information at time step t to predict the price movements ending from 0 to 30 days in the future. The relative plot is depicted in Fig. 5.

The outcome of the experiments show that the simple rule classifier, based on the 'POSITIVE' and 'NEGATIVE' categories of the LoughranMcDonald dictionary has a unexpected performance regarding the stock movements ending now y_t and starting yesterday y_{t+21} (marked as 0 and 3, respectively). In particular it is able to reach 60% of accuracy, 0.15 MCC in predicting the past movements y_t. This is probably a sign of the correlation between some of the categories of the dictionary and the current trend of the market. Nevertheless as soon as we move 1 day further and we start predicting price movements starting now y_{t+28} performances come closer to random guessing.

The predictions of both ANW and SNW seem strongly correlated with the past price movements. The highest performance is reached by ANW (62.8% accuracy, 0.18 MCC) in predicting the trend ending in the current moment. ANW is in fact able to predict better than other models, and shows a significant MCC even for predictions of movements far in the future.

Considering the accuracy, the plot can be divided in two sub parts: predicting trends related to the past (before day

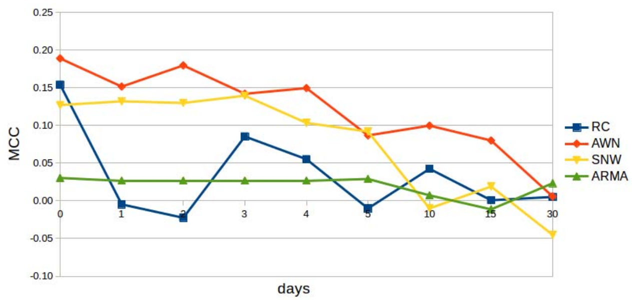

Fig. 5. Matthews correlation coefficients

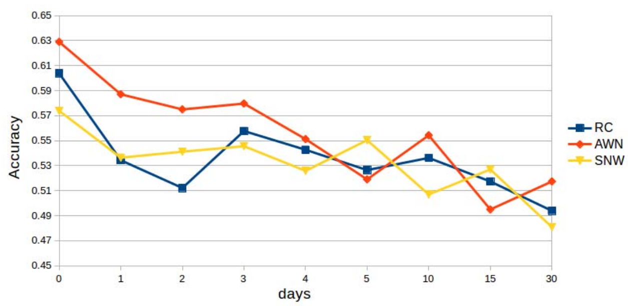

Fig. 6. Accuracies

4) and forecasting the future. In the predictions related to the past, ANW performs significantly better than SNW. A possible explanation can be derived looking more carefully at our news dataset. The collected news are written by various authors, come from different sources and in some moments the frequency of the news is so high that their order in a window can be inaccurate. Predicting the stock market movements starting from the content of the single news is likely to be subject to noise from the content and from the temporal point of view. ANW instead solves this problems by considering since the beginning an aggregate representation of the news at fixed time steps so that the possible noise on the content and on the time is smoothed.

Instead, considering the future predictions, results look different. ANW is still able to show a correlation with the price movements of the near future (0.15 MCC) but, if the accuracy is considered none of our models is able to perform significantly different than random guessing.

Table III reports accuracies of the models. Table IV summarizes some details relative to the predictions of single stocks performed by ANW.

TABLE II
MATTHEWS CORRELATION COEFFICIENT

	Days								
MCC	**0**	**1**	**2**	**3**	**4**	**5**	**10**	**15**	**30**
ARMA	0.030	0.027	0.027	0.027	0.027	0.029	0.007	-0.011	0.023
RC	0.154	-0.004	-0.022	0.085	0.055	-0.010	0.042	0.001	0.005
SNW	0.127	0.132	0.130	0.139	0.103	0.092	-0.010	0.019	-0.045
AWN	0.189	0.152	0.180	0.142	0.150	0.087	0.100	0.080	0.006

TABLE III
ACCURACIES

	Days								
MCC	**0**	**1**	**2**	**3**	**4**	**5**	**10**	**15**	**30**
RC	0.604	0.534	0.512	0.558	0.543	0.526	0.536	0.517	0.494
AWN	0.629	0.587	0.575	0.580	0.551	0.519	0.554	0.495	0.517
SNW	0.574	0.536	0.541	0.546	0.526	0.550	0.507	0.527	0.481

TABLE IV
ANW PREDICTIONS - SINGLE STOCKS DETAILS

		Days								
		0	**1**	**2**	**3**	**4**	**5**	**10**	**15**	**30**
AAPL	tn	120	85	139	131	118	108	138	96	148
AAPL	fp	85	120	66	74	87	97	67	108	61
AAPL	fn	76	61	107	152	135	103	176	116	198
AAPL	tp	203	218	172	127	143	174	99	156	73
AAPL	accuracy	0.67	0.63	0.64	0.53	0.54	0.59	0.49	0.53	0.46
AAPL	precision	0.70	0.64	0.72	0.63	0.62	0.64	0.60	0.59	0.54
AAPL	recall	0.73	0.78	0.62	0.46	0.51	0.63	0.36	0.57	0.27
AAPL	F1	0.72	0.71	0.67	0.53	0.56	0.64	0.45	0.58	0.36
AAPL	MCC	0.32	0.21	0.29	0.09	0.09	0.15	0.03	0.04	-0.02
INTC	tn	134	157	188	172	194	155	229	200	166
INTC	fp	114	91	60	73	49	88	10	39	73
INTC	fn	86	113	163	141	158	99	256	262	244
INTC	tp	169	142	92	118	104	163	12	6	24
INTC	accuracy	0.60	0.59	0.56	0.58	0.59	0.63	0.48	0.41	0.37
INTC	precision	0.60	0.61	0.61	0.62	0.68	0.65	0.55	0.13	0.25
INTC	recall	0.66	0.56	0.36	0.46	0.40	0.62	0.04	0.02	0.09
INTC	F1	0.63	0.58	0.45	0.52	0.50	0.64	0.08	0.04	0.13
INTC	MCC	0.20	0.19	0.13	0.16	0.21	0.26	0.01	-0.25	-0.27

VI. Conclusion

According to the results of this paper, the information contained in news articles can be used as an accurate predictor of past stock price movements. However, the forecast performance decreases rapidly as soon as we move our predictions forward. We have investigated two different agents that can possibly drive stock market movements: aggregate news and new information. In the results, we have shown how in both cases even if the proposed models achieve good performances in predicting the past price, when we move to the future we are not able to perform significantly better than random guessing.

In our future work, we aim to investigate which is the behavior of other kinds of representations for news articles in stock market prediction. We also aim to understand if and in which way we can add features derived from technical analysis to improve our forecasts of the future.

References

[1] C.-H. Park and S. H. Irwin, "The profitability of technical analysis: A review," *AgMAS Project Research Report 2004-04*, 2004.

[2] E. Cambria and B. White, "Jumping NLP curves: A review of natural language processing research," *IEEE Computational Intelligence Magazine*, vol. 9, no. 2, pp. 48–57, 2014.

[3] P. Chikersal, S. Poria, E. Cambria, A. Gelbukh, and C. E. Siong, "Modelling public sentiment in twitter: Using linguistic patterns to enhance supervised learning," in *Computational Linguistics and Intelligent Text Processing*. Springer, 2015, pp. 49–65.

[4] 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.

[5] J. Bollen, H. Mao, and A. Pepe, "Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena." in *International AAAI Conference on Weblogs and Social Media*, 2011.

[6] Y. Xu and S. B. Cohen, "Stock movement prediction from tweets and historical prices," in *Annual Meeting of the Association for Computational Linguistics*, vol. 1, 2018.

[7] T. H. Nguyen and K. Shirai, "Topic modeling based sentiment analysis on social media for stock market prediction," in *Annual Meeting of the Association for Computational Linguistics and the International Joint Conference on Natural Language Processing*, 2015.

[8] V. S. Pagolu, K. N. Reddy, G. Panda, and B. Majhi, "Sentiment analysis of twitter data for predicting stock market movements," in *International Conference on Signal Processing, Communication, Power and Embedded System*, 2016.

[9] Y. Peng and H. Jiang, "Leverage financial news to predict stock price movements using word embeddings and deep neural networks," *arXiv preprint arXiv:1506.07220*, 2015.

[10] F. Xing, E. Cambria, and R. Welsch, "Natural language based financial forecasting: A survey," *Artificial Intelligence Review*, vol. 50, no. 1, pp. 49–73, 2018.

[11] E. F. Fama, "Efficient capital markets: A review of theory and empirical work," *The journal of Finance*, vol. 25, no. 2, pp. 383–417, 1970.

[12] B. Qian and K. Rasheed, "Stock market prediction with multiple classifiers," *Applied Intelligence*, vol. 26, no. 1, pp. 25–33, 2007.

[13] G. P. C. Fung, J. X. Yu, and W. Lam, "News sensitive stock trend prediction," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 2002.

[14] P. C. Tetlock, "Giving content to investor sentiment: The role of media in the stock market," *The Journal of finance*, vol. 62, no. 3, pp. 1139–1168, 2007.

[15] X. Ding, Y. Zhang, T. Liu, and J. Duan, "Using structured events to predict stock price movement: An empirical investigation," in *Conference on Empirical Methods in Natural Language Processing*, 2014.

[16] B. Xie, R. J. Passonneau, L. Wu, and G. G. Creamer, "Semantic frames to predict stock price movement," in *Annual Meeting of the Association for Computational Linguistics*, 2013.

[17] A. A. Adebiyi, A. O. Adewumi, and C. K. Ayo, "Comparison of arima and artificial neural networks models for stock price prediction," *Journal of Applied Mathematics*, vol. 2014, 2014.

[28] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[18] K. J. Kim and H. Ahn, "Simultaneous optimization of artificial neural networks for financial forecasting," *Applied Intelligence*, vol. 36, no. 4, pp. 887–898, 2012.

[19] J. L. Ticknor, "A bayesian regularized artificial neural network for stock market forecasting," *Expert Systems with Applications*, vol. 40, no. 14, pp. 5501–5506, 2013.

[20] J. Patel, S. Shah, P. Thakkar, and K. Kotecha, "Predicting stock market index using fusion of machine learning techniques," *Expert Systems with Applications*, vol. 42, no. 4, pp. 2162–2172, 2015.

[21] R. Akita, A. Yoshihara, T. Matsubara, and K. Uehara, "Deep learning for stock prediction using numerical and textual information," in *IEEEInternational Conference on Computer and Information Science*, 2016.

[22] Z. Hu, W. Liu, J. Bian, X. Liu, and T. Y. Liu, "Listening to chaotic whispers: A deep learning framework for news-oriented stock trend prediction," in *ACM International Conference on Web Search and Data Mining*, 2018.

[23] R. P. Schumaker and H. Chen, "Textual analysis of stock market prediction using breaking financial news: The azfin text system," *ACM Transactions on Information Systems*, vol. 27, no. 2, p. 12, 2009.

[24] Y. Nakajima, M. Ptaszynski, H. Honma, and F. Masui, "A method for extraction of future reference sentences based on semantic role labeling," *IEICE TRANSACTIONS on Information and Systems*, vol. 99, no. 2, pp. 514–524, 2016.

[25] ——, "Future reference sentence extraction in support of future event prediction," *International Journal of Computational Linguistics Research*, vol. 9, no. 1, pp. 27–41, 2018.

[26] T. Loughran and B. McDonald, "When is a liability not a liability? textual analysis, dictionaries, and 10-ks," *The Journal of Finance*, vol. 66, no. 1, pp. 35–65, 2011.

[27] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.

[29] B. LeBaron, W. B. Arthur, and R. Palmer, "Time series properties of an artificial stock market," *Journal of Economic Dynamics and control*, vol. 23, no. 9-10, pp. 1487–1516, 1999.

[30] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.