



# Discovering Bayesian Market Views for Intelligent Asset Allocation

Frank Z. Xing<sup>1</sup>(✉), Erik Cambria<sup>1</sup>, Lorenzo Malandri<sup>2</sup>, and Carlo Verzellis<sup>2</sup>

<sup>1</sup> School of Computer Science and Engineering, Nanyang Technological University, Singapore, Singapore

{zxing001,cambria}@ntu.edu.sg

<sup>2</sup> Data Mining and Optimization Research Group, Politecnico di Milano, Milan, Italy  
{lorenzo.malandri,carlo.verzellis}@polimi.it

**Abstract.** Along with the advance of opinion mining techniques, public mood has been found to be a key element for stock market prediction. However, how market participants' behavior is affected by public mood has been rarely discussed. Consequently, there has been little progress in leveraging public mood for the asset allocation problem, which is preferred in a trusted and interpretable way. In order to address the issue of incorporating public mood analyzed from social media, we propose to formalize public mood into market views, because market views can be integrated into the modern portfolio theory. In our framework, the optimal market views will maximize returns in each period with a Bayesian asset allocation model. We train two neural models to generate the market views, and benchmark the model performance on other popular asset allocation strategies. Our experimental results suggest that the formalization of market views significantly increases the profitability (5% to 10% annually) of the simulated portfolio at a given risk level.

**Keywords:** Market views · Public mood · Asset allocation

## 1 Introduction

Sales and macroeconomic factors are some of the driving forces behind stock movements but there are many others. For example, the subjective views of market participants also have important effects. Along with the growing popularity of social media in the past decades, people tend to rapidly express and exchange their thoughts and opinions [21]. As a result, the importance of their views has dramatically risen [6]. Currently, stock movements are considered to be essentially affected by new information and the beliefs of investors [17].

Meanwhile, sentiment analysis has emerged as a new tool for analyzing the opinions shared on social media [7]. It is a branch of affective computing research that aims to classify natural language utterances as either positive or negative,

---

A public available dataset is released at <https://github.com/fxing79/ibaa>.

© Springer Nature Switzerland AG 2019

U. Brefeld et al. (Eds.): ECML PKDD 2018, LNAI 11053, pp. 120–135, 2019.

[https://doi.org/10.1007/978-3-030-10997-4\\_8](https://doi.org/10.1007/978-3-030-10997-4_8)

but sometimes also neutral [9]. In the financial domain, sentiment analysis is frequently used to obtain a data stream of public mood toward a company, stock, or the economy. Public mood is the aggregation of individual sentiments which can be obtained and estimated from various sources, such as stock message boards [2, 19], blogs, newspapers, and really simple syndication (RSS) feeds [34].

Recently, Twitter has become a dominant microblogging platform on which many works rely for their investigations, such as [20, 23, 27]. Many previous studies support the claim that public mood helps to predict the stock market. For instance, the fuzzy neural network model considering public mood achieves high directional accuracy in predicting the market index. The mood time series is also proved a Granger cause of the market index [4]. Si et al. build a topic-based sentiment time series and predict the market index better with a vector autoregression model to interactively link the two series [26]. The Hurst exponents also suggest a long-term dependency for time series of mood extracted from financial news, similar to many market indices [8].

Despite the important role in stock market prediction, we assume that public mood does not directly effect the market: it does *indirectly* through market participants' views. The actions taken by market participants as agents, are dependent on their own views, and their knowledge about other agents' views. The changes of asset prices are the consequences of such actions. These assumptions are very different from econometric research using productivity, equilibrium, and business cycle models [1], but closer to agent-based models [14]. However, the mechanism of how market views are formed from public mood is heavily overlooked even in the latter case. An intuitive hypothesis could be: the happier the public mood, the higher the stock price. In the real-world market, however, this relationship is far more complicated. Therefore, existing superficial financial applications of AI do not appear convincing to professionals.

In this paper, we attempt to fill this gap by proposing a method for incorporating public mood to form market views computationally. To validate the quality of our views, we simulate the trading performance with a constructed portfolio. The key *contributions* of this paper can be summarized as follows:

1. We introduce a stricter and easier-to-compute definition of the market views based on a Bayesian asset allocation model. We prove that our definition is compatible, and has the equivalent expressiveness as the original form.
2. We propose a novel online optimization method to estimate the expected returns by solving temporal maximization problem of portfolio returns.
3. Our experiments show that the portfolio performance with market views blending public mood data stream is *better* than directly training a neural trading model without views. This superiority is robust for different models selected with the right parameters to generate market views.

The remainder of the paper is organized as follows: Sect. 2 explains the concept of Bayesian asset allocation; following, we describe the methodologies developed for modeling market views in Sect. 3; we evaluate such methodologies by running trading simulations with various experimental settings in Sect. 4 and show the

interpretability of our model with an example in Sect. 5; finally, Sect. 6 concludes the paper and describes future work.

## 2 Bayesian Asset Allocation

The portfolio construction framework [18] has been a prevalent model for investment for more than half a century. Given the an amount of initial capital, the investor will need to allocate it to different assets. Based on the idea of trading-off between asset returns and the risk taken by the investor, the mean-variance method proposes the condition of an efficient portfolio as follows [18, 29]:

$$\begin{aligned}
 & \text{maximize} && \overbrace{\sum_{i=1}^N \mu_i w_i}^{\text{return item}} - \frac{\delta}{2} \overbrace{\sum_{i=1}^N \sum_{j=1}^N w_i \sigma_{ij} w_j}^{\text{risk item}} && (1) \\
 & \text{subject to} && \sum_{i=1}^N w_i = 1, \quad i = 1, 2, \dots, N. \quad w_i \geq 0.
 \end{aligned}$$

where  $\delta$  is an indicator of risk aversion,  $w_i$  denotes the weight of the corresponding asset in the portfolio,  $\mu_i$  denotes the expected return of asset  $i$ ,  $\sigma_{ij}$  is the covariance between returns of asset  $i$  and  $j$ . The optimized weights of an efficient portfolio is therefore given by the first order condition of Eq. 1:

$$w^* = (\delta \Sigma)^{-1} \mu \tag{2}$$

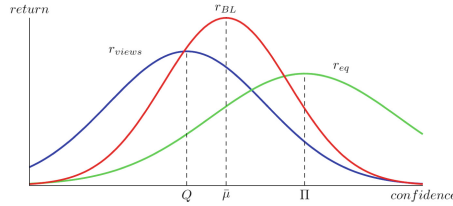
where  $\Sigma$  is the covariance matrix of asset returns and  $\mu$  is a vector of expected returns  $\mu_i$ . At the risk level of holding  $w^*$ , the efficient portfolio achieves the maximum combinational expected return.

However, when applying this mean-variance approach in real-world cases, many problems are faced. For example, the two moments of asset returns are difficult to estimate accurately [25], as they are non-stationary time series. The situation is worsened by the fact that, the Markowitz model is very sensitive to the estimated returns and volatility as inputs. The optimized weights can be very different because of a small error in  $\mu$  or  $\Sigma$ . To address the limitation of the Markowitz model, a Bayesian approach that integrates the additional information of investor’s judgment and the market fundamentals was proposed by Black and Litterman [3]. In the Black-Litterman model, the expected returns  $\mu_{BL}$  of a portfolio is inferred by two antecedents: the equilibrium risk premiums  $\Pi$  of the market as calculated by the capital asset pricing model (CAPM), and a set of views on the expected returns of the investor.

The Black-Litterman model assumes that the equilibrium returns are normally distributed as  $r_{eq} \sim \mathcal{N}(\Pi, \tau \Sigma)$ , where  $\Sigma$  is the covariance matrix of asset returns,  $\tau$  is an indicator of the confidence level of the CAPM estimation of  $\Pi$ . The market views on the expected returns held by an investor agent are also normally distributed as  $r_{views} \sim \mathcal{N}(Q, \Omega)$ .

Subsequently, the posterior distribution of the portfolio returns providing the views is also Gaussian. If we denote this distribution by  $r_{BL} \sim \mathcal{N}(\bar{\mu}, \bar{\Sigma})$ , then  $\bar{\mu}$  and  $\bar{\Sigma}$  will be a function of the aforementioned variables (see Fig. 1).

$$[\bar{\mu}, \bar{\Sigma}] = f(\tau, \Sigma, \Omega, \Pi, Q) \quad (3)$$



**Fig. 1.** The posterior distribution of the expected returns as in the Black-Litterman model, which has a mean between two prior distributions and a variance less than both of them.

The function can be induced from applying Bayes' theorem on the probability density function of the posterior expected returns:

$$pdf(\bar{\mu}) = \frac{pdf(\bar{\mu}|\Pi) pdf(\Pi)}{pdf(\Pi|\bar{\mu})} \quad (4)$$

Then, the optimized Bayesian portfolio weights have a similar form to Eq. 2, only substituting  $\Sigma$  and  $\mu$  by  $\bar{\Sigma}$  and  $\bar{\mu}$ :

$$w_{BL}^* = (\delta \bar{\Sigma})^{-1} \bar{\mu}. \quad (5)$$

The most common criticism of the Black-Litterman model is the subjectivity of investor's views. In other words, the model resorts to the good quality of the market views, while it leaves the question of how to actually form these views unanswered. In Sect. 3, we will investigate the possibility of automatically formalizing the market views from public mood distilled from the Web and the maximization of portfolio returns for each time period.

## 3 Methodologies

### 3.1 Modeling Market Views

The Black-Litterman model defines a view as a statement that the expected return of a portfolio has a normal distribution with mean equal to  $q$  and a standard deviation given by  $\omega$ . This hypothetical portfolio is called a *view portfolio* [13]. In practice, there are two intuitive types of views on the market, termed

*relative views* and *absolute views*, that we are especially interested in. Next, we introduce the formalization of these two types of views.

Because the standard deviation  $\omega$  can be interpreted as the confidence of expected return of the view portfolio, a *relative view* takes the form of “I have  $\omega_1$  confidence that asset  $x$  will outperform asset  $y$  by  $a\%$  (in terms of expected return)”; an *absolute view* takes the form of “I have  $\omega_2$  confidence that asset  $z$  will outperform the (whole) market by  $b\%$ ”. Consequently, for a portfolio consisting of  $n$  assets, a set of  $k$  views can be represented by three matrices  $P_{k,n}$ ,  $Q_{k,1}$ , and  $\Omega_{k,k}$ .

$P_{k,n}$  indicates the assets mentioned in views. The sum of each row of  $P_{k,n}$  should either be 0 (for relative views) or 1 (for absolute views);  $Q_{k,1}$  is a vector comprises expected returns for each view. Mathematically, the confidence matrix  $\Omega_{k,k}$  is a measure of covariance between the views. The Black-Litterman model assumes that the views are independent of each other, so the confidence matrix can be written as  $\Omega = \text{diag}(\omega_1, \omega_2, \dots, \omega_n)$ . In fact, this assumption will not affect the expressiveness of the views as long as the  $k$  views are compatible (not self-contradictory). Because when  $\Omega_{k,k}$  is not diagonal, we can always do spectral decomposition:  $\Omega = V\Omega^A V^{-1}$ . Then we write the new mentioning and new expected return matrices as  $P^A = V^{-1}P$ ,  $Q^A = V^{-1}Q$ , where  $\Omega^A$  is diagonal. Under these constructions, we introduce two important properties of the view matrices in Theorems 1 and 2.

**Theorem 1 (Compatibility of Independent Views).** *Any set of independent views are compatible.*

*Proof.* Compatible views refer to views that can hold at the same time. For example, {asset  $x$  will outperform asset  $y$  by 3%, asset  $y$  will outperform asset  $z$  by 5%, asset  $x$  will outperform asset  $z$  by 8%} is compatible. However, if we change the third piece of view to “asset  $z$  will outperform asset  $x$  by 8%”, the view set becomes self-contradictory. Because the third piece of view is actually a deduction from the former two, the view set is called “not independent”.

Assume there is a pair of incompatible views  $\{p, q\}$  and  $\{p, q'\}$ ,  $q \neq q'$ . Both views are either explicitly stated or can be derived from a set of  $k$  views. Hence, there exist two different linear combinations, such that:

$$\begin{aligned} \sum_{i=1}^k a_i p_i &= p & \sum_{i=1}^k a_i q_i &= q \\ \sum_{i=1}^k b_i p_i &= p & \sum_{i=1}^k b_i q_i &= q' \end{aligned}$$

where  $(a_i - b_i)$  are not all zeros.

Thus, we have  $\sum_{i=1}^k (a_i - b_i)p_i = \mathbf{0}$ , which means that matrix  $P$  is rank deficient and the  $k$  views are not independent. According to the law of contrapositive, the statement “all independent view sets are compatible” is true.  $\square$

**Theorem 2 (Universality of Absolute View Matrix).** *Any set of independent relative and absolute views can be expressed with a non-singular absolute view matrix.*

*Proof.* Assume a matrix  $P$  with  $r$  relative views and  $(k - r)$  absolute views.

$$P_{k,n} = \begin{bmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{r,1} & p_{r,2} & \cdots & p_{r,n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{k,1} & p_{k,2} & \cdots & p_{k,n} \end{bmatrix}$$

The corresponding return vector is  $Q = (q_1, q_2, \dots, q_k)$ , the capital weight vector for assets is  $w = (w_1, w_2, \dots, w_k)$ . Hence, we can write  $(r + 1)$  equations with regard to  $r$  new variables  $\{q'_1, q'_2, \dots, q'_r\}$ , where  $j = 1, 2, \dots, r$ :

$$1 + q'_j = \sum_{i \neq j}^r (1 + q'_i) \frac{w_i}{\sum_{s \neq j} w_s} (1 + q_j)$$

$$\sum_{i=1}^r q'_i w_i + \sum_{i=r+1}^k q_i w_i = Q w^\top$$

If we consider  $\{asset_{r+1}, \dots, asset_k\}$  to be one asset, return of this asset is decided by  $P_{r,n}$ . Hence,  $r$  out of the  $(r + 1)$  equations above are independent.

According to Cramer’s rule, there exists a unique solution  $Q' = (q'_1, q'_2, \dots, q'_r, q_{r+1}, \dots, q_k)$  to the aforementioned  $(r + 1)$  equations, such that view matrices  $\{P', Q'\}$  is equivalent to view matrices  $\{P, Q\}$  for all the assets considered, where

$$P'_{k,n} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & p_{r,r} = 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ p_{k,1} & p_{k,2} & \cdots & p_{k,n} \end{bmatrix}.$$

Now,  $P'_{k,n}$  only consists of absolute views. By deleting those dependent views, we can have a non-singular matrix that only consists of absolute views and is compatible.  $\square$

Given Theorems 1 and 2, without loss of generality, we can use the following equivalent yet stricter definition of market views to reduce computational complexity.

**Definition 1.** *Market views on  $n$  assets can be represented by three matrices  $P_{n,n}$ ,  $Q_{n,1}$ , and  $\Omega_{n,n}$ , where  $P_{n,n}$  is an identity matrix;  $Q_{n,1} \in \mathbb{R}^n$ ;  $\Omega_{n,n}$  is a nonnegative diagonal matrix.*

### 3.2 The Confidence Matrix

In the most original form of the Black-Litterman model, the confidence matrix  $\Omega$  is set manually according to investors' experience. Whereas in the numerical example given by [13], the confidence matrix is derived from the equilibrium covariance matrix:

$$\hat{\Omega}_0 = \text{diag}(P(\tau\Sigma)P') \quad (8)$$

This is because  $P(\tau\Sigma)P'$  can be understood as a covariance matrix of the expected returns in the views as well. Using our definition, it is easier to understand this estimation, because  $P$  is an identity matrix,  $P(\tau\Sigma)P'$  is already diagonal. The underlying assumption is that the variance of an absolute view on asset  $i$  is proportional to the volatility of asset  $i$ . In this case, the estimation of  $\Omega$  utilizes past information of asset price volatilities.

### 3.3 Optimal Market Views

We obtain the optimal market views  $\{P, Q, \Omega\}$  in a hybrid way, first we adopt the confidence matrix  $\hat{\Omega}_0$ , then  $Q$  can be derived from the inverse optimization problem using the Black-Litterman model.

We start from the optimal portfolio weights that maximize the portfolio returns for each period  $t$ . Obviously, without short selling and transaction fees, one should re-invest his whole capital daily to the fastest-growing asset in the next time period.

The optimal holding weights for each time period  $t$  thus take the form of a one-hot vector, where  $\oslash$  and  $\odot$  denote element-wise division and product:

$$w_t^* = \operatorname{argmax} w_t \oslash \text{price}_t \odot \text{price}_{t+1} \quad (9)$$

Let this  $w_t^*$  be the solution to Eq. 1, we will have:

$$w_t^* = (\delta \bar{\Sigma}_t)^{-1} \bar{\mu}_t \quad (10)$$

where the Black-Litterman model gives<sup>1</sup>:

$$\bar{\Sigma}_t = \Sigma_t + [(\tau\Sigma_t)^{-1} + P' \hat{\Omega}_t^{-1} P]^{-1} \quad (11)$$

$$\bar{\mu}_t = [(\tau\Sigma_t)^{-1} + P' \hat{\Omega}_t^{-1} P]^{-1} [(\tau\Sigma_t)^{-1} \Pi_t + P' \hat{\Omega}_t^{-1} Q_t] \quad (12)$$

According to Eqs. 10, 11, and 12, the optimal expected returns for our market views for each period  $t$  is:

$$\begin{aligned} Q_t^* &= \hat{\Omega}_{0,t} \{ [(\tau\Sigma_t)^{-1} + P' \hat{\Omega}_{0,t}^{-1} P] \bar{\mu}_t - (\tau\Sigma_t)^{-1} \Pi_t \} \\ &= \delta [ \hat{\Omega}_{0,t} (\tau\Sigma_t)^{-1} + \mathbb{I} ] \bar{\Sigma}_t w_t^* - \hat{\Omega}_{0,t} (\tau\Sigma_t)^{-1} \Pi_t \\ &= \delta [ \hat{\Omega}_{0,t} (\tau\Sigma_t)^{-1} + \mathbb{I} ] [ \Sigma_t + [(\tau\Sigma_t)^{-1} + \hat{\Omega}_t^{-1}]^{-1} ] w_t^* \\ &\quad - \hat{\Omega}_{0,t} (\tau\Sigma_t)^{-1} \Pi_t \end{aligned} \quad (13)$$

<sup>1</sup> The proof of Eqs. 11 and 12 can be found from the appendix of [24].

### 3.4 Generating Market Views with Neural Models

Equation 13 provides a theoretical perspective on determining the expected return of optimal market views. However, computing  $w_t^*$  requires future asset prices, which is not accessible. Therefore, the feasible approach is to learn approximating  $Q_t^*$  with historical data and other priors as input. We use the time series of asset prices, trading volumes, and public mood data stream to train neural models ( $nn$ ) for this approximation problem of optimal market views:

$$\hat{Q}_t = nn(prices, volumes, sentiments; Q_t^*) \tag{14}$$

We denote the time series of asset prices  $price_{t-k}, price_{t-k+1}, \dots, price_t$  by a lag operator  $\mathcal{L}^{0\sim k}price_t$ . The notation of trading volumes follows a similar form. Then the model input at each time point:  $[\mathcal{L}^{0\sim k}price_t, \mathcal{L}^{0\sim k}volume_t, sentiment_t, capital_t]$  can be denoted by  $[p, v, s, c]$  in short.

Two types of neural models, including a neural-fuzzy approach and a deep learning approach are trained for comparison. Figure 2 provides an illustration of the online training process using a long short-term memory (LSTM) network, where  $\hat{Q}$  is the output.

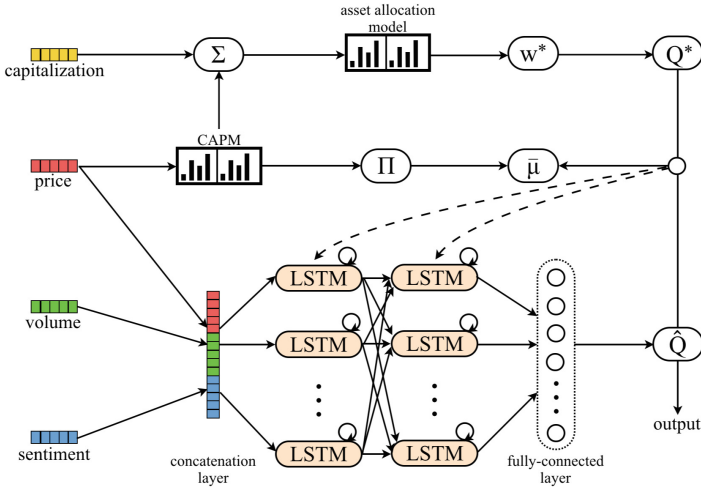


Fig. 2. Model training process (LSTM) with/without sentiment information.

**Dynamic Evolving Neural-Fuzzy Inference System (DENFIS)** is a neural network model with fuzzy rule nodes [16]. The partitioning of which rule nodes to be activated is dynamically updated with the new distribution of incoming data. This evolving clustering method (ECM) features the model with stability and fast adaptability. Comparing to many other fuzzy neural networks, DENFIS performs better in modeling nonlinear complex systems [32].



Considering the financial market as a real-world complex system, we learn the first-order Takagi-Sugeno-Kang type rules online. Each rule node has the form of:

$$\begin{aligned} & \text{IF } \mathcal{L}^{0 \sim k} \text{ attribute}_{t,i} = \text{pattern}_i, i = 1, 2, \dots, N \\ & \text{THEN } \hat{Q}_t = f_{1,2,\dots,N}([p, v, s]_t) \end{aligned}$$

where we have 3 attributes and  $(2^N - 1)$  candidate functions to activate. In our implementation of the DENFIS model, all the membership functions are symmetrical and triangular, which can be defined by two parameters  $b \pm d/2$ .  $b$  is where the membership degree equals to 1;  $d$  is the activation range of the fuzzy rule. In our implementation,  $b$  is iteratively updated by linear least-square estimator of existing consequent function coefficients.

**LSTM** is a type of recurrent neural network with gated units. This unit architecture is claimed to be well-suited for learning to predict time series with an unknown size of lags and long-term event dependencies. Early attempts, though not very successful [11], have been made to apply LSTM to time series prediction. It is now recognized that though LSTM cells can have many variants, their performance across different tasks are similar [12].

Therefore, we use a vanilla LSTM unit structure. Our implementation of LSTM cells follows the update rules of the input gate, forget gate, and output gate as in Eq. 15:

$$\begin{aligned} i_t &= \sigma(W_i \cdot [h_{t-1}, [p, v, s]_t] + b_i) \\ f_t &= \sigma(W_f \cdot [h_{t-1}, [p, v, s]_t] + b_f) \\ o_t &= \sigma(W_o \cdot [h_{t-1}, [p, v, s]_t] + b_o) \end{aligned} \quad (15)$$

where  $\sigma$  denotes the sigmoid function,  $h_{t-1}$  is the output of the previous state,  $W$  is a state transfer matrix, and  $b$  is the bias.

The state of each LSTM cell  $c_t$  is updated by:

$$\begin{aligned} c_t &= f_t \odot c_{t-1} + i_t \odot (W_c \cdot [h_{t-1}, [p, v, s]_t] + b_c) \\ h_{t-1} &= o_t \odot \tanh(c_{t-1}) \end{aligned} \quad (16)$$

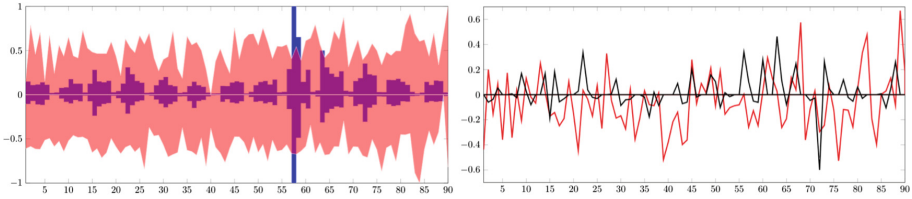
We make the training process online as well, in a sense that each time a new input is received, we use the previous states and parameters of LSTM cells  $[c_{t-1}, \mathbf{W}, \mathbf{b}]$  to initialize the LSTM cells for period  $t$ .

## 4 Experiments

To evaluate the quality and effectiveness of our formalization of market views, we run trading simulations with various experimental settings.

## 4.1 Data

The data used in this study are publicly available on the Web. We obtain the historical closing price of stocks and daily trading volumes from the Quandl API<sup>2</sup>; the market capitalization data from Yahoo! Finance; the daily count and intensity of company-level sentiment time series from PsychSignal<sup>3</sup>. The sentiment intensity scores are computed from multiple social media platforms using NLP techniques. Figure 3 depicts a segment example of the public mood data stream. The market is closed on weekends, so a corresponding weekly cycle of message volume can be observed.



**Fig. 3.** The volume of daily tweets filtered by cashtag AAPL (blue, left); average sentiment intensity (red, left); net sentiment polarity (red, right); daily returns (black, right) in a time period of 90 days (2017-03-04 to 2017-06-04). All the series are normalized. (Color figure online)

We investigate a window of around 8 years (2800 days). All the time series are trimmed from 2009-10-05 to 2017-06-04. For missing values such as the closing prices on weekends and public holidays, we fill them with the nearest historical data to train the neural models. The lagged values we use for both price and trading volume consist of 4 previous days and a moving average of the past 30 days, that is, the input of our neural models takes the form of Eqs. 17 and 18:

$$\mathcal{L}^{0\sim k} price_t = (p_t, p_{t-1}, p_{t-2}, p_{t-3}, \frac{\sum_{i=1}^{30} p_i}{30}) \quad (17)$$

$$\mathcal{L}^{0\sim k} volume_t = (v_t, v_{t-1}, v_{t-2}, v_{t-3}, \frac{\sum_{i=1}^{30} v_i}{30}) \quad (18)$$

## 4.2 Trading Simulation

We construct a virtual portfolio consisting of 5 big-cap stocks: Apple Inc (AAPL), Goldman Sachs Group Inc (GS), Pfizer Inc (PFE), Newmont Mining Corp (NEM), and Starbucks Corp (SBUX). This random selection covers both the NYSE and NASDAQ markets and diversified industries, such as technology, financial services, health care, consumer discretionary etc. During the period investigated,

<sup>2</sup> <http://www.quandl.com/tools/api>.

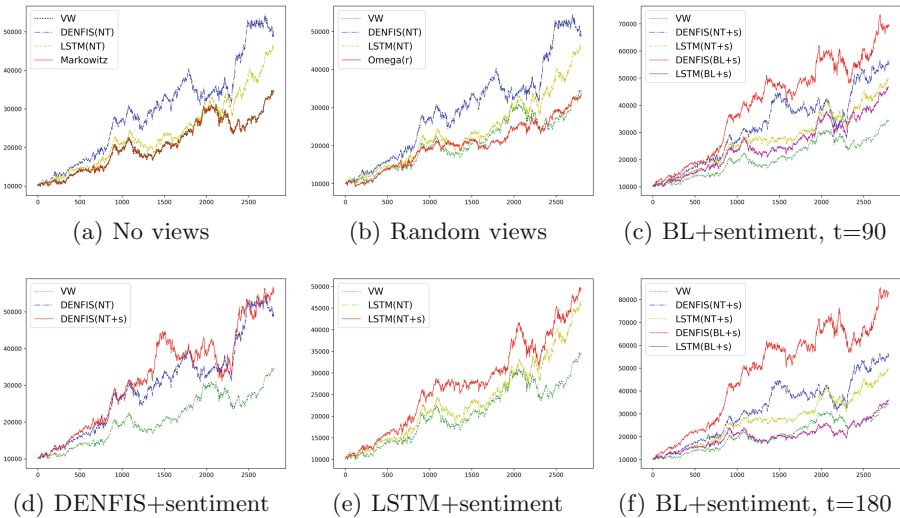
<sup>3</sup> <http://psychsignal.com>.

there were two splits: a 7-for-1 split for AAPL on June 9th 2014, and a 2-for-1 split for SBUX on April 9th 2015. The prices per share are adjusted according to the current share size for computing all related variables, however, dividends are not taken into account. We benchmark our results with two portfolio construction strategies:

(1) **The value-weighted portfolio (VW):** we re-invest daily according to the percentage share of each stock’s market capitalization. In this case, the portfolio performance will be the weighted average of each stock’s performance. This strategy is fundamental, yet empirical study [10] shows that beating the market even before netting out fees is difficult.

(2) **The neural trading portfolio (NT):** we remove the construction of market views and directly train the optimal weights of daily position with the same input. For this black-box strategy, we can not get any insight on how this output portfolio weight comes about.

In the simulations, we assume no short selling, taxes, or transaction fees, and we assume the portfolio investments are infinitely divisible, starting from 10,000 dollars. We construct portfolios with no views ( $\Omega_\emptyset$ , in this case the degenerate portfolio is equivalent to Markowitz’s mean-variance portfolio using historical return series to estimate covariance matrix as a measure of risk), random views ( $\Omega_r$ ), the standard views using the construction of Black-Litterman model ( $\Omega_0$ ), with and without our sentiment-induced expected returns ( $s$ ). The trading performances are demonstrated in Fig. 4.



**Fig. 4.** Trading simulation performance with different experimental settings: (x-axis: number of trading days; y-axis: cumulative returns). In particular, we use a timespan of 90 and 180 days for our approach. The performance of neural trading is independent from timespan, accordingly the two neural models are compared in (d) and (e) respectively for better presentation.

Following the previous research [13], we set the risk aversion coefficient  $\delta = 0.25$  and confidence level of CAPM,  $\tau = 0.05$ . Let the activation range of fuzzy membership function  $d = 0.21$ , we obtain 21 fuzzy rule nodes from the whole online training process of DENFIS. This parameter minimizes the global portfolio weight error. For the second neural model using deep learning, we stack two layers of LSTMs followed by a densely connected layer. Each LSTM layer has 3 units; the densely connected layer has 50 neurons, which is set times larger than the number of LSTM units. We use the mean squared error of vector  $Q$  as the loss function and the rmsprop optimizer [30] to train this architecture. We observe fast training error convergence in our experiments.

### 4.3 Performance Metrics

Diversified metrics have been proposed to evaluate the performance of a given portfolio [5, 15, 31]. We report four metrics in our experiments.

Root mean square error (RMSE) is a universal metric for approximation problems. It is widely used for engineering and data with normal distribution and few outliers. We calculate the RMSE of our realized portfolio weights to the optimal weights:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n \|w_i - \hat{w}_i\|^2} \quad (19)$$

Annualized return (AR) measures the profitability of a given portfolio. We calculate the geometric mean growth rate per year, which is also referred to as compound annual growth rate (CAGR) for these 2800 days.

Sharpe ratio (SR) is a risk-adjusted return measure. We choose the value-weighted portfolio as a base, consequently the Sharpe ratio of VW will be 1:

$$\text{SR} = \frac{\mathbb{E}(R_{portfolio}/R_{VW})}{\sigma(R_{portfolio})/\sigma(R_{VW})} \quad (20)$$

SR uses the standard deviation of daily returns as the measure of risk. Note that to distinguish between good and bad risk, we can also use the standard deviation of downside returns only [28]. Our results suggest that the Sortino ratios, which are not reported due to page limit, are very close to SRs and lead to the same conclusion.

The maximum drawdown (MDD) measures the maximum possible percentage loss of an investor:

$$\text{MDD} = \max_{0 < t < \tau} \left\{ \frac{\text{Value}_t - \text{Value}_\tau}{\text{Value}_t} \right\} \quad (21)$$

Asset allocation strategies with large MDD are exposed to the risk of withdrawal. Table 1 presents the metrics.

**Table 1.** Performance metrics for various portfolio construction strategies, timespan = 90 and 180 days. Top three metrics are in bold.

	RMSE	SR	MDD(%)	AR(%)
VW	0.8908	1.00	25.81	17.49
Markowitz90( $\Omega_\emptyset$ )	0.9062	1.00	25.81	17.51
Markowitz180( $\Omega_\emptyset$ )	0.8957	1.00	25.82	17.45
BL90( $\Omega_r$ )	0.9932	0.90	<b>23.47</b>	17.17
BL180( $\Omega_r$ )	0.9717	1.06	<b>20.59</b>	22.31
DENFIS(NT)	0.9140	<b>2.94</b>	29.84	23.09
DENFIS(NT+s)	0.9237	<b>4.35</b>	<b>23.07</b>	<b>25.16</b>
DENFIS(BL90+s)	0.9424	1.52	24.44	<b>28.69</b>
DENFIS(BL180+s)	0.9490	<b>1.58</b>	24.19	<b>29.49</b>
LSTM(NT)	<b>0.8726</b>	1.38	25.68	22.10
LSTM(NT+s)	0.8818	1.42	25.96	23.21
LSTM(BL90+s)	<b>0.8710</b>	1.34	25.90	22.33
LSTM(BL180+s)	<b>0.8719</b>	1.07	24.88	17.68

#### 4.4 Findings

We have some interesting observations from Fig. 4 and Table 1. SR and AR are usually considered as the most important, and besides, RMSE and MDD are all very close in our experiments. The correlation between RMSE and the other three metrics is weak, though it is intuitive that if the realized weights are close to the optimal weights, the portfolio performance should be better. On the contrary, the LSTM models seem to overfit as they are trained on the mean squared error of weights or expected return of views [22]. However, as mentioned in Sect. 1, the relationship between weights and daily returns is nonlinear. Therefore, *holding portfolio weights that are close to the optimal weights does not necessarily mean that the AR must be higher*. In fact, it is dangerous to use any seemingly reasonable metrics outside the study of asset allocation, such as directional accuracy of price change prediction [4, 33], to evaluate the expected portfolio performance.

The Markowitz portfolio ( $\Omega_\emptyset$ ) displays a very similar behavior to the market-following strategy. This is consistent with the inefficacy of the mean-variance approach in practice mentioned by previous studies: holding the Markowitz portfolio is holding the market portfolio. In fact, if the CAPM holds, the market portfolio already reflects the adjustments to risk premiums, that is, fewer market participants will invest on highly risky assets, for this reason their market capitalization will be smaller as well.

However, the Black-Litterman model does not always guarantee better performance over the Markowitz portfolio. “Garbage in, garbage out” still holds for this circumstance. Given random views ( $\Omega_r$ ), it can be worse than market-following in terms of both SR and AR. The lesson learned is that *if the investor*

*knows nothing, it is better to hold no views and follow the market than pretending to know something.*

In our experiments, DENFIS generally performs better than LSTM models, achieving higher SRs and ARs. The reason may be LSTM models adapt faster to the incoming data, whereas financial time series are usually very noisy. The ECM mechanism provides DENFIS models with converging learning rates, which may be beneficial to the stability of memorized rules. However, it is important to note that *the ARs for both neural models improve with the blending of sentiments*. The timespan used to estimate correlation and volatility of assets seems not that critical. DENFIS models perform better with longer timespan, while LSTM models perform better with shorter timespan. The Markowitz portfolio is less affected by timespan.

## 5 A Story

One of the main advantages of our formalization and computing of market views is that some *transparency* is brought to the daily asset reallocation decisions. In most cases, a stock price prediction system based on machine learning algorithms cannot justify “why he thinks that price will reach that predicted point”. Unlike these systems, our method can tell a story of the portfolio to professional investors and advice seekers. Take June 1st 2017 as an example:

“On June 1st 2017, we observe 164 positive opinions of polarity +1.90, 58 negative opinions of polarity -1.77 on AAPL stock; 54 positive opinions of polarity +1.77, 37 negative opinions of polarity -1.53 on GS stock; 5 positive opinions of polarity +2.46, 1 negative opinion of polarity -1.33 on PFE stock; no opinion on NEM stock; and 9 positive opinions of polarity +1.76, 5 negative opinions of polarity -2.00 on SBUX stock. Given the historical prices and trading volumes of the stocks, we have 6.29% confidence that AAPL will outperform the market by -70.11%; 23.50% confidence that GS will outperform the market by 263.28%; 0.11% confidence that PFE will outperform the market by -0.50%; 1.21% confidence that SBUX will outperform the market by 4.57%. Since our current portfolio invests 21.56% on AAPL, 25.97% on GS, 29.43% on PFE, and 23.04% on SBUX, by June 2nd 2017, we should withdraw all the investment on AAPL, 2.76% of the investment on GS, 81.58% of the investment on PFE, and 30.77% of the investment on SBUX, and re-invest them onto NEM.”

## 6 Conclusion and Future Work

In previous studies which have considered sentiment information for financial forecasting, the role of the investor as a market participant is often absent. In this paper, we present a novel approach to incorporate market sentiment by fusing public mood data stream into the Bayesian asset allocation framework.

This work is pioneering in formalizing sentiment-induced market views. Our experiments show that the market views provide a powerful method to asset

management. We also confirm the efficacy of public mood data stream based on social media for developing asset allocation strategies.

A limitation of this work is that we fixed a portfolio with five assets, though in practice the portfolio selection problem is of equal importance. How to assess the quality of sentiment data is not discussed in this paper as well. We are not at the stage to distinguish or detect opinion manipulation though concern like the open networks are rife with bots does exist. Another limitation is that survivor bias is not taken into account: the risk that assets selected in the portfolio may quit the market or suffer from a lack of liquidity. This problem can be alleviated by only including high quality assets. In the future, we will study examining the quality of sentiment data obtained using different content analysis approaches. We also plan to develop a Bayesian asset allocation model that can deal with market frictions.

## References

1. Angelatos, G., La'O, J.: Sentiments. *Econometrica* **81**(2), 739–779 (2013)
2. Antweiler, W., Frank, M.Z.: Is all that talk just noise? The information content of internet stock message boards. *J. Finance* **59**(3), 1259–1294 (2004)
3. Black, F., Litterman, R.: Asset allocation: combining investor view with market equilibrium. *J. Fixed Income* **1**, 7–18 (1991)
4. Bollen, J., Mao, H., Zeng, X.: Twitter mood predicts the stock market. *J. Comput. Sci.* **2**(1), 1–8 (2011)
5. Brandt, M.W.: Portfolio choice problems. In: *Handbook of Financial Econometrics*, vol. 1, chap. 5, pp. 269–336. Elsevier B.V., Oxford (2009)
6. Cambria, E.: Affective computing and sentiment analysis. *IEEE Intell. Syst.* **31**(2), 102–107 (2016)
7. Cambria, E., Das, D., Bandyopadhyay, S., Feraco, A. (eds.): *A Practical Guide to Sentiment Analysis*. Springer International Publishing, Switzerland (2017). <https://doi.org/10.1007/978-3-319-55394-8>
8. Chan, S.W., Chong, M.W.: Sentiment analysis in financial texts. *Decis. Support Syst.* **94**, 53–64 (2017)
9. Chaturvedi, I., Ragusa, E., Gastaldo, P., Zunino, R., Cambria, E.: Bayesian network based extreme learning machine for subjectivity detection. *J. Frankl. Inst.* **355**(4), 1780–1797 (2018)
10. Fama, E.F., French, K.R.: Luck versus skill in the cross-section of mutual fund returns. *J. Financ.* **65**(5), 1915–1947 (2010)
11. Gers, F.A., Eck, D., Schmidhuber, J.: Applying LSTM to time series predictable through time-window approaches. In: Dorffner, G., Bischof, H., Hornik, K. (eds.) *ICANN 2001. LNCS*, vol. 2130, pp. 669–676. Springer, Heidelberg (2001). [https://doi.org/10.1007/3-540-44668-0\\_93](https://doi.org/10.1007/3-540-44668-0_93)
12. Greff, K., Srivastava, R.K., Koutnik, J., Steunebrink, B.R., Schmidhuber, J.: LSTM: a search space odyssey. *IEEE Trans. Neural Netw. Learn. Syst.* **28**(10), 2222–2232 (2017)
13. He, G., Litterman, R.: The intuition behind black-litterman model portfolios. Goldman Sachs Working Paper (1999). <https://doi.org/10.2139/ssrn.334304>
14. Hommes, C.: *The New Palgrave Dictionary of Economics. Interacting Agents in Finance*, 2nd edn. Palgrave Macmillan, Basingstoke (2008)

15. Hyndman, R.J., Koehler, A.B.: Another look at measures of forecast accuracy. *Int. J. Forecast.* **22**(4), 679–688 (2006)
16. Kasabov, N.K., Song, Q.: Denfis: dynamic evolving neural-fuzzy inference system and its application for time-series prediction. *IEEE Trans. Fuzzy Syst.* **10**, 144–154 (2002)
17. Li, Q., Jiang, L., Li, P., Chen, H.: Tensor-based learning for predicting stock movements. In: *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, pp. 1784–1790 (2015)
18. Markowitz, H.: Portfolio selection. *J. Finance* **7**, 77–91 (1952)
19. Nguyen, T.H., Shirai, K.: Topic modeling based sentiment analysis on social media for stock market prediction. In: *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pp. 1354–1364 (2015)
20. Nofer, M., Hinze, O.: Using twitter to predict the stock market: where is the mood effect? *Bus. Inf. Syst. Eng.* **57**(4), 229–242 (2015)
21. O’Connor, B., Balasubramanyan, R., Routledge, B.R., Smith, N.A.: From tweets to polls: linking text sentiment to public opinion time series. In: *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, pp. 122–129 (2010)
22. Pant, P.N., Starbuck, W.H.: Innocents in the forest: forecasting and research methods. *J. Manag.* **16**(2), 433–460 (1990)
23. Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., Mozetič, I.: The effects of twitter sentiment on stock price returns. *PLoS One* **10**(9), 1–21 (2015)
24. Satchell, S., Scowcroft, A.: A demystification of the black-litterman model: managing quantitative and traditional portfolio construction. *J. Asset Manag.* **1**(2), 138–150 (2000)
25. Shen, W., Wang, J.: Portfolio selection via subset resampling. In: *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, pp. 1517–1523 (2017)
26. Si, J., Mukherjee, A., Liu, B., Li, Q., Li, H., Deng, X.: Exploiting topic based twitter sentiment for stock prediction. In: *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pp. 24–29 (2013)
27. Smailović, J., Grčar, M., Lavrač, N., Žnidaršič, M.: Predictive sentiment analysis of tweets: a stock market application. In: Holzinger, A., Pasi, G. (eds.) *HCI-KDD 2013. LNCS*, vol. 7947, pp. 77–88. Springer, Heidelberg (2013). [https://doi.org/10.1007/978-3-642-39146-0\\_8](https://doi.org/10.1007/978-3-642-39146-0_8)
28. Sortino, F.A., Price, L.N.: Performance measurement in a downside risk framework. *J. Invest.* **3**, 59–64 (1994)
29. Steinbach, M.C.: Markowitz revisited: mean-variance models in financial portfolio analysis. *SIAM Rev.* **43**(1), 31–85 (2001)
30. Tieleman, T., Hinton, G.E.: Lecture 6.5-RMSProp: divide the gradient by a running average of its recent magnitude. *COURSERA: Neural Netw. Mach. Learn.* **4**, 26–31 (2012)
31. Xing, F.Z., Cambria, E., Welsch, R.E.: Natural language based financial forecasting: a survey. *Artif. Intell. Rev.* **50**(1), 49–73 (2018)
32. Xing, F.Z., Cambria, E., Zou, X.: Predicting evolving chaotic time series with fuzzy neural networks. In: *International Joint Conference on Neural Networks*, pp. 3176–3183 (2017)
33. Yoshihara, A., Seki, K., Uehara, K.: Leveraging temporal properties of news events for stock market prediction. *Artif. Intell. Res.* **5**(1), 103–110 (2016)
34. Zhang, W., Skiena, S.: Trading strategies to exploit blog and news sentiment. In: *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, pp. 375–378 (2010)