

Intelligent Asset Allocation via Market Sentiment Views

Frank Z. Xing School of Computer Science and Engineering, Nanyang Technological University, Singapore, SINGAPORE

Erik Cambria School of Computer Science and Engineering, Nanyang Technological University, Singapore, SINGAPORE

Roy E. Welsch MIT Sloan School of Management, Massachusetts Institute of Technology, Cambridge, USA

Abstract—The sentiment index of market participants has been extensively used for stock market prediction in recent years. Many financial information vendors also provide it as a service. However, utilizing market sentiment under the asset allocation framework has been rarely discussed. In this article, we investigate the role of market sentiment in an asset allocation problem. We propose to compute sentiment time series from social media with the help of sentiment analysis and text mining techniques. A novel neural network design, built upon an ensemble of evolv-

Digital Object Identifier 10.1109/MCI.2018.2866727 Date of publication: 15 October 2018 ing clustering and long short-term memory, is used to formalize sentiment information into market views. These views are later integrated into modern portfolio theory through a Bayesian approach. We analyze the performance of this asset allocation model from many aspects, such as stability of portfolios, computing of sentiment time series, and profitability in our simulations. Experimental results show that our model outperforms some of the most successful forecasting techniques. Thanks to the introduction of the evolving clustering method, the estimation accuracy of market views is significantly improved.

Corresponding Author: Erik Cambria (Email: cambria@ntu.edu.sg)

1556-603X/18©2018IEEE

NOVEMBER 2018 | IEEE COMPUTATIONAL INTELLIGENCE MAGAZINE 25

I. Introduction

inancial markets are among the most complex and chaotic dynamic systems in human society. Numerous factors could contribute to the fluctuation of market prices through the bids and offers on assets. In this price formation mechanism, the psychology and behavior of market participants have an important role to play. Public mood is a very efficient and universal variable that reflects the attitudes of market participants. Furthermore, the rise of Web 2.0 applications and the growing popularity of social media have accelerated the spread of information, which brings more importance to the subjective views on the market. Empirical study [1] suggests that current stock price movements in major markets are essentially affected by new information and the beliefs of investors.

Another reason that we believe incorporating the public mood would be beneficial to the stock market prediction task is that this approach brings in public yet incremental information. In contrast, many technical analysts rely solely on mining of the patterns of past price series. In recent trends of applying artificial intelligence techniques, especially machine learning and deep neural networks to stock market prediction, a large part of the computer science community has the same limitations. However, as chaos theory and many cases in [2] suggest, there are no "detectable patterns" as time evolves even for deterministic systems. This does not necessarily mean the current prices reflect all the past information as the efficient-market hypothesis (EMH) suggests, but as the prices are driven by new information, the past patterns fade quickly away. Consequently, the pattern-chasers are always one step behind if they simply build the model with past prices.

Other than the price series, there are models in literature that include macroeconomic variables, such as a company's book value and investment suggested by multi-factor models [3]. However, the problem of these models is that updates of these factors are usually slow. Unlike many economic factors, the public mood can be instantaneously monitored, and estimated as an aggregation of the market sentiments of individuals. Previous studies have investigated various sources of public mood, such as stock message boards [4], microblogging platforms [5], newspapers [6], Really Simple Syndication (RSS) feeds [7] and more [8]. Wuthrich et al. [6] used occurrencesweighted keyword tuples from an expert system to measure the public mood; Zhang and Skiena [7] leveraged word-level positive and negative counts to derive polarity and subjectivity for specific companies; Antweiler and Frank [4] manually labelled some messages to train a Naive Bayes classifier that predicts bullish, bearish, or neither based on the bag-of-words representation of messages; Smailović et al. [5] trained an Support Vector Machine (SVM) based on a large tweet data collection classified by emoticons. Recently, Weichselbraun et al. [8] proposed sentiment analysis of social media stream based on mining knowledge base enriched dependency trees. Though using different techniques from knowledge engineering to machine learning, many of them have reported correlations between public mood

and price movements. Statistical test and simulation results also manifested the predictive power of public mood [9, 10].

Despite the important role in stock market prediction, it is not sufficient or straightforward for an individual to make his investment decision based on a set of public mood data and predicted prices. Because public mood does not directly affect the market: it does indirectly through market participants' views and their consequent behavior. The interaction is often referred as higher order beliefs in game theory. Then, a question that naturally arises is about bridging public mood with market views [11]. However, discussion about the mechanism of how market views are formed from public mood is heavily overlooked in specific scenarios. In this article, we address the problem of incorporating public mood to the asset allocation framework. The market views are formed computationally from the sentiment time series as a prior belief of the investor. Trading simulation and experiments prove the high quality of our approach of formulating market sentiment views. The informational enhancement using this sentiment prior leads to more than 10% annualized portfolio yield on average when compared to various state-of-the-art asset allocation strategies.

The remainder of the article is organized as follows: the next section provides the background of modern portfolio theory, and explains the concept of Bayesian asset allocation; following, we describe modeling market sentiment views and the optimization objectives; next, we present the method for generating sentiment time series; later, we evaluate our methodologies by running trading simulations with various experimental settings; in the end, we discuss our findings and propose concluding remarks.

II. The Asset Allocation Problem

A. The Mean-Variance Method

The portfolio construction paradigm has been prevalent for investment for more than half a century. Given the total amount of capital available as a constraint, the investor will need to allocate it to different assets. Generally, assets that generate higher returns also bear more risk. Based on the idea of trading off between asset returns and the risk taken by the investor, the concept of an "efficient portfolio" was proposed by Markowitz [12, 13]. Consider a one-period model. Assume n assets are selected and the *i*-th is assigned a weight w_i , then the portfolio return will be the weighted mean of expected return for each asset, portfolio risk can be measured by the variance of return vector. Therefore, an "efficient portfolio" meets the following condition:

maximize
$$\sum_{i=1}^{n} \mu_i w_i = \frac{\delta}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} w_i \sigma_{ij} w_j$$
subject to $w_i \ge 0$ and $\sum_{i=1}^{n} w_i = 1, i = 1, 2, ..., n$ (1)

where δ is a risk aversion coefficient, μ_i denotes the expected return on asset *i*, σ_{ij} is the covariance between asset *i* and *j*, *n* is the number of assets.

The optimized weights of an efficient portfolio are then given by the first order condition of equation 1:

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

where $\mu_{(n \times 1)}$ is a vector consisting of expected returns μ_i , and $\Sigma_{(n \times n)}$ is the covariance matrix of asset returns. At the risk level of holding w^* , the efficient portfolio achieves the maximum expected return among all other alternatives, and for all portfolios that have an expected return equal to holding w^* , the efficient portfolio has the minimum risk measure [14].

However, both μ and Σ are unknown in practice. The traditional approach to this problem is to use their estimation $\hat{\mu}$ and $\hat{\Sigma}$ instead, based on the observed past asset prices. Using a time window of length *T*, we can calculate the two maximum likelihood estimators as:

$$\hat{\mu} = \frac{1}{T} \sum_{t=1}^{T} \frac{p_t - p_{t-1}}{p_{t-1}}$$
(3)

and

$$\hat{\Sigma} = \frac{1}{T} \sum_{t=1}^{T} \left(\frac{p_t - p_{t-1}}{p_{t-1}} - \hat{\mu} \right) \left(\frac{p_t - p_{t-1}}{p_{t-1}} - \hat{\mu} \right)'$$
(4)

where p_t is the price vector of length *n* for time point *t*. Noted that the price series is often non-stationary in the realworld, as a result, the estimators are very sensitive to the choice of *T*. The situation is worsened by the fact that, the Markowitz model per se is not stable for the estimators of return and volatility as inputs, because errors propagate during the multiplication of matrices. Consequently, the Markowitz model often delivers many zero positions and an imbalanced portfolio [15].

B. Bayesian Asset Allocation

Many theoretical approaches have been developed to incorporate Bayesian priors [16]. Unlike the original Markowitz model, the Bayesian perspective treats μ and Σ not as fixed numbers, but as random variables. One can only infer their probability distribution function (pdf). Intuitively, the observed sample size T can be included as a so-called "diffuse prior" to indicate the uncertainty of parameter estimation. By doing so, the pdf will be flatter for a smaller T, indicating the wider confidence interval with fewer samples. Therefore, the assets are riskier in a Bayesian framework since parameter uncertainty can be an additional source of risk that always exists. However, with the aforementioned diffuse prior, the optimized vector of weights is just a scalar adjustment of the Markowitz model, which makes little difference in terms of information leveraged. To exhibit the decisive advantage of the Bayesian approach, it is crucial to elicit informative variables [15], which in our case is a sentiment prior.

To elegantly combine the sentiment prior with other market fundamentals, we resort to a specific form of the Bayesian approach proposed by Black and Litterman [17]. In the Black-Litterman model, the probability distribution of portfolio returns is inferred by two antecedents: the equilibrium risk premiums (Π), and a set of views on the expected returns of the investor. Usually, the equilibrium risk premiums are calculated as in the capital asset pricing model (CAPM). CAPM states that for asset *i*, the equilibrium risk premium is proportional to the market premium:

$$\Pi_i = \mu_i - \mu_f = \beta_i (\mu_m - \mu_f) \tag{5}$$

where μ_m is the market expected return, and μ_f is risk-free interest rate.

The Black-Litterman model assumes that the equilibrium returns are normally distributed as $r_{eq} \sim \mathcal{N}(\Pi, \tau \Sigma)$, where Σ is the covariance matrix of asset returns, and τ is an indicator of the confidence level of the CAPM estimation of Π . The market views on the expected returns are also normally distributed as $r_{rieus} \sim \mathcal{N}(Q, \Omega)$. We denote the posterior distribution of the portfolio returns providing the views by r_{BL} , where the subscript stands for the Black-Litterman model. Then, it is mathematically clear that r_{BL} can also be written as a normal distribution $\mathcal{N}(\mu_{BL}, \Sigma_{BL})$, where both μ_{BL} and Σ_{BL} can be derived from the Bayes' theorem:

$$pdf(\mu_{BL}) = \frac{pdf(\mu_{BL} \mid \Pi) \ pdf(\Pi)}{pdf(\Pi \mid \mu_{BL})}.$$
 (6)

With μ_{BL} and Σ_{BL} , the optimized Bayesian portfolio has a similar form to equation 2:

$$w_{BL}^{*} = (\delta \Sigma_{BL})^{-1} \mu_{BL}.$$
 (7)

The next section will discuss how the Black-Litterman model presents the market views in a more natural and human-understandable way, instead of explicitly giving the *pdf* of Bayesian posterior returns.

III. Market Views

Starting from the physical meaning of Q and Ω , the Black-Litterman model defines two types of market views [18]. A relative view takes the form of "I have ω_1 confidence that asset x will outperform asset γ by a% (in terms of expected return)"; an absolute view takes the form of "I have ω_2 confidence that asset z will outperform the market by b%". Consequently, we obtain the definition of market views as below.

Definition 1

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* indicates the assets mentioned in views. The sum of each row of *P* should either be 0 (for relative views) or 1 (for absolute views); *Q* is the expected return for each view; and the confidence matrix Ω 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 $\Omega = diag(\omega_1, \omega_2, ..., \omega_n)$. Following the steps described in [19], it can be further derived from equation 6 and definition 1 that:

$$\mu_{BL} = [(\tau \Sigma)^{-1} + P' \hat{\Omega}^{-1} P]^{-1} [(\tau \Sigma)^{-1} \Pi + P' \hat{\Omega}^{-1} Q]$$
(8)

$$\Sigma_{BL} = \Sigma + [(\tau \Sigma)^{-1} + P' \hat{\Omega}^{-1} P]^{-1}.$$
 (9)

In practice, an easier-to-compute definition of market views is more frequently used as below.

Definition 2

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.

It can be mathematically proved that the two definitions are equivalent in terms of expressiveness. However, definition 2 is more intuitive, since matrix P can be eliminated. And only when definition 2 holds, we can use the Black-Litterman assumption that the views can be described using a multivariate normal distribution. Finally, our task can be restated as estimating the variables in equation 8 and 9 with the help of a sentiment prior.

A. Estimating volatility, confidence, and return

We adopt the calculation of the equilibrium risk premiums (Π) using CAPM. It follows that the estimation of parameters of posterior distribution of the expected portfolio returns as in the Black-Litterman model depends on three factors: the equilibrium volatility as a covariance matrix (Σ), the investor's confidence of his own views (Ω), and the investor's expected returns as in his views (Q).

Our method uses the past *k*-days observed returns to calculate the covariance matrix. For asset *i* and asset *j*, the element σ_{ij} as in covariance matrix Σ is estimated as follows:

$$\hat{\sigma}_{ij} = k^{-1} \sum_{n=1}^{k} (r_{i-n} \cdot r_{j-n}) - k^{-2} \sum_{n=1}^{k} r_{i-n} \sum_{n=1}^{k} r_{j-n}$$
(10)

where $r_{i-n} = (p_{i-n} - p_{i-n-1})/(p_{i-n-1})$.

In the most original form of the Black-Litterman model, the confidence matrix Ω is set manually according to investors' experience. Whereas in the numerical example given by [18], the confidence matrix is derived from the covariance matrix:

$$\Omega = diag(P(\tau\Sigma)P').$$
(11)

This is because $P(\tau \Sigma) P'$ can be understood as a covariance matrix of the expected returns in the views as well. Using definition 2, 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*. If the past return series of asset *i* implies high risk, then no matter how the market views are formed, the investor is less confident in it. In this initial case, the estimation of Ω utilizes past information of asset price volatilities.

The expected return has the most salient relation to the market sentiment. Our hypothesis is that there exists a responding strategy to surf market sentiment that statistically makes profits (generates alpha). Assuming the Black-Litterman agent uses the past price series $(p_{t,k})$ and trading volumes $(v_{t,k})$ to empirically form and update the expected return of their views, we further use the current time market sentiment on assets (s_t) as a prior. We learn this time-varying strategy using a novel deep recurrent neural network (RNN) design that is based on evolving clustering method (ECM) and long short-term memory (LSTM) network and, hence, termed ECM-LSTM:

$$\hat{Q}_t = \text{ECM-LSTM} \left(\hat{Q}_{BL}(p_{t,k}, \nu_{t,k}); s_t \right).$$
(12)

ECM [20] is usually used for on-line systems, in which it performs a one-pass, maximum distance-based clustering process without any optimization. The method is very fast due to its nature of efficiently recording and updating the centroids and clustering radiuses. LSTM is a special type of RNN with gated units. The LSTM unit often includes an input gate, a forget gate, and an output gate. All the gates are updated with the current input and previous output state. 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 [21, 22].

ECM-LSTM is inspired by the observation that forecasts made by simply applying LSTM adapt to the incoming data too fast. Whereas real-world financial time series are usually very noisy, which will cause over-fitting to meaningless signals if used for an off-line training. The ECM mechanism was first proposed for partitioning of the input space to learn rules for fuzzy inference systems. Similarly, we can endow the LSTM model with stability by only learning from critical new incoming data, namely when the old clustering pattern is updated. [22] shows that none of its variants can improve upon the standard LSTM architecture significantly on various tasks. Therefore, we implement the vanilla LSTM unit as described in [22].

The ECM-LSTM training and forecasting procedure is depicted in algorithm 1, where σ denotes the sigmoid function, \hat{Q}_{t-1} is the model forecasting of the previous state, while Q_{t-1}^* is the last observable ground truth, or a guideline for the investor's expected returns. Activation functions of input gate, forget gate, and output gate are denoted by **i**, **f**, and **o**. **W** are state transfer matrices, and **b** are the bias vectors. The state of each LSTM cell for time point *t* is updated by the current period information on its previous state c_{t-1} . (C_i, \mathcal{R}_i) are the clustering centroids and corresponding radii for the input vector space.

B. The Optimal Market Sentiment Views

We derive the optimal market views $[P, Q^*, \hat{\Omega}]$ as in definition 2 with the sentiment conditioned expected returns using the inverse optimization problem of the Black-Litterman model. Consider a multi-period model of the portfolio, our objective is to maximize the amount of capital at period (t + 1):

$$Capital_{t+1} = Capital_t \times w_t \odot \frac{p_{t+1}}{p_t}.$$
 (13)

Because w_t is independent from *Capital*_t, for each period *t* the optimal portfolio weights are thus:

$$w_t^* = \operatorname{argmax} \ w_t \oslash p_t \odot p_{t+1} \tag{14}$$

where \oslash and \odot are element-wise operators. Obviously, the solution of equation 14 is a one-hot vector representation where the weight of the asset with the maximum p_{t+1}/p_t equals 1. The interpretation can be without short selling and transaction fees, one should reinvest his whole capital daily to the fastest-growing asset in the next time period. Let this w_t^* be w_{BL}^* in equation 7, we will have:

$$w_t^* = (\delta \Sigma_{BL,t})^{-1} \mu_{BL,t}$$
(15)

substituting $\Sigma_{BL,t}$ and $\mu_{BL,t}$ with equations 8 and 9 for period *t*, we will have:

$$w_{t}^{*} = \left[\delta\left(\Sigma_{t} + \left[(\tau\Sigma_{t})^{-1} + P'\hat{\Omega}_{t}^{-1}P\right]^{-1}\right]\right]^{-1} \\ \left[(\tau\Sigma_{t})^{-1} + P'\hat{\Omega}_{t}^{-1}P\right]^{-1}\left[(\tau\Sigma_{t})^{-1}\Pi_{t} + P'\hat{\Omega}_{t}^{-1}Q_{t}^{*}\right].$$
(16)

Therefore, the optimal expected returns for our market views for each period t can be solved from equation 16:

$$Q_{t}^{*} = \hat{\Omega}_{t} [\delta[(\tau \Sigma_{t})^{-1} + P' \hat{\Omega}_{t}^{-1} P] \\ \times [\Sigma_{t} + [(\tau \Sigma_{t})^{-1} + P' \hat{\Omega}_{t}^{-1} P]^{-1}] w_{t}^{*} - (\tau \Sigma_{t})^{-1} \Pi_{t}] \\ = \delta[\hat{\Omega}_{t} (\tau \Sigma_{t})^{-1} + \mathbb{I}] [\Sigma_{t} + [(\tau \Sigma_{t})^{-1} + \hat{\Omega}_{t}^{-1}]^{-1}] w_{t}^{*} \\ - \hat{\Omega}_{t} (\tau \Sigma_{t})^{-1} \Pi_{t}.$$
(17)

IV. Sentic Computing

The market views require summarizing sentiment from a great deal of textual data. The quality of sentiment time series is obviously critical, because the data is later employed in the model training of estimating expected return Q. This is a nontrivial sentiment analysis task that involves other natural language processing techniques, such as named-entity recognition, word polarity disambiguation, sarcasm detection, and aspect extraction [23]. Sentic computing [24] is the state-of-the-art framework that enables sentiment analysis of text not only at document or paragraph level, but also at sentence, clause, and concept level. In contrast to the statistical approaches, sentic computing combines both knowledge-based polarity inference and a backup machine learning technique. A basic statistical approach counts the positive and negative words in a sentence; however, the sentence structure is not taken into account. By averaging the word polarities, positive and negative words will nullify each other, which brings about difficulties for analyzing sentiment in complicated contexts.

Sentic computing mainly leverages a concept-level knowledge base termed SenticNet¹ [25], a commonsense knowledge base of 100,000 concepts, and sentic patterns [26], a group of linguistic rules to explicitly catch the long-term dependency in texts. First, multiple relation tuples are extracted from the sentence with the Stanford typed dependency parser [27]. Then, a semantic parser further extracts concepts. We look up the concepts from SenticNet, and trigger sentic patterns to process the relations and intrinsic polarities of these concepts. If the concepts are not in SenticNet, the method resorts to a classifier built by machine learning. Figure 1 depicts this sentence-level polarity detection process.

Sentic computing embraces high interpretability required by most of financial applications and is powerful in many tricky cases. The rest of this section provides real-world examples from social media where sentic computing outperforms many other techniques.

A. Examples of Applying Sentic Patterns

□ Example 1: I had a feeling \$AAPL would go down, but this is stupid

The preprocessing of this sentence will need to know that "\$AAPL" refers to "Apple company" and completes the missing period at the end of the sentence. However, the interesting part is that by denying his own previous opinion, the speaker actually advocates his bullish mood of Apple company and labels this sentence as positive. The bag-of-words model would

Algori	thm 1 ECM-LSTM training and forecasting procedure.
Dat	a: Incoming data stream p, v, s
res	un: Expected return estimation Q_t
1	initialize LSTW parameters W, D;
2	$\int C = (\mathbf{p} \cdot \mathbf{y} \cdot \mathbf{c})$
2	$C_0 = (p_{t,k},v_{t,k},s_{t}),$
4	$\kappa_0 = 0;$
5	Go to line 15;
6	
7	$D_{\min} = \min((\mathbf{p}_{t,k}, \mathbf{v}_{t,k}, \mathbf{s}_{t}) - C_{i});$
8	if $\mathcal{A}(\mathcal{R}_i \geq D_{\min})$ then
9	Add ($p_{t,k}$, $v_{t,k}$, s_t) to C_i where D_{min} holds;
10	Go to line 24;
11	else
12	$(S_{\min}, i) = \min((\mathbf{p}_{t,k}, \mathbf{v}_{t,k}, \mathbf{s}_t) - C_i + \mathcal{R}_i);$
13	if $S_{\min} > 2\mathcal{R}_i$ then
14	Add $(p_{t,k_r}v_{t,k_r}s_t)$ to C;
15	$\mathbf{i}_{t} = \sigma (\mathbf{W}_{i} \cdot [\hat{\mathbf{Q}}_{t-1}, \mathbf{p}_{t,k}, \mathbf{v}_{t,k}, \mathbf{s}_{t}] + \mathbf{b}_{i});$
16	$f_{t} = \sigma (W_{f} \cdot [\hat{Q}_{t-1}, p_{t,k}, v_{t,k}, s_{t}] + b_{f});$
17	$\mathbf{o}_{t} = \sigma \left(W_{o} \cdot \left[\hat{Q}_{t-1}, p_{t,k}, v_{t,k}, s_{t} \right] + b_{o} \right);$
18	$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot (\mathbf{W}_c \cdot [\hat{\mathbf{Q}}_{t-1}, \mathbf{p}_{t,k}, \mathbf{v}_{t,k}, \mathbf{s}_t] + \mathbf{b}_c);$
19	$\hat{Q}_t = o_t \odot \tanh(c_t);$
20	Update W, b with $\frac{\partial (Q_{t-1} - Q_{t-1})}{\partial (i, t, o)}$;
21	else
22	Add $(p_{t,k}, v_{t,k}, s_t)$ to C_i where S_{min} holds;
23	Update (C_i, \mathcal{R}_i) ;
24	$\hat{Q}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_{t-1});$
25	end
26	end
27	end
28	return \hat{Q}_t ;

¹https://github.com/yurimalheiros/senticnetapi/



FIGURE 1 The sentic computing algorithm working at sentence level, adapted from [26].

identify "down" and "stupid", both are negative and concludes the whole sentence as negative. A machine learning based sentiment analysis model, for instance provided by Google Cloud Natural Language API (Google SA)², also fails for this example.

Sentic computing would first identify the concept "go_ down" from SenticNet, which is negative. This polarity will be passed through a nominal subject relation to "Apple company", and the relative clause modifier relation to "feeling". Note that this whole structure and "stupid" are linked by an adversative but-conjunction, thus the sentic pattern "negative but negative \Rightarrow positive" is triggered, giving the overall sentence a positive polarity. This process is further elaborated with Figure 2.

Example 2: \$AAPL will be down today again but the down draft is slowing. By end of next week I think it's getting bought back

In addition to example 1, the preprocessing will have to convert the ASCII based encoding for the apostrophe, and recognize

²https://cloud.google.com/natural-language/ [accessed on 2017-12-16]

two sentences: "Apple will be down today again but the down draft is slowing." and "By end of next week I think it's getting bought back." Google SA assigns the first sentence with sentiment score -0.20 and the second sentence 0.0, thus the overall sentiment is averaged as -0.10. However, the user labels the message

as positive and many would agree on the obviousness that both the two sentences are positive. Sentic computing does not provide the correct score for the first sentence but gets the correct overall label as positive. First, a but-conjunction has the highest priority such that the polarity of the first sentence is consistent with "the down draft is slowing". Concept "down_draft" is not in SenticNet, hence it inherits the polarity of "down": -0.31. Although slowing of down draft is positive in the stock market, the concept "is_slowing" is neutral in the general domain knowledge base SenticNet. Therefore, the negative score passes through the whole sentence, giv-

ing the first sentence sentiment score of -0.31. The concept "bought_back" carries a sentiment score of 0.82. The sentiment score of "next" -0.56 is passed through an adjective modifier relation to "next week", and because of the noun modifier relation between "end" and "week", the polarity of "by end of next week" is inverted to a slightly positive 0.02, The overall polarity of the second sentence is thus 1 - (1 - 0.82)(1 - 0.02) = 0.82. The entire message has a sentiment score of ((0.82 + (-0.31))/2 = 0.26.

□ Example 3: \$AAPL moment of silence for the 180 call gamblers. lol.

This message contains an "lol", probably acronym for "laughing out loud". The user expresses his negative mood by indicating there is evidence that the stock price of Apple would not reach 180 and derogating those hold the optimistic opinion as "gamblers". Google SA gives the sentence a positive sentiment score of 0.30, and the same as most of machine learning based methods, it is difficult to analyze where the error does come from.



FIGURE 2 Sentiment score propagates from low-level concepts to sentence-level polarity via dependency tree.

Sentic computing takes the polarity of "silence" 0.11 as the sentiment score for "moment of silence" since it is a noun modification relation. However, the case mark "for" updates the overall polarity to depend on the latter part "the 180 call gamblers". Since "moment of silence" is positive but "gambler" has a negative score of -0.74 in SenticNet, the pattern triggered a more intense negativity for "moment of silence for gamblers" as $-\sqrt{|-0.74|} = -0.86$.

By applying sentic computing to the message data stream from social media, we can count the daily positive and negative messages and compute the average sentiment score for a specific asset and hence form sentiment time series. In the section of experiments, we can observe the magic power of agglomerating individual level sentiment of messages as a prior for market prediction.

V. Experiments

In this section, we evaluate the quality and effectiveness of our formalization of market sentiment views. First, we compare the result of sentic computing with labels given by users themselves. Next, we run trading simulations with the intelligent Bayesian asset allocation model and benchmark on several portfolio construction strategies. Finally, we discuss our findings.

A. Data and Sentiment Time Series

In this study, we collect the opinion messages from StockTwits, which is a popular social network for investors and traders to share financial information. Besides, we obtain the historical closing price of stocks and the daily trading volumes from the Quandl API; the market capitalization data from Yahoo! Finance. We investigate a time period of 3 months from 2017-08-14 to 2017-11-16. For missing values, such as the closing prices on weekends and public holidays, we fill the gap with the closest historical data.

Our dataset comprises 38,414 messages for Apple, 4,298 messages for Goldman Sachs, 2,157 messages for Pfizer, 1,094 messages for Newmont Mining, 2,847 messages for Starbucks, and 76,553 messages for other tickers. Table 1 provides the confusion matrix between user labeling and sentic computing results on Apple Inc. In total, 14,524 user labeled messages that mentioned Apple are analyzed in the investigated period. From Table 1 the accuracy of polarity detection can be easily calculated as 59.8%. Given the noise in raw data and the general domain knowledge base we use, this is a fairly acceptable result. Another issue to note is that user labeling cannot be fully understood as a ground truth, but only as a reference. Because only a small portion (usually less than 20%) of users will label their messages regardless of the sentiment expressed, the subset that has user labels may not be an unbiased sample.

However, the visualization of the two time series of positive and negative message counts that expands on time axis can better exhibits the consistency between two sources (Figure 3).

We take the trader mood index from a third commercial product (PsychSignal) for comparison. Their message data stream gathers various sources including Stocktwits and others. Their sentiment analysis engine is not disclosed as well. We calculate the correlation of two time series as:

Correlation
$$(s_1, s_2) = \frac{\mathbb{E}\left((s_1 - \bar{s}_1)(s_2 - \bar{s}_2)\right)}{\sigma_{s_1}\sigma_{s_2}}.$$
 (18)

Table 2 reports the significant and positive correlation between the time series from three sources.

B. Trading Simulation

We construct a portfolio by randomly selecting stocks of big companies. The portfolio consists of 5 stocks: Apple Inc (AAPL), Goldman Sachs Group Inc (GS), Pfizer Inc (PFE), Newmont Mining Corp (NEM), and Starbucks Corp (SBUX).

 TABLE 1 Confusion matrix between user labeling and sentic computing results.

		SENTIC COMPUTING		
		POSITIVE	NEGATIVE	TOTAL
USER LABELING	POSITIVE	7234	3748	10982
	NEGATIVE	2097	1445	3542
	TOTAL	9331	5193	14524



FIGURE 3 The time series of positive and negative message counts from two sources (x-axis: days, y-axis: counts).

TABLE 2 Correlation of message time series between user	
labeling, sentic computing, and PsychSignal.	

	POSITIVE MESSAGES	NEGATIVE MESSAGES
USER-SENTIC	+0.964	+0.795
USER-PSYCH	+0.185	+0.449
SENTIC-PSYCH	+0.276	+0.282

This selection covers both the NYSE and NASDAQ markets and diversified industries, such as technology, financial services, healthcare, consumer discretionary. The social media post frequencies also vary to a large extent among these companies. Traditional industries generally get less attention. The prices per share are adjusted according to the stock split history for computing all related variables, however, dividends are not taken into account. In the simulations, we assume no short selling,



FIGURE 4 Performance of intelligent Bayesian portfolios with different sources of sentiment time series (x-axis: days, y-axis: thousand dollars).

taxes, or transaction fees, and we assume the portfolio investments are infinitely divisible, starting from 10,000 dollars.

We benchmark our portfolio performance with the following three asset allocation strategies:

1) **The equal-weighted portfolio (EW):** we hold equal weights (20%) for the five stocks in our portfolio throughout the

period investigated. In this case, the portfolio performance averages the price movement of five stocks. This strategy is fundamental and minimum information is required. However, in forecasting of complex systems such as stock market, this effortless strategy performs better than many more complicated strategies.

- 2) The ARIMA portfolio (ARIMA): we re-invest daily according to the forecasted prices. The forecasting is produced by an ARIMA (p, d, q) model (autoregressive integrated moving average) for each stock and parameters are inferred from historical data as follows. First, increase d until the differentiated time series is stationary. Then, set the maximum p and q as the order of the last significant partial autocorrelation and autocorrelation. Finally, choose (p, q) that produces the minimum Akaike information criterion (AIC). In fact, except the ARIMA(0, 1, 2) model for PFE, other prices exhibit random walk behavior (ARIMA(0, 1, 0)).
- 3) The Holt-Winters portfolio (HW): we re-invest daily according to the one-step-forward price forecasts. The forecasting is produced by a Holt-Winters additive smoothing method with time-varying parameters. The model HW $(\hat{\alpha}_t, \hat{\beta}_t, \hat{\gamma}_t)$ is specified at each time point t by minimizing the root mean square error (RMSE) of simulated time series in a sliding window (t k, t).

Note that ARIMA and HW portfolios do not require any prior, however, they are considered to be among the most effective forecasting techniques across different tasks when informative data from other sources are not available [28].

We further construct intelligent Bayesian portfolios with sentiment time series from different sources using the Black-Litterman approach. Following the previous research [18], we set $\delta = 0.25$ and $\tau = 0.05$. For the simplicity of neural network structure, we use a single layer of 64 LSTM units followed by a densely connected layer of the size of the number of assets. The hidden LSTM layer uses 20% dropout. The trading performances are demonstrated in Figure 4, where unit of portfolio capital is 1,000 dollars.

C. Evaluation

Diversified metrics have been proposed to evaluate the performance of portfolios [29–31]. The difficulty of evaluation is partially due to the fact that there does not exist a simple way to calculate the distance to the gold standard portfolio. The same deviation in portfolio weights may be amplified or narrowed by volatility of asset prices. Therefore, we abandon metrics on

The improvement of introducing an ECM mechanism is confirmed by the fact that in terms of all these metrics, the ECM-LSTM portfolios systematically outperform their counterparts using the same source of sentiment views.

weights and report four direct financial metrics: annualized return (AR), Sharpe ratio (Sh.R), Sortino ratio (So.R), and the maximum drawdown (MDD).

Annualized return measures the profitability of a given portfolio. We assume in a natural year, the portfolio keeps a constant compound growth rate as in the investigated period *t*. Let T = 365.25, we have:

$$AR = \left(\frac{Capital_t}{Capital_0}\right)^{\frac{T}{t}}.$$
(19)

Sharpe ratio is a risk-adjusted return measure. We choose the equal-weighted portfolio as a base, so that the Sh.R of EW will be 1.00:

Sh.R =
$$\frac{\mathbb{E} \left(R_{\text{portfolio}} / R_{\nu\nu} \right)}{\sigma \left(R_{\text{portfolio}} \right) / \sigma \left(R_{\nu\nu} \right)}.$$
 (20)

Sh.R 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 [32]. So.R is calculated in this manner.

MDD measures the maximum possible percentage loss of an investor:

$$MDD = \max_{0 \le t \le \tau} \left\{ \frac{Value_t - Value_\tau}{Value_t} \right\}.$$
 (21)

Asset allocation strategies with large MDD tend to give rise to panic and impatience among investors and expose the portfolio to the risk of withdrawal. Table 3 presents these metrics.

TABLE 3 Performance metrics with the top 3 in bold.								
	AR(%)	SH.R	SO.R	MDD(%)				
EW	23.07	1.00	1.00	1.76				
ARIMA	10.72	0.56	0.61	3.79				
HW	13.03	0.34	0.36	6.16				
LSTM(PSYCH)	33.52	0.71	0.79	3.84				
LSTM(SENTIC)	27.21	0.61	0.68	5.05				
LSTM(USER)	24.82	0.64	0.68	4.61				
ECM-LSTM(PSYCH)	45.51	0.74	0.82	3.45				
ECM-LSTM(SENTIC)	35.45	0.66	0.73	2.89				
ECM-LSTM(USER)	37.53	0.71	0.87	3.40				

D. Findings

Figure 4 shows that, regardless of the sentiment source and network implementation details for estimating market views, the intelligent Bayesian portfolios exhibit similar moving patterns. These patterns can be seen as intrinsic to the model and portfolio selection. In addition, in two of three portfolios using ECM-LSTM, the crash observed elsewhere between 2017-09-15 and 2017-09-25 is effectively corrected.

EW is the most stable strategy in the experiments. In Table 3, EW also has the best Sh.R, So.R, and minimum MDD.ARIMA and HW are more volatile than EW. This is because after forecasting of next-day prices, the whole capital is invested to the only winning asset, thus the risk is not well diversified. The cumulative return of these two strategies cannot compare EW in this period as well, resulting in very small Sh.R and So.R. These two strategies would not be preferred.

All the portfolios that have taken market sentiment into account achieve higher AR than the three basic strategies discussed above. The improvement of introducing an ECM mechanism is confirmed by the fact that in terms of all these metrics, the ECM-LSTM portfolios systematically outperform their counterparts using the same source of sentiment views.

In the experiments, So.Rs are greater than Sh.Rs, indicating that the market trend in this period is going up. However, all the strategies have So.Rs and Sh.Rs less than 1. This observation holds in most well-formed markets, because seeking for a higher AR inevitably causes the investor to take greater unit risk.

The quality of the source of sentiment time series should be important, though the difference between the three sources we examined is not very clear. It seems that PsychSignal provides the most accurate sentiment data stream, in terms of both the volume of social media data collected and the portfolio performance. However, using just the user labeled message counts sometimes also achieved a balanced and advantageous result.

VI. Conclusion

Market sentiment has attracted a great deal of attention in the computational intelligence and econometrics communities. However, the problem is often formulated as a price forecasting task rather than asset allocation task. In this work, we proposed a sophisticated approach to compute the asset-level market sentiment from social media data stream, and integrate it to the state-of-the-art asset allocation method using market views. Cross validation experiments suggest that the sentiment time series obtained using sentic computing is comparable to some commercial tools. Considering its transparency and good interpretability, sentic computing is of great potential for broader financial applications that require natural language processing.

Another important contribution is made to the problem "how to deal with noisy financial data when applying machine learning techniques". By introducing ECM as a screening mechanism for LSTM, the learned market views are smoothed and portfolio crash is effectively reduced. This novel method improves the AR of our asset allocation strategies by circa 10% on average. Other metrics experimented with, such as the Sharpe ratio and MDD, are improved as well when compared to the vanilla LSTM estimation based strategies.

References

 Q. Li, L. Jiang, P. Li, and H. Chen, "Tensor-based learning for predicting stock movements," in *Proc. 29th Conf. Association Advancement Artificial Intelligence*, Austin, TX, USA, Jan. 25–30, 2015, pp. 1784–1790.

[2] F. Z. Xing, E. Cambria, and X. Zou, "Predicting evolving chaotic time series with fuzzy neural networks," in *Proc. Int. Joint Conf. Neural Networks*, Anchorage, AK, USA, May 14–19, 2017, pp. 3176–3183.

[3] E. F. Fama and K. R. French, "A five-factor asset pricing model," J. Financial Econ., vol. 116, no. 1, pp. 1–22, Apr. 2015.

[4] W. Antweiler and M. Z. Frank, "Is all that talk just noise? The information content of internet stock message boards," J. Finance, vol. 59, no. 3, pp. 1259–1294, June 2004.

[5] J. Smailović, M. Grčar, N. Lavrač, and M. Žnidaršič, "Stream-based active learning for sentiment analysis in the financial domain," *Inf. Sci.*, vol. 285, pp. 181–203, Nov. 2014.

[6] B. Wuthrich, V. Cho, S. Leung, D. Permunetilleke, K. Sankaran, and J. Zhang, "Daily stock market forecast from textual web data," in *Proc. IEEE Int. Conf. Systems Man and Cybernetics*, San Diego, CA, USA, Oct. 14, 1998, vol. 3, pp. 2720–2725.

[7] W. Zhang and S. Skiena, "Trading strategies to exploit blog and news sentiment," in Proc. 4th Int. Association Advancement Artificial Intelligence Conf. Weblogs and Social Media, Washington, DC, May 23–26, 2010, pp. 375–378.

[8] A. Weichselbraun, S. Gindl, F. Fischer, S. Vakulenko, and A. Scharl, "Aspect-based extraction and analysis of affective knowledge from social media streams," *IEEE Intell. Syst.*, vol. 32, no. 3, pp. 80–88, May 2017.

[9] J. Bollen, H. Mao, and A. Pepe, "Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena," in *Proc.5th Int. Association Advancement Artificial Intelligence Conf. Weblogs and Social Media*, Barcelona, Spain, July 17–21, 2011, pp. 450–453.

[10] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," J. Comput. Sci., vol. 2, no. 1, pp. 1–8, Mar. 2011.

[11] F. Z. Xing, E. Cambria, L. Malandri, and C. Vercellis, "Discovering Bayesian market views for intelligent asset allocation," in *Proc. European Conf. Machine Learning and Principles and Practice Knowledge Discovery Databases*, Dublin, Ireland, Sept. 10–14, 2018.

[12] H. Markowitz, "Portfolio selection," J. Finance, vol. 7, pp. 77-91, Mar. 1952.

[13] M. C. Steinbach, "Markowitz revisited: Mean-variance models in financial portfolio analysis," Soc. Ind. Appl. Math. Rev., vol. 43, no. 1, pp. 31–85, Mar. 2001.

[14] R. Ruiz-Torrubiano and A. Suarez, "Hybrid approaches and dimensionality reduction for portfolio selection with cardinality constraints," *IEEE Comput. Intell. Mag.*, vol. 5, no. 2, pp. 92–107, Apr. 2010.

[15] D. Avramov and G. Zhou, "Bayesian portfolio analysis," Annu. Rev. Financial Econ., vol. 2, pp. 25–47, Dec. 2010.

[16] T. Bodnar, S. Mazur, and Y. Okhrin, "Bayesian estimation of the global minimum variance portfolio," *Eur. J. Oper. Res.*, vol. 256, no. 1, pp. 292–307, Jan. 2017.

[17] F. Black and R. Litterman, "Asset allocation: Combining investor view with market equilibrium," J. Fixed Income, vol. 1, no. 2, pp. 7–18, July 1991.

[18] G. He and R. Litterman, The intuition behind Black-Litterman model portfolios, 1999.
[Online]. Available: http://dx.doi.org/10.2139/ssrn.334304. Accessed on August 27, 2018.
[19] S. Satchell and A. Scowcroft, "A demystification of the Black-Litterman model: Managing quantitative and traditional portfolio construction," *J. Asset Manage.*, vol. 1, no. 2,

pp. 138–150, Sept. 2000. [20] N. K. Kasabov and Q. Song, "Denfis: Dynamic evolving neural-fuzzy inference system and its application for time-series prediction," *IEEE Trans. Fuzzy Syst.*, vol. 10, no. 2, pp. 144–154, Aug. 2002.

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

[22] K. Greff, R. K. Srivastava, J. Koutnik, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, July 2017.

[23] E. Cambria, S. Poria, A. Gelbukh, and M. Thelwall, "Sentiment analysis is a big suitcase," IEEE Intell. Syst., vol. 32, no. 6, pp. 74–80, Nov. 2017.

[24] E. Cambria and A. Hussain, Sentic Computing: A Common-Sense-Based Framework for Concept-Level Sentiment Analysis, ser. Socio-Affective Computing. New York, NY, USA: Springer, 2015.
[25] E. Cambria, S. Poria, D. Hazarika, and K. Kwok, "SenticNet 5: Discovering conceptual primitives for sentiment analysis by means of context embeddings," in *Proc. 32nd* Association Advancement Artificial Intelligence Conf., New Orleans, LA, USA, Feb. 2–7, 2018, pp. 1795–1802.
[26] S. Poria, E. Cambria, A. Gelbukh, F. Bisio, and A. Hussain, "Sentiment data flow analysis by means of dynamic linguistic patterns," *IEEE Comput. Intell. Mag.*, vol. 10, no. 4, pp. 26–36, Oct. 2015.
[27] M.-C. de Marneffe and C. D. Manning, "The Stanford typed dependencies representation," in *Proc. Workshop Cross-Framework and Cross-Domain Parser Evaluation*, 2008, pp. 1–8.
[28] S. Makridakis and M. Hibon, "The m3-competition: results, conclusions and implications," Int. J. Forecasting, vol. 16, no. 4, pp. 451–476, Oct. 2000.

[29] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," Int. J. Forecasting, vol. 22, no. 4, pp. 679-688, Oct. 2006.

[30] M. W. Brandt, Chapter 5: Portfolio Choice Problems. Handbook of Financial Econometrics. New York, NY, USA: Elsevier, 2010.

[31] F. Z. Xing, E. Cambria, and R. E. Welsch, "Natural language based financial forecasting: A survey," Artif. Intell. Rev., vol. 50, no. 1, pp. 49–73, June 2018.

[32] F. A. Sortino and L. N. Price, "Performance measurement in a downside risk framework," J. Investing, vol. 3, no. 3, pp. 59–64, July 1994.