Cognitive Computation (2019) 11:799–808
https://doi.org/10.1007/s12559-019-9626-9

Determination of Temporal Stock Investment Styles via Biclustering Trading Patterns

Jianjun Sun [1] · Qinghua Huang [2] · Xuelong Li [3]

Received: 19 June 2018 / Accepted: 16 January 2019 / Published online: 8 March 2019

Abstract

Due to the effects of many deterministic and stochastic factors, it has always been a challenging goal to gain good profits from the stock market. Many methods based on different theories have been proposed in the past decades. However, there has been little research about determining the temporal investment style (i.e., short term, middle term, or long term) for the stock. In this paper, we propose a method to find suitable stock investment styles in terms of investment time. Firstly, biclustering is applied to a matrix that is composed of technical indicators of each trading day to discover trading patterns (regarded as trading rules). Subsequently a *k*-nearest neighbor (KNN) algorithm is employed to transform the trading rules to the trading actions (i.e., the buy, sell, or no-action signals). Finally, a min-max and quantization strategy is designed for determination of the temporal investment style of the stock. The proposed method was tested on 30 stocks from US bear, bull, and flat markets. The experimental results validate its usefulness.

Keywords Machine learning · Biclustering · Technical analysis · Trading rules · Investment styles

Introduction

Ever since the appearance of stock market, it is an everlasting challenging task to predict the trading point of the stock. Despite the volatile nature of stock market, researchers still find some hidden rules about the fluctuations of the stock price. Now, there are two popular methods for stock analysis [1], namely fundamental analysis and technical analysis. Fundamental analysis focuses on the study of economic factors that influence the fluctuation of stock price. The factors include the basic information of the companies such as profitability and industry prospects. All the factors would be taken into consideration when buying or selling a stock (https://en.wikipedia.org/wiki/Technical_analysis). The technical analysis makes use of user-designed technical indicators to reflect the moving trend of stock price.

The technical analysis is frequently used in the stock investment, as it is easy to understand and operate with the indicators that can discover some signs of the price change. It has been empirically and theoretically demonstrated that learning technical indicators jointly always gains better performance than learning each indicator independently [2, 3].

Several intelligent systems have shown excellent performance in such field [3–11], for example, taking advantage of the strong learning ability in nonlinear systems; the neural network has widely been applied to the analysis of stock data. Schierholt et al. [6] used many kinds of neural networks to predict the fluctuation of stock price, achieving excellent performance. Based on the historical price and transaction volume, Potvin et al. [12] generated short-term trading rules by genetic programming. The trading rules provide better performance over the Buy-and-Hold [13] approach only when the market does not rise. Due to the excellent inference ability of fuzzy system, Chang and Liu [14] developed a fuzzy rule–based system for the fluctuating prediction of stock price. Using the

✉ Qinghua Huang
enicarwhw@qq.com

Jianjun Sun
jjsun@mail.nwpu.edu.cn

Xuelong Li
xuelong_li@nwpu.edu.cn

1 School of Computer Science, and Center for OPTical IMagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an 710072, Shaanxi, People's Republic of China

2 School of Mechanical Engineering, and Center for OPTical IMagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an 710072, Shaanxi, People's Republic of China

3 School of Computer Science and Center for OPTical IMagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an 710072, Shanxi, People's Republic of China

trading signals from the combination of several technical indicators, Briza and Naval [15] proposed a trading system with multi-objective particle swarm optimization. Considering the advantage of "bull flag" technical charting heuristic, Leigh et al. [16] proposed a recognizer. Experimental results show the validity of trading rules mined by the recognizer. The SVM was applied to the stock price movement by Huang et al. [17]. It is superior in accuracy to linear/quadratic discriminant analysis and BP neural networks. Besides, other technologies, e.g., Bayesian network [18–20] and logistic regression [21] are also frequently used in this field.

In a review of existing algorithms, it is apparent that those methods take into account many factors to obtain maximum profit in varied investment terms. However, stocks always show quite different behaviors in the price change. Some stock may fluctuate greatly in a short investment term, while some shows smoother change at the same time. In other words, it is appropriate to hold some stocks for a long investment time but may be inappropriate for other stocks. This phenomenon tells that there should be a temporal investment style for every stock, and the discovery of those styles is necessary before selecting a trading strategy.

However, little attention has been paid to mining the temporal investment styles. Some researches [22–24] just considered the short-term investment, ignoring the middle- and long-term trading strategies. In this paper, we propose to make use of a biclustering-based intelligent trading strategy to determine the temporal investment styles for a set of stocks. The biclustering method is applied to mine the trading rules from the data matrix composed of the technical indicators and future return rates. Each rule is the combination of technical indicators with different time lengths. The trading rules are used to discover the trading signals (i.e., buy, sell, or no action). In the training periods, we can find the trading rules corresponding to three investment terms, i.e., the short term, middle term, and long term, respectively. Then these rules are applied in the validation periods, and the temporal investment style can be determined for each stock according to the profit rates gained in the validation period. Finally, we use testing periods to further evaluate the performance of the proposed method in discovering the investment styles for the stocks.

The remaining sections of this paper are organized as follows. The "Methods" section describes the proposed method in detail. The "Experiments and Results" section presents the experimental method and the results. In the "Conclusions and Future Work" section, we summarize this work and propose several future extensions.

Methods

As mentioned above, our objective is to find the temporal investment styles for various stocks. If a stock has a specific style in terms of the investment term, investors can have a prediction on how long they will hold it and hence avoid missing important timing of transaction. Accordingly, given a stock, the whole processing procedures are as follows, we firstly discover the trading rules that can make good profits in different investment terms (i.e., short term, middle term, and long term) using its historical training data. Then we evaluate the performance of the trading rules in different terms in the validation data sets. If the trading rules could show good performance for predicting the price trend and make good profits in a specific investment term, we consider that the stock may be of the investment style corresponding to the same term. Furthermore, we apply the same trading rules to testing historical trading data. If the similar performance could be achieved in the testing data, we can then confirm the temporal investment style for the stock; otherwise, the style for the stock cannot be determined.

Data Preparation

The historical data was downloaded from (http://finance.yahoo.com/). Figure 1 shows an example in the price domain including the opening, highest, lowest and closing prices, and the volume for every trading day. The rows denote the trading days, and the columns denote the date, opening, highest, lowest and closing prices, and the transaction volume. As mentioned above, we need to discover the trading rules based on the technical indicators using the historical data. Therefore, with respect to the method used in (https://en.wikipedia.org/wiki/Category:Technical_indicators), the historical data is transformed to a new data matrix M in which the technical indicators are set as the columns.

Table 1 gives 8 indicators selected for the computation of M. A brief introduction and the time length used in this study follow each indicator. A total of 31 technical indicators can therefore be obtained as shown in Fig. 2. Given a data element (i, j) in M, it corresponds to the value of the jth indicator on the ith trading day. For example, let us take into account the indicator ROC:

$$\mathrm{ROC}(n) = \frac{\mathrm{cp}(t) - \mathrm{cp}(t-n)}{\mathrm{cp}(t-n)} \times 100\% \tag{1}$$

where cp(t) is the closing price of the tth trading day and cp(t-n) is the closing price of the n days before the tth trading day. If $n = 6$, the ROC(n) should be assigned with the element (t, 14) in M. For the computation of the other indicators, please refer to (https://en.wikipedia.org/wiki/Category:Technical_indicators).

Because the most important factor of stock analysis is the price trend of the stock, we use additional columns to indicate

Fig. 1 Original data structure of stock NVDA

Date	Open	Highest	Lowest	Close	Volume
2011/11/3	13.97	14.68	13.60	14.65	1873430
2011/11/4	14.52	14.87	14.29	14.82	1637100
2011/11/7	14.71	14.96	14.40	14.74	1533840
2011/11/8	14.93	15.17	14.69	15.08	1374320

the price trend in M. Like (https://en.wikipedia.org/wiki/Category:Technical_indicators), we compute the future return rates (FRR) to reflect the fluctuation of stock price and add it to M. To avoid the effect of noise, for the ith trading day, we use the averaged closing price (ACP), i.e., the mean price of its following n daily closing prices, as shown below:

$$\mathrm{ACP}_i = \frac{1}{n}\sum_{m=i+1}^{i+n}\mathrm{CP}_m \tag{2}$$

where CP_m denotes the closing price of the mth trading day, n denotes the length of the future trading days to be taken into account. We empirically set n to be 2, 5, and 22, for short-term, middle-term, and long-term price trends, respectively. Consequently, the FRR corresponding to the ith trading day is computed with the following formula:

$$\mathrm{FRR}_j = \frac{\mathrm{ACP}_j - \mathrm{CP}_j}{\mathrm{CP}_j} \times 100\% \tag{3}$$

Accordingly, we can have 3 additional columns corresponding to the FRRs for three investment terms. Figure 2 shows the construction of M, where the rows of the data matrix are the trading days, the left 31 columns are the 31 technical indicators, and the rightmost 3 columns (columns 32–34) are the three FRRs for short-term, middle-term, and long-term investments, respectively.

Because the data scales are different for the technical indicators, we perform the data normalization under every column of indicator. The normalizing method adopted in this study is the min-max strategy as follows:

$$D_{\mathrm{new}}(i,j) = \frac{D_{\mathrm{old}}(i,j) - D_{\min}(:,j)}{D_{\max}(:,j) - D_{\min}(:,j)} \tag{4}$$

where $D_{\mathrm{old}}(i, j)$ is the raw value of the technical indicator at the ith row and jth column, $D_{\max}(:, j)$ is the maximum value of the jth column, $D_{\min}(:, j)$ is the minimum value of the jth column, and $D_{\mathrm{new}}(i, j)$ is the new value of $D_{\mathrm{old}}(i, j)$. After the normalization, all the values are within the range [0, 1].

With the normalized M, we can discover the trading patterns each of which can be regarded as a trading signal. With respect to the FRR, the trading signals can be grouped into three categories, i.e., buy, sell, and no action. In other words, the FRR plays the role of class label. Therefore, we further quantize the FRR values in M using the following rule:

- 1: If FRR(i,j) is greater than a predefined threshold T_{t}, set FRR(i,j) to be 1, indicating that in the future trading days, the stock price rises.
- 2: If FRR(i,j) is less than $-T_{\mathrm{t}}$, set FRR(i,j) to be − 1, indicating that in the future trading days, the stock price falls.
- 3: Otherwise, set FRR(i,j) to be 0, indicating that in the future trading days, the stock price looks to be stable.

In this paper, the threshold T_{t} is empirically set to be 0.005. With the labels denoted by the FRR, we can have a supervised mining of trading patterns.

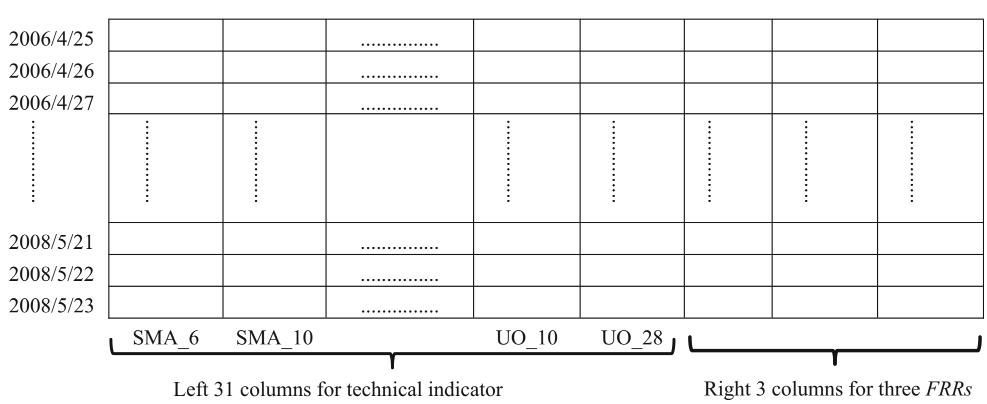

Fig. 2 The data matrix incorporating technical indicatorsand future returns

Table 1 Technical indicators used as input variables

Indicators	Description	Variation of n days
SMA	It computes the average price of a security over a number of periods.	SMA_6, SMA_10, SMA_12, SMA_24, SMA_30
RSI	It measures the speed and change of price movements.	RSI_6, RSI_12, RSI_18, RSI_24, RSI_30
%R	It reflects the level of the close relative to the highest high for the look-back period.	%R_6, %R_10, %R_20
ROC	It measures the percent change in price from one period to the next.	ROC_6, ROC_12, ROC_24, ROC_36
CCI	It could be used to identify a new trend or warn against extreme conditions.	CCI_6, CCI_12, CCI_14, CCI_28
EMV	It fluctuates above and below the zero line.	EMV_6, EMV_12, EMV_14, EMV_28
SO	It shows the location of the close relative to the high-low range over a set number of periods.	SO_14, SO_28, SO_42
UO	It designs to capture momentum across three different time frames.	UO_14, UO_28, UO_42

Biclustering-Based Trading Pattern Discovery

From a long-time perspective, the fluctuating rule of stock price may appear repeatedly with high possibility [25]. Therefore, we aim to find frequent patterns of the technical indicators from the data matrix M mentioned above. Before discovering the trading pattern in M, one may ask if a traditional clustering [26–31] works well. Note that in real trading, the investor often make decisions according to a smaller number of technical indicators, implying some useful trading rules do not include all technical indications. It makes us try to find the patterns under a subset of technical indicators.

Fortunately, biclustering is a two-way clustering that clusters the data from two dimensions (the row and the column) at the same time [22, 32]. It can discover the local coherent pattern in a two-dimensional data matrix. Thus, we apply the biclustering method to M to find local coherent patterns (called biclusters), each of which contains a subset of technical indicators and a subset of trading days. Each bicluster represents a trading pattern in this study and can be grouped into short term, middle term, or long term category [25, 33–35]. In our biclustering method, the minimum entropy score (MES) is used to evaluate the biclusters. The computation is shown below:

$$e_j = -\sum_{i\in N_{\mathrm{cl}}} \frac{p(i)}{n}\log_2\frac{p(i)}{n} = \log_2(n) - \frac{1}{n}\sum_{i\in N_{\mathrm{cl}}} p(i)\log_2 p(i) \quad (5)$$

$$\mathrm{MES} = \frac{1}{|C|}\sum_{j\in C} e_j \quad (6)$$

In Eq. 5, n denotes the number of the elements in the column, N_{cl} the number of the clusters in the column after agglomerative hierarchical clustering, and $p(i)$ the number of the elements belonging to cluster i in the column. In Eq. 6, e_j denotes the entropy score of each column, and MES denotes the mean entropy score of all columns. If MES is smaller than a threshold δ, the bicluster is regarded as being valid. Except the criterion for evaluating the bicluster as shown in Eq. 6, we follow the same biclustering method used in [36]. It is worth noting that to remove noisy biclusters, we also set the minimum rows of a bicluster to be 5, and the other parameters are the same as in [36].

Figure 3 shows an example of the biclustering procedures in this study. Given an original data matrix, the agglomerative clustering is first applied to each column to find the clusters, each of which may be part of a bicluster and is called bicluster seed. The seed is then expanded to the whole columns to form a very large bicluster. We thereafter delete the columns or rows iteratively, until the residual score is not greater than the predefined threshold. Finally, the bicluster with nearly constant column values can be output as a trading pattern. Each column in the final bicluster is averaged, and the mean values form a vector called trading pattern vector. More technical details can be found in [36].

A trading pattern can be deemed as a valid trading signal if its corresponding FRRs show similar values with a confidence value larger than 0.65. The label of the majority of FRRs is assigned to the trading pattern as its label. We only consider valid trading signals after the biclustering and use them to generate trading rules and actions. In this study, a trading rule denotes that if an input trading day is satisfied with a trading pattern, then a trading signal occurs and the corresponding trading action can be executed.

Determination of Trading Actions

In [25], the trading actions are made with respect to a method incorporating the K-nearest neighbor (KNN) and the root-mean-square (RMS) distance [25] between the values of the technical indicators in the trading pattern and the values of the technical indicators in an input trading day. The maximum distance from the input trading day and a valid pattern is

0.68	0.19	......	0.51	1
0.31	0.11	......	0.84	2
0.69	0.20	......	0.50	2
0.68	0.35	......	0.69	1
0.33	0.48	......	0.06	2
0.34	0.21	......	0.31	1
0.67	0.20	......	0.49	1

(a)

0.68	0.19	......	0.51	1
0.31	0.11	......	0.84	2
0.69	0.20	......	0.50	2
0.68	0.35	.....	0.69	1
0.33	0.48	......	0.06	2
0.34	0.21	......	0.31	1
0.67	0.20	......	0.49	1

(b)

0.68	**0.69**	**0.68**	**0.67**
0.31	0.33	0.34	
0.19	0.20	0.21	0.20
....	....	....	....
0.51	0.50	0.49	

(c)

0.68	**0.19**	**......**	**0.51**	**1**
0.31	0.11	......	0.84	2
0.69	**0.20**	**......**	**0.50**	**2**
0.68	**0.35**	**......**	**0.69**	**1**
0.33	0.48	......	0.06	2
0.34	0.21	......	0.31	1
0.67	**0.20**	**......**	**0.49**	**1**

(d)

0.68	0.19	......	0.51	1
0.69	0.20	......	0.50	2
0.68	0.35	......	0.69	1
0.67	0.20	......	0.49	1

(e)

0.68	0.19		0.51	1
0.69	0.20		0.50	2
0.67	0.20		0.49	1

(f)

0.68	0.197		0.50	1

(g)

Fig. 3 An example for finding trading rule. (**a**) The normalized data matrix M (for simplicity only one of the three *FRRs* is shown). (**b**) Apply agglomerative hierarchical clustering to each column of *M* to achieve many bicluster seeds. (**c**) Store the bicluster seeds in a matrix whose row is a bicluster seed. (**d**) Expand the first bicluster seed to the whole columns. (**e**) Select submatrix containing the rows in bold. (f) Delete the rows or columns till the *MES* is not greater than predefined threshold, the columns and rows in gray are deleted. (g) Average each column to obtain the trading pattern

limited to 0.1. We use a voting method that selects the trading action according to the majority of the pattern labels.

In this study, the trading strategy is based on the assumption that each time we buy the stock, the number of shares is the same. It should be noted that the transaction fee is not taken into account in this study. The trading strategy [37] is as follows:

1: If there is no position, open a position on the next trading day with the opening price when current trading day meets a buy signal.
2: If a position exists, close it on the next trading day with the opening price when meeting a sell signal.
3: If there is buy signal, but no sell signal, during the whole period, sell the stock in the last trading day with closing price.
4: In other cases, there would not be transactions, so the profit rate is set as 0.
5: Finally the profit rate is calculated as Eq. 8:

$$\mathrm{PR} = \sum_{i=1}^{n} \frac{\mathrm{ops}(i) - \mathrm{opb}(i)}{\mathrm{opb}(i)} \tag{8}$$

where ops(i) is the opening price of the trading day when selling the stock, opb(i) is the opening price of the trading day when buying the stock, and n is the number of transactions composed of a pair of buy and sell operations.

Determination of Temporal Stock Investment Style

In the above sections, we have introduced the method for automatic trading based on biclustering mining. Given a stock, the trading patterns discovered in the training period can be divided into three groups, i.e., short term, middle term, and long term. More specifically, if a trading pattern shows its validity with short-term FRR labels, it is regarded as a short-term trading pattern. Likewise, we use a triplet to represent the trading pattern as Tp(Short, Mid, Long). Suppose that the triplet of a trading pattern is Tp(1, 0, 0), it implies that the pattern is suitable for short-term investment but not suitable for mid-term and long-term investments.

Given a stock, we can find many trading patterns from the training periods. Because those patterns cannot always work well for different markets, we further divide the training periods

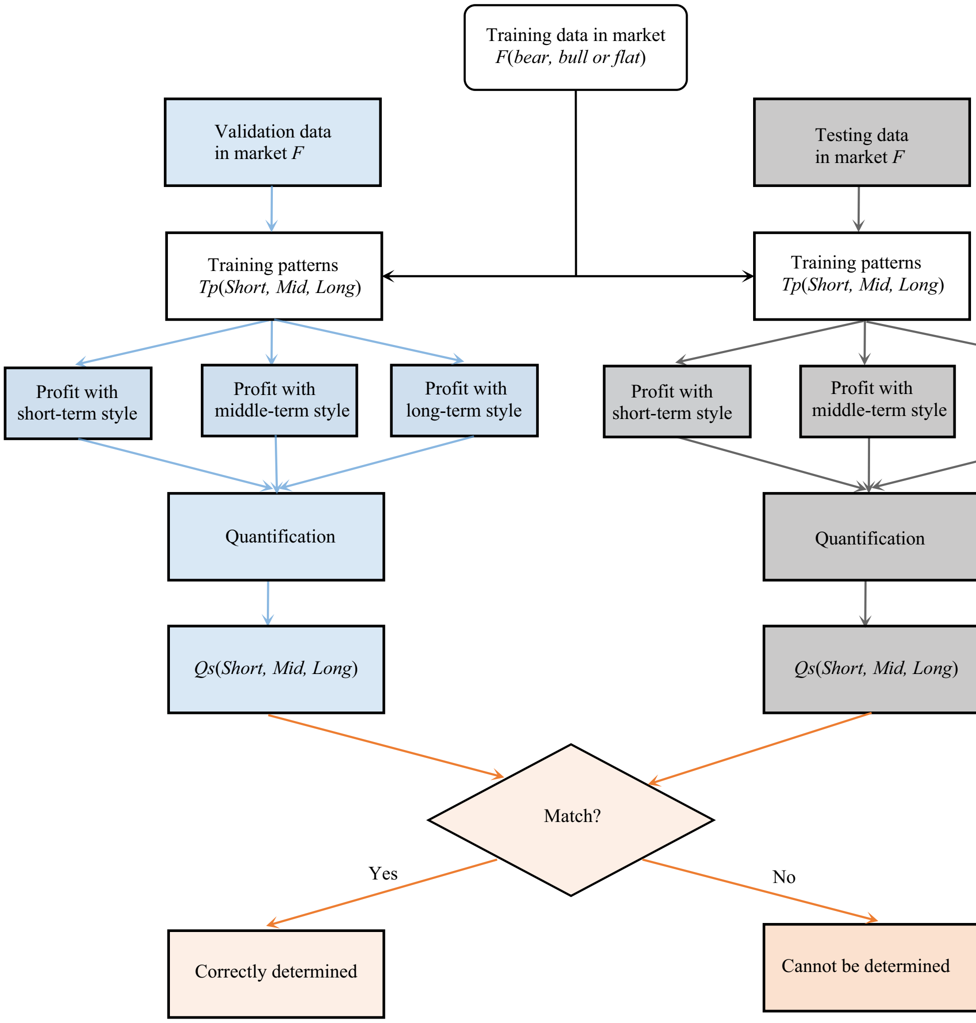

Fig. 4 The flowchart for determining investment style

into three classes with respect to the flags of trend, i.e., the bull, bear, and flat markets, which can be easily recognized from the price curve. This implies that we only investigate the stock's temporal investment style within the same flag of trend.

Consequently, we select another historical period with the same flag (called validation period) to validate whether trading the stock using the discovered patterns can make good profit in a specific term. For example, if the long-term patterns work well and can make good profit in the validation period, we can preliminarily regard the stock as a good long-term investment objective. Accordingly, a quadruple for stocks is defined as Qs(Short, Mid, Long, flag). A stock may be a good investment

Table 2 Description of 10 stocks during bear market

Stock name	Training period	Validation period	Testing period
Fossil Group (FOSL)	2015/2/5–2015/11/23	2015/11/24–2016/4/20	2016/4/21–2016/9/13
Cognizant Technology Solutions (CTSH)	2007/4/17–2008/2/4	2008/2/5–2008/6/27	2008/6/30–2008/11/19
Bed Bath & Beyond (BBBY)	2015/3/30–2016/1/18	2016/1/19–2016/6/10	2016/6/14–2016/11/3
Verizon communications (VZ)	2000/6/6–2001/3/26	2001/3/27–2001/8/17	2001/8/20–2001/8/20
AT&T (T)	2001/1/10–2001/11/4	2001/11/5–2002/4/2	2002/4/3–2002/8/23
Merck & Co (MRK)	2001/1/17–2001/11/8	2001/11/9–2002/4/8	2002/4/9–2002/8/29
International Business Machines (IBM)	2014/6/30–2015/4/20	2015/4/21–2015/9/11	2015/9/14–2016/2/5
HP (HPQ)	2010/6/8–2011/3/25	2011/3/28–2011/8/18	2011/8/19–2012/1/12
3D Systems (DDD)	2014/3/12–2014/12/29	2014/12/30–2015/5/26	2015/5/27–2015/10/16
VanEck Vectors High Inc. Infras MLP ETF((YMLI)	2014/5/12–2015/3/2	2015/3/3–2015/7/24	2015/7/27–2015/12/16

Table 3 Description of the stocks during flat market

Stock name	Training period	Validation period	Testing period
Caterpillar. (CAT)	1998/2/11–1998/11/30	1998/12/1–1999/4/27	1999/4/28–1999/9/20
Home Depot (HD)	2003/9/15–2004/7/6	2004/7/7–2004/11/26	2004/11/29–2005/4/22
Walt Disney (DIS)	2001/11/26–2002/9/16	2002/9/17–2003/2/10	2003/2/11–2003/7/7
Chevron (CVX)	1998/5/26–1999/3/15	1999/3/16–1999/8/6	1999/8/9–1999/12/30
Equinix (EQIX)	2003/9/29–2004/7/20	2004/7/21–2004/12/10	2004/12/13–2005/5/6
NVIDIA (NVDA)	2011/12/5–2012/9/24	2012/9/25–2013/2/21	2013/2/22–2013/7/17
Paychex (PAYX)	2009/2/11–2009/11/30	2009/12/1–2010/4/27	2010/4/28–2010/9/20
Travelers (TRV)	2007/6/18–2008/4/7	2008/4/8–2008/8/28	2008/8/29–2009/1/23
Tractor Supply Company (TSCO)	2004/5/3–2005/2/18	2005/2/22–2005/7/15	2005/7/18–2005/12/7
Walmart (WMT)	2012/10/1–2013/7/24	2013/7/25–2013/12/16	2013/12/17–2014/5/13

objective for more than one investment term. Likewise, a stock may not be a good investment objective for all investment terms.

In order to determine whether the stock is suitable for short-term, middle-term, or long-term investment with a market flag, we randomly select a number of stocks and compare their profit rates using the patterns corresponding to the three investment terms in the validation periods. To simplify the problem, we just normalize the profits gained for all stocks into the range of [0, 1]. Then, we evenly divide the range into thirds and set the stocks whose profits fall into [0, 0.33] to be − 1, those whose profits fall into [0.34, 0.66] to be 0, and those whose profits fall into [0.67, 1] to be 1. After this quantization procedure, every stock can have a label corresponding to each investment term for a specific flag. The quadruples for all stocks can be preliminarily constructed.

However, the quadruples may not always be correct for any trading periods and we need to further testify them. To this end, we select a new trading period called testing period for each stock. If one of the term labels in the quadruple computed from the testing periods is the same as that computed from the validation periods, we consider that the stock's temporal investment style as for the corresponding term can be eventually determined. In contrast, if no term label is matched between the validation and testing periods, the style cannot be determined, implying that the stock may have greatly varied the fluctuation patterns in various investment terms. Please note that we consider the temporal investment style of a stock can be determined if at least one term label can be matched between the validation and testing periods. This is the key idea proposed in this study and is summarized in Fig. 4.

Experiments and Results

Dataset

To evaluate the performance of the proposed method, three experiments were designed. As shown in Tables 2, 3, and 4, we have selected 30 stocks from the US stock markets from various time points during 1998–2018. The stocks were selected from different indices and different industries. As mentioned above, the situation of stock market can be categorized into three groups, namely bear market, flat market, and bull market. There is a detailed description on the training period, validation period, and testing period in the datasets during the bear market in

Table 4 Description of the stocks during bull market

Stock name	Training period	Validation period	Testing period
Alphabet (GOOG)	2012/3/27–2013/1/16	2013/1/17–2013/6/12	2013/6/13–2013/11/4
Exxon Mobil Corporation (XOM)	2003/8/28–2004/6/18	2004/6/21–2004/11/10	2004/11/11–2005/4/7
Apple (AAPL)	2013/7/1–2014/4/21	2014/4/22–2014/9/12	2014/9/15–2015/2/6
Check Point Software technologies(CHKP)	2013/6/10–2014/3/28	2014/3/31–2014/8/21	2014/8/22–2015/1/15
Dollar Tree (DLTR)	2010/8/23–2011/6/10	2011/6/13–2011/11/2	2011/11/3–2012/3/29
Expeditors International of Washington (EXPD)	2016/3/15–2016/12/30	2017/1/3–2017/5/26	2017/5/30–2017/10/19
Fiserv (FISV)	2013/3/26–2014/1/13	2014/1/14–2014/6/9	2014/6/10–2014/10/30
SBA Communications (SBAC)	2011/11/18–2012/9/10	2012/9/11–2013/2/6	2013/2/7–2013/7/2
DENTSPLY SIRONA (XRAY)	2003/6/17–2004/4/5	2004/4/6–2004/8/30	2004/8/31–2005/1/24
Boeing (BA)	2016/8/9–2017/5/30	2017/5/31–2017/10/20	2017/10/23–2018/3/19

Table 2. Likewise, Tables 3 and 4 present the detailed information about the datasets corresponding to flat market and bull market, respectively. The durations for the training, validation, and testing of 30 stocks are the same, i.e., 203, 101, and 101 trading days, respectively.

The program was implemented with MATLAB (MatlabR2014a) and run on Windows10 64 bits Platform with a configuration of Core i5-8250 U and 8 GB memory. After training for each stock, many trading rules corresponding to short-term, middle-term, and long-term investment styles could be obtained.

Experimental Results

In the validation and testing experiments, the results are obtained using the period setting in Tables 2, 3, and 4. On average, the runtime for training the trading rules is 16 min for each stock, and the memory space occupied is about 600 MB. Tables 5, 6, and 7 collect the results obtained over all runs on the validation and testing experiments. Please note that we evenly select 10 stocks for the bull, bear, and flat markets, respectively, in order to avoid occasionality.

The results in Table 5 show that there are 8 stocks each of which has at least one term-wise match in the bear market. Table 6 shows that there are 8 stocks each of which has at least one term-wise match in the flat market. Table 7 shows that there are 9 stocks each of which has at least one term-wise match in the bull market. Therefore, the total accuracy in the 30 stocks is 83.34%. From the results, we can see that not all stocks can have temporal investment style and almost no stock can have the same behaviors in the validation and testing periods for all of the three investment terms. It indicates that the great randomness of the price change exists for all stocks. However, we can still find some stocks that have similar behaviors in the validation and testing periods for one or two specific terms. This can provide a useful hint for the investors to make corresponding allocations of the capital for different investment terms.

Table 5 Results in bear market

Stock name	Validation Qs(s, m, l, f)	Test Qs(s, m, l, f)	Match
FOSL	(1, 0, 1, bear)	(1, 1, 1, bear)	Y(short, long)
CTSH	(− 1, − 1, 1, bear)	(1, 1, 1, bear)	Y(long)
BBBY	(− 1, 0, 0, bear)	(0, 1, 1, bear)	N
VZ	(− 1, 1, 1, bear)	(1, 1, 1, bear)	Y(mid, long)
T	(− 1, − 1, − 1, bear)	(0, − 1, 0, bear)	Y(mid)
MRK	(− 1, − 1, 1, bear)	(0, 1, 1, bear)	Y(long)
IBM	(− 1, − 1, 0, bear)	(0, 0, 1, bear)	N
HPQ	(− 1, 0, − 1, bear)	(− 1, 1, − 1, bear)	Y(short, long)
DDD	(0, 1, 0, bear)	(0, − 1, 1, bear)	Y(short)
YMLI	(− 1, 0, 0, bear)	(0, 1, 0, bear)	Y(long)

Table 6 Results in flat market

Stock name	Validation Qs(s, m, l, f)	Test Qs(s, m, l, f)	Match
CAT	(1, 1, − 1, flat)	(1, 1, 0, flat)	Y(short, mid)
HD	(− 1, 0, 0, flat)	(0, 0, 0, flat)	Y(mid, long)
DIS	(− 1, − 1, − 1, flat)	(0, 1, 0, flat)	N
CVX	(0, 1, 0, flat)	(− 1, 1, 0, flat)	Y(mid, long)
EQIX	(1, 0, 1, flat)	(1, 1, 1, flat)	Y(short, long)
NVDA	(− 1, 0, 0, flat)	(1, 0, 1, flat)	Y(mid)
PAYX	(− 1, − 1, − 1, flat)	(− 1, − 1, − 1, flat)	Y(short, mid, long)
TRV	(− 1, − 1, 0, flat)	(1, 1, 0, flat)	Y(long)
TSCO	(− 1, 1, 0, flat)	(0, 1, 1, flat)	Y(mid)
WMT	(− 1, − 1, − 1, flat)	(0, 1, 0, flat)	N

Conclusions and Future Work

In this paper, a novel idea is put forward to analyze the temporal investment style for stocks. The key steps include biclustering the technical indicators of the trading data to discover trading patterns and trading rules, determining the temporal investment style by validating the trading rules in various trading terms and testifying the styles in the testing periods. If the profit rates can be matched in the validation and testing periods, the corresponding term-wise investment style is eventually determined. Experimental results show that the proposed method could determine the investment style for some US stocks.

However, this study is quite preliminary because the discovered investment styles for those stocks were not fully verified using more trading data. In addition, the trading rules and strategy may not be the best choice in real financial markets. The data matrix for mining the trading patterns is relatively small and more technical indicators should be taken into account in our future work. The three investment terms were empirically selected, and we will test more possible parameters. More importantly, the current method for the

Table 7 Results in bull market

Stock name	Validation Qs(s, m, l, f)	Test Qs(s, m, l, f)	Match
GOOG	(− 1, − 1, − 1, bull)	(− 1, − 1, − 1, bull)	Y(short, mid, long)
XOM	(0, − 1, − 1, bull)	(− 1, − 1, 1, bull)	Y(mid)
AAPL	(1, 0, 0, bull)	(− 1, 0, − 1, bull)	Y(mid)
CHKP	(− 1, − 1, − 1, bull)	(− 1, 0, 0, bull)	Y(short)
DLTR	(1, 1, 1, bull)	(− 1, 1, 1, bull)	Y(mid, long)
EXPD	(− 1, − 1, − 1, bull)	(− 1, 0, 0, bull)	Y(short)
FISV	(− 1, − 1, − 1, bull)	(− 1, 0, 0, bull)	Y(short)
SBAC	(− 1, 0, 0, bull)	(− 1, 0, 0, bull)	Y(short, mid, long)
XRAY	(− 1, − 1, − 1, bull)	(1, 0, 0, bull)	N
BA	(0, 1,1, bull)	(− 1, 1, 1, bull)	Y(mid, long)

determination of the temporal investment style can be improved by incorporating more reasonable inference and multi-view learning methods [38, 39] in the future study.

Funding Information This work was financially supported in part by the National Natural Science Foundation of China under Grant 61571193; in part by the Natural Science Foundation of Guangdong Province, China under Grant 2017A030312006 and Grant 2017A030330247; in part by the Fundamental Research Funds for the Central Universities under Grant 2017MS064; in part by the Science and Technology Program of Guangzhou under Grant 201704020134; and in part by the Project of Science and Technology Department of Guangdong province (Nos. 2016A010101021, 2016A010101022, and 2016A010101023).

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Wang F, Zhang Y, Rao Q, Li K, Zhang H. Exploring mutual information-based sentimental analysis with kernel-based extreme learning machine for stock prediction. Soft Comput. 2017;21(12):3193–205.
2. Chavarnakul T, Enke D. Intelligent technical analysis based equivolume charting for stock trading using neural networks. Expert Syst Appl. 2008;34(2):1004–17.
3. Chang PC, Fan CY, Liu CH. Integrating a piecewise linear representation method and a neural network model for stock trading points prediction. IEEE Trans Syst Man Cybern Part C Appl Rev. 2009;39(1):80–92.
4. Li Y, Liu W, Li X, Huang Q, Li X. GA-SIFT: a new scale invariant feature transform for multispectral image using geometric algebra. Inf Sci. 2014;281:559–72.
5. Li Y, Liu W, Huang Q. Traffic anomaly detection based on image descriptor in videos. Multimed Tools Appl. 2016;75(5):2487–505.
6. Wang YH. Using neural network to forecast stock index option price: a new hybrid GARCH approach. Qual Quant. 2009;43(5):833–43.
7. Tilakaratne CD, Morris SA, Mammadov MA, Hurst C P. Predicting stock market index trading signals using neural networks. In: Proceedings of the 14th annual global finance conference (GFC'07); 2007. pp. 171-179.
8. White H. Economic prediction using neural networks: the case of IBM daily stock returns. In: IEEE International Conference on Neural Networks; 1988. vol. 2, pp451–458.
9. Wu Y, Zhu X, Li L, et al. Mining dual networks: models, algorithms, and applications. ACM Trans Knowl Discov Data. 2016;10(4):40.
10. Zhang D, Jiang Q, and Li X (2004) Application of neural networks in financial data mining. In: International Conference on Computational Intelligence, pp. 392-395.
11. Zhang S, Li X, Zong M, et al. Learning k for knn classification. ACM Trans Intell Syst Technol. 2017;8(3):43.
12. Potvin JY, Soriano P, Vallée M. Generating trading rules on the stock markets with genetic programming. Comput Oper Res. 2004;31(7):1033–1047.4.
13. Hilliard JE, Hilliard, J. Rebalancing versus buy and hold: theory, simulation and empirical analysis. Rev Quant Finan Acc. 2018;50(1):1–32.
14. Chang PC, Liu CH. A TSK type fuzzy rule based system for stock price prediction. Expert Syst Appl. 2008;34(1):135–44.
15. Briza AC, Naval PC Jr. Stock trading system based on the multi-objective particle swarm optimization of technical indicators on end-of-day market data. Appl Soft Comput. 2011;11(1):1191–201.
16. Leigh W, Modani N, Purvis R, Roberts T. Stock market trading rule discovery using technical charting heuristics. Expert Syst Appl. 2002;23(2):155–9.
17. Huang W, Nakamori Y, Wang SY. Forecasting stock market movement direction with support vector machine. Comput Oper Res. 2005;32(10):2513–22.
18. Yang F, Huang Q, Jin L, Wee-Chung Liew A. Segmentation and recognition of multi-model photo event. Neurocomputing. 2016;172:159–67.
19. Zarandi MHF, Rezaee B, Turksen IB, et al. A type-2 fuzzy rule-based expert system model for stock price analysis. Expert Syst Appl. 2009;36(1):139–54.
20. Zhang H, Su J. Naive bayesian classifiers for ranking. In: European conference on machine learning. Berlin: Springer; 2004. p. 501–12.
21. Upadhyay A, Bandyopadhyay G, Dutta A. Forecasting stock performance in Indian market using multinomial logistic regression. J Bus Stud Q. 2012;3(3):16.
22. Asghar MZ, Khan A, Bibi A, Kundi FM, Ahmad H. Sentence-level emotion detection framework using rule-based classification. Cogn Comput. 2017;9(6):868–94.
23. Sun X, Peng X, Ding S. Emotional human-machine conversation generation based on long short-term memory. Cogn Comput. 2018;10(3):389–97.
24. Ding S, Zhang J, Jia H, Qian J. An adaptive density data stream clustering algorithm. Cogn Comput. 2016;8(1):30–8.
25. Huang Q, Wang T, Tao D, Li X. Biclustering learning of trading rules. IEEE Trans Cybern. 2015;45(10):2287–98.
26. Hussain A, Cambria E. Semi-supervised learning for big social data analysis. Neurocomputing. 2018;275:1662–73.
27. Mahmud M, Kaiser MS, Rahman MM, Rahman MA, Shabut AM, Mamun SA, et al. A brain-inspired trust management model to assure security in a cloud based iot framework for neuroscience applications. Cogn Comput. 2018;9:1–10.
28. Malik ZK, Hussain A, Wu QMJ. Extracting online information from dual and multiple data streams. Neural Comput & Applic. 2018;30(1):87–98.
29. Gong C, Tao D, Maybank SJ, Liu W, Kang G, Yang J. Multi-modal curriculum learning for semi-supervised image classification. IEEE Trans Image Process. 2016;25(7):3249–60.
30. Gong C, Liu T, Tao D, Fu K, Tu E, Yang J. Deformed graph Laplacian for semisupervised learning. IEEE Trans Neural Netw Learn Syst. 2015;26(10):2261–74.
31. Gong C, Liu T, Tang Y, Yang J, Yang J, Tao D. A regularization approach for instance-based superset label learning. IEEE Trans Cybern. 2018;48(3):967–78.
32. Wang M, Shang X, Li X, Liu W, Li Z. Efficient mining differential co-expression biclusters in microarray datasets. Gene. 2013;518(1):59–69.
33. Huang Q. A biclustering technique for mining trading rules in stock markets. Appl Inform Commun. 2011;224:16–24.
34. Huang Q, Huang X, Kong Z, Li X, Tao D. Bi-phase evolutionary searching for biclusters in gene expression data. IEEE Trans Evol Comput. 2018. https://doi.org/10.1109/TEVC.2018.2884521.

35. Huang Q, Chen Y, Liu L, Tao D, Li X. On combining biclustering mining and AdaBoost for breast tumor classification. IEEE Trans Evol Comput. 2019. https://doi.org/10.1109/TKDE.2019.2891622.
36. Huang Q, Kong Z, Li Y, Yang J, Li X. Discovery of trading points based on Bayesian modeling of trading rules. World Wide Web. 2018;21(6):1473–90.
37. Ratner M, Leal RPC. Tests of technical trading strategies in the emerging equity markets of Latin America and Asia. J Bank Financ. 1999;23(12):1887–905.
38. Xu C, Tao D, Xu C. A survey on multi-view learning. Comput Sci. 2013.
39. Zhang L, Zhang Q, Zhang L, Tao D, Huang X, Du B. Ensemble manifold regularized sparse low-rank approximation for multiview feature embedding. Pattern Recogn. 2015;48(10):3102–12.