Support vector machine for stock movement direction prediction with sparsity for feature selection

Maoxuan Miao[a], Jinran Wu [*b], Fengjing Cai [†a], Liya Fu [‡c] and You-Gan Wang [§b]

[a]**College of Mathematics and Physics, Wenzhou University, Wenzhou, PR China**

[b]**School of Mathematical Sciences, Queensland University of Technology, Brisbane, Australia**

[c]**School of Mathematics and Statistics, Xi'an Jiaotong University, China**

June 27, 2021

[*]jinran.wu@hdr.qut.edu.au; orcid: 0000-0002-2388-3614

[†]caifj7704@wzu.edu.cn

[‡]fuliya@mail.xjtu.edu.cn; orcid: 0000-0003-0981-391

[§]you-gan.wang@qut.edu.au; orcid: 0000-0003-0901-4671

Abstract

In the stock market, accurate prediction of stock price movement direction can effectively increase the profits for investors. However, the stock price is an extremely complex dynamic system with strong fluctuation, proper selection of technical indicators can potentially improve the accuracy of the direction prediction. We propose a novel sparse least squares support vector machine (LSSVM) by combining recursive feature elimination (RFE) and Relief via a weight parameter. Specially, the benefit if this hybrid is three fold: (1) accounting for any intrinsic correlations among the features, (2) more effective prediction due to the sparse framework capable of removing some "noise" features completely; and (3) simultaneously select technical indicators according to the feature ranking and accounts for possible interactions and possible non-linear effects among the features. Three stock datasets from the liquor and spirits concept are analyzed to demonstrate the superiority of our proposed new framework providing sparse solutions resulting in more accurate predictions and higher returns among all seven considered classifiers.

Keywords: Variable selection; Sparse method; Machine learning; Stock movement direction forecasting.

1 Introduction

Stock investment can be regarded as one of the main choices to increase the profits for individuals and companies. However,the stock price usually is influenced by some uncertainties from economic conditions, social factors, and political events, etc. Thus, highly accurate stock prediction is a very challenging project in such an uncertain stock market (Z. Li et al., 2019). Technical indicators, which can provide an amount of related information, are developed as technical analysis for the price prediction. Now, many studies on direction prediction, which incorporate technical indicators in machine learning methods, have been developed (Gandhmal & Kumar, 2019; Gunduz et al., 2017; Q. Wang et al., 2018). Generally, plenty of technical indicators can be designed according to different prospects; meanwhile, any specific stock usually is only sensitive to indicators. Specially, in this paper, we focus on a sparse machine learning framework, which can remove redundant indicators during model training and provide good stock movement direction predictions.

The work on stock market prediction generally can be divided into two groups: stock price prediction and stock price movement direction. The first group is to predict the stock price value directly by machine learning methods based on stock price series modeling. For example, P. Yu & Yan (2020) proposed to predict stock price by combining time series phase-space method with deep neural networks. Xiao et al. (2020) hybridized least squares support vector machine with an auto-regressive moving average for stock forecasting. On the other hand, the stock price movement direction can benefit decision-making directly, thus such direction prediction has been popularly researched. One of the effective modeling methods is the artificial neural networks (ANNs) (Hu et al., 2018), which usually is modeled based on some novel predictors. For instance, J. Long et al. (2020) incorporated stock market information and public market information into a deep neural network to improve the prediction performance. Chen et al. (2021) integrated graph convolutional features with convolutional neural networks to increase prediction accuracy. Similarly, Ismail et al. (2020) presented a persistent homology method to obtain some invariant topological features as input variables of ANNs to improve the prediction performance. Here, we should mention that some problems of local minima and overfitting during ANNs training often lead to degradation of the predicting performance (W. Long et al., 2019). Different from ANNs methods, support vector machine

(SVM) is also very popular in stock price movement direction forecasting, which is based on a very solid statistical foundation. For example, Hao et al. (2021) developed a fuzzy twin support vector machine to improve the performance of SVM, which is robust to outliers. Notice that compared with ANNs, SVM not only can obtain the global optimum solution but also can prevent over-fitting (L. Yu et al., 2008). However, among SVM methods for stock price movement direction prediction, the work on variable selection (i.e., removing redundant and irrelevant features) is very limited. Therefore, to improve the performance of SVM for predicting stock price movement direction, the feature selection method is incorporated into SVM to eliminate some unimportant features. Simultaneously, due to the low computational cost of the least squares support vector machine (LSSVM) (Suykens & Vandewalle, 1999), the LSSVM is employed for our stock price movement direction prediction, and we name our new SVM framework as sparse least squares support vector machines (Sparse LSSVM) in the paper.

In variable selection fields, sparse methods mainly can be divided into filter methods, wrapper methods,and embedded methods (Zhang et al., 2019), which usually generate different quality of feature subsets. The first methods, filter methods, rely on data characteristics to eliminate redundant, irrelevant, and noise features. Many recent studies have incorporated filter methods in SVM, such as fuzzy filter (Roy et al., 2016), information gain (Kurniawati & Pardede, 2018), and correlation-based feature selection (Khaokaew & Anusas-amornkul, 2016), to enhance its forecasting performance. It should be noted that filter methods ignore interaction with a learning algorithm, and most filter methods suffer from the problem of setting up the default threshold to distinguish influential features from redundant features (Cherrington et al., 2019). The second methods are the wrapper methods, which search for the optimal feature subset in the whole feature space based on the performance of learning algorithms. One of the most obvious cases is the genetic algorithm-based SVM (Tao et al., 2019), whose main advantage is to utilize crossover and mutation to control the balance between exploitation and exploration. However, when dimensions of data are extremely high,the wrapper algorithm has expensive computational cost. Moreover, overuse of classifier performance in feature subset selection would result in overfitting in the feature subset space (Kohavi & John, 1997). Different two mentioned methods before, embedded methods is defined as embedding the feature selection into the model training process, which commonly utilizes a penalty-based method to select feature subsets, such as elastic net-SVM (Lorbert & Ramadge, 2013) and l_p-norm SVM

(Nie et al., 2017). The most typical embedded-based SVM framework is the SVM with the least absolute shrinkage and selection operator (Lasso-SVM) (K. Wang et al., 2020). However, when there are several highly correlated variables, Lasso-SVM tends to pick only a few of them and remove the rest, which may remove important features. Moreover, the essence of penalty-based method is to force coefficients to be zero by controlling hyperparameters. Some useful features may be eliminated, leading to degradation of prediction performance, or all feature is retained, leading to model's interpret-ability unchanged.

Later, to overcome above shortcoming of the feature selection method, SVM-Recursive Feature Elimination (RFE) was proposed by Guyon et al. (2002). It has three main advantages: 1) compared with the wrapper method, its computational cost is low. It only needs to train the classifier once to obtain feature subset and is more robust in overfitting data than the wrapper method (Guyon et al., 2002); 2) The less informative feature is eliminated based on changes in loss function; and 3) it interacts with the learning algorithm without any setting threshold. However, as pointed out by Guyon et al. (2002), its feature evaluation criteria do not consider redundancy between features and relevance between features and target variable, so its feature evaluation result is unreliable.

Motivated by the above, to further improve the stock price movement direction prediction, we present a new sparse SVM framework based on the RFE method and the ReliefF method to select the best technical indicator subsets. Furthermore, three stock datasets from the liquor and spirits concept are investigated to validate the performance of our proposed new sparse framework. Feature evaluation criterion in the new sparse SVM framework takes into account not only feature evaluation criteria based on changes in the loss function, but also conditional dependencies between features to enhance reliability in the evaluation of feature quality. It can therefore eliminate more effectively some unimportant technical indicators and obtain optimal technical indicators subset; In the application to the stock data from three Chinese Liquors, the accuracy of the new framework can achieve as much as 84.71%, 85.29%, and 84.12%, with the corresponding F-measures as 0.8687, 0.8663, and 0.8439, respectively.

The rest of the paper is organized as follows. Section 2 reviews the theoretical background of least squares support vector machines. In Section 3, we illustrate the LSSVM-RFE and ReliefF method; and the proposed LSSVM-RFE with ReliefF framework is described. Then, Section 4 report the case studies of three stock datasets and compare our method with other forecasting frameworks. Finally, Section 5

concludes the paper.

2 Least Square Support Vector Machines

Support Vector Machines, proposed by Cortes & Vapnik (1995), where a hyperplane with maximal margin can separate well positive instances and negative instances. Given a training set $\{(x_1, y_1), \ldots, (x_n, y_n)\}$, where $x_i \in R^{p\times 1}$ is the p-dimensional input vector, and $y_i \in \{-1, 1\}$ is a label. The model can be formulated as $y = \text{sign}(w^T x + b)$ with a scalar b and $w = [w_1; \ldots; w_p] \in R^{p\times 1}$, which determines direction of hyperplane. Based on the structural risk minimization principle, the objective function can be formulated as:

$$\begin{aligned} &\min \frac{1}{2} \parallel w \parallel^2 + C \sum_{i=1}^{n} \xi_i \\ &s.t. \quad y_i(w^T x_i + b) \geq 1 - \xi_i, i = 1, 2, \ldots, n, \\ &\qquad \xi_i \geq 0, i = 1, \ldots, n, \end{aligned} \tag{1}$$

where ξ_i is a tolerable classification error of the i-th sample, and C is a penalty parameter that strikes a balance between structural risk and empirical risk. When C is large, the empirical risk will be emphasized. To speed up the training process for SVM, Suykens & Vandewalle (1999) modified the Eq. (1) and proposed the LSSVM, which can solve a linear equation to obtain the solution, as:

$$\begin{aligned} &\min \frac{1}{2} \parallel w \parallel^2 + C \sum_{i=1}^{n} \xi_i^2 \\ &s.t. \quad y_i(w^T x_i + b) \geq 1 - \xi_i, i = 1, 2, \ldots, n, \\ &\qquad \xi_i \geq 0, i = 1, \ldots, n. \end{aligned} \tag{2}$$

To solve Eq. (2), a Lagrangian function is constructed as:

$$\mathcal{L}(w, b, \xi_i, \alpha_i) = \frac{1}{2} w^T w + C \sum_{i=1}^{n} \xi_i^2 - \sum_{i=1}^{n} \alpha_i \Big[y_i(w^T x_i + b) - 1 + \xi_i \Big], \tag{3}$$

with Lagrange multipliers α_i.

Then, with the derivative of Eq. (3) with respect to w, b, ξ_i, α_i, respectively, and simple substitutions, the

solution can be obtained as:

$$\begin{cases} \sum_{i=1}^{n} \alpha_i y_i = 0, \\ \sum_{i=1}^{n} \alpha_i y_i y_j x_i x_j + \frac{1}{C}\alpha_j + b y_j = 1, j = 1, \ldots, p. \end{cases} \tag{4}$$

3 Sparse LSSVM

In this section, we first review the basic framework of LSSVM-RFE by Guyon et al. (2002) and the basic idea of ReliefF (Robnik-Šikonja & Kononenko, 2003). Then, we present our new sparse LSSVM framework based on the LSSVM-RFE and ReliefF.

3.1 LSSVM-RFE

LSSVM-Recursive feature elimination (LSSVM-RFE) is that nested subsets of features are selected in a sequential backward elimination manner, which starts with all the features and each time removes one feature with the smallest ranking score (Guyon et al., 2002). It inherits the two advantages of the wrapper and filter algorithms: (i) the learning of feature subset interact with learning algorithm; and (ii) the computational cost is low, and it only trains LSSVM classifier once at a time to obtain feature subsets. The evaluation of feature quality for LSSVM-RFE is based on changes in objective function by removing a given feature. These changes can be approximated by the optimal brain damage (OBD) algorithm (LeCun et al., 1990) as:

$$\triangle\mathcal{L}(j) = \frac{\partial \mathcal{L}}{\partial w_j}\triangle w_j + \frac{1}{2}\frac{\partial^2 \mathcal{L}}{\partial w_j^2}(\triangle w_j)^2, \tag{5}$$

where $\triangle w_j = w_j$ corresponds to removing the j-th feature, and $\triangle\mathcal{L}(j)$ is changes in the objective function caused by removing the j-th feature with the objective function $\mathcal{L}$ as Eq. (3). Here, at the optimum of $\mathcal{L}$, $\frac{\partial \mathcal{L}}{\partial w_j}$ is set as 0. Then, Eq. (5) becomes

$$\triangle\mathcal{L}(j) = \frac{1}{2}w_j^2. \tag{6}$$

Therefore, evaluation of the j-th feature quality is w_j^2. Apparently, the larger the w_j^2 is, the more important the j-th feature is.

3.2 ReliefF

ReliefF algorithm not only takes into account conditional dependencies between features given the predicted value and correlation between feature and class but also is more robust in noisy data compared with Relief (Robnik-Šikonja & Kononenko, 2003). The key idea of the ReliefF is to estimate the quality of features according to how well their values distinguish between instances that are near to each other. For that purpose, assume that the training set $\{(x_1, y_1), \ldots, (x_n, y_n)\}$, where $x_i = (x_i^{(1)}, \cdots, x_i^{(p)}) \in R^{p\times 1}$, $y_i \in \{-1, 1\}$; and all feature have been normalized. The evaluation of feature quality is as follows:

1) Suppose instance (x_i, y_i) is selected from training set $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ $(i = 1, 2, \ldots, n)$;

2) Searching for K of its nearest neighbors from the same class as $H = [H_1, \ldots, H_K]$, and K nearest neighbors from the other class as $M = [M_1, \ldots, M_K]$ by using Algorithm 1;

3) The quality estimation W_j for the j-th feature is updated, $j = 1, \ldots, p$, by using Eq. (7) as

$$W_j = W_j - \sum_{H_k \in H} (H_k^{(j)} - x_i^{(j)})/(m \cdot K) + \left[\frac{P(C)}{1 - P(y_i)} \sum_{M_k \in M} (M_k^{(j)} - x_i^{(j)})\right]/(m \cdot K), \tag{7}$$

   where m is the number of iterations; y_i is the class of instance x_i; C is instance class, which is different from the class of instance x_i; $P(C)$ is the probability of an instance being from the class C; $P(y_i)$ is the probability of an instance being from the class y_i; and

4) Repeat (1)-(3) process for m times.

Here, for big data modeling, m instances are randomly sampled from a training set and m is a user-defined parameter (Robnik-Šikonja & Kononenko, 2003). Since the selected stock dataset size in the paper is not very large, the number of iterations m is set as the training size (i .e., all instances in training set $\{x_i, y_i\}$ is selected). The specific algorithm is showed in Algorithm 2.

3.3 LSSVM-RFE with ReliefF

Before the presentation of our new LSSVM-RFE with ReliefF, we should mention two points again: 1) the LSSVM-RFE does not comprehensively take into account the redundancy among technical indicators and

Algorithm 1 Searching for nearest hits set H and nearest misses set M for the $i-$th instance

Input: Instance (x_i, y_i); distance vector of x_i, i.e, $D(i,:) = (d_{i1}, \ldots, d_{in})$, with $d_{ic} = \sum_{j=1}^{p} |x_i^{(j)} - x_c^{(j)}|$, the nearest hits set $H = [\quad]$, and the nearest misses set $M = [\quad]$;
Each element d_{ic} in $D(i,:)$ is ranked according to increasing sort as $D(i,:) = [d_{iq_1}, \ldots, d_{iq_n}]$, where q_i represents instance index of the i-th small distance d_{iq_i} between x_i and x_{q_i};
for $q = q_1, q_2, \cdots, q_n$ **do**
  if $x_i \neq x_q$ and $y_i = y_q$ **then**
    $H = [x_q, H]$;
  else if $x_i \neq x_q$ and $y_i \neq y_q$ **then**
    $M = [x_q, M]$;
  if $|H| = K$ and $|M| = K$ **then**
    Break;
Output: The nearest hits set $H = [H_1, \ldots, H_K]$, and the nearest misses set $M = [M_1, \ldots, M_K]$.

Algorithm 2 ReliefF

Input: Training data $\{x_i, y_i\}, i = 1, \ldots, n$;
Initialization: $W_j = 0$, $j = 1, \ldots, p$;
for $i = 1, \ldots, n$ **do**
  Choose an instance x_i in the training set $\{(x_1, y_1), \ldots, (x_n, y_n)\}$;
  Find the nearest hits set H of x_i and the nearest misses set M of x_i by using algorithm 1;
  for $j = 1, \ldots, p$ **do**
    $W_j = W_j - \sum_{H_k \in H} (H_k^{(j)} - x_i^{(j)})/(m \cdot K) + \left[\frac{P(C)}{1-P(y_i)} \sum_{M_k \in M} (M_k^{(j)} - x_i^{(j)})\right]/(m \cdot K)$;
Output: Update the estimators W_j of the qualities of feature j, $j = 1, \ldots, p$

relevancy between target variable and technical indicators, which may not yield optimal feature subset; and 2) the ReliefF algorithm sometimes ignores the interaction with the learning algorithm. Therefore, we incorporate the ReliefF method into LSSVM-RFE to develop our more effective sparse LSSVM.

In our proposed framework of LSSVM-RFE with reliefF, the quality of each feature is evaluated based on Eq. (8) as

$$C_j = \alpha W_j + (1 - \alpha) w_j^2, \tag{8}$$

where parameter $\alpha \in [0, 1]$ determines the trade-off between ReliefF ranking and LSSVM-RFE ranking. W_j is obtained by using iteration of Eq. (7), while w_j^2 is obtained by using LSSVM-RFE algorithm. Then the feature corresponding to the smallest C_j is removed, and the quality of the feature subset composed of the rest features is evaluated by classifier performance, such as average accuracy.

Now, the procedure of obtaining optimal feature subset in LSSVM-RFE with reliefF is given as:

Step 1 Initialization. Import original feature set $S = [1, \ldots, p]$;

Step 2 Process. Score W_j of feature j on the feature set S and score w_j^2 of feature j on the feature set S is

calculated by reliefF method and LSSVM-RFE method, respectively;

Step 3 Evaluation. Utilize Eq. (8) to obtain a quality of feature j on the feature set S;

Step 4 Remove. Find a feature j corresponding to the smallest C_j on the feature set S. Then, feature j is removed on feature set S, i.e. $S = S \setminus j$;

Step 5 Evaluation. The quality of removed feature set S is evaluated by classifier performance. Calculate average accuracy A of LSSVM on training sets with features in removed feature sets S by using cross-validation. The average accuracy A is considered as quality of removed feature set S;

Step 6 Repeat Step 2-5 until feature set S is empty; and

Step 7 Output. The feature set S corresponding to the highest average accuracy A is found and output.

Furthermore, the flowchart of our new sparse LSSVM is shown in Fig. 1.

4 The case study

In this section, three stock datasets are used to validate the performance of stock price movement direction forecasting with our proposed LSSVM-RFE with ReliefF. Specifically, we employ the sparse LSSVM for one-day-ahead stock price movement direction forecasting.

4.1 Data gathering and preparation

In this subsection, three stock datasets from 18 April 2016 to 30 December 2020 in Liquor and Spirits Concept, including Jinhui Liquor Co., Ltd (603919), King's Luck Brewery Joint-Stock Co., Ltd. (603369), V V Food & Beverage Co., Ltd. (600300), are selected for our demonstration, which are archived from Hithink RoyalFlush Information Network Co., Ltd. Here, each stock has 49 predictors (which are reported in Tab. 1), and their assigned label is calculated with Eq. (9) as,

$$\text{target}_{t+1} = \begin{cases} 1 & \text{open}_{t+1} \geq \text{open}_t, \\ -1 & \text{else}, \end{cases} \tag{9}$$

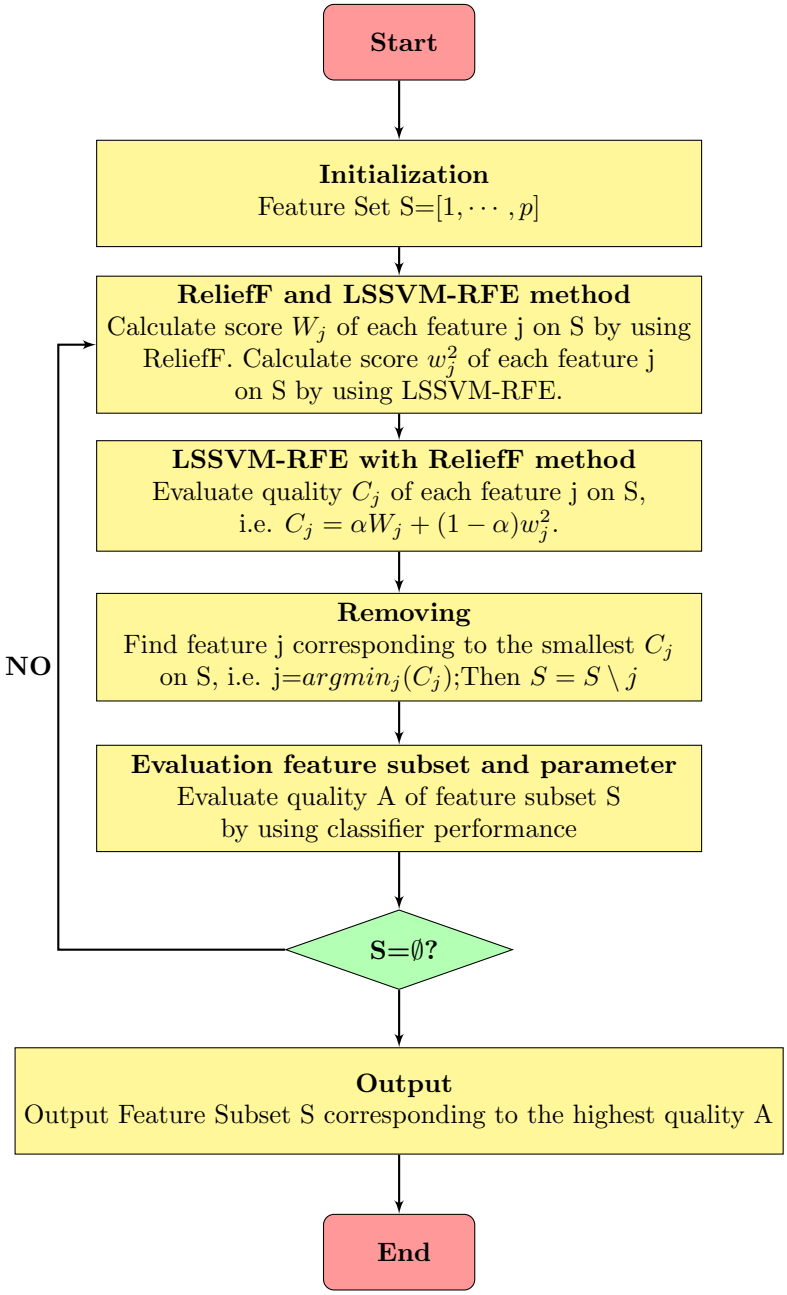


Figure 1: The flowchart of the proposed RFE-LSSVM with ReliefF training

where open_t represents the opening price at the t-th day.

Table 1: The technical indicators

Indicator types	Technical indicators
Turnover Index	Quantity Relative Ratio (QRR) Total Amount Weighted Stock Index (TAPI) Volume Moving Average (VMA) Volume Oscillator (VOSC) Volume Standard Deviation (VSTD) Vol Moving Average Convergence and Divergence (VMACD)
Energy Index	Energy Index (CR) Psychological Line (PSY) Volume Ratio (VR) Popularity Index (AR) Willingness Indicator (BR) Williams Accumulation/Distribution (WAD)
Volume Price Index	Modified On balance volume (MOBV) Price and Volume Trend (PVT) William's Variable Accumulation Distribution (WVAD)
Directional Movement Index	Bull And Bear Index (BBI) Moving average (MA) Exponential Moving Average (EXPMA) Moving Average Convergence and Divergence (MACD) Momentom Index (MTM) Price Oscillator (PRICEOSC) Triple Exponential Smoothed Average (TRIX)
Inverse Directional Movement Index	Bias Ratio (BIAS) Commodity Channel Index (CCI) Deviation Bias Ratio Convergence and Divergence (DBCD) Detrended Price Oscillator (DPO) Stochastic K% (K) Stochastic %D (D) Stochastic %J (J) Relative Strength Index (RSI) Swing Ratio of Dynamic Movement (SRDM) Volume Rate of Change (VROC) Volume Relative Strength Index (VRSI) Williams&Rate (WR)
OverBought & OverSold Pressure and Support index (SSL)	Adaptive Medias (ADTM)
	Contrarian Operation (CDP)
Relative Strength Index	Market Synchronization Index (DPTB) Stage Strength Indicator (JDQS) Stage Weakness Indicator (JDRS)
Oscillator	Momentum Index (MI) Momentum Index Convergence and Divergence (MICD) Rate of Change (RC) Rate of Change Convergence and Divergence (RCCD) Momentum Index Correction Indicator (SRMI)
Volatility Indicator	Average True Range (ATR) Chaikin Volatility (CVLT) Mass Index (MASS) Standard Deviation (STD) Vertical Horizontal Filter (VHF)

In addition, all three datasets are divided into two parts: training set from 18 April 2016 to 21 April 2020 and test set from 22 April 2020 to 30 December 2020. The label distribution for three stocks in training and test sets are plotted in Fig. 2.

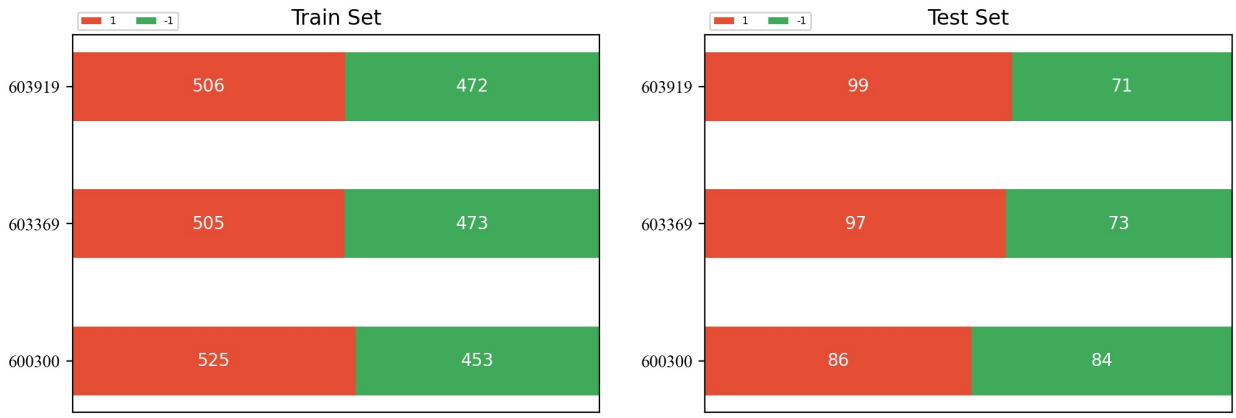


Figure 2: The label distributions of training and test sets for three investigated stocks

4.2 Performance metrics

In this subsection, based on True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN), we calculate six performance metrics (Accuracy, Recall, Specificity, Precision, and F-measure) to measure the prediction performance of stock price movement price direction as,

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \tag{10}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{11}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}, \tag{12}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{13}$$

and

$$\text{F-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{14}$$

4.3 Correlation Analysis

In the subsection, the correlation analysis on training sets is conducted to explore the correlation among all considered variable. Here, we plot the correlation between technical indicators and technical indicators and technical indicators and target variable (target) in Fig. 3.

As illustrated in Fig. 3, for all stocks, the target variable is strongly correlated with MOBV, while the

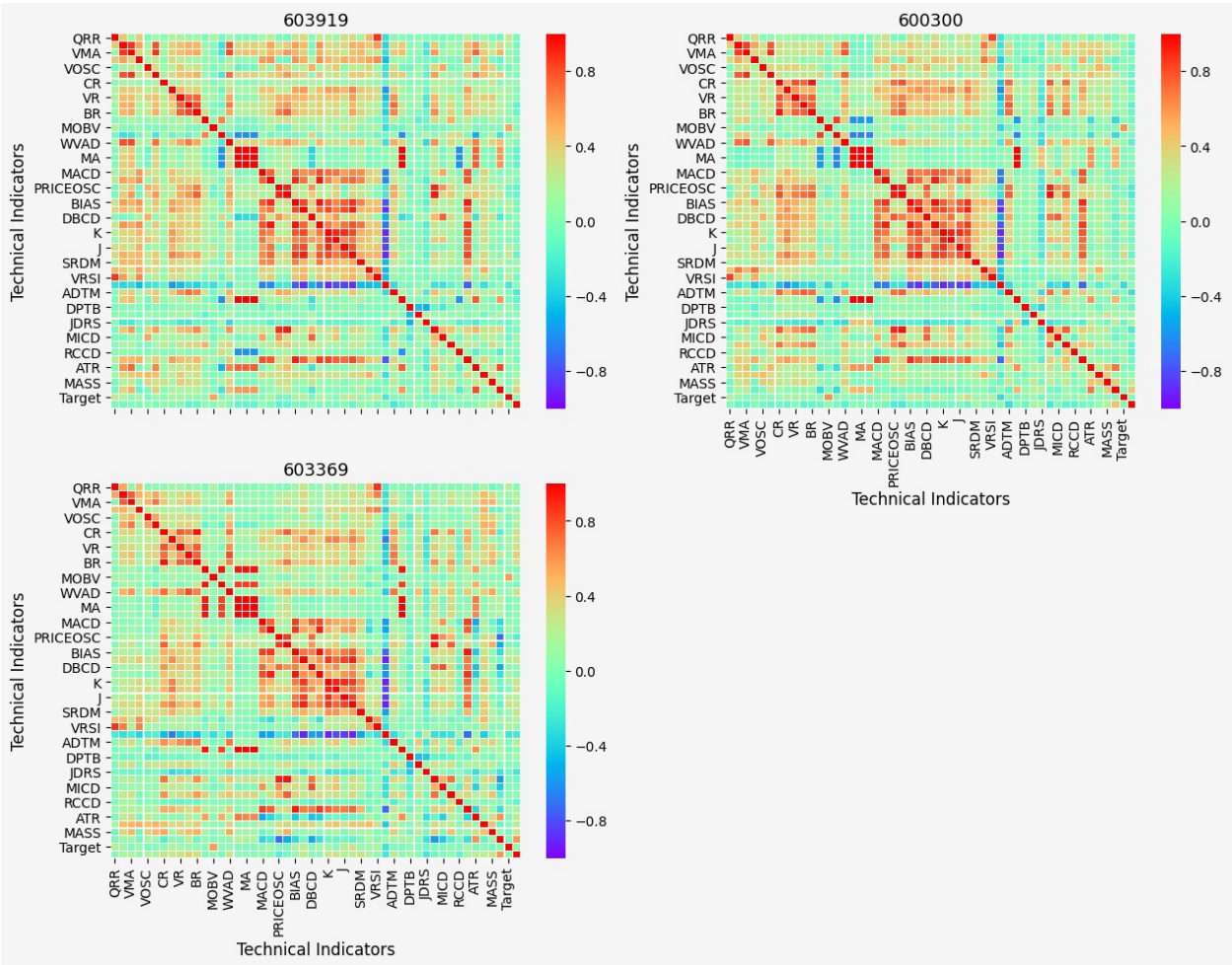


Figure 3: The correlation analysis of technical indicators and response for three stocks

target variable is weakly correlated with other predictors. Potentially, the weak correlated technical indicators would provide a little useful information for model establishment. Thus, the feature selection method is promising to eliminate redundant features to improve the accuracy of stock price movement direction prediction.

4.4 The proposed LSSVM-RFE with ReliefF training

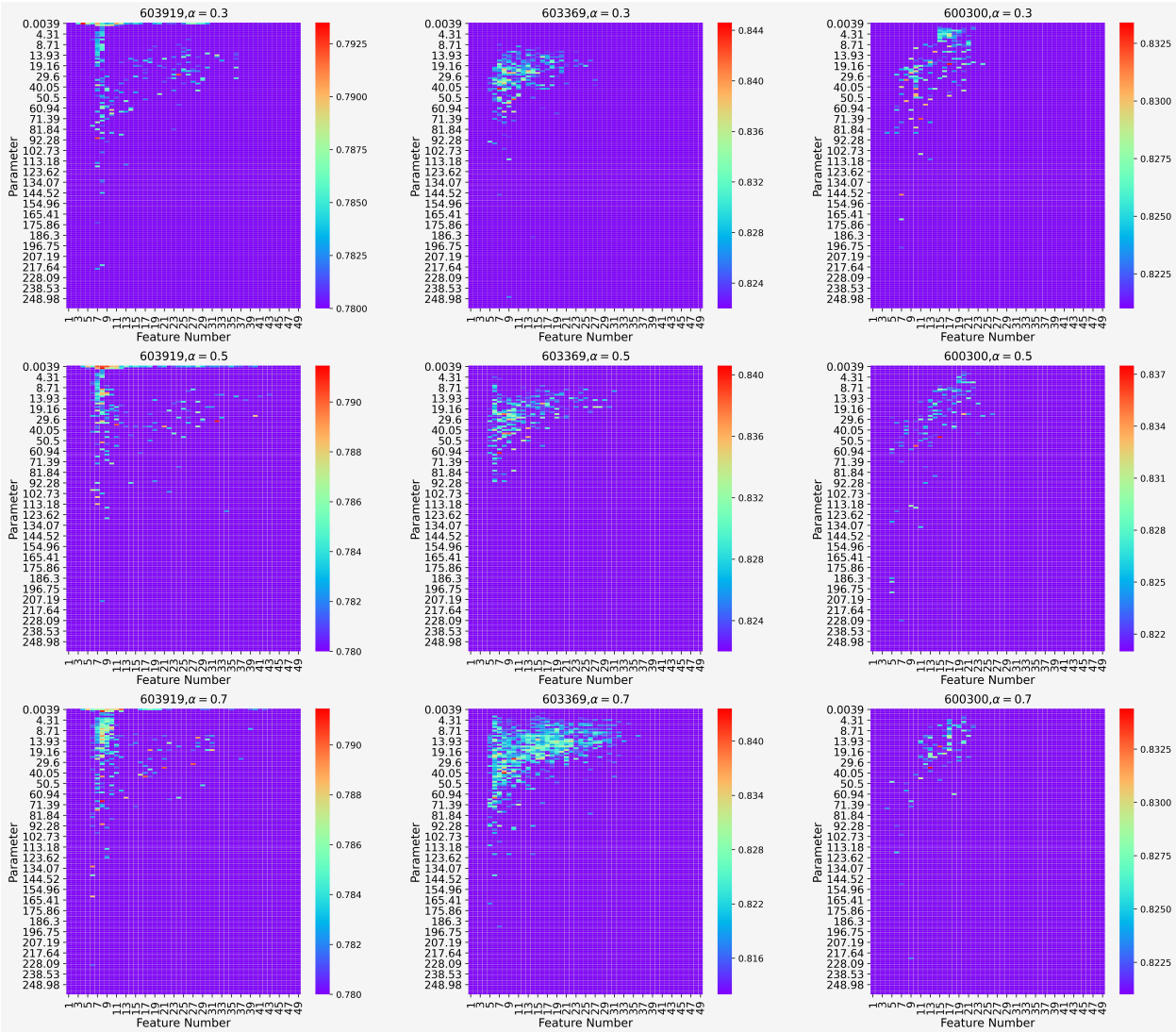
We normalize all 49 technical indicators to eliminate the impact from scale of each indicator. Furthermore, according to the average accuracy in validation sets, we use 5-fold cross-validation method in training sets to search for the two hyper-parameters α in Eq. (8) and C in Eq. (2). In particular, we present three alternative weight parameters as 0.3, 0.5, 0.7. Meanwhile, we first do a coarse search for C from the sequence $\left[2^{-8}, 2^{8}\right]$ with step size $2^{0.8}$; then, we do the fine search for C from the preset sequence $\left[2^{-4}, 2^{4}\right]$ with step size $2^{0.5}$ (Yuanyuan et al., 2017). In addition,When size of nearest hits set H and nearest misses set M is reasonably small,important features can be better separated from unimportant features(Robnik-Šikonja & Kononenko (2003)).Therefore,the size of their sets is set 5.The linear kernel is chosen for our proposed sparse LSSVM. The results of our 5-fold cross-validation are detailed in Fig. 4. According to the best average accuracy in Fig. 4, we report the optimal hyper-parameter settings and corresponding accuracy and features subsets with the LSSVM-RFE with ReliefF for three investigated stocks in Tab. 2.

Table 2: The training results for our LSSVM-RFE with ReliefF

Stock	α	C	Accuracy	Feature subset
603919	0.3	0.0039	0.7935	MOBV, D, RSI, VRSI
603369	0.3	41.7903	0.8446	MOBV, BBI, MA, EXPMA, BIAS, DPO, CDP
600300	0.5	47.0136	0.8375	QRR, TAPI, MOBV, WVAD, BBI, EXPMA, DPO, D, RSI, VRSI, WR, ADTM, CDP, MI, CVLT

4.5 The experiment results

We retrain the LSSVM-RFE with ReliefF by using the tuning hyper-parameters and feature subset in Tab. 2 on the whole training sets for three stock datasets, respectively. Then, the forecasting performances for stock price movement direction with five error indexes (Accuracy, Recall, Specificity, Precision, and



n

Figure 4: The average accuracy with different pre-set parameters and feature number for three stocks

F-measure) are reported in Tab. 3.

Table 3: The performance of LSSVM-RFE with ReliefF

Stock	Accuracy	Recall	Specificity	Precision	F-measure
603919	0.8471	0.8687	0.8169	0.8687	0.8687
603369	0.8529	0.8351	0.8767	0.9000	0.8663
600300	0.8412	0.8488	0.8333	0.8391	0.8439

As illustrate in Tab. 3, our proposed sparse LSSVM, LSSVM-RFE with ReliefF, can provide very excellent forecasting performance in the test set, where all three Accuracy indexes are larger than 84%; furthermore, the three F-measure indexes are also very high as 0.8687, 0.8663, and 0.8439, for Stocks 603919, 603369, and 600300, respectively.

4.6 The model comparison

To further demonstrate the capacity of our proposed LSSVM-RFE with ReliefF framework, we investigate seven classifiers: LSSVM (Razavi et al., 2019), CFS-based LSSVM, Backward Elimination-LSSVM, ReliefF-based LSSVM (Jin et al., 2019), LASSO-LSSVM (Roth, 2004), Elastic Net-LSSVM, and LSSVM-RFE (Guyon et al., 2002), where the details of the CFS-based LSSVM, Elastic Net-LSSVM, and the Backward Elimination-LSSVM are illustrated in Appendix A. All LSSVM models are considered with linear kernel. Additionally, based on the cross-validation method on training sets, the hyper-parameter settings for all considered seven comparative models are recorded in Appendix B. Finally, the forecasting performances for three stock datasets are reported in Tab. 4, and the corresponding feature subsets with all considered sparse LSSVM models are displayed in Tab. 5.

4.6.1 The comparison of error indexes

This subsection discusses the results of model comparison as following four aspects:

1) *Sparse LSSVM vs. LSSVM:*

According to Tab. 4 and Tab. 5, the MOBV is the most technical indicator for stock price movement direction prediction. Interestingly, the CFS-based LSSVM which excludes the MOBV indicator is remarkably worse than the basic LSSVM without variable selection. Another point also can be found

Table 4: Forecasting performance for the three stock datasets

Panel A: Stock 603919					
Model	Accuracy	Recall	Specificity	Precision	F-measure
LSSVM	0.7765	0.7778	0.7746	0.8280	0.8021
CFS-based LSSVM	0.6588	0.7172	0.5775	0.7030	0.7100
ReliefF-based LSSVM	0.8059	0.7778	0.8451	0.8750	0.8235
Backward Elimination-LSSVM	0.8176	0.7778	**0.8732**	**0.8953**	0.8324
LASSO-LSSVM	0.8176	0.8081	0.8310	0.8696	0.8377
Elastic Net-LSSVM	0.8059	0.7778	0.8451	0.8750	0.8235
LSSVM-RFE	0.8235	0.8283	0.8169	0.8632	0.8454
LSSVM-RFE with ReliefF	**0.8471**	**0.8687**	0.8169	0.8687	**0.8687**
Panel B: Stock 603369					
Model	Accuracy	Recall	Specificity	Precision	F-measure
LSSVM	0.7824	0.6907	0.9041	0.9054	0.7836
CFS-based LSSVM	0.5588	0.5670	0.5479	0.6250	0.5946
ReliefF-based LSSVM	0.8176	0.8247	0.8082	0.8511	0.8377
Backward Elimination-LSSVM	0.8000	0.7113	**0.9178**	**0.9200**	0.8023
LASSO-LSSVM	0.8235	0.7938	0.8630	0.8851	0.8370
Elastic Net-LSSVM	0.8000	0.8144	0.7808	0.8316	0.8229
LSSVM-RFE	0.8118	0.8247	0.7945	0.8421	0.8333
LSSVM-RFE with ReliefF	**0.8529**	**0.8351**	0.8767	0.9000	**0.8663**
Panel C: Stock 600300					
Model	Accuracy	Recall	Specificity	Precision	F-measure
LSSVM	0.8059	0.8140	0.7976	0.8046	0.8092
CFS-based LSSVM	0.5941	0.7326	0.4524	0.5780	0.6462
ReliefF-based LSSVM	0.8118	0.8372	0.7857	0.8000	0.8182
Backward Elimination-LSSVM	0.8176	0.8256	0.8095	0.8161	0.8208
LASSO-LSSVM	0.8059	**0.8488**	0.7619	0.7849	0.8156
Elastic Net-LSSVM	0.8118	0.8372	0.7857	0.8000	0.8182
LSSVM-RFE	0.8000	**0.8488**	0.7500	0.7766	0.8111
LSSVM-RFE with ReliefF	**0.8412**	**0.8488**	**0.8333**	**0.8391**	**0.8439**

that some weakly-correlated technical indicators also determine the forecasting performance. Particularly, for Stock 603369, the Specificity with LSSVM is 0.9041 while that with the Elastic Net-LSSVM is 0.7808.

2) *LSSVM-RFE with ReliefF vs. Backward Elimination-LSSVM:*

Both the LSSVM-RFE with ReliefF and the Backward Elimination-LSSVM adopt the backward elimination method. Moreover, in Tab. 4, two indexes (Accuracy and F-measure) of the LSSVM-RFE with ReliefF are higher than those of Backward Elimination LSSVM in two stock datasets. We can find that compared with the Backward Elimination Method based greedy way, backward elimination method based feature evaluation criterion can select a better technical indicator subset.

3) *LSSVM-RFE with ReliefF vs. ReliefF-based LSSVM and LSSVM-RFE:*

In Tab. 4, compared with the LSSVM-RFE, all five indexes (Accuracy, Recall, Specificity, Precision and F-measure) of the proposed LSSVM are the most outstanding in all stock datasets. This means the evaluation of technical indicators quality is more reliable with the proposed method, which can more accurately eliminate unimportant technical indicators. Compared with ReliefF-based LSSVM, the performance of the LSSVM-RFE with ReliefF outperforms the ReliefF-based LSSVM in all datasets, which shows that the feature selection method interacts with the learning algorithm, which can improve the quality of the technical indicator subset.

4) *LSSVM-RFE with ReliefF vs. Embedded-based LSSVM:*

In Tab. 4, the performance of the LSSVM-RFE with ReliefF outperforms the LASSO-LSSVM and the Elastic Net-LSSVM in all stock datasets according to accuracy and F-measure. It shows that compared with eliminating some unimportant technical indicators by controlling hyper-parameters, using improved feature evaluation criterion can effectively eliminate some unimportant technical indicators.

Table 5: Technical indicator subset with different spares LSSVMs on three stock datasets

Panel A: Stock 603919	
Model	Technical Indicator Subset
CFS-based LSSVM	VOSC,BR,D,J,TAPI,VMA
ReliefF-based LSSVM	**MOBV**,VRSI,QRR,RSI,WR,DPO,BIAS,TAPI,RC,MTM,WVAD,VOSC,VSTD,VMA, SRDM
Backward Elimination-LSSVM	All Technical Indicators Except For QRR,VOSC,CR,AR,MACD,PRICEOSC,DPO, D,SRDM,ADTM,DPTB,SRMI
LASSO-LSSVM	All Technical Indicators Except For VSTD,BR,WAD,PVT,EXPMA,PRICEOSC,BIAS, D,SRMI,ATR,STD
Elastic Net-LSSVM	All Technical Indicators Except For CR,PRICEOSC,STD
LSSVM-RFE	TAPI,VMA,**MOBV**,BBI,MA,MACD,DPO,K,RSI,WR,CDP,MI
LSSVM-RFE with ReliefF	**MOBV**,D,RSI,VRSI
Panel B: Stock 603369	
Model	Technical Indicator Subset
CFS-based LSSVM	CR,TAPI,VMA,VMACD,VOSC,VSTD
ReliefF-based LSSVM	**MOBV**,RSI,VRSI,QRR,WR,TAPI,SRDM,D,PSY,BIAS
Backward Elimination-LSSVM	All Technical Indicators Except For QRR,VMA,VOSC,PSY,BR,WVAD,EXPMA,TRIX, CCI,CDP,SRMI
LASSO-LSSVM	ALL Technical Indicators Except For TAPI,VSTD,BR,WAD,WVAD,BBI,MA,EXPMA, PRICEOSC,TRIX,CCI,DPO,J,VROC,JDQS,MI,MICD,SRMI,CVLT,MASS,VHF
Elastic Net-LSSVM	QRR,VMA,VOSC,CR,PSY,**MOBV**,TRIX,D,RSI,WR,JDRS,RC
LSSVM-RFE	VMACD,CR,PSY,**MOBV**,MACD,MTM,BIAS,K,D,RSI,VRSI,WR
LSSVM-RFE with ReliefF	**MOBV**,BBI,MA,EXPMA,BIAS,DPO,CDP
Panel C: Stock 600300	
Model	Technical Indicator Subset
CFS-based LSSVM	TAPI,VMA,VMACD.VOSC,PRICEOSC,CCI
ReliefF-based LSSVM	**MOBV**,RSI,WR,VRSI,QRR,BIAS,DPO,PSY,VROC,AR
Backward Elimination-LSSVM	All Technical Indicators Except For TAPI,VMACD,CR,VR,BR,PVT,PRICEOSC,BIAS, K,SRDM,ADTM,MI,RC,ATR,MASS
LASSO-LSSVM	QRR,**MOBV**,MACD,D,RSI
Elastic Net-LSSVM	**MOBV**,RSI
LSSVM-RFE	**MOBV**,MACD,DPO,K,D,RSI,WR,MICD
LSSVM-RFE with ReliefF	QRR,TAPI,**MOBV**,WVAD,BBI,EXPMA,DPO,D,RSI,VRSI,WR,ADTM,CDP,MI,CVLT

4.6.2 Comparison of cumulative return rate

To further explore the superiority of our proposed sparse LSSVM, we design a stock investment strategy based on all forecasting results. Generally, investors prefer to focus on the cumulative return rate from a classifier. Regardless of transaction costs such as commission, cumulative return rate with investment strategy showed in Eq. (15) as,

$$R_{T+t} = \frac{\sum_{i=1}^{t} P_{T+i}}{O_T}$$
$$\text{where } P_{T+i} = \begin{cases} O_{T+i} - O_{T+i-1}, & Y_{T+i} = 1, \\ 0, & \text{else}, \end{cases} \tag{15}$$

where R_{T+t} denotes cumulative rate of return from the T-th day to the $(T+i)$-th day and O_T denotes opening price of a stock on the T-th day. P_{T+i} denotes that if the classifier's prediction result is 1 on the $(T+i)$-th day, that is, $Y_{T+i} = 1$, the increase in profit on the $(T+i)$-th day is $O_{T+i} - O_{T+i-1}$. Otherwise, the increase in profit on the $(T+i)$-th day is 0.

Additionally, the cumulative return rate without investment strategy is calculated as Eq. (16),

$$R_{T+t} = \frac{\sum_{i=1}^{t}(O_{T+i} - O_{T+i-1})}{O_T}, \tag{16}$$

where R_{T+t} denotes the cumulative rate of return from the T-th day to the $(T+i)$-th day, and O_T denotes the opening price of stock on the T-th day.

In the experiment, we buy these stocks on the opening price of 22 April 2020, respectively. All cumulative return rates for all investigated forecasting models are calculated from 23 April 2020 to 31 December 2020, which are reported in Fig. 5.

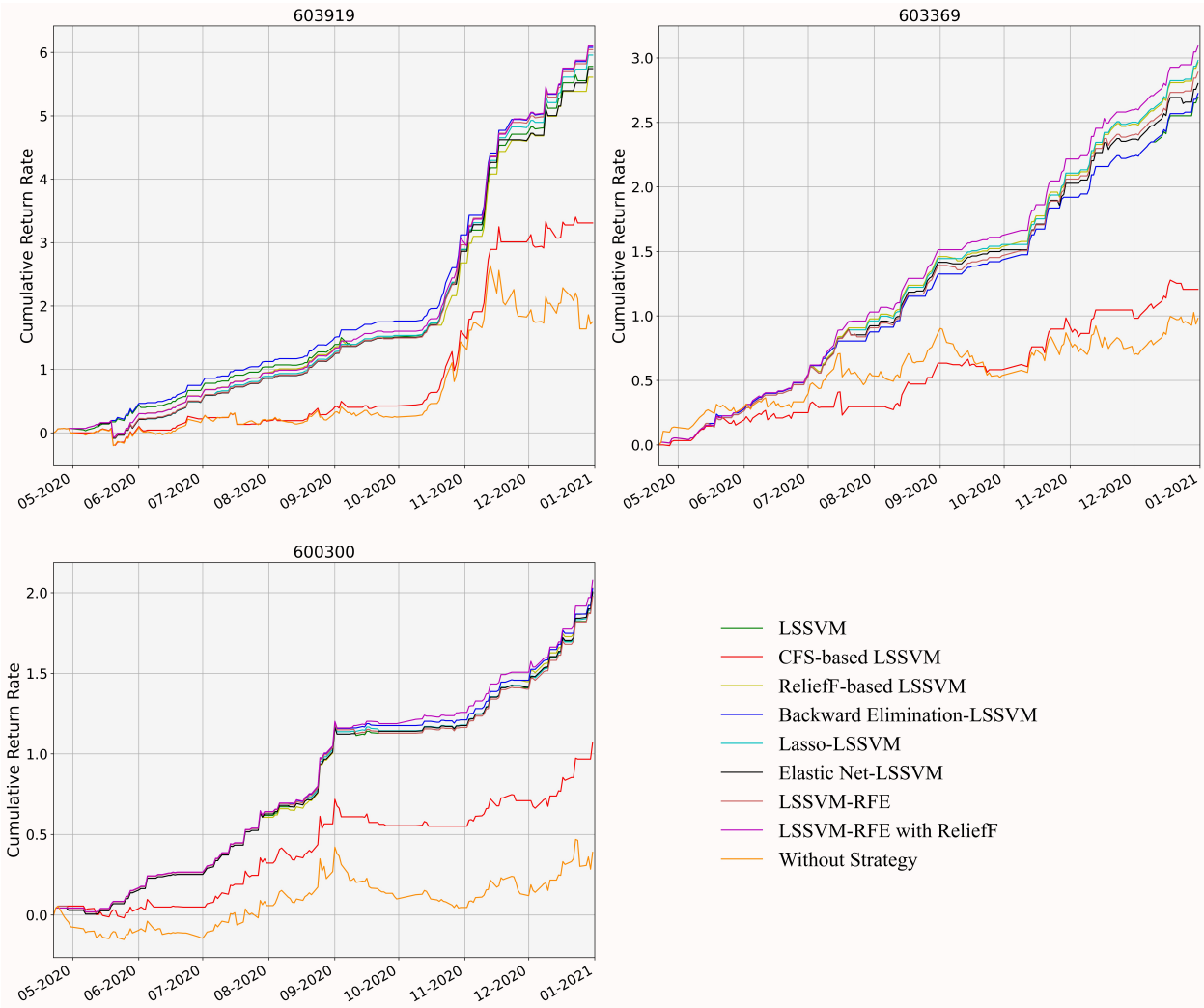


Figure 5: Cumulative return rate for three stocks

According to Fig. 5, the purple line (LSSVM-RFE with ReliefF) is beyond other remaining lines; this means the forecasting results of the LSSVM-RFE with ReliefF can bring more profits than those of other eight investment approaches. Moreover, the final cumulative return rates for all considered approaches are reported in Tab. 6, where the final cumulative return rates with our proposed sparse LSSVM are 6.1008, 3.0926, and 2.076 for Stock 603919, 603369, and 600300, respectively.

Table 6: The final cumulative return rates with different strategies

	Strategy*								
Stock	S1	S2	S3	S4	S5	S6	S7	S8	S9
603919	5.7782	3.3100	5.6088	6.0798	5.9587	5.7439	6.0448	1.7509	**6.1008**
603369	2.6970	1.2062	2.9654	2.7246	2.9808	2.8025	2.8892	0.9818	**3.0926**
600300	1.9795	1.0733	2.0293	2.0264	1.9971	2.0059	1.9765	0.3900	**2.0762**

* S1: LSSVM; S2: CFS-based LSSVM; S3: ReliefF-based LSSVM; S4: Backward Elimination-LSSVM; S5: LASSO-LSSVM; S6: Elastic Net-LSSVM; S7: LSSVM-RFE; S8: Without strategy; S9: LSSVM-RFE with ReliefF.

5 Conclusions

In this present paper, a new sparse LSSVM framework, LSSVM-RFE with ReliefF, has been proposed for stock price movement direction prediction. Different from the ReliefF-based LSSVM, the learning of feature subset is related to the learning algorithm. Furthermore, our feature evaluation criteria are based on the changes in loss function and the conditional dependencies between features and correlation between features and class, which greatly enhances reliability in the evaluation of feature quality. Therefore, our framework can effectively eliminate unimportant technical indicators and obtain a better technical indicators subset. Moreover, by using three stock datasets in the Liquor and Spirits Concept, we have demonstrated that our new sparse LSSVM is more superior to the other considered seven forecasting models for stock price movement direction prediction according to both error indexes and cumulative return rates.

Acknowledgements

This work was supported in part by the Australian Research Council Centre of Excellence for Mathematical and Statistical Frontiers (ACEMS), under grant number CE140100049. The demo of the

proposed sparse LSSVM in this paper can be obtained by sending a request to the authors (you-gan.wang@qut.edu.au; wujrtudou@gmail.com).

Credit authorship contribution statement

Maoxuan Miao: Software, Visualization, Formal analysis, Writing-original draft; **Jinran Wu**: Visualization, Formal analysis, Writing-review & editing; **Fengjing Cai**: Investigation, Project administration; **Liya Fu**: Writing-review & editing; **You-Gan Wang**: Supervision, Project administratio, Investigation, Writing-review & editing,

Declaration of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

A Appendix A

A.1 Correlation Feature Selection(CFS)-based LSSVM

CFS-based LSSVM is composed of CFS and LSSVM model. Its algorithm procedure is divided into two steps: 1) the first step: CFS is utilized to select feature subset; and 2) the second step: feature subset is considered as input variables of LSSVM model. The basic idea of CFS (J. Li et al., 2017) is to use a correlation-based heuristic to evaluate the worth of a feature subset $\mathcal{S}$:

$$\text{CFS\_SCORE}(\mathcal{S}) = \frac{k\bar{r}_{cf}}{\sqrt{k + k(k-1)\bar{r}_{ff}}}, \tag{17}$$

where the CFS score shows the heuristic"merit" of the feature subset $\mathcal{S}$ with k features. $\bar{r}_{cf}$ is the mean feature class correlation and $\bar{r}_{ff}$ is the average feature-feature correlation. Because searching globally optimal subset is NP-hard problem, best first search is utilized to search local optimal feature subset.It starts with empty set. Then, a feature corresponding to the highest score calculated by Eq. (17) is

introduced into the set at a time. Meanwhile, in order to avoid searching the entire feature subspace, the method uses a stopping criterion of five consecutive fully expanded non-improving subsets.

A.2 Backward Elimination-LSSVM

Backward Elimination-LSSVM is that backward elimination feature selection is introduced into LSSVM.It starts with full feature set. Then a feature is removed on feature set in a greedy way,which can improve performance of LSSVM at a time until performance of LSSVM can not be improved.

A.3 Elastic Net-LSSVM

The Elastic Net-LSSVM is composed of LSSVM and elastic-net penalty, a mixture of the L_1-norm and the L_2-norm penalties. Elastic net penalty term not only simultaneously does automatic variable selection and continuous shrinkage, but also can select groups of correlated variables (Zou & Hastie, 2005). Its objective function can be formulated as:

$$\min_{w_0,w} \sum_{i=1}^{n} [1 - y_i(w_0 + \sum_{i=1}^{q} w_i x_i)]_+^2 + \lambda_1 ||w||_1 + \lambda_2 ||w||_2, \tag{18}$$

with two tuning parameters λ_1, and $\lambda_2 \geq 0$.

B Appendix B

References

Chen, W., Jiang, M., Zhang, W.-G., & Chen, Z. (2021). A novel graph convolutional feature based convolutional neural network for stock trend prediction. *Information Sciences*, *556*, 67–94.

Cherrington, M., Thabtah, F., Lu, J., & Xu, Q. (2019). Feature selection: filter methods performance challenges. In *2019 international conference on computer and information sciences (iccis)* (pp. 1–4).

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297.

Table 7: Parameter setting of sparse LSSVM for three stock datasets

Panel A: Stock 603919				
Model	C	λ_1	λ_2	α
LSSVM	50	-	-	-
CFS-based LSSVM	0.01	-	-	-
LASSO- LSSVM	-	6.2170×10^{-3}	0	-
Elastic Net-LSSVM LSSVM	-	2.4559×10^{-3}	0.01	-
Backward Elimination-LSSVM	1	-	-	-
LSSVM-RFE	0.05	-	-	-
LSSVM-RFE with ReliefF	0.0039	-	-	0.3
ReliefF-based LSSVM	50	-	-	-
Panel B: Stock 603369				
Model	C	λ_1	λ_2	α
LSSVM	0.1	-	-	-
CFS-based LSSVM	2	-	-	-
LASSO- LSSVM	-	2.7611×10^{-2}	0	-
Elastic Net-LSSVM LSSVM	-	6.3785×10^{-2}	0.01	-
Backward Elimination- SSVM	0.08	-	-	-
LSSVM-RFE	0.01	-	-	-
LSSVM-RFE with ReliefF	41.7903	-	-	0.3
ReliefF-based LSSVM	0.01	-	-	-
Panel C: Stock 600300				
Model	C	λ_1	λ_2	α
LSSVM	0.01	-	-	-
CFS-based LSSVM	0.01	-	-	-
LASSO- LSSVM	-	7.9258×10^{-2}	0	-
Elastic Net-LSSVM LSSVM	-	2.2337×10^{-1}	0.01	-
Backward Elimination-LSSVM	0.01	-	-	-
LSSVM-RFE	0.01	-	-	-
LSSVM-RFE with ReliefF	47.0136	-	-	0.5
ReliefF-based LSSVM	0.1	-	-	-

Gandhmal, D. P., & Kumar, K. (2019). Systematic analysis and review of stock market prediction techniques. *Computer Science Review*, *34*, 100190.

Gunduz, H., Yaslan, Y., & Cataltepe, Z. (2017). Intraday prediction of borsa istanbul using convolutional neural networks and feature correlations. *Knowledge-Based Systems*, *137*, 138–148.

Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning*, *46*(1), 389–422.

Hao, P.-Y., Kung, C.-F., Chang, C.-Y., & Ou, J.-B. (2021). Predicting stock price trends based on financial news articles and using a novel twin support vector machine with fuzzy hyperplane. *Applied Soft Computing*, *98*, 106806.

Hu, H., Tang, L., Zhang, S., & Wang, H. (2018). Predicting the direction of stock markets using optimized neural networks with google trends. *Neurocomputing*, *285*, 188–195.

Ismail, M. S., Noorani, M. S. M., Ismail, M., Razak, F. A., & Alias, M. A. (2020). Predicting next day direction of stock price movement using machine learning methods with persistent homology: Evidence from kuala lumpur stock exchange. *Applied Soft Computing*, *93*, 106422.

Jin, L., Zeng, Q., He, J., Feng, Y., Zhou, S., & Wu, Y. (2019). A relieff-svm-based method for marking dopamine-based disease characteristics: A study on swedd and parkinson's disease. *Behavioural Brain Research*, *356*, 400–407.

Khaokaew, Y., & Anusas-amornkul, T. (2016). A performance comparison of feature selection techniques with svm for network anomaly detection. In *2016 4th international symposium on computational and business intelligence (iscbi)* (pp. 85–89).

Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence*, *97*(1-2), 273–324.

Kurniawati, I., & Pardede, H. F. (2018). Hybrid method of information gain and particle swarm optimization for selection of features of svm-based sentiment analysis. In *2018 international conference on information technology systems and innovation (icitsi)* (pp. 1–5).

LeCun, Y., Denker, J. S., & Solla, S. A. (1990). Optimal brain damage. In *Advances in neural information processing systems* (pp. 598–605).

Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., & Liu, H. (2017). Feature selection: A data perspective. *ACM Computing Surveys (CSUR)*, *50*(6), 1–45.

Li, Z., Yang, D., Zhao, L., Bian, J., Qin, T., & Liu, T.-Y. (2019). Individualized indicator for all: Stock-wise technical indicator optimization with stock embedding. In *Proceedings of the 25th acm sigkdd international conference on knowledge discovery & data mining* (pp. 894–902).

Long, J., Chen, Z., He, W., Wu, T., & Ren, J. (2020). An integrated framework of deep learning and knowledge graph for prediction of stock price trend: An application in chinese stock exchange market. *Applied Soft Computing*, *91*, 106205.

Long, W., Lu, Z., & Cui, L. (2019). Deep learning-based feature engineering for stock price movement prediction. *Knowledge-Based Systems*, *164*, 163–173.

Lorbert, A., & Ramadge, P. J. (2013). The pairwise elastic net support vector machine for automatic fmri feature selection. In *2013 ieee international conference on acoustics, speech and signal processing* (pp. 1036–1040).

Nie, F., Wang, X., & Huang, H. (2017). Multiclass capped p-norm svm for robust classifications. In *Proceedings of the aaai conference on artificial intelligence* (Vol. 31).

Razavi, R., Bemani, A., Baghban, A., Mohammadi, A. H., & Habibzadeh, S. (2019). An insight into the estimation of fatty acid methyl ester based biodiesel properties using a lssvm model. *Fuel*, *243*, 133–141.

Robnik-Šikonja, M., & Kononenko, I. (2003). Theoretical and empirical analysis of relieff and rrelieff. *Machine Learning*, *53*(1), 23–69.

Roth, V. (2004). The generalized lasso. *IEEE Transactions on Neural Networks*, *15*(1), 16–28.

Roy, A., Singha, J., Devi, S. S., & Laskar, R. H. (2016). Impulse noise removal using svm classification based fuzzy filter from gray scale images. *Signal Processing*, *128*, 262–273.

Suykens, J. A., & Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural Processing Letters*, *9*(3), 293–300.

Tao, Z., Huiling, L., Wenwen, W., & Xia, Y. (2019). Ga-svm based feature selection and parameter optimization in hospitalization expense modeling. *Applied Soft Computing*, *75*, 323–332.

Wang, K., Liu, L., Yuan, C., & Wang, Z. (2020). Software defect prediction model based on lasso–svm. *Neural Computing and Applications*, 1–11.

Wang, Q., Xu, W., & Zheng, H. (2018). Combining the wisdom of crowds and technical analysis for financial market prediction using deep random subspace ensembles. *Neurocomputing*, *299*, 51–61.

Xiao, C., Xia, W., & Jiang, J. (2020). Stock price forecast based on combined model of ari-ma-ls-svm. *Neural Computing and Applications*, 1–10.

Yu, L., Chen, H., Wang, S., & Lai, K. K. (2008). Evolving least squares support vector machines for stock market trend mining. *IEEE Transactions on Evolutionary Computation*, *13*(1), 87–102.

Yu, P., & Yan, X. (2020). Stock price prediction based on deep neural networks. *Neural Computing and Applications*, *32*(6), 1609–1628.

Yuanyuan, S., Yongming, W., Lili, G., Zhongsong, M., & Shan, J. (2017). The comparison of optimizing svm by ga and grid search. In *2017 13th ieee international conference on electronic measurement & instruments (icemi)* (pp. 354–360).

Zhang, R., Nie, F., Li, X., & Wei, X. (2019). Feature selection with multi-view data: A survey. *Information Fusion*, *50*, 158–167.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *67*(2), 301–320.