SVM APPLICATION OF FINANCIAL TIME SERIES FORECASTING USING EMPIRICAL TECHNICAL INDICATORS

Jianxue Chen
College of Electrical and Electronic Engineering
Shanghai University of Engineering Science
Shanghai, P. R. China
cjx@sues.edu.cn

***Abstract*—Support vector machines (SVMs) are promising methods of pattern recognition in financial time series because they use a risk function consisting of the empirical error and a regularized term which is derived from the structural risk minimization principle. This study applies SVM to pattern recognition in the financial engineering domain. Compared with present machine learning methods in financial forecasting, this study does not simply work on the original time data series, but some interesting and empirical technical indicators. The experimental results show that transforming the input data space of SVM can bring good performance in finance engineering.**

Keywords: Support Vector Machines; Financial time series

I. INTRODUCTION

Over the past years, there has been a lot of buzz in the industry around the idea of recognize the invisible patterns of time series data, along with the data mining technology and time series analysis theory. This subject is of so much economic significance that everyone wants to understand the market and want to make profit. Till now, this topic is still regarded as one of the most challenging applications of modern time series forecasting.

Quantitative trading groups say that in the near future they will look to develop models that use the historical time series on stock performance to predict the effect real-time consequence may have on future performance. And, potentially, research group could develop programs that automatically follow the trading records and time series to discover some patterns for prediction. In fact, some pioneers have developed various methods for volatility prediction of financial market, including statistics, data fitting, machine learning and data mining. Related report mainly includes genetic algorithm, artificial neural networks, rule-based technology, fuzzy logic, wavelets, chaos theory, and the state-of-the-art support vector machines.

Already, some researchers have reported experimental results of SVM application in finance engineering field. Mukherjee et al. [1] showed the applicability of SVM to time-series forecasting. Tay and Cao [2] examined the predictability of financial time series including time series data with SVMs. They showed that SVMs outperformed the BP networks on the criteria of normalized mean square error, mean absolute error, directional symmetry and weighted directional symmetry. They estimated the future value using the theory of SVM in regression approximation.

However, most results are based on this belief that the market wave patterns lies in the time series data. Take a close look at the market wave, obviously most of the time the market is randomly walking, together with some regular predictable movements, which might be discovered not in time series data form.

In this paper, a novel pattern recognition method is developed to discover the mid-term volatility of stock price, which is different to classical historical data analysis methods, used by many people in considerable long period. The fact behind the idea – trading on the price trend acceleration, not the price itself, is somehow obvious. And more important is, the chosen reference is the regular technical indicators, which is well-known to all individuals in the market. One fact is the people, always the majority, rely on these common indicators to make decisions. And actually, the other fact is the market behavior depends more on peoples' opinion, reaction, and feeling of specific indicators, not the stock itself. So as an assumption, common indicators may be good candidate features for machine learning technologies, e.g. SVM. So in this paper, traditional time series data are replaced by technical indicators.

This paper gives an attempt on machine learning application on stock indicators, fully using SVM technology to discover the patterns in history data of both SSE (Shanghai Security Exchange) and SZSE (ShenZhen Security Exchange). The details of the methodology and the system construction are described in the following sections.

II. BACKGROUND

A. *Machine Learning Methods*

Candidates of the learning machine could be the neural network, fuzzy logic, rough set, and genetic algorithm. Various machine learning technologies are proper to the classifier problem. Among them, kernel methods are chosen for the learning solution [3].

Support Vector Machines(SVM)is a relatively new class of machine learning techniques introduced by Vapnik, and firstly acquired by Joachimes [4] with great success in text processing field. Based on the structural risk minimization principle from the computational learning theory, SVM seeks a decision surface to separate the training data points into two classes and makes decisions based on the support vectors that are selected as the only effective elements from the training set.

978-1-4244-8106-4/$26.00 © 2010 IEEE

For binary classification, it is assumed that the training set is $(X_1,y_1),(X_2,y_2),...,(X_n,y_n), y_i \in \{1,-1\}(i=1,2,...,n)$. A separating hyper-plane divides it into two sides, each side containing points with the same class label only. The goal of the SVM learning is to find the optimal separating hyper-plane (OSH) that has the maximal margin to both sides. This can be formularizes as:

$$\min \phi(W) = \frac{1}{2}\|W\|^2 = \frac{1}{2}(W \cdot W) \quad (1)$$
$$s.t.\ \ y_i\left[(W \cdot X_i)+b\right]-1 \geq 0$$

The dual problem is:

$$\min Q(\alpha) = \tfrac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j (X_i X_j) - \sum_{i=1}^{n}\alpha_i$$
$$s.t.\quad \alpha_i \geq 0, i=1,2,...,n \quad (2)$$
$$\sum_{i=1}^{n} y_i\alpha_i = 0$$

The discriminate function is:

$$f(X) = sign\left\{\sum_{i=1}^{n}\alpha_i y_i (X_i, X) + b\right\} \quad (3)$$

The nonlinear support vector machine maps the input variable into a high dimensional (often infinite dimensional) feature space, and applies the linear support vector machine in the feature space. It turns out that the computation of this linear SVM in the feature space can be carried out in the original space through a (reproducing) kernel trick. Therefore we do not really need to know the feature space and the transformation to the feature space. Commonly used kernels include:

1) Polynomial kernels: $K(X, X_i) = (X \cdot X_i + 1)^d$

2) RBF kernels: $K(X, X_i) = \exp(-\frac{|X - X_i|^2}{\sigma^2})$

3) Sigmoid kernels: $K(X, X_i) = \tanh\{C_1[(X \cdot X_i) + C_2]\}$

Thus the problem can be converted like that with kernel K:

$$\min Q(\alpha) = \tfrac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j K(X_i \cdot X_j) - \sum_{i=1}^{n}\alpha_i$$
$$s.t.\quad 0 \leq \alpha_i \leq C, i=1,2,...,n \quad (4)$$
$$\sum_{i=1}^{n} y_i\alpha_i = 0$$

The optimal separating hyper-plane is:

$$f(X) = \sum_{i=1}^{n}\alpha_i y_i K(X, X_i) + b = 0 \quad (5)$$

Given any X_j, we can determinate b from:

$$y_j\left(\sum_{i=1}^{n}\alpha_i y_j K(X_j \cdot X_i) + b\right) - 1 = 0 \quad (6)$$

SVM has shown to yield very good generalization performance in both text classification problems and hypertext categorization tasks [5]. Thus it is taken as the learning machine in this paper. The popular neural network method is also a candidate learning machine, however, it is not chosen in this paper due to various work has shown that SVMs beat neural network in most of the application domains.

B. Technical Indicators

Various technical indicators are developed to explore the market, especially in security close price and deal volume. Here list some indicators that will be used in the experiments, in the following section.

TABLE I. TECHNICAL INDICATORS

Indicator	Formula	Description
MA(n)	$\frac{\sum_{i=t-n}^{t} C_i}{n}$	Moving average of a security's close price.
%K	$\frac{C_t - LL_{t-n}}{HH_{t-n} - LL_{t-n}} \times 100$	Stochastic %K. It composes where a security's price closed relative to its price range over a given time period.
%D	$\frac{\sum_{i=0}^{n-1} \%K_{t-i}}{n}$	Stochastic %D, moving average of %K.
%J	$\frac{\sum_{i=0}^{n-1} \%D_{t-i}}{n}$	Stochastic slow D%, moving average of %D.
WMR	$\frac{H_t - C_t}{H_n - L_n} \times 100$	Larry William's %R. It is a momentum indicator that measures overbought/oversold levels.
OBV	$Obv_n = Obv_{n-1} + vol_n, C_n \geq C_{n-1}$ $Obv_n = Obv_{n-1} - vol_n, C_n \leq C_{n-1}$	On balance volume. It is a complex indicator including both price and volume.
RSI	$100 - \frac{100}{1 + \frac{\sum_{i=0}^{n-1} \frac{U_{Pt-i}}{n}}{\sum_{i=0}^{n-1} \frac{D_{Wt-i}}{n}}}$	Relative strength index. It is a price following an oscillator that ranges from 0 to 100. Here Upt means upward-price-change and Dwt means downward-price-change at time t.

III. EXPERIMENTS

A. Setup

The system collects all valid daily trading data of SSE and SZSE, from 2000-01-01 to 2010-03-01, which adopt classical OHLC format (open price, highest price, lowest price, close price), except less than 200 trading days security from opening day.

Not like others researchers' work, original time series data is not chosen in this experiment, but the transformed technical indicator data of the trigger point. In this paper, all samples are collected at MACD bar reverse day, that is, buy at the negative bar length short point, and sell at the positive bar length short point. The behind idea is MACD is proved to be a relative effective indicator of trend prediction. So the time series length is not identical, decided by the sell point and buy point. And the input vector is composed by some historical information of the buy point, mainly including some empirical technical indicators.

At each buy point, some aspects are considered. Obviously the market itself should be an important factor, classical MA system is used here. Here, to smooth the absolute price gap of different security, the MA ratio is used, that is, not take the MA value directly, but the different MA division ratios, e.g. 13 days close price MA divide 55 days close price MA. Actually to explore the MA system relationship, in this paper, 3 MA reference is adopted. Then KDJ indicators are considered also, together with William's %R indicator and OBV indicator, which represent the security trend and trading volume reference. Finally the regular RSI indicator is also adopted as part of input vector. Generally speaking, the indicators' choice is based on three principles: global trend, mid-term trend following, and mean reversion.

TABLE II. TABLE II. SELECTED FEATURES

Feature Index	Feature Name	Description
1	MA13/MA55	Ratio of short MA and mid MA
2	MA13/MA144	Ratio of short MA and long MA
3	MA55/MA144	Ratio of mid MA and long MA
4	%K	Stochastic %K
5	%D	Stochastic %D
6	%J	Stochastic %J
7	WMR	William's %R
8	OBV/Caps	Ratio of OBV and valid capital
9	RSI	RSI indicator

For the test strategy, the random train/test set split method is chosen as the basic benchmark. According to Cao's suggestion [2], the training set size is four times than the test set size. So, first result is about random choice over the whole samples.

Secondly, it is interesting to explore the effect of taking advantage of the training set as much as possible, that is, training set includes 2000 to 2008, then take 2009 as test set. Since the training set has no prior information of year 2009, it is considered that the algorithm evaluation is fair, and close to practice usage.

Finally, the inter-security relationship is considered, which is the well-known board or group effect. Some broads are chosen to perform the test, in which the securities are split into training set and test set. The SVM leaning and validation is restricted in the board range to investigate the correlation between securities.

B. Result

As the benchmark base, the train set and test set are split randomly, totally ignoring the time series attribute of the market. According to Cao's suggestion [2], the train/test size ratio is 4:1. Following the traditional evaluation method, the confusion matrix is taken as the experimental result, in which the precision and recall can be easily calculated. The precision is used to show the prediction accuracy, and the recall is used to show the prediction coverage. Not like in information retrieval research domain, the F1 index is not used, and the precision/recall index are not cared for the negative prediction, because the positive prediction is significant for the market prediction in China stock market.

Furthermore, among the three positive categories, the most important index is win rate, which is practicable for real operation in the market.

For the SVM tool, the libsvm toolbox is chosen. And the grid searching method is used to determinate the optimized model of SVM, in which the best C parameter and gamma parameter are given.

The experiments are scheduled as the following table.

TABLE III. TEST ROUNDS SCHEDULE

Round	Sampling method	Description
1	Random	Random chosen in full market data, from 2000 to 2009, split in 4:1 ratio (training/test)
2	Full	Full training samples from 2000 to 2008, full test samples of 2009
3	Group	Full training samples from 2000 to 2008, and test samples of 2009, both in nonferrous metal group
4	Group	Full training samples from 2000 to 2008, and test samples of 2009, both in estate group
5	Group	Full training samples from 2000 to 2008, and test samples of 2009, both in telecommunication group

The following figures and tables show the experimental results of the 5 round tests.

1) 1st Round:

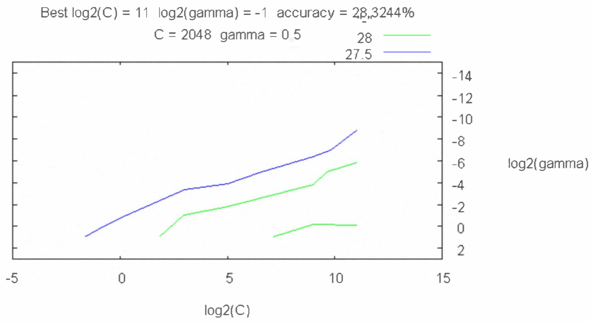

Figure 1. Model Selection of Random Sampling

TABLE IV. CONFUSION MATRIX OF RANDOM SAMPLING

Prediction / Label	<-10 %	-10% ~-3%	-3%~0 %	0%~3 %	3%~10 %	>10%	*Recall (%)*
<-10%	24	4	1	12	381	25	*5.37*
-10%~-3%	6	5	10	12	419	25	*1.05*
-3%~0%	6	4	8	17	308	16	*2.23*
0%~3%	4	4	6	28	444	11	*5.63*
3%~10%	16	5	4	27	818	47	*89.20*
>10%	14	3	4	12	429	95	*17.06*
accuracy (%)	**34.29**	**20.0**	**24.24**	**25.93**	**29.22**	**43.38**	-
win rate(%)	-			**56.78**	**60.41**	**69.86**	-
Overall win	-			*61.13%*			-

Basically, the raw random sampling method already shows fairly good result, compared with traditional time series prediction methods. For those over 10% profit prediction, the win rate is satisfying for practice.

2) 2nd Round:

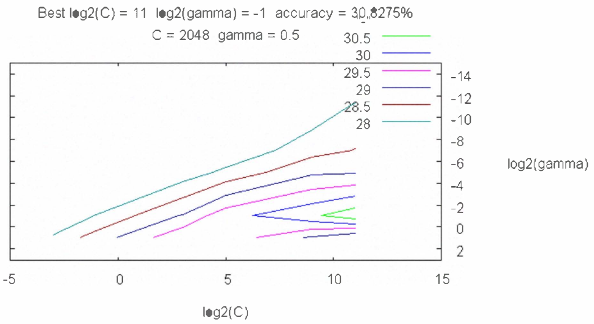

Figure 2. Model Selection of Max Training Split

TABLE V. CONFUSION MATRIX OF MAX TRAINING SPLIT

Prediction / Label	<-10 %	-10%~-3%	-3%~0 %	0%~3 %	3%~10 %	>10%	Recall (%)
<-10%	38	9	9	3	278	85	*9.00*
-10%~-3%	66	6	12	10	468	91	*0.92*
-3%~0%	41	19	23	9	528	40	*3.48*
0%~3%	75	15	32	17	1078	67	*1.32*
3%~10%	188	30	36	26	2406	150	*84.84*
>10%	112	14	25	8	1274	135	*8.61*
accuracy (%)	**7.31**	**6.45**	**16.79**	**23.29**	**39.89**	**23.77**	-
Win rate (%)		-		**69.86**	**78.88**	**61.97**	-
Overall win		-			*77.34%*		-

When the training set contains more samples, the result shows the performance is significantly improved.

3) 3rd Round:

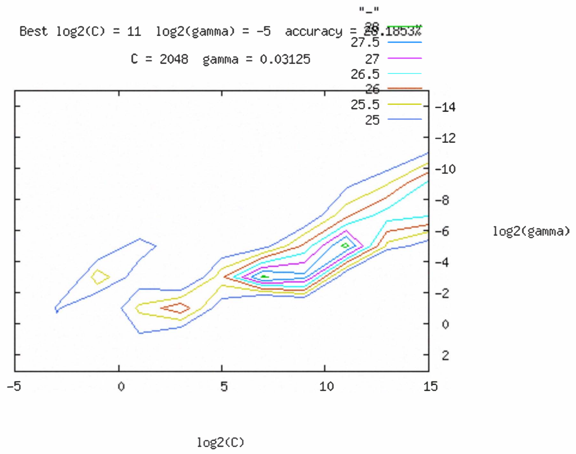

Figure 3. Model Selection of Nonferrous Metal Group

TABLE VI. CONFUSION MATRIX OF NONFERROUS METAL GROUP

Prediction / Label	<-10 %	-10%~-3%	-3%~0 %	0%~3 %	3%~10 %	>10%	Recall (%)
<-10%	1	0	0	3	7	9	*5.00*
-10%~-3%	3	1	1	4	7	9	*4.00*
-3%~0%	0	1	2	1	5	3	*16.67*
0%~3%	5	2	6	11	5	10	*28.21*
3%~10%	3	4	4	18	29	26	*34.52*
>10%	12	1	1	17	23	58	*51.79*
accuracy (%)	**4.17**	**11.11**	**14.29**	**20.37**	**38.16**	**50.43**	-
Win rate (%)		-		**85.19**	**75.00**	**81.74**	-
Overall win		-			*80.41%*		-

When the test runs only in nonferrous metal group, the win rates are better than full market test. The behavior behind the figures implies that nonferrous metal group is relatively active in whole market.

4) 4th Round:

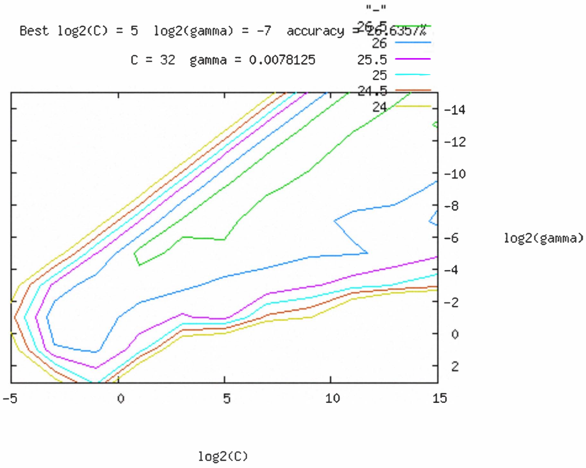

Figure 4. Model Selection of Estate Group

TABLE VII. CONFUSION MATRIX OF ESTATE GROUP

Prediction / Label	<-10 %	-10%~-3%	-3%~0 %	0%~3 %	3%~10 %	>10%	Recall (%)
<-10%	0	0	0	0	70	31	*0.00*
-10%~-3%	0	0	0	0	39	15	*0.00*
-3%~0%	0	0	0	0	23	4	*0.00*
0%~3%	0	0	0	0	76	4	*0.00*
3%~10%	0	0	0	0	191	13	*93.63*
>10%	0	0	0	0	186	29	*13.49*
Accuracy(%)	**0.00**	**0.00**	**0.00**	**0.00**	**32.65**	**30.21**	-
Win rate (%)		-		**0.00**	**77.44**	**47.92**	-
Overall win		-			*73.27%*		-

When test runs in estate group, it is difficult to explain the prediction is heavily biased. All predictions are positive, and over 10% profit predictions are not as accurate as previous tests. This behavior may due to the unsteady attribute of estate companies in the domestic economy.

5) 5th Round:

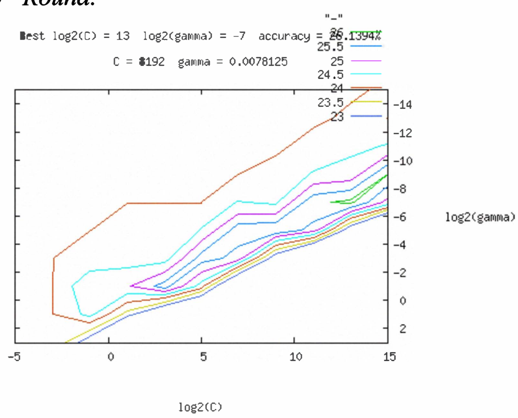

Figure 5. Model Selection of Telecommunication Group

TABLE VIII. CONFUSION MATRIX OF TELECOMMUNICATION GROUP

Prediction / Label	<-10 %	-10% ~-3%	-3%~0 %	0%~3 %	3%~10 %	>10%	*Recall (%)*
<-10%	0	1	0	1	7	1	*0.00*
-10%~-3%	0	2	0	1	5	2	*20.00*
-3%~0%	0	5	0	2	12	0	*5.00*
0%~3%	0	11	1	2	18	0	*6.45*
3%~10%	3	34	0	9	42	3	*46.15*
>10%	1	12	0	5	23	3	*6.82*
accuracy (%)	**0.00**	**3.08**	**100.00**	**10.00**	**39.25**	**33.33**	-
Win rate (%)	-			**80.00**	**77.57**	**66.67**	-
Overall win	-			*77.21%*			-

While test runs in telecommunication group, the result is fairly stable and good.

Among these experiments, it is concluded that more historical information can help to improve the prediction performance, and the performance is different on the groups, depends on the group's inner attributes. Anyway, SVM-based method is proved to be a powerful tool for time series data prediction candidate, by using some feature selection skills, e.g. incorporating the empirical technical indicators.

IV. CONCLUSION

In this paper, driven by the idea – introducing machine learning method into financial time series forecasting, the experimental system has given a try for operation in Chinese stock exchange market. By using support vector machines technology and empirical technical indicators of stock market, some raw but important conclusions are discussed then.

Here leaves some key problems to be further study in the future, mainly including exit strategy, position management strategy, and the learning machine performance in different time scales. Still a lot of work need to do to make the system practical and stable, and the prototype system is a bit far away from real financial marketing product, but the first step has clearly shown the capability of machine learning in financial engineering domain.

REFERENCES

[1] Mukherjee, S., E. Osuna and F. Girosi, *Nonlinear Prediction of Chaotic Time Series Using Support Vector Machines.* In NNSP'97: Neural Networks for Signal Processing VII: Proc. Of the IEEE Signal Processing Society Workshop, 1997, Amelia Island, FL, USA, pp. 511-520.

[2] L. J. Cao and Francis E. H. Tay, *Support Vector Machine with Adaptive Parameters in Financial Time Series Forecasting.* In IEEE Transactions on Neural Networks, Vol. 14, No. 6, November 2003, pp 1506-1518.

[3] Kim, Kyoung-jae, *Financial time series forecasting using support vector machines.* In Neurocomputing, Volume 55, Issues 1-2 (September 2003), Pages 307-319.

[4] Joachims, T. 1998. *Text Categorization with Support Vector Machines: Learning with Many Relevant Features.* Proceedings of ECML-98, 10th European Conference on Machine Learning (Chemnitz, Germany, 1998), pp. 137-142

[5] T. Joachims, *A Statistical Learning Model of Text Classification with Support Vector Machines,* Conference on Research and Development in Information Retrieval (SIGIR), ACM, 2001.