Using support vector machine with a hybrid feature selection method to the stock trend prediction

Ming-Chi Lee(李明錡)

Department of Computer science and Information Engineering, National Pingtung Institute of Commerce, No. 51 Minsheng E. Rd., Pingtung 900, Taiwan, ROC

Expert Systems with Applications 36 (2009), pp. 10896-10904

Presenter: Hung-Hsi Chen
Date: Nov. 17, 2009

Abstract

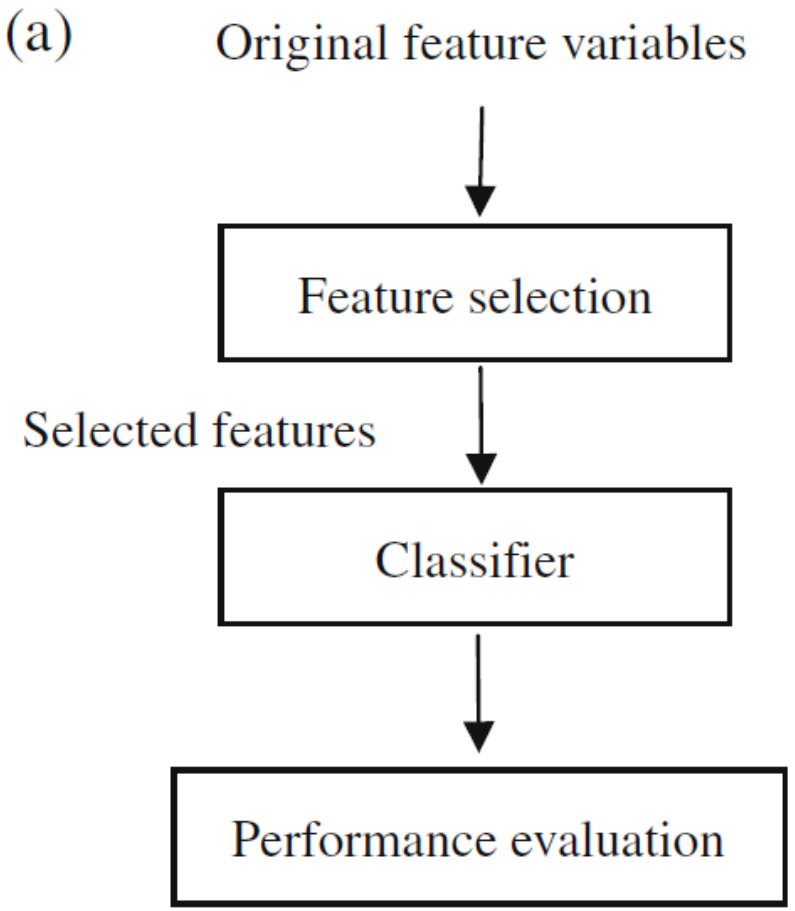
In this paper, we developed a prediction model based on support vector machine(SVM) with a hybrid feature selection method to predict the trend of stock markets. This proposed hybrid feature selection method, named F-score and Supported Sequential Forward Search (F_SSFS), combines the advantages of filter methods and wrapper methods to select the optimal feature subset from original feature set. To evaluate the prediction accuracy of this SVM-based model combined with F_SSFS, we compare its performance with back-propagation neural network(BPNN) along with three commonly used feature selection methods including Information gain, Symmetrical uncertainty, and Correlation-based feature selection via paired t-test. The grid-search technique using 5-fold cross-validation is used to find out the best parameter value of kernel function of SVM.

Abstract

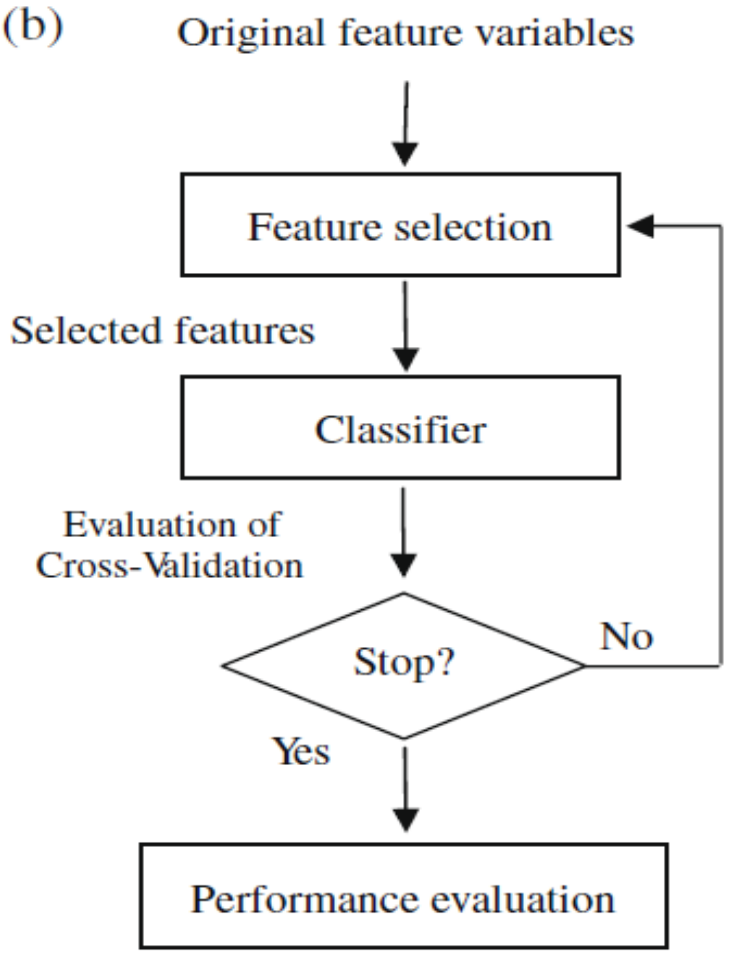
In this study, we show that SVM outperforms BPNN to the problem of stock trend prediction. In addition, our experimental results show that the proposed SVM-based model combined with F-SSFS has the highest level of accuracies and generalization performance in comparison with the other three feature selection methods. With these results, we claim that SVM combined with F_SSFS can serve as a promising addition to the existing stock trend prediction methods.

Feature selection

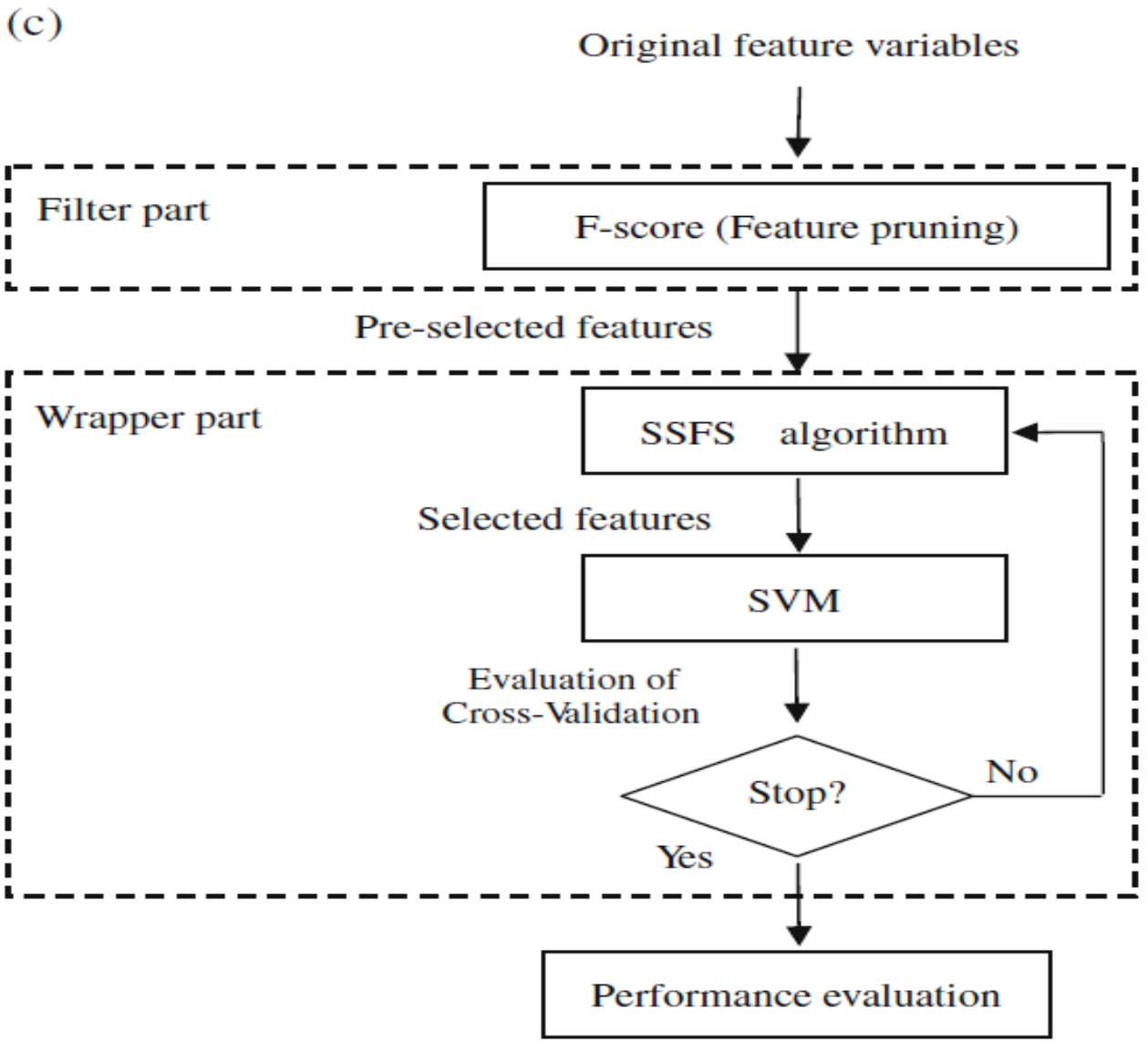
Filter method



Wrapper method



Hybrid method



Filter method： F-score

F-score for i-th feature：

$$F(i) \equiv \frac{(\bar{x}_i^{(+)} - \bar{x}_i)^2 + (\bar{x}^{(-)} - \bar{x}_i)^2}{\frac{1}{n_+ - 1}\sum_{k=1}^{n_+}(x_{k,i}^{(+)} - \bar{x}_i^{(+)})^2 + \frac{1}{n_- - 1}\sum_{k=1}^{n_-}(x_{k,i}^{(-)} - \bar{x}_i^{(-)})^2},$$

where $\bar{x}_i, \bar{x}_i^{(+)}$, and $\bar{x}^{(-)}$ are the averages of the ith feature of the whole, positive, and negative data sets, respectively; $x_{k,i}^{(+)}$ is the ith feature of the kth positive instance, and $x_{k,i}^{(-)}$ is the ith feature of the kth negative instance.

The larger the F-score is, the more likely this feature is more discriminative.

feature	59	51	61	78	12
class	-1	-1	+1	+1	-1

$X_i = (59+51+61+78+12)/5=52.2$

$X^{(+)}_i = (61+78)/2 = 69.5$

$X^{(-)}_i = (59+51+12)/3=40.67$

$var(x^{(+)}) = 144.5$

$var(x^{(-)}) = 632.33$

$F = 432.23/776.83 = 0.56$

Wrapper method： Supported sequential forward search(SSFS)

Binary classification scenario：
Class label Y={+1, -1}
Feature set F = {f_1, f_2....f_k}
Dataset S = {X(L), Y(L) | L=1,2...N}
={[x_1(L), x_2(L)...x_k(L)], Y(L) | L = 1, 2, ...N}

First, choose the best single features among the k choices. SSFS trains SVM k times, each of which use all the training samples available but with only one feature f_i.

The initial feature combination set & active training set：

$$F_1^i = f_i, \quad f_i \in F$$

$$V_1^i = \{1, 2, \ldots, N\}$$

After the training, each single-feature combination F^i_1 is associated with a value M^i_1, which is the minimum of the object function, and a group of support vectors v_i.

$$j = \arg\min_{i \in \{1,2,\ldots,N\}} M_1^i$$

Thus SSFS obtains the initial feature combination $F_1=\{f_j\}$, and active training set $v_1 = \{v_j\}$

At step n :

$$F_{n+1}^i = F_n \cup \{f_i\} \text{ for } f_i \in F_n^{av}, \qquad (8)$$
$$V_{n+1}^i = V_n \cup \{v_i\},$$

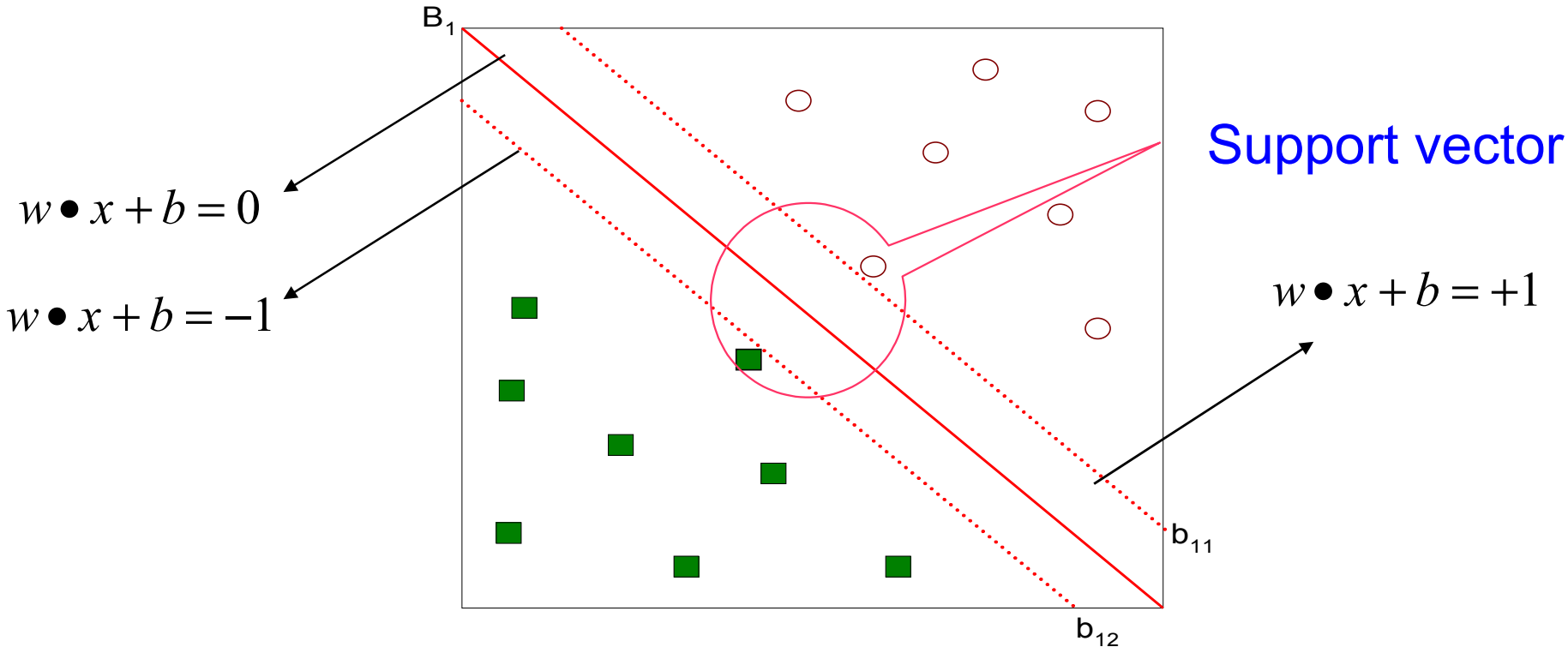
where $F_n^{av} = \{f_r | f_r \in F \text{ and } f_r \notin F_n\}$ is the collection of the available features to be selected from.

$$j = \arg\min_{f_i \in F_n^{av}} M_{n+1}^i$$

For each F_{n+1}^i, SSFS trains SVM just by using the samples in V_{n+1}^i. The resulting minimum of the objective functions and the collection of the support vectors are denoted as M_{n+1}^i and SV_{n+1}^i, respectively.

$$F_{n+1} = F_{n+1}^j,$$
$$V_{n+1} = SV_{n+1}^j$$

Support Vector Machines



We want to maximize: $\text{Margin} = \frac{2}{\| w \|^2}$

Related methods of feature selection

1、Information gain

$$InfoGain = H(Y) - H(Y|X),$$

where X and Y are features and

$$H(Y) = -\sum_{y \in Y} p(y)\log_2(p(y)),$$

$$H(Y|X) = -\sum_{x \in X} p(y) \sum_{y \in Y} p(y|x)\log_2(p(y|x)).$$

The higher score represent the higher importance of a feature

temperature(X1)	High	Low	Medium	Low	High
Wind(X2)	Medium	Weak	Strong	Strong	Weak
class(Y)	+1	-1	-1	-1	+1

H(Y) = -2/5*log(2/5) – 3/5*log(3/5) = 0.62
H(Y|X1) = -2/5*H(X1=High) – 1/5*H(X1=Medium) – 2/5*H(X1=Low) = 0+0+0=0
H(Y|X2) = -2/5*H(X2=Strong) -1/5*H(X2=Medium)– 2/5*H(X2=Weak) = 0+0+0.4 = 0.4
InfoGain(X1) = 0.62- 0 = 0.62
InfoGain(X2) = 0.62-0.4 = 0.22

2、Symmetrical uncertainty：
(balances for Information gain bias)

$$SU = 2.0 \times \frac{InfoGain}{H(Y) + H(X)}.$$

SU(X1) = 2.0 * 0/(0.62+0) = 0
SU(X2) = 2.0 * 0.22/(0.62+0.4) = 0.43

3、Correlation-based feature selection：

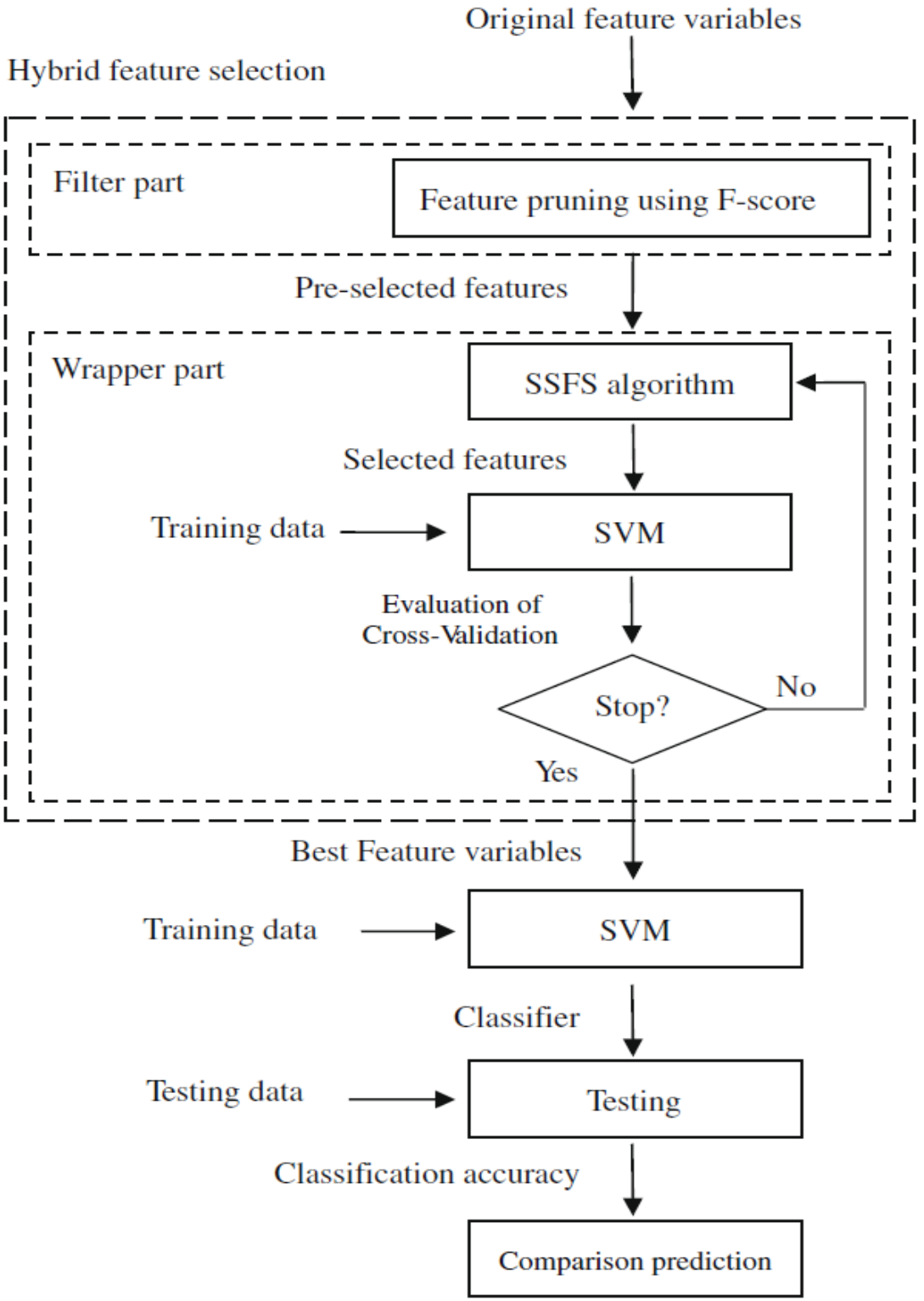
$$CFS_s = \frac{k\bar{r}_{cf}}{\sqrt{k + k(k-1)\bar{r}_{ff}}}$$

where CFS_s is the score of a feature subset S containing k features, $\bar{r}_{cf}$ is the average feature to class correlation ($f \in S$), and $\bar{r}_{ff}$ is the average feature to feature correlation.

Experiment

Market： NASDAQ Index(USA)
Period： Nov. 8, 2001 ~ Nov. 8, 2007
Data： Taiwan Economic Journal database
Features(total 30)：

- 20 future s contracts
  - 9 commodities(silver, platinum, palladium(鈀), heating oil, copper, gold, crude oil, coal, natural gas)
  - 11 foreign currencies(Swiss frank, yen, mark, Canadian dollar, British pound, Euro dollar, Renminbi(人民幣), Australia dollar, South Korea dollar, Hong Kong dollar, Singapore dollar).
- 9 technical indices
  - DJIA index, NYSE Composite Index, Philadelphia Semiconductor Index, UTIL Index, DJCOMP Index, TRAN Index, AMEX Composite Index, Russell Index, S&P 500 Index
- 1-day lagged NASDAQ index



Filter part :

Sort F-score and select K(threshold) highest scored features

Wrapper part :

Each feature fi does 5-fold cross-validation and calculates the average accuracy.

The highest average accuracy of 5-fold cross-validation is the least minimum of the objective functions

Table 1
Comparison of the prediction accuracy with different values of the threshold K.

	Prediction accuracy	
	Training (%)	Testing (%)
$K = \left\lfloor \frac{\lvert F \rvert}{4} \right\rfloor = 7$	76.1	75.5
$K = \left\lfloor \frac{\lvert F \rvert}{2} \right\rfloor = 15$	86.8	84.5
$K = \left\lfloor \frac{3\lvert F \rvert}{4} \right\rfloor = 22$	88.0	87.5
$K = \lvert F \rvert = 30$	88.4	87.5

The results of *F*-score of selected features and average accuracy rate.

No	Feature variables	*F*-score	Average accuracy rate (%)
1	Yen	14.70	62.3
2	Philadelphia Semiconductor Index	12.30	67.2
3	Renminbi	8.87	71.5
4	Swiss frank	3.41	75.6
5	Australia dollar	2.90	77.4
6	British pound	2.13	78.5
7	Nasdaq 1-day-lagged	1.91	80.0
8	Gold	0.90	81.3
9	DJCOMP Index	0.90	82.1
10	Russell 2000 Index	0.86	83.2
11	Silver	0.76	84.4
12	Palladium	0.70	85.2
13	Coal	0.54	85.8
14	Platinum	0.53	86.4
15	Singapore dollar	0.42	86.8
16	Hong Kong dollar	0.40	87.3
17	Canadian dollar	0.22	87.7

accumulated

Table 4
The performance of SVM model with four different feature selection methods using 5-fold cross-validation.

Classifier + feature selection methods	Accuracy of 5-fold cross-validation					
	Set 1	Set 2	Set 3	Set 4	Set 5	Average
SVM + F_SSFS	85.5	87.0	87.0	88.5	88.5	87.3
SVM + Information Gain	78.0	79.0	79.0	81.5	83.0	80.1
SVM + Symmetrical uncertainty	78.0	79.5	79.5	81.0	82.5	80.1
SVM + CFS	75.0	75.0	76.5	77.0	78.0	76.3
SVM	83.2	85.3	86.2	85.3	87.2	85.4

The performance of BPNN model with four different feature selection methods using 5-fold cross-validation.

Classifier + feature selection methods	Accuracy of 5-fold cross-validation					
	Set 1	Set 2	Set 3	Set 4	Set 5	Average (%)
BPNN + F_SSFS	71.5	71.5	72.0	73.5	74.0	72.5
BPNN + Information Gain	65.0	65.0	67.0	68.5	68.5	66.8
BPNN + Symmetrical uncertainty	65.0	65.0	66.5	68.5	69.0	66.8
BPNN + CFS	62.0	62.5	63.0	63.5	64.0	63.0
BPNN	70.2	71.1	70.7	72.2	73.1	71.5

Table 6
Paired *t*-test and Mann–Whitney nonparametric test comparison between BPNN and SVM.

Feature selection methods	Classifier	Accuracy (%)	Paired *t*-test		Mann–Whitney nonparametric test	
			T statistic	*p* (two-tailed)	*Z* statistics	*p* (two-tailed)
F_SSFS	SVM	87.3	19.268	0.001**	−3.496	0.001**
	BPNN	72.5				
Information Gain	SVM	80.1	10.951	0.001**	−3.046	0.002**
	BPNN	66.8				
Symmetrical uncertainty	SVM	80.1	11.665	0.001**	−3.046	0.002**
	BPNN	66.8				
CFS	SVM	76.3	19.504	0.001**	−2.922	0.003**
	BPNN	63.0				

Thank you!!!