Prediction of First-Day Returns of Initial Public Offering in the US Stock Market Using Rule Extraction from Support Vector Machines

Rolf Mitsdorffer[1] and Joachim Diederich[2]

[1] Queensland Education, Brisbane, Australia
[2] American University of Sharjah, UAE and University of Queensland, Australia

Summary. Artificial neural networks (ANNs) and support vector machines have successfully improved the quality of predicting share movements in relation to statistically based counterparts. However, it has not been feasible to gain insight into the reasons why a certain prediction is made. Due to this limitation, the use of machine learning techniques in the capital market has met a critical hurdle. This chapter outlines a method based on pedagogical learning for extracting rules from support vector machines. To the best of our knowledge, the experiments reported here are the first attempt to utilize learning based rule extraction from support vector machines for financial data mining.

The experiments use predictions from support vector machines for extracting rules associated with the first-day returns of “initial public offerings” (IPOs) in the US stock market. A novel feature of the experiments is the simultaneous application of *fundamental* and *technical analysis* in the context of predicting the success of IPOs. Cross-industry IPOs covering the period from 1974 to 1984 and software and services IPOs launched between 1996 and 2000 are utilized.

1 Motivation

Predictions of share prices in the capital market are said to be inconsistent with the theory underlying the “capital asset pricing method” (CAPM). CAPM is based on the random walk hypothesis which assumes linearity of the data. Unfortunately, the statistically based linearity assumption is frequently invalid. Marginal improvements in the prediction of share prices have been obtained based on non-linear models. Forecasting limitations are also imposed by the difficulty of understanding the time dependent dynamics of the share market. Models based on symbolic manipulation never quite capture the dynamic essence of the market. This problem results in a “knowledge acquisition bottleneck” and limits human understanding of the multi-dimensional problem.

R. Mitsdorffer and J. Diederich: *Prediction of First-Day Returns of Initial Public Offering in the US Stock Market Using Rule Extraction from Support Vector Machines*, Studies in Computational Intelligence (SCI) **80**, 185–203 (2008)
www.springerlink.com

The experiments reported here are aimed at addressing the current forecasting limitations at the birth of the capital market: initial public offerings.

2 Introduction

2.1 Financial Data Mining

Machine learning techniques are increasingly being adopted by capital market analysts and have been used in numerous studies. It is likely that only a minority of projects have been published due to the commercial nature of the applications.

A number of authors have used ANNs to model stock returns. Refenes and Zapranis (1995) describe an experiment using a set of unspecified factors extracted from the balance sheets of companies in the universe of UK stocks. Mitsdorffer et al. (2001, 2002) employed ANNs and other machine learning techniques for the predictions of first-day returns of initial public offerings and found significant market inefficiencies.

Comparing results from classical statistical techniques with simple neural learning procedures, Mitsdorffer et al. (2001, 2002) concluded that predictions derived from ANNs outperform current best practice and have a significantly better generalisation capability.

Similar to artificial neural networks, the knowledge embedded in support vector machines is opaque in that it cannot easily be made comprehensible to a human user. While rule extraction from ANNs is now established (Andrews et al. 1995), there have not been any attempts to extract rules from support vector machines prior to 2002. Mitsdorffer et al. (2001, 2002) report first experiments including learning based rule extraction from support vector machines.

2.2 IPOs as a Case Study

This research explores the application of machine learning techniques in the field of initial public offerings, a subset of the capital market. In this section, the purpose of IPOs and processes leading up to them are outlined.

A private company can be converted to a public corporation by raising funds for expansion, product development or the restructuring of debt. IPOs reduce the dependence of companies on bank credit, a notoriously unstable way to allocate capital. IPOs are often used by venture capitalists as an “exit route”.

The most important and time consuming task facing the IPO preparation team is the development of the prospectus, a business document that serves as a brochure for the company. Since the US Securities and Exchange Commission (SEC) imposes a “quiet period” on companies once they file for an

IPO until 25 days after the stock starts trading, the prospectus is the primary source of information for the investor.

At the heart of a prospectus is the company's financial position and past financial performance. Financial reports are essential for informed decision making. Relevant financial information may be found in several sections of a prospectus – including the balance sheet, the profit and loss statement, cash flow analysis and the accountant's report and notes.

2.3 The Valuation of IPOS

According to Ritter (1991), IPOs are not different from other stocks, where discounted cash flow (DCF) analysis and the comparable firm analysis are used. Numerous studies have investigated the "short run under-pricing" of IPOs and the "hot issue market" phenomenon (Ibbotson & Jaffe 1975; Ritter (1984)). Under pricing of IPOs is an internationally widespread phenomenon. Ritter (1991, p. 3) observed that "investors are periodically overoptimistic about the earning potential of young growth companies and firms take advantage of these windows of opportunity".

There is a presumption that many young firms issuing new shares have growth potential, which is difficult to forecast using one-year-ahead earnings projections. Moonchul and Ritter (1999) tested this idea by using a sample of young and older firms. Consistent with the assumption that younger firms are more difficult to value; the authors determined that the valuation errors were noticeably smaller for older firms.

3 Overview of the Chapter

The rule extraction experiments described in this chapter focus on two different datasets: "cross-industry" and "single-industry" IPOs. Following trials to investigate the overall ability of SVMs to predict first-day IPO returns, support vector machine predictions are obtained from test datasets. The predictions associated with these test datasets are then used to extract rules representing what the SVM has learned using pedagogical rule extraction techniques. Finally, statistical tests are utilized to establish that the extracted rules represent what the SVM has learned.

4 Methodology

The project includes the collection of data from company reports and stock market indices. Following a pre-processing phase to make the data amenable to support vector and decision tree learning, the data serves as training input to machine learning techniques.

For each of the cross-industry and single-industry datasets:

1. SVM training and prediction using the *leave-one-out cross-validation method* is conducted by use of SVMlight (Joachims, 1999). Different parameters are used to establish the effect on prediction and generalisation quality.
2. SVM predictions are transcribed into the test dataset. This dataset represents what the SVM has learned.
3. Rules are extracted from the transcribed dataset using See5, C4.5, Ripper and Rulex.
4. An overall analysis of the prediction, generalisation and rule extraction procedure is performed.

The performance of individual machine learning techniques is evaluated in terms of precision, and recall as well as the f-value:

- Precision = true positive/(true positive + false positive)
- Recall = true positive/(true positive + false negative)
- F-value = (2*precision*recall)/(precision + recall)

The experiments result in several competing models of the stock market dynamics governing returns of IPOs on the first trading day based on the extracted rules.

4.1 Statistical Tests

In order to establish that rules represent what the SVM has learned, the McNemar Test (Gardner & Altman 1989) is used to test whether combinations between two dichotomous variables are equally likely. The exact p-value is determined using the binomial distribution as described in Gardner and Altman (1989) and implemented in the statistical analysis tool *Analyse-it*. Based on the p-value (p-value < critical value) the null hypothesis of inequality is accepted. The rejection of the null hypothesis leads to the conclusion that rules from a given technique significantly represent what the SVM has learned at a level of confidence of 95%.

4.2 Data

Cross-Industry IPOs

An extensive search was conducted to locate data in the public domain suitable for this research. The search established IPO data sources in the public domain of R. J. Ritter, University of Florida that are freely available for academic research.[1] Ritter (1991) has used this data to investigate the long-term performance of IPOs.

[1] `http://bear.cba.ufl.edu/ritter/ipodata.htm`.

The dataset includes 2,609 firms with common stock initial public offerings in the period between 1974 and 1984. Companies included in the dataset have used S1 or S18 registration statements. The primary source of information is the direct inspection of the prospectuses.

For each IPO, the following aspects are relevant:

- First-day trading data, including the date the company went public, open and closing prices
- Fundamental data to enable the valuation of companies, such as assets and liabilities as well as shareholder equity and dilution
- Past performance data, including sales, the cost of sales, expenditure for R&D, etc.
- Proxies for market sentiment

In order to reduce extreme outliers, the following selection criteria are used to exclude IPOs with the following attribute values:

- Minimum number of shares $< 2,000$
- Market capitalisation $< \$1\,\mathrm{m} > \$1,000\,\mathrm{m}$
- Offer price $> \$30$
- First-day variation $< -80\% > 200\%$
- Offer fraction < 0.05.

The cross-industry attributes are shown in Table 1.

Market sentiment data is represented by wins and losses of the NASDAQ computer index in the 100 trading days immediately preceding the IPO. The period of 100 trading days is condensed into five separate 20-day periods.

Input values are scaled in the range from -1 to $+1$. Outliers are replaced by maximum and minimum values established according to the table of bin attribute values.

Next, the target values are established. The analysis of first-day gains identified about 24% of the 1,841 IPOs with gains of over 18.1%. Using binary classification, IPOs in the above 18.1% bracket are classified as positive while the remainders are classified as negative.

Computer Software and Services IPOs

This section describes the data used for predicting first-day returns of single industry IPOs. The rationale for selecting a single industry is to ascertain if financial ratios representing underlying company fundamentals within a single industry are more comparable than cross-industries data and thus provide more plausible explanations of first-day IPO returns.

IPO data related to the computer software and services categories were retrieved from the Hoover service. Companies in this sector are involved in the design and marketing of all types of software and the provision of computer services, such as mainframe and system integration.

Table 1. Attribute summary of cross-industry IPOs

Attribute name	Attribute description
BOOK.CAPZTN	Book value/capitalization ratio
BOOK.VALUE	Book value (absolute)
CAPITALIZTN	Capitalization (absolute)
EXPENSES	Expenses (absolute)
GAIN.LOSS.1	First-day IPO gain/loss 1–20 days prior to first trading day
GAIN.LOSS.2	First-day IPO gain/loss 21–40 days prior to first trading day
GAIN.LOSS.3	First-day IPO gain/loss 41–60 days prior to first trading day
GAIN.LOSS.4	First-day IPO gain/loss 61–80 days prior to first trading day
GAIN.LOSS.5	First-day IPO gain/loss 81–100 days prior to first trading day
NO.IPOS.1	Number of IPOs released 1–20 days prior to the first trading day
NO.IPOS.2	Number of IPOs released 21–40 days prior to the first trading day
NO.IPOS.3	Number of IPOs released 41–60 days prior to the first trading day
NO.IPOS.4	Number of IPOs released 61–80 days prior to the first trading day
NO.IPOS.5	Number of IPOs released 81–100 days prior to the first trading day
OFFER.FRACT	Offer fraction
OFFER.PR	Offer price (absolute)
REV.CAP	Revenue vs. capitalization ratio
REVENUE	Revenue absolute
RISKS	Number of risk factors
SEL.DAYS	Number of days selling days
SP.DIFF.1	Absolute difference of the S&P Index 1–20 days prior to the first trading day
SP.DIFF.2	Absolute difference of the S&P Index 21–40 days prior to the first trading day
SP.DIFF.3	Absolute difference of the S&P Index 41–60 days prior to the first trading day
SP.DIFF.4	Absolute difference of the S&P Index 61–80 days prior to the first trading day
SP.DIFF.5	Absolute difference of the S&P Index 81–100 days prior to the first trading day
UW.DISCOUNT.CAP	Underwriter discount in relation to the market capitalization
YR.FOUND	Year the company was founded
BOOK.CAPZTN	Book value/capitalization ratio
BOOK.VALUE	Book value (absolute)
CAPITALIZTN	Capitalization (absolute)
EXPENSES	Expenses (absolute)

IPOs satisfying certain criteria were selected. These criteria include a minimum of $10 million in sales and 80 employees or more. At the time of data collection, nearly all IPOs between 1996 and 1999 were considered.

The aim of the next stage of the data collection is the acquisition of balance sheet and income data from company IPO prospectuses launched with the SEC.

The following aspects are included in the model:

- IPO specific data, such as the date of listing, the date the company went public, offer price, offer fraction, post offering shares
- First-day closing prices
- Fundamental data to enable the valuation of companies, such as assets and liabilities as well as shareholder equity and dilution
- Past performance data, including sales, the cost of sales, expenditure for R&D, etc.
- Daily NASDAQ computer index values for 100 days preceding each IPO

The single-industry attributes considered for the analysis are shown in Table 2.

Each IPO is represented by a vector with 27 input features and one output or predictor attribute. Input features include attributes calculated from balance sheets and income data and those constructed from the NASDAQ computer index. Attributes are scaled which requires knowledge of the statistical properties of attributes to eliminate the effects of outliers.

The decision was made to consider the influence of the NASDAQ computer index on 100 trading days preceding the IPO. Considering the small dataset of 182 IPOs, the number of attributes representing the index was reduced by dividing the 100 index values into 10 time periods and forming the absolute difference of the index for each 10-day period. Generating attributes based on the absolute difference also has the effect of removing the time dependency of index values which in turn enables the formation of time independent rules.

The output attribute represents first-day gains or losses of an IPO. Since the aim of the research is the prediction of first-day gains and the extraction of rules, a binary model was built where IPOs exceeding a certain percentage of first-day gains are classified as positive and others as negative.

4.3 Machine Learning Techniques Used in This Study

Support Vector Machines

Support vector machines are an alternative to neural networks as tools for solving pattern recognition problems. SVMs have a major advantage over neural networks in that they formulate the learning problem as a quadratic optimization problem whose error surface is free of local minima and has a unique global optimum.

Table 2. Attribute summary of single-industry IPOs

Attribute name	Attribute description
ActualOffer	Actual offer price (absolute)
CashLiab	Cash vs. liabilities (%)
CurAssLiab	Current assets vs. liabilities (%)
EquityShare	Equity per share (%)
GrMarginShare	Gross margin per share (%)
GrossMargin	Gross margin (%)
IncGrowth	Income growth (%)
MarketCap	Market capitalization (absolute)
NASDAQ.Period.1	NASDAQ computer index 10 days prior to first trading day
NASDAQ.Period.2	NASDAQ computer index 20 days prior to first trading day
NASDAQ.Period.3	NASDAQ computer index 30 days prior to first trading day
NASDAQ.Period.4	NASDAQ computer index 40 days prior to first trading day
NASDAQ.Period.5	NASDAQ computer index 50 days prior to first trading day
NASDAQ.Period.6	NASDAQ computer index 60 days prior to first trading day
NASDAQ.Period.7	NASDAQ computer index 70 days prior to first trading day
NASDAQ.Period.8	NASDAQ computer index 80 days prior to first trading day
NASDAQ.Period.9	NASDAQ computer index 90 days prior to first trading day
NASDAQ.Period.10	NASDAQ computer index 100 days prior to first trading day
NetIncShare	Net income per share (%)
OfferOutst	Offer vs. outstanding shares (%)
PropActOffer	Proposed vs. actual offer price (%)
RDRev	Research & development vs. revenue (%)
RegDays	Registration days (absolute)
Revenue	Revenue (absolute)
RevGrowth	Revenue growth (%)
RevShare	Revenue per share (%)
SGRev	Sales and general expenses vs. revenue (%)

SVMs are based on some simple ideas and provide a clear intuition of what learning from examples is all about. More importantly, they also show high performance in practical applications. SVMs correspond to a linear method in a very high dimensional feature space that is non-linearly related to the input space. Even though SVMs implement a linear algorithm in a high dimensional feature space, in practice they do not involve any computations in that high dimensional space. By use of kernels, all necessary computations are performed directly in input space. Data vectors nearest to the separating hyperplane in the transformed space are called support vectors. Classification as well as regression can be learned by SVMs.

Joachims (1998) reported that SVMs are well suited to learn in high dimensional spaces (>10,000 inputs). They achieve substantial improvements over currently best performing methods, reducing the need for feature selection.

Rule Extraction from Neural Networks (Rapid Backpropagation)

Rule extraction from neural networks is used for benchmark purposes in this context. The extraction of symbolic knowledge from ANNs and the direct encoding of partial knowledge into ANNs before training are important issues. They allow the exchange of information between symbolic and neural network knowledge representation. ANNs store knowledge in a completely numerical form, which is not open to explanation, a situation similar to SVMs.

Rule extraction from local function networks employs decompositional algorithms that directly decompile weights to generate rules. The underlying network is formed by "Rapid Back Propagation" (RBP), a three layer architecture similar to radial base function networks. The network consists of an input layer, a hidden layer of locally responsive basis function nodes, and an output node. The network is suitable for binary classification tasks as well as function approximation.

Rulex, a tool used in this project and described by Andrews and Geva (1996), is a program that converts the numeric weights of RBP networks into symbolic IF THEN rules that explain the decisions made by the network.

Other Machine Learning Techniques

Classification techniques such as decision trees play a major role in machine learning and knowledge based systems. These learning methods have been successfully applied to a large range of tasks, from learning medical diagnostics to credit risks assessment and are used in this research.

See5/C5 is a system commercialized by Rulequest Research (1997) for analysing data and generating classifiers in the form of decision trees and/or rule sets. C4.5 is the program used in our experiments. Quinlan's work (1986,1993,2001) on C4.5 is widely acknowledged as a major contribution to the development of classifier systems. Examples include a mixture of nominal and numeric properties that are analysed to allow the discrimination of classes. The patterns are expressed in the form of a decision tree or a set of IF THEN rules. The rules can be used to classify new cases.

The Ripper rule learner is a system for inducing classification rules from a set of pre-classified examples and has been used in this project for benchmark purposes. Ripper (Repeated Incremental Pruning to Produce Error Reduction) is an efficient, noise tolerant propositional rule learning algorithm based on the separate and conquer strategy. The basic strategy used by Ripper is to find an initial model and then to iteratively improve that model using an optimisation procedure described in Cohen (1995).

5 Results

For each of the datasets, cross-industry and single industry IPOs, the following results are reported:

- SVM training and prediction results using the leave-one-out method built-in to SVMlight (Joachims, 1999).
- SVM prediction results using explicit test sets. SVM prediction results are then transcribed into the test dataset. This dataset in essence represents what the SVM has learned.
- Rules extracted from the transcribed dataset using See5, C4.5, Ripper and Rulex.

5.1 Results of Rule Extraction from SVM for Cross-Industry IPOs

Leave-one-out Cross-Validation Results

At first glance, the learning and generalisation ability of support vector machines is sufficient as indicated by an error rate of 22.38% (Table 3).

SVM Prediction Using a Test Set

Results from a randomly created test set are shown below.

As is evident from Table 4, the quality of the SVM prediction is not satisfactory with a precision of 0.33, a recall 0.35 and an f-value 0.34. The problem is obviously a confusion of the positive class. This may be due to lack of data.

Table 3. Leave-one-out training results of cross-industry IPOs (rbf kernel)

Trade-off between training error and margin (c)	Cost factor (j)	Parameter in gamma rbf kernel (g)	Test error %
100	0.12	0.5	24.12
50	0.30	0.2	22.87
Default	0.1	0.1	23.85
Default	0.90	0.1	**22.38**

Table 4. Confusion matrix for SVM predictions

(a)	(b)	<-classified as
38	76	(a): class positive
72	275	(b): class negative

Rules Extracted by Use of See5, C4.5, Ripper and Rulex

This part of the experiment is based on transcribing the SVM prediction results into the training sets for rule learners. The results of training the four rule learners with the SVM test predictions as target output are shown below.

In order to establish if See5 rules represented what the SVM has learned the McNemar test (Gardner & Altman 1989) was used to investigate whether combinations between two dichotomous variables shown in the confusion matrix in Table 5 are equally likely.

The exact p-value was computed using the binomial distribution as described by Gardner and Altman (1989) and implemented in the statistical analysis tool *Analyse-it*. Based on the p-value (p-value $<$ critical value) the null hypothesis of inequality is accepted. The acceptance of the null hypothesis leads to the conclusion that See5 rules fail to represent what the SVM has learned at a level of confidence of 95%. Consequently the resulting rules are not discussed here.

The results of extracting rules by use of C4.5 are shown in Table 6.

Similar to See5, C4.5 did not represent what the SVM has learned at a level of confidence of 95%.

Ripper extracted *one* rule from the dataset representing what the SVM has learned (Table 7).

Table 5. See5 confusion matrix for cross-industry IPOs

Evaluation on training data (460 cases):		
(a)	(b)	<-classified as
44	30	(a): class positive
3	383	(b): class negative

Table 6. Decision tree (C4.5) results

Classification	C4.5	
SVM	Positive	Negative
Positive	48	26
Negative	2	384

Table 7. Ripper rules of what the SVM has learned for cross-industry IPOs

Final hypothesis is:
positive :-
SP_DIFF_1>= 1, SP_DIFF_2>= -0.422777, YR_FOUND>= 0.317073 (40/30)
Default negative (328/62)
Train error rate: 20.00% 1.87% (460 data points) $\ll$
Hypothesis size: 1 rule, 4 conditions

Table 8. RBP/Rulex results of what the SVM has learned for cross-industry IPOs

Classification	Rulex	
	Positive	Negative
Positive	23	51
Negative	12	374

The precision for Ripper is 0.57, recall 0.38 and the f-value 0.46.

The null hypothesis of inequality (McNemar test) was rejected (p-value > critical value) and the alternative hypothesis of equality of what the SVM and Ripper have learned was accepted at a level of confidence of 95%. The success of this test is the extraction of a minimal rule set representing SVM learning results.

The RBP/Rulex results are shown in Table 8.

The McNeamar Test leads to the conclusion that Rulex failed to significantly represent what the SVM has learned at a level of confidence of 95% and the resulting rules are not discussed here.

5.2 Rule Extraction from SVM Results for Single-Industry IPOs

SVM Training and Prediction Using the Leave-one-out Method

Support vector machines are used to explore to what extent the upper 25% of first-day returns of "Software and Services IPOs" can be predicted and what are the rules governing these predictions are.

Results from SVM leave-one-out predictions (Table 9) including an error rate as low as 18.13% are an indication that market inefficiencies exist.

SVM Training and Prediction Using Test Sets

Similar to the earlier approach, a learning-based method for rule extraction from support vector machines is used. Software and Services IPOs are randomly split into 122 training and 60 test cases. Repeated random selection is performed until the test set contains about 25% positive cases (22), the same proportion as in the total dataset.

As is evident from Table 10, the quality of the SVM predictions is insufficient, with a precision of 0.3, recall of 0.5 and an f-value of 0.38.

To determine why SVM learning has failed, SVM prediction results are transcribed into the training sets for rule learners to establish what the SVM learned or failed to learn.

Rules Extracted by Use of See5, C4.5, Ripper and Rulex

The results of training the four rule learners with the SVM test predictions as target output are shown below.

Table 9. Leave-one-out results for single-industry IPOs (Linear SVMs)

C value	Test error %	Test recall %	Precision %
Default	19.23	57.14	74.42
0.1	19.23	57.14	74.42
0.2	29.12	5.36	100
1	18.68	60.71	73.91
2	18.13	62.50	74.47
4	21.98	54.17	66.67

Table 10. SVM prediction using a test set for single-industry IPOs

Classed as positive	Classed as negative	
7	16	Positive
7	32	Negative

The rejection of the null hypothesis in the McNemar Test leads to the conclusion that the See5 rules significantly represented what the SVM has learned at a level of confidence of 97.5%. Rule precision was established as 0.86, recall as 1 and the f-value as 0.92.

The rules extracted from C4.5 are shown in Table 12. Evaluation of the rules yields the following results: Since the confusion matrix in Table 13 is identical to the See5 learning results (Table 11) the conclusions are identical. The results of what Ripper has learned are shown in Table 14: Ripper established just one rule with a precision of 0.8, recall 0.57 and f-value 0.67. The McNemar Test established that Ripper significantly represents what the SVM has learned. The results of using the local functions network RBP and the rule extraction technique Rulex are shown in Table 15. The confusion matrix based on RBP and Rulex is shown in Table 16. RBP/Rulex precision is 0.93, recall 0.93 and f-value 0.93. This concludes the rule extraction from SVM experiments, leading to the interpretation of results. The McNemar Test established that RBP/Rulex significantly represents what the SVM has learned.

6 Discussion of Results

The knowledge stored in support vector machines is opaque and cannot easily be extracted. The aim of this research is the extraction of rules from support vector machines in the context of initial public offerings in the US stock market as well as the evaluation of the quality of the rules.

The results from these experiments show how pedagogical techniques using cross-industry and single-industry IPO datasets successfully extract rules from support vector machines.

Table 11. See5 rules of what the SVM has learned for single-industry IPOs

Extracted rules:		
Rule 1: (cover 8)		
CurAssLiab > 0.6863084		
RevGrowth > -0.4512843		
$->$ class positive (0.900)		
Rule 2: (cover 4)		
CurAssLiab $<= -0.1916766$		
MarketCap > 0.3253333		
$->$ class positive (0.833)		
Rule 3: (cover 38)		
CurAssLiab $<= 0.6863084$		
MarketCap $<= 0.3253333$		
$->$ class negative (0.950)		
Rule 4: (cover 16)		
RevGrowth $<= -0.4512843$		
$->$ class negative (0.944)		
Rule 5: (cover 25)		
CurAssLiab > -0.1916766		
CurAssLiab $<= 0.6863084$		
$->$ class negative (0.926)		
Default class: negative		
(a)	(b)	<-classified as
12	2	(a): class positive
0	48	(b): class negative

Table 12. C4.5 rules

Rule 1:
CurAssLiab > 0.686308
RevGrowth > -0.407389
$->$ class positive (84.1%)
Rule 2:
CurAssLiab $<= -0.191677$
OfferOutst $<= -0.761158$
$->$ class positive (70.7%)
Rule 3:
CurAssLiab $<= 0.686308$
OfferOutst > -0.761158
$->$ class negative (93.2%)
Rule 4:
RevGrowth $<= -0.407389$
$->$ class negative (92.6%)
Rule 5:
CurAssLiab > -0.191677
CurAssLiab $<= 0.686308$
$->$ class negative (89.8%)
Default class: negative

Table 13. C4.5 rule evaluation

Rule	Size	Error	Used	Wrong	Advantage	
1	2	15.9%	8	0 (0.0%)	8 (8|0)	positive
2	2	29.3%	4	0 (0.0%)	4 (4|0)	positive
3	2	6.8%	38	1 (2.6%)	0 (0|0)	negative
4	1	7.4%	6	0 (0.0%)	0 (0|0)	negative
5	2	10.2%	6	1 (16.7%)	0 (0|0)	negative
		(a)	(b)	<-classified as		
		12	2	(a): class positive		
		0	48	(b): class negative		

Table 14. Ripper results

Final hypothesis is:
Positive: – CurAssLiab >= 1 (8/2)
Default negative (46/6)
Train error rate: 12.90% 4.29% (62 data points) ≪
Hypothesis size: 1 rules, 2 conditions

The conclusions drawn from the cross-industry experiments are:

- There is an indication of market inefficiencies, however, the learning results are insufficient.
- SVM learning from randomly selected data points to a "hard to learn" dataset. Precision is established as 0.33, while recall is 0.5 and the f-value 0.38.
- The subsequently extracted rules from the SVM using the rule learners See5, C4.5 and RBP/Rulex did not significantly represent what the SVM has learned.
- The one rule extracted from the SVM using Ripper significantly represents what the SVM has learned. Rule precision was established as 0.57, recall as 0.38 and the f-value as 0.46.

The Ripper rule uses technical and company specific attributes: a steep increase in the S&P index in the last two periods combined with more established companies (year founded). The rule is economically plausible since it points to a "hot issue market" phenomenon (Ibbotson & Jaffe 1975; Ritter 1984) and the age of the firms is found to be significant by Moonchul and Ritter (1999). The issue that accounting ratios did not feature in the one Ripper rule may point to the difficulty of comparing companies across different industries.

Table 15. RBP/Rulex results

Number of rules = 2	Number of antecedents = 24
RULE 1	
IF CashLiab	IS BETWEEN −0.0357694 AND 1
AND ActualOffer	IS BETWEEN −1 AND 0.484134
AND PropActOffer	IS BETWEEN −1 AND 0.603759
AND RegDays	IS BETWEEN −1 AND 0.0335572
AND GrossMargin	IS BETWEEN −1 AND 0.598339
AND GrMarginShare	IS BETWEEN −1 AND −0.26638
AND RevGrowth	IS BETWEEN −0.605131 AND 1
AND Revenue	IS BETWEEN −1 AND 0.381687
AND NASDAQ_Period_5	IS BETWEEN −1 AND 0.286907
AND NASDAQ_Period_4	IS BETWEEN −0.67481 AND 1
AND NASDAQ_Period_3	IS BETWEEN −0.183016 AND 1
AND NASDAQ_Period_1	IS BETWEEN −1 AND 0.0780277
THEN >50%	
RULE 2	
IF MarketCap	IS BETWEEN 0.131663 AND 1
AND PropActOffer	IS BETWEEN −1 AND 0.563242
AND OfferOutst	IS BETWEEN −1 AND 0.780495
AND RegDays	IS BETWEEN −1 AND 0.655597
AND GrossMargin	IS BETWEEN −0.380277 AND 1
AND RevGrowth	IS BETWEEN −0.418135 AND 1
AND NASDAQ_Period_10	IS BETWEEN −0.502027 AND 1
AND NASDAQ_Period_9	IS BETWEEN −1 AND 0.342823
AND NASDAQ_Period_7	IS BETWEEN −0.446386 AND 1
AND NASDAQ_Period_5	IS BETWEEN −1 AND 0.334158
AND NASDAQ_Period_3	IS BETWEEN −1 AND 0.770492
AND NASDAQ_Period_1	IS BETWEEN −0.906461 AND 1
THEN >50%	

The conclusions drawn from the single-industry experiments are:

- The overall error rate of 18.13% achieved by SVM learning using cross-validation points to market inefficiencies, but not to the specific factors responsible for these inefficiencies.
- SVM learning from randomly selected data points to a “hard to learn” dataset. Precision was established as 0.3, recall as 0.5 and the f-value as 0.46.
- The subsequently extracted rules from the SVM by all rule learners (See5, C4.5, Ripper and RBP/Rulex) significantly represent what the SVM has learned.

Table 16. Performance summary of RBP/Rulex

Evaluation on test data (62 items):					
Rule	Size	Used	Correct	Wrong	Certainty
1	12	10	10	0	1.00
3	12	6	5	1	0.83
		Performance summary			
		predicted			
		0	1	No classification	
Class	0	47	1	0	
	1	1	13	0	

- The attributes of See5 rules exclusively focus on accounting ratios. Higher asset to liabilities ratios and higher market capitalisation feature as making a positive contribution to higher first-day IPO returns. See5 rule accuracy is very high and rules are economically plausible. Precision was established as 0.86, recall as 1 and the f-value as 0.92.
- Similarly C4.5 focuses exclusively on accounting ratios, but adds the rule element of offered vs. outstanding shares. C4.5 rule accuracy is also very high and rules are economically plausible. Precision was established as 0.86, recall as 1 and the f-value as 0.92.
- The one rule Ripper generates, including the asset vs. liabilities ratio, is precise and economically plausible. Precision was established as 0.8, recall as 0.57 and the f-value as 0.67.
- In contrast RBP/Rulex yield two rules consisting of a mix of company specific and market sentiment elements. RBP/Rulex rules are very precise. Again RBP/Rulex rules are economically plausible, adding the "hot issues market" theme. Precision was established as 0.93, recall as 0.93 and the f-value as 0.93. The RBP/Rulex rules are therefore by a slim margin superior to See5 and C4.5 rules.

In summary, all rule learners found it substantially easier to extract rules from the dataset representing what the SVM has learned in comparison to the original data. This points to a filtering or smoothing effect as a result of SVM learning. Two major competing rule models emerge, one that exclusively focuses on accounting ratios and one that combines accounting ratios with market sentiment.

7 Conclusions

The experiments have shown how the ability of SVMs to solve problems can be combined with the benefits of extracting the symbolic representation of the knowledge contained in SVMs.

The experiments have shown that small pockets of predictability exist in the IPO market. The economic plausibility of the rule attributes associated with the predictions has been confirmed. These results are not only relevant for investment decisions in the capital market, but may be of benefit to other applications, such as software verification and safety applications.

References

Andrews R, Diederich J, Tickle AB (1995) A Survey and Critique of Techniques For Extracting Rules From Trained Artificial Neural Networks, Knowledge Based Systems, 8, pp. 373–389

Andrews R, Geva S (1996) Rules and local function networks, In Rules and Networks, R. Andrews, R. & J. Diederich (eds), Queensland University of Technology, Neurocomputing Research Centre

Cohen WW (1995) Fast effective rule induction (Ripper), AT&T Laboratories, New York: Proceedings of 12th International Conference of Machine Learning, Lake Tahoe, California, July 9–12

Gardner MJ, Altman DG (1989) Statistics with confidence. London BMG Books

Ibbotson RG, Jaffe JF (1975) Hot issue markets, Journal of Finance, vol. 30, pp. 1027–1032

Joachims T (1998) Categorization with support vector machines, learning with many relevant features', European Conference of Machine Learning. Chemnitz, Germany, April 21–23, Proceedings: Springer (1998)

Joachims T (1999) Making large-Scale SVM learning practical. In: Advances in kernel methods – Support Vector Learning. Schölkopf B, Burges C, Smola A (eds), MIT Press, Cambridge, Massachusetts

Mitsdorffer R, Diederich J, Tan TW (2001) Predicting first-day returns of cross-industry initial public offerings in the US stock market', ANNES 2001, the biennial International New Zealand Conference on Artificial Neural Networks and Expert Systems, University of Otago, Dunedin, New Zealand

Mitsdorffer R, Diederich J, Tan TW (2002) Rule-extraction from technology IPOs in the US Stock Market: The Ninth International Conference on Neural Information Processing (ANNES 02), November 2002, Singapore

Quinlan JR (1986) Induction of decision trees, Machine Learning, Vol. 1(1), pp. 81–106

Quinlan JR (1993) C4.5: Programs for machine learning. San Mateo: CA, Morgan Kaufmann

Quinlan JR (2001) Bagging and boosting, and C4.5, University of Sydney, Sydney

Moonchul K, Ritter JR (1999) Valuing IPOs, Journal of Financial Economics, Vol. 53, pp. 409–437

Refenes AN, Zapranis AD (1995) Modeling stock returns in the framework of APT: A comparative study with regression models, Neural networks in the capital market. pp. 138–161. John Wiley & Sons, Chichester

Ritter JR (1984) The "Hot Issue" market of 1980, Journal of Business, Vol. 32, pp. 215–240

Ritter JR (1991) The long-run performance of initial public offerings, Journal of Finance, vol. 46, pp. 3–27

Rulequest Research (1997) Data Mining Tools See5 and C5.0, Available at: http://www.rulequest.com/see5-info.html, viewed: 2003, October 28