Detecting Concept Drift in Financial Time Series Prediction using Symbolic Machine Learning

Detecting Concept Drift in Financial Time Series Prediction using Symbolic Machine Learning

Michael Harries
mbh@cse.unsw.edu.au
Department of Artificial Intelligence
School of Computer Science and Engineering
University of NSW

Kim Horn
kimh@ags.com.au
Predictive Strategies Unit
Australian Gilt Securities
Limited

Keywords: Supervised Concept Learning, Machine Learning, C4.5, ID3, Concept Drift, Forecasting, Noise, Divide and Conquer, Decision Trees, Finance.

Abstract

This paper investigates the use of strategies to enhance an existing machine learning tool, C4.5, to deal with concept drift and non-determinism in a time series domain. Temporal prediction is a difficult problem faced in most human endeavours. While many specialised time series prediction techniques have been developed, these techniques have limitations. Most are restricted to modeling whole series rather than extracting predictive features and are difficult for domain experts to understand. Symbolic machine learning promises to address these limitations. Symbolic machine learning has been very successful on a broad range of complex problems. To date, few attempts have been made to apply symbolic machine learning directly to temporal prediction. This has resulted in systems that cannot explicitly represent temporally ordered examples or handle changing target concepts.

Financial prediction is a challenging target domain, which is temporally ordered, has target concepts that change over time and exhibits a high level of non-determinism. Financial markets are considered to be unpredictable by many academics and thus any improvement on chance is interesting. An aim of this study is to demonstrate that machine learning is capable of providing useful predictive strategies in financial prediction. For short term financial prediction a successful prediction rate of 60% is considered the minimum useful to domain experts. Our results imply that machine learning can exceed this target with the use of new techniques. By trading off coverage for accuracy, we were able to minimise the effect of both noise and concept drift.

The work reported was in collaboration with and funded by Australian Gilt Securities Limited. Michael Harries was supported by an Australian Postgraduate Award (Industrial).

1. Introduction

Time series prediction is the task of forecasting future values from a temporally ordered sequence of data. Such prediction is required in many fields, for example, the prediction of, weather, sun spot activity, floods, heart beat irregularities in medicine, network load, and future sales. This paper examines the prediction problem in the financial markets.

Time series prediction, in the financial markets is extremely controversial. On the one hand there are market analysts who claim that the market is deterministic; on the other, there are those who claim that it is random. A theory called the efficient markets hypothesis states that all information received by the market is immediately reflected in the price. As there is no unused information, it is impossible to use past prices to predict future prices. The market is then best described as a random walk. This description implies that a price change is merely the sum of a series of random steps and that the best possible prediction for any future price is the current price. There is, however, considerable evidence to suggest that the market is not totally random. Peters (1991,1994) provides evidence to suggest that the markets have a fractal structure. Many studies attempt to find directly predictive rules. A survey of statistical approaches can be found in Granger (1992). Neural networks are also well represented in this quest, for example, Kimoto et al. (1993), Trippi & DeSieno (1992) and Bergerson & Wunsch (1993).

We now proceed as follows. Section 2 reviews prediction techniques currently used in finance and their limitations. Section 3 briefly reviews a class of machine learners. Section 4 looks at the issues involved with applying machine learning to temporal domains. Some challenging features of the financial domain are identified. Strategies to deal with these challenges are presented in Section 5. Section 6 introduces the application domain and task. Section 7 looks at the changes to accuracy estimating techniques required in a temporal domain. Section 8 shows the results of applying machine learning to the application domain. We conclude that the concept drift and noise detection strategies were successful and that machine learning is a useful technology for this domain.

2. Approaches to Time Series Analysis

There are a broad range of approaches to time series prediction. These range from naive methods, such as using the most recent observation as a forecast to highly complex non-linear systems approaches (Weigend, 1992). The most common "classical" approaches applied in the financial markets include: Autoregressive (AR), Moving Average (MA) and combined approaches: ARMA, ARIMA and ARCH (Peters, 1994). Neural networks are a more recent approach that are encouraging considerable interest. All of these approaches have achieved some success but are limited in many respects.

The common "classical" approaches mentioned above require considerable expert knowledge to apply, are time consuming to build, and are limited by attempting to model the whole market in a single, succinct package. These approaches usually require the data to be sampled at a specified time interval and cannot be directly applied to data based on transactions with no reliable frequency. Neural networks are also restricted by requiring considerable expert knowledge to use; both to set parameters and to structure the network. They are very slow to train which limits the maximum practical training set size. In all cases, the primary limitation from our perspective is that the systems built are not easily understood by domain experts. Neural networks tend to be even more impenetrable than classical models. As a result traders are hesitant to use them and have little understanding of how they work, or more importantly, whether they are reliable.

3. Machine Learning Approaches

Machine learning focuses on systems that can learn a symbolic description of a concept. The goals for this concept are that it be accurate, simple, general and that it be readable by domain experts. In the machine learning literature divide-and-conquer approaches have become popular and have demonstrated success in many real world applications. CART (Brieman et al., 1984), C4.5 (Quinlan, 1993), M5 (Quinlan, 1992) are examples of this approach. These approaches work by successively dividing the problem space into nested rectangular regions and fitting surfaces to these regions. Divide-and-conquer approaches offer many advantageous features. These techniques can cope with noisy, missing and misclassified data. They are also very fast to train which permits the use of extremely large data sets. The C4.5 system can produce descriptions that are either in the form of rules or as a decision tree. Probabilistic information about error can be obtained for individual tree nodes or rules. In this study we decided to use C4.5 as it is well known, well documented, is familiar to the authors and the source code is available.

4. Limitations of Machine Learning Approaches In Temporal Domains

To date, machine learning has largely been directed toward static problems. The extension of the existing paradigm to encompass time series type problems is not trivial but should provide considerable advantages both to the field and in applications to real problems. To be successful machine learning systems should be able to capitalise on the information that attribute values may be ordered and that separate attributes may be related. Some interesting work has been done applying machine learning to temporal domains. Dietterich & Michalski (1986) built a system to learn to predict sequences in a card game called Eleusis. Of particular interest is the fact that Eleusis learns non-deterministic sequencing rules. Laird (1992) describes discrete sequence prediction as a fundamental problem largely ignored outside the data compression community. He uses a Transition Directed Acyclic Graph (TDAG) as a tool to learn to predict sequences of signals.

Temporally ordered data poses a considerable problem for machine learning systems like C4.5 that cannot utilise inherent ordering information and are not capable of learning new relationships between attributes. To use C4.5 in temporal domains it is often required that new attributes be constructed apriori to represent informative temporal relationships. Rather than investigate new higher order learning algorithms we decided to use and extend C4.5 for success in this domain.

Prediction tasks in real world domains are faced with the additional problem that the underlying phenomena itself may change over time. For instance, in the financial domain, market behaviour can change dramatically with changes in contract prices, interest rates, inflation rates, budget announcements, and political and world events. This kind of phenomena is termed "concept drift". Another problem for a prediction system is that the distribution of classes changes over time and it is not possible to know whether the distribution used for training is characteristic of the current domain. The machine learning system thus needs to be able to detect when its knowledge is out of date and needs to be updated. Some work has been done in this area. Schimmler and Granger (1986) built a system to learn simple boolean concepts in the presence of noise and concept drift. They use weighted multiple models and reinforce the most correct one over time.

Another problem with prediction in the financial markets is the high level of non-determinism (or for our purposes, noise). While C4.5 is extremely robust to high levels of noise, it appears to be near the limit of it's robustness in this domain. Additional techniques are required to use decision trees in highly non-deterministic domains.

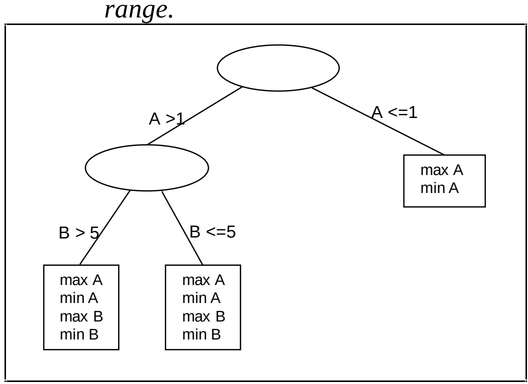
5. Concept Drift And Non-Determinism

We investigated the use of two strategies to improve C4.5 for this domain. Both strategies trade off breadth of coverage for accuracy. The aim is to make a smaller number of strong predictions rather than many weaker predictions. In the financial domain every trade has an associated transaction cost. Predictions resulting in trades that do not cover the transaction cost should be avoided if possible. The result is that conservative prediction strategies are more likely to be profitable.

The first strategy we investigated was aimed at alleviating the noise problem. After being trained, C4.5 provides an expected misclassification rate for each leaf. Rather than associate a confidence with each prediction, we instead used only those predictions with an expected misclassification rate below a given threshold. We examined thresholds of 20, 5, 4, 3 and 2 %. This strategy is different to pruning in that we make no prediction when the misclassification rate is above the threshold, whereas with pruning we would utilise the majority class from higher in the tree.

The second strategy we investigated attempted to minimise the effects of concept drift. The goal was to provide a classification only when an unseen instance is 'like' the training data. We did this by associating a set of permitted attribute ranges with each leaf of the decision tree. To provide a classification, the instance's attributes must be within the ranges recorded during training for that leaf. For each leaf, only the attributes used to reach that leaf are used in the range test. Figure 1 shows a case in which two leaves use ranges on both attributes A and B, and one leaf uses ranges only on attribute A.

Figure 1: Detecting concept drift using leaf range.



Attribute range is a fairly simplistic and statistically naive approach to determine when a case is like another. The major assumption is that when concept drift has occurred, the tails of the distribution of cases for a given attribute (reflected by range) will have changed. In the financial markets this is reasonable assumption as prices naturally stay within certain bounds for periods of time then move on to new bounds when things change. We could have employed other approaches such as nearest neighbour, as used in instance based learning schemes (Aha et al., 1994). However, these methods would significantly slow down testing and increase storage requirements. The aim, in this first study, was for speed and simplicity.

We can expect three possible scenarios when taking concept drift into consideration. The first is when the test set is not different in nature to the training set and we have no concept drift. The second is when there is some concept drift but with a majority of cases still being like the training cases. In this situation there will be a reduced number of cases classified but the accuracy should approximate the first scenario. The third scenario occurs when there is significant concept drift, here we would expect a large reduction in the number of cases classified with a concomitant fall in accuracy.

Finally, we investigated a combined approach using both noise and concept drift approaches together.

6. Share Price Index Prediction Domain

Australian Gilt Securities is an active trader of Share Price Index (SPI) futures. SPI futures are futures sold on the Sydney Futures Exchange (SFE) based on the All Ordinaries Index. The company has considerable interest in the ability to predict the movement of the SPI and provided the SPI data for this work. The data set contains tick by tick (transaction by transaction) data for all of 1993. Each SPI transaction contains a time-stamp, the price and the volume traded. Transactions were recorded as they occurred, which means that they are sequentially ordered, but that the time elapsed between each is not fixed. SPI futures contracts are available in the market three years prior to their expiry but are only heavily traded in their last three months. New contracts are issued every three months giving four separate three month contracts each year. These contracts mature in March, June, September and December.

The dataset in its raw form does not afford the use of traditional time series techniques. This is because of the assumption that each time step is of the same size. As converting the information to a summarised form must lead to a loss of information, we chose to maintain the data in largely its initial form. Each transaction was augmented with attributes based on past data (shown in Table 1) and within case normalisation was used on the price based attributes. A limitation with this data relates to the separate nature of each contract as in each year there are four relatively independent data sets rather than one single sequence. A number of splicing techniques are available but each demands somewhat artificial assumptions. Market traders are aware of the separate nature of contracts and trade them separately and so we sought to achieve a similar approach.

Table 1: Attributes used.

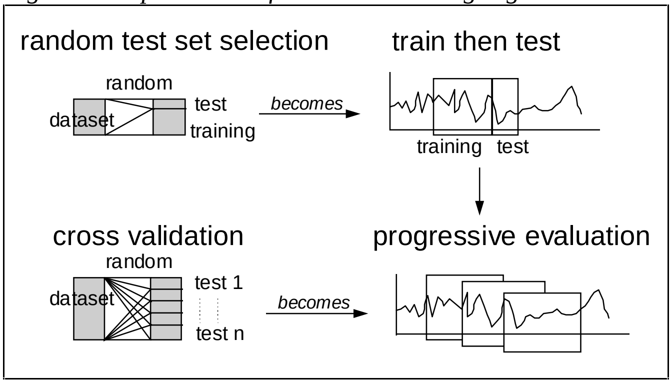
seconds of trading this contract	days to end of contract
price	how close to start of day
volume	how close to lunch starting
price five minutes ago	how close to lunch ending
price ten minutes ago	how close to end of day
price twenty minutes ago	volume last hour
price thirty minutes ago	volume last two hours
price one hour ago	average price yesterday
price two hours ago	average price last 7 days,
price three hours ago	average price last 20 days
price four hours ago	ratio 7 day av. price to 20 day av. Price
price five hours ago	

The aim of this study was to demonstrate that machine learning is capable of providing useful predictive strategies. As a relatively simple initial task, the direction of market movement, either up or down, over the next half hour was chosen as a target concept. A minimum acceptable accuracy of 60% on this task would be directly useful in trading (Private communication with AGS traders). Less than this amount would not be useful due to transaction costs and unavoidable delays in trading.

7. Testing Regime

When using machine learning in a static domain, the understandable assumption is made that the distribution of instances is uniform. This implies that a randomly selected test set can be used to determine classifier accuracy. When dealing with temporal data and prediction, the only way to get a fair estimate of classifier accuracy is to select a test set that occurs after the training set.

Figure 2: Implications of domain on testing regime.



However, using a single test set is limiting as the estimated results can vary widely depending on how fortuitous the training set is. In the machine learning literature, this problem is overcome using a technique called cross validation. This involves the dataset being randomly split into *n* sets and repeated train and test cycles being carried out on different combinations of these sets. This technique is not valid in this domain as the results from randomly selected sets cannot provide an accurate estimate of predictive accuracy when concept drift is present. An alternate approach, called progressive evaluation, is to move a window across the data. Within this window the data is split consecutively into a train and test set, as shown in figure 2. The results for the whole dataset can be determined by averaging across all of the windows. Progressive evaluation was used in this study. Consecutive pairs of SPI contracts were employed as testing and training sets.

8. Results

Table 2: Naive prediction and data set sizes.

Contract	N	Correct
Mar	18091	48.91%
Jun	17857	50.67%
Sep	20781	50.46%
Dec	32971	51.81%
Total	89700	

When machine learning techniques are applied in non-temporal domains, the first criterion for accuracy is that the classifier must perform better than a naive predictor based on the default class. In this domain however, the default class is different in each subsequent quarter of 1993. Because we cannot rely on a fixed default class, it is not possible to simply choose the past highest frequency class and expect this to provide accurate forecasts. A more temporally aware naive predictor is to use what ever happened in the last 30 minute period as the forecast for the next 30 minute period. Table 2 shows that with this method, only the December quarter was significantly different ($p<0.01$) to chance. However, the percentage accuracy was only 51.81%. All significance tests in this paper used Press's q statistic. As this, and most other statisitics are sensitive to large sample sizes, care must be taken in drawing conclusions about the significance of very small classification rates.

Table 3 shows that C4.5 with no restrictions provides a classification accuracy of 52.31% when averaged over the 3 test periods, this is significantly better than 50% ($p<0.01$). As we altered our noise reduction setting, we were able to achieve further improvements in prediction. Tables 3 and 4 show that reducing the estimated misclassification threshold progressively

Table 3: Individual and average accuracy's with noise reduction strategy.

description		**% correct**					
train	test	no restriction	20% est	5% est	4% est	3% est	2% est
Mar	Jun	54.06	51.84	51.24	51.24	64.05	100.00
Jun	Sep	47.38	47.85	58.68	63.51	100.00	
Sep	Dec	54.48	55.02	56.62	59.49		
	% average	52.31	52.42	55.66	58.01	64.13	100.00

Table 4: Number of transactions classified with noise reduction strategy.

description		**number in range**					
train	test	no restriction	20% est	5% est	4% est	3% est	2% est
Mar	Jun	17857	8298	6175	6175	2320	27
Jun	Sep	20781	16194	593	507	5	0
Sep	Dec	32971	30406	26614	26424	0	0
	total	71609	54898	33382	33106	2325	27

increases the accuracy of prediction and reduces the number actually classified.

Table 5: Percent accuracy with concept drift detection.

description		**% correct**		
train	test	no restriction	range	range+20%
Mar	Jun	54.06	58.43	67.91
Jun	Sep	47.38	59.47	60.83
Sep	Dec	54.48	51.22	50.46
	% average	52.31	57.47	60.62

Table 6: Number of transactions classified with concept drift detection and mis-classification thresholding

description		**Number classified**		
train	test	no restriction	range	range+20%
Mar	Jun	17857	3914	1770
Jun	Sep	20781	2319	2045
Sep	Dec	32971	1349	1312
	total	71609	7582	5127

Tables 5 and 6 show the results of using concept drift detection. Range detection alone substantially improves the results but in combination with a 20% misclassification threshold the results are further improved. For two out of three periods the prediction accuracy was better than our target of 60% and significantly better than chance ($p<0.01$). The third test period (December) does not perform better than chance. In this period the proportion of cases classified by the system also dropped significantly ($p<0.001$), from 9.84 to 3.98 %, as seen in table 7. This drop suggests that the market has changed. Considering this, we should not be surprised that our system cannot perform well in December as it was trained in quite a different market.

Table 7: Percentage of allowed transactions in range to total

description		**% classified**
train	test	range+20%
Mar	Jun	9.91
Jun	Sep	9.84
Sep	Dec	3.98

In retrospect we discovered that a major change *did* occur in the Share Price Index in the 4th quarter of 1993. The SPI contract price changed from $100 to $25 in October. The SFE reduced the contract purchase price in order to generate interest from smaller investors. The result was that a whole new range of players entered the market which concomitantly affected market behaviour. Contract price change was not an attribute included in the data, nor would it have been any use including it as the change did not occur within a training period. Despite this, we were able to detect this market change. The results, in table 7, indicate that attribute range is useful in detecting concept drift when a known change does occur in the market.

9. Conclusion

By using simple strategies to detect concept drift and non-determinism we significantly improved the results of C4.5 alone. The prediction results meet the target set by the trading experts at AGS and should provide a useful prediction tool. The concept drift technique provides a useful indicator as to whether the world has changed. In real time trading, a significant drop in the frequency of classified cases would indicate a drop in reliability. At this stage a new system should be induced to cover the changed market.

Future work will involve additional testing on more recent futures data and examine the applicability of these techniques to other domains. We will also test the system within a trading system to simulate the actual buying and selling of futures to determine if an actual profit can be realised. The examination of synthetic data sets is also relevant as the level of concept drift can be controlled to explore different detection techniques. Other methods to detect concept drift, such as nearest neighbour and model reinforcement, will be explored and the computational cost-benefits compared.

References

Aha, D.W., Kibler, D. & Albert, M.K., Instance-Based Learning Algorithms. *Machine Learning,* 6, 37-66, 1991.

Bergerson, K. & Wunsch II, D. C., A Commodity Trading Model Based on a Neural Network-Expert System Hybred, in *Neural Networks in Finance and Investing*, eds Trippi, R. R. & Turban, E., 1993.

Breiman, L, Friedman, J, Olshen, R, & Stone, C, *Classification and Regression Trees*, Wadsworth and Brooks, 1984.

Dietterich, T.G & Michalski, R.S., Learning to predict sequences, *Machine Learning: An Artificial Intelligence Approach*: Volume II, 1986.

Granger, C.W.J., Forecasting stock market prices: lessons for forecasters, *International Journal of Forecasting* 8 (1992) 3-13, 1992.

Hair JR, J.F., Anderson, R.E., Tatham, R.L., & Black, W.C., *Multivariate Data Analysis: with Readings*, Macmillan Publishing, 1992.

Kimoto, T., Asakawa, K., Yoda, M., & Takeoka, M., Stock Market Prediction System with Modular Neural Networks, *Proceedings of the IEEE International Joint Conference on Neural Networks*, 1990.

Laird, P., Discrete Sequence Prediction and its Applications, AAAI-92 *Proceedings of the Tenth National Conference on Artificial Intelligence*, 1992.

Peters, E. E., *Chaos and Order in the Capital Markets*, John Wiley and Sons Inc, 1991.

Peters, E. E., *Fractal market Analysis*, John Wiley and Sons Inc, 1994.

Quinlan, J. R, *C4.5: Programs for Machine Learning*, Morgan Kaufmann Publishers Inc., 1993.

Quinlan, J. R, Learning with Continuous Classes, In *Proceedings AI-92,* eds Admas and Sterling, 343-348, Singapore: World Scientific, 1992.

Schlimmer, J.C. & Granger Jr, R. II, Beyond incremental processing: Tracking concept drift, *AAAI-86*, 1986.

Trippi, R. R. & DeSieno, D., Trading Equity Index Futures with a Neural Network, *The Journal of Portfolio Management,* Fall 1992 p27-33, 1992.

Weigend, A. S. & Gershenfeld, N. A, *Time Series Prediction*, 1992.