DATA MINING OF MULTIPLE NONSTATIONARY TIME SERIES

RICHARD J. POVINELLI AND XIN FENG
Department of Electrical and Computer Engineering
Marquette University, P.O. Box 1881, Milwaukee, WI 53201-1881, USA
E-mail: Richard.Povinelli@marquette.edu
Ph: 414.288.7088
Fx: 414.288.5579

ABSTRACT:
A data mining method for synthesizing multiple time series is presented. Based on a single time series algorithm, the method embeds multiple time series into a phase space. The reconstructed state space allows temporal pattern extraction and local model development. Using an *a priori* data mining objective, an optimal local model is chosen for short-term forecasting. For the same sampling period, multiple time series embedding produces better temporal patterns than single time series embedding. The method is applied to a financial time series.

Keywords: Temporal Pattern Identification, Data Mining, Time Series Analysis

Time series analysis is fundamental to engineering and scientific endeavors. Researchers study nature as it evolves through time, hoping to develop models useful for predicting or controlling it. Data mining is the analysis of data with the goal of uncovering hidden patterns. It encompasses a set of methods that automate the scientific discovery process. Its uniqueness is found in the types of problems addressed - those with large data sets and complex, hidden relationships.

Traditional time series analysis methods such as the Box-Jenkins method (Bowerman and O'Connell 1993) are limited by the requirement of stationarity of the time series and normality and independence of the residuals. However, for most real world time series, these conditions are not met. One of the most severe drawbacks of this approach is its inability to identify complex characteristics. These drawbacks of the Box-Jenkins method occur because it tries to characterize and predict all points in a time series.

This work applies data mining concepts to time series analysis. In particular, it establishes a method that uncovers hidden patterns in time series data. This novel approach overcomes the limitations of previous time series analysis methods by finding temporal patterns. Previous work (Povinelli and Feng 1998) found optimal local models for event (important occurrence localized in time) characterization and prediction, using a single time series to recreate an attractor. This paper looks at embedding multiple time series – a process analogous to having multiple sensors on a system.

The method is capable of handling nonstationary, nonperiodic, irregular time series, including chaotic deterministic time series. It is applicable to time series that appear stochastic, but occasionally (though not necessary periodically) contain distinct patterns that are characteristic and predictive of the desired events. This might include predicting when a droplet from a welder will release, when a device will fail, when a stock price will drop, when an earthquake will strike, or when the heart will go into fibrillation.

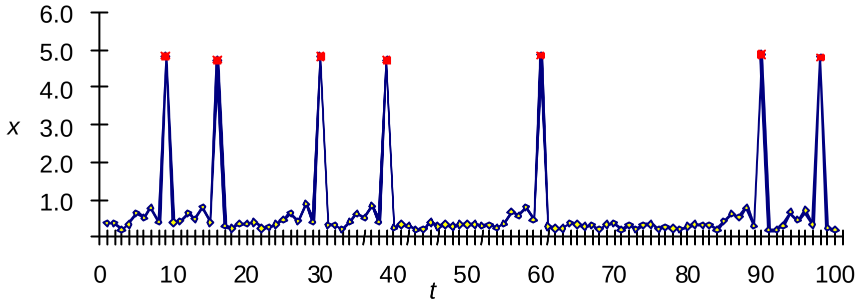

Figure 1- Training Time Series

This research is the combination of ideas from several fields. It incorporates concepts from data mining, time series analysis, wavelets, genetic algorithms, and dynamical systems. From wavelets comes the idea of a temporal pattern. From genetic algorithms comes a robust and easily applied optimization method (Goldberg 1989, pp. 106-120). From the study of dynamical systems comes the theoretical justification of the method, specifically Takens' Theorem (Takens 1980) and Sauer's extension (Sauer et al. 1991).

PROBLEM STATEMENT

Graphically, the problem can be understood by studying Figure 1, which shows a synthetic, nonstationary, non-periodic time series.

The squares indicate observations that are deemed important – events. The goal is to characterize and predict events in the time series. To make the time series more concrete, consider it a measure of seismic activity with events being those points over 4.5. The method aims to characterize when peak seismic activity (earthquakes) occurred and then use the characterizations for prediction.

Mathematically the problem is defined as follows. For a training time series $X=\{x_t,t=1,\ldots,N\}$ and a testing time series $Y=\{x_t,t=R,\ldots,S\}$ $R>N$, define, *a priori*, an application dependent event function, $g(x_t)$. The goal is to identify and characterize a subset X_{events} of X where, X_{events} has the following properties. First, a characteristic function exists which allows the separation of X_{events} from X^c_{events}, the complement of X_{events}. The characteristic function encapsulates the features or patterns that distinguish events from non-events. Second, the mean of $g(X_{events})$ is greater than the mean of $g(X^c_{events})$. Third, X_{events} is statistically different from X^c_{events} and X.

The method seeks to apply the features to predicting events in Y, the test time series. The aim is to predict a subset Y_{events} of Y, where the mean of $g(Y_{events})$ is greater than the mean of $g(Y^c_{events})$, and Y_{events} is statistically different from Y^c_{events} and Y.

SOLUTION PROCEDURE

The key to the new approach is that it forgoes the need to characterize and predict at all times for the advantages of being able to identify the "optimal" local model for characterizing and predicting important events. The solution procedure involves a transformation from the time series space to a phase space. In the phase space, a clustering algorithm is applied. The goal is to find the optimal temporal pattern that characterizes and predicts events.

The key components of the method are the event function, time series space, phase space, temporal pattern, and temporal pattern cluster. The event function assigns an "eventness" to every observation. The time series space is a two-dimensional space of time and magnitude.

The phase space is a mapping, $\mathbb{R}^2 \to \mathbb{R}^{m \cdot Q}$, from the time series space into an $m \cdot Q$ dimensional metric space. At the core of the method is the temporal pattern, an $m \cdot Q$ dimensional vector. It is the data mining indicator, allowing the characterization and prediction of events. The final component, the temporal pattern cluster, is defined using the phase space metric, the temporal pattern, and a threshold d. The temporal pattern cluster is a hypersphere centered on the temporal pattern with a radius d.

Given m training time series $X_1, X_2, \ldots, X_m$ where each time series takes the form $X_m=\{x_{m,t}, t=1,\ldots,N\}$, and m test time series $Y_1, Y_2, \ldots, Y_m$ which take the form $Y_m=\{x_{m,t}, t=R,\ldots,S\}$ $R>N$, a phase space is formed. The phase space is an $m \cdot Q$ dimensional real space $P \subseteq \mathbb{R}^{m \cdot Q}$ with metric d. The time series are embedded into the phase space yielding $\mathbf{x}_t$ as follows

$$\mathbf{x}_t^T = \left(x_{1,t-t_{Q-1}}, \ldots, x_{N,t-t_{Q-1}}, \ldots, x_{1,t-t_1}, \ldots, x_{N,t-t_1}, x_{1,t}, \ldots, x_{N,t} \right),\ t = t_{Q-1}+1, \ldots, N-1 ,$$

where $t_1 < t_2 < \ldots < t_{Q-1}$. In a similar manner, the test time series are embedded yielding $\mathbf{y}_t$.

The time series magnitudes may be normalized to assist the optimization routines. Normalization does not change the topology of the phase space, but giving each time series the same range allows similar search step sizes for each phase space dimension. The normalization constant is retained for use in predicting events in the testing time series.

An *a priori*, application dependent event function, $g(x_t)$, is defined. The method finds a pattern cluster, defined by the temporal pattern $\mathbf{p} \in P$ and $d \in \mathbb{R}$, that has the following two properties. Several definitions are needed to describe the properties. The first is the index sets M_{train} and M_{test} which are the times t when $\mathbf{x}_t$ and $\mathbf{y}_t$, respectively, are within the temporal pattern cluster.

$$M_{train} = \{ t : d(\mathbf{p}, \mathbf{x}_t) \le d \}, \qquad t = t_{Q-1}+1, \ldots, N-1 .$$
$$M_{test} = \{ t : d(\mathbf{p}, \mathbf{y}_t) \le d \}, \qquad t = t_{Q-1}+R, \ldots, S-1 .$$

The second definition is the average event value of the times in M.

$$m_M = \frac{1}{c(M)} \sum_{t \in M} g(x_t) ,$$

where $c(M)$ is the cardinality of M. The average event values of all times m_X and m_Y are defined as follows:

$$m_X = \frac{1}{N - t_{Q-1} - 2} \sum_{t=t_{Q-1}+1}^{N-1} g(x_t) \,.$$

$$m_Y = \frac{1}{S - t_{Q-1} - R - 1} \sum_{t=t_{Q-1}+R}^{S-1} g(x_t) \,.$$

Property 1, which is for the training time series, requires that $m_{M_{train}} > m_X$ and that the set $\{g(x_t)$: $t \in M_{train}\}$ is statistically different from the set $\{g(x_t)$: $t = t_{Q-1}+1,\ldots,N-1\}$. Property 2, which is for the test time series, requires that $m_{M_{test}} > m_X$ and that the set $\{g(x_t)$: $t \in M_{test}\}$ is statistically different from the set $\{g(x_t)$: $t = R \ldots S\text{-}Q+1\}$.

This is all accomplished by

$$\begin{aligned} &\underset{\mathbf{p},d}{\text{maximize}} \quad f(\mathbf{p}, d, X, g) \equiv m_{M_{train}} = \frac{1}{c(M_{train})} \sum_{t \in M_{train}} g(x_t) \,. \\ &\text{subject to} \quad c(M_{train}) > b\, N, \;\; 0 < b \le 1. \end{aligned}$$

The "subject to" allows a b to be selected so that $c(M)$ is non-trivial, i.e., so that the neighborhood around **p** includes some percentage of the total embedded time series. If $b = 0$ then $c(M) = 1$ or all elements of M are identical.

APPLICATIONS AND RESULTS

The method is applied to daily open price and volume data of ICN, a NYSE traded stock, from 1990 and 1991 as shown in Figure 2. The time series were filtered to provide the daily percentage change in open price and in volume. For this type of time series, the filtering facilitates temporal pattern identification. The training results are shown in Table 1 and Table 3. The test results are shown in Table 2 and Table 4.

The statistical test used to show the significance of the results is the runs test. The test hypothesis is

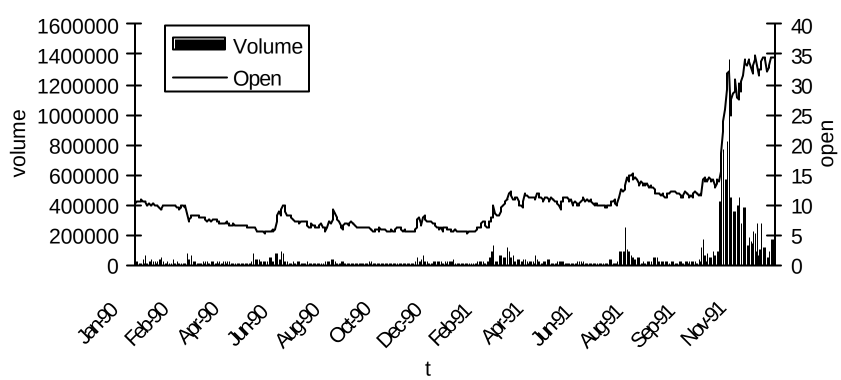

Figure 2 – ICN Daily Open Price and Volume for 1990 – 1991

H_0: There is no difference between the matched time series and the remaining time series.

H_A: There is significant difference between the matched time series and the remaining time series.

The test uses a 1% probability of Type I error (a = 0.01). The null hypothesis can be rejected in almost all cases.

Table 1 – Training Results for ICN, First Half 1990

				Open Price				Open Price and Volume			
Q	***c*(x$_t$)**	m_X	s_X	***c*(*M*)**	m_M	s_M	a_{runs}	***c*(*M*)**	m_M	s_M	a_{runs}
1	123	-0.17%	4.3%	8	5.4%	8.7%	1.4×10^{-26}	6	7.2%	9.5%	1.5×10^{-17}
3	121	-0.18%	4.3%	10	3.5%	7.0%	3.4×10^{-14}	7	4.9%	7.7%	5.9×10^{-3}
5	119	-0.18%	4.4%	7	6.5%	7.5%	6.5×10^{-3}	6	8.0%	7.2%	7.0×10^{-4}

Table 1 provides the training results for the first half of the trading days in 1990. For the training phase, all temporal pattern clusters are significant. By comparing the m_M of both the single and multiple methods, it can be seen that the multiple method has identified better patterns in all cases. The requirements set forth by Property 1 have been met.

Table 2 – ICN Testing Results, First Half 1990 Patterns Applied to Second Half 1990

				Open Price				Open Price and Volume			
Q	***c*(x$_t$)**	m_X	s_X	***c*(*M*)**	m_M	s_M	a_{runs}	***c*(*M*)**	m_M	s_M	a_{runs}
1	124	-0.10%	5.6%	13	4.2%	9.6%	2.0×10^{-27}	12	5.2%	9.1%	2.4×10^{-27}
3	122	-0.04%	5.7%	16	1.0%	8.4%	7.8×10^{-5}	7	3.1%	10.7%	5.6×10^{-3}
5	120	-0.07%	5.7%	12	2.0%	9.6%	1.7×10^{-2}	6	4.4%	12.6%	5.5×10^{-1}

Table 2 shows the application of the patterns learned from the first half of 1990 time series to the prediction of events in the second half of 1990. All but two temporal pattern clusters meet Property 2 – the predicted events are greater than the average and the sets are statistically different. Again, the multiple time series method outperforms the single time series method.

Table 3 – Training Results for ICN, First Half 1991

				Open Price				Open Price and Volume			
Q	***c*(x$_t$)**	m_X	s_X	***c*(*M*)**	m_M	s_M	a_{runs}	***c*(*M*)**	m_M	s_M	a_{runs}
1	122	0.62%	4.8%	7	4.6%	3.6%	1.5×10^{-15}	7	5.8%	5.0%	6.6×10^{-6}
3	120	0.63%	4.9%	8	4.4%	9.5%	1.9×10^{-4}	7	10.5%	6.9%	6.3×10^{-10}
5	118	0.68%	4.9%	6	5.5%	10.1%	8.4×10^{-8}	6	9.9%	7.9%	7.5×10^{-4}

Table 3 provides the training results for the first half of 1991. All temporal pattern clusters meet Property 1. The multiple time series method outperforms the single time series method.

Table 4 – ICN Testing Results, First Half 1991 Patterns Applied to Second Half 1991

				Open Price				Open Price and Volume			
Q	***c*(x$_t$)**	m_X	s_X	***c*(*M*)**	m_M	s_M	a_{runs}	***c*(*M*)**	m_M	s_M	a_{runs}
1	125	1.1%	5.7%	13	2.1%	7.2%	$2.4x10^{-25}$	9	5.1%	8.0%	$3.7x10^{-1}$
3	123	1.2%	5.8%	7	0.5%	5.1%	$3.0x10^{-1}$	6	1.3%	15.1%	$5.4x10^{-4}$
5	121	1.0%	5.7%	7	0.9%	11.9%	$3.0x10^{-1}$	4	6.4%	11.9%	$6.9x10^{-1}$

Table 4 shows the application of the patterns learned from the first half of 1991 time series to the prediction of events in the second half of 1991. All temporal pattern clusters outperform the average “eventness”. Two temporal pattern clusters are statistically significant. The multiple time series method outperforms the single time series method.

CONCLUSIONS

In this paper, the time series data mining method is extended to multiple time series. Using temporal pattern clusters from multiple time series as a data mining tool has yielded meaningful results. Instead of modeling the time series everywhere, the method matches only when there is a high similarity between the temporal pattern cluster and the time series. To find such temporal pattern clusters, a genetic algorithm is used. Even with a complex, non-stationary time series like stock price and volume, the method uncovered patterns that were both characteristic and predictive.

REFERENCES

Bowerman, B. L., and O'Connell, R. T. (1993). *Forecasting and Time Series: An Applied Approach*, Duxbury Press, Belmont, California.

Goldberg, D. E. (1989). *Genetic algorithms in search, optimization, and machine learning*, Addison-Wesley, Reading, Mass.

Povinelli, R. J., and Feng, X. “Temporal Pattern Identification of Time Series Data using Pattern Wavelets and Genetic Algorithms.” *Artificial Neural Networks in Engineering*, St. Louis, MO, 691-696.

Sauer, T., Yorke, J. A., and Casdagli, M. (1991). “Embedology.” *Journal of Statistical Physics*, 65(3/4), 579-616.

Takens, F. “Detecting strange attractors in turbulence.” *Dynamical Systems and Turbulence*, Warwick, 366-381.