21st International Conference on Pattern Recognition (ICPR 2012)
November 11-15, 2012. Tsukuba, Japan

Statistical Modeling and Signal Selection in Multivariate Time Series Pattern Classification

Ruoqian Liu Shen Xu Chen Fang Yung-wen Liu Yi L Murphey
University of Michigan-Dearborn
{*rqliu, shenxu, chfang, ywliu, yilu*}*@umich.edu*

Dev S. Kochhar
Ford Motor Company
dkochhar@ford.com

Abstract

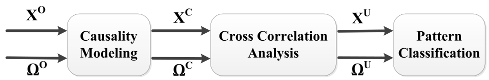
This paper presents an algorithm for selecting a compact subset of relevant signals for pattern classification problems involving multivariate time series (MTS) data. The algorithm uses a statistical causality modeling method to select relevant signals, and a correlation analysis method to remove redundant signals. The MTS signal selection algorithm along with the statistical modeling methods was evaluated through a case study of real-world driving data. From a set of 20 time series signals, the signal selection algorithm selected a subset of 9 signals that are independent and most relevant to the pattern class. We trained a driver state classification system using Random Forest(RF) with the input of 20 original signals, and another system with the selected 9 signals. The experimental results show that the system with 9 selected signals consistently performed better than the system with the original set of 20 signals.

1. Introduction

Time series data are ubiquitous and are of importance to many application problems in engineering, science, medicine, economics and entertainment [4]. Many real-world pattern classification problems involve the process and analysis of multiple variables in temporal domain. This type of problems is referred to as Multivariate Time Series(*MTS*) problems.

It has been well understood that pattern classification systems benefit from a compact set of non-redundant features that are most relevant to a given problem. Many feature selection algorithms have been developed in the research community [4], but few methods explicitly address issues involving *MTS* signal selections [3][2]. An effective approach for solving a *MTS* problem is to first select a subset of signal variables that are independent and most relevant to a given pattern classification problem, and then extract and select features generated from the selected signals.

This paper addresses the first issue: selecting a subset of signals that are independent and most relevant to a given pattern classification problem. The second issue, regarding feature extraction, can be seen in Xu et.al [8]. In the current work, we propose to use statistical causality modeling to measure signal relevance and correlation analysis to measure signal dependence. We select a subset of time series signals that have the most significant causality to the pattern class, and then eliminates redundant signals based on correlation analysis. The effectiveness of these algorithms is analyzed through a case study of driving data. From a



$\mathbf{X^O}$: Set of original signals; $\mathbf{X^C}$: Set of signals having high causality to the pattern class; $\mathbf{X^U}$: Set of mutually uncorrelated and causal signals to the pattern class.
$\mathbf{\Omega^O}$: training data for all signals in $\mathbf{X^O}$; $\mathbf{\Omega^C}$: training data for all signals in $\mathbf{X^C}$; $\mathbf{\Omega^U}$: training data for all signals in $\mathbf{X^U}$.

Figure 1. MTS signal selection system based on causality and correlation modeling

set of 20 time series signals, the causality modeling determines 12 signals that are most relevant to the driver state, and the correlation analysis eliminated 3 more signals. Experiments show that using the subset of 9 signals performed consistently better than using the original set of 20 signals. This paper is organized as follows. In Section 2 the algorithms for *MTS* causality and correlation analysis, and signal selection are presented. Section 3 presents the case study and Section 4 presents conclusions of this research.

2. Statistical Modeling and Signal Selection in *MTS* Pattern Classification

Fig. 1 illustrates the major components in the proposed signal selection system. Our focus is in the first two blocks: selecting independent and relevant signals for pattern classification.

MTS pattern classification is formally described as follows. Let $X^U(t) = \{< x_i^U(1), x_i^U(2), \ldots, x_i^U(t) > | i = 1, 2, \ldots, N^U\}$, where $x_i^U(\tau)$ is a random variable in temporal domain $\tau \in [0, t]$. The N^U random variables are synchronized in time domain. A *MTS* pattern classification problem is to build a system F such that $X^U(t) \xrightarrow{F} \Phi$ where Φ is the pattern space. The research issue in this paper is to find $X^U(t)$, a set of mutually uncorrelated and strong causal signals to the pattern class, and $X^U(t) \subset X^O(t)$, where $X^O(t)$ is the set of original signals. Under this definition, we model the output from the *MTS* pattern classification system F as a time series signal, $P(t) = \{p(1), p(2), \ldots, p(t)\}$ where $p(*) \in \Phi$. We also assume that there is no prior knowledge about the interrelationship among the signals in $X^O(t)$ and all variables are all continuous.

The proposed algorithm consists of two major components: selecting signals with highest causality with the pattern class, and eliminating redundant signals using correlation analysis.

2.1. Causality Analysis and Signal Selection

Causality Analysis is a statistical tool to analyze the one-to-one or many-to-one causal relationships in the time series data. The most popular method for testing the causality is called Granger Causality. A variable is Granger causal for another variable if the information in the former helps to improve the predictions of the latter. In this research we use Granger Causality to select time series signals that have high causal relationship with the target patterns.

For a pair of signals X and Y with length N, we use Autoregressive and Moving Average (ARMA) to model these two signals with the assumption that both X and Y are covariance stationary. The main focus is to investigate whether Y or Ys scalars can help forecast/predict X scalar.

F test is used for the statistical significance test for this comparison. In F test, after specifying a significant level, the p-value is used to evaluate the significance of causality relationship. The smaller the p-value the more significant the causality relationship is [5].

We model the problem as follows. The study is conducted on supervised learning. Assume we are given a set of training data, $\Omega = \{(X^O; P)^j\}$, where X^O is a vector of N^O available signals, P is a vector of target pattern class at each time instance within a temporal domain t, and $j = 1, \ldots, N$, N is the number of training samples in Ω. We apply Granger test to X^O and P to select N^C signals that have the most causal relation to the target class signal, P. The output from this algorithm is X^c, a set of signals that are most relevant to the pattern class P.

2.2. Correlation Analysis and Signal Selection

Cross-correlation is a well-known statistical method for estimating the strength of correlation between two sets of random data that are correlated. Let the two sets be x_i and y_i, $i =$

$1, 2, \ldots, N$. The cross-correlation of X and Y is defined as

$$r = \frac{\sum_{i=1}^{N}[(x_i - \mu_x)\cdot(y_i - \mu_y)]}{\sqrt{\sum_{i=1}^{N}(x_i - \mu_x)^2}\sqrt{\sum_{i=1}^{N}(y_i - \mu_y)^2}} \quad (1)$$

where μ_x and μ_y are the means of the corresponding sets. The cross-correlation r has values between -1 and +1, where the +1 indicates a strong positive correlation and -1 a strong negative correlation.

However a time signal can be autocorrelated and this autocorrelation can inflate the variance of cross-correlation to make the above measure no longer applicable [6]. In the past, researchers have proposed methods like resampling [7]. Similarly, we develop a data preprocessing procedure, Moving Window Average (*MWA*), to transform the original signal to the form that is no longer autocorrelated. The *MWA* transforms original x_i to a new series $\ddot{x}_j$, where $\ddot{x}_j = \frac{1}{w}\sum_{i=w(j-1)}^{jw} x_i$ for $j \leq N/w$ where w is the number of samples in a moving window.

We apply the above *MWA* process to every signal sample in Ω^C to obtain the transformed signals, and then apply pairwise correlation to the transformed signal samples. The correlation of every signal pair in X^C is measured by taking the average of correlations calculated over all the respective data samples in Ω^C.

If a signal pair has a correlation measure greater than a preset threshold, then the two signals are considered highly correlated. Correlated signals are grouped so signals in the same group are mutually correlated. From each group, we only select one signal that has the highest causality measure to add to the final signal set X^U. In addition, the signals in X^C that have no correlation to any other signals are all included in X^U.

3. Case Study

In this case study we use driving data that was collected in three concurrent streams. Driving related vehicle data was collected from the vehicle's controller area network (CAN). The second stream comprised data on the driving environment. The third data stream comprised the driver's physiological data. The objective of this case study was to use and validate a signal selection algorithm as it reflected the 'state' of alertness of the driver.

Table 1. The three classes of signals

$X_1 \sim X_8$	P	physiological
$X_9 \sim X_{12}$	E	environmental
$X_{13} \sim X_{20}$	V	vehicle

3.1. Data

The data used in this set of experiments consisted of 20 signals acquired in a vehicle driven on the highway and city streets. For ease of categorization, 8 of the 20 signals were deemed to be vehicle (V) signals, e.g. steering angle, accelerator position, etc.;4 of the 20 signals were environmental (E) signals, e.g. lane width, ambient light, etc, and the remaining 8 were physiological (P) signals, e.g. heart rate, skin conductance, etc. The list in TABLE.1 is denoted as $X^O = \{X_i^O(\tau) | i = 1, .., 20$, and $\tau = 1, \ldots, t\}$. The training data, Ω^O, contains data samples of all 20 signals acquired from 6 drivers. The test data set used to analyze system performances contains data samples acquired from 3 drivers. Every signal sample is about 1.5 hour long and the sampling rate is 10 Hz. Each data sample is labeled with an index to represent driver state $A(t)$ which reflects a drivers alertness level in time domain. This index is used as the target value in machine learning for driver state classification.

3.2. Signal selection

We first select the signals in X^O that are most relevant to the target signal $A(t)$. The selection is based on the causality measures calculated from training samples contained in Ω^O. For a signal pair, we calculate the average causality measure over all their respective data samples in Ω^O.

We select the signals that the p-values in their average causality measures with the target signals $A(t)$ are less than 0.5. As a result the following 12 signals are selected. These are listed in descending strength of causality: $X^C = \{X_{11}, X_{13}, X_{12}, X_1, X_{20}, X_7, X_{16}, X_3, X_2, X_{10}, X_5, X_8\}$.

With these 12 signals, we calculate averaged pairwise cross-correlation analysis over all data samples in Ω^O With a threshold of 0.5, we found three groups that contain highly correlated signals: $(X_7, X_8$ with r = -0.957, (X_{13}, X_{16}) with r = 0.990, and X_{11}, X_{20} with r = -0.523. From each group, we select the one signal that has a higher causality measure. This resulted in three signals, X_7, X_{13}, X_{11}, being selected.

The final set of the selected signals in X^U is $= \{X_{11}, X_{13}, X_{12}, X_1, X_7, X_3, X_2, X_{10}, X_5\}$.

3.3. Classification of Driver State

In order to show the effectiveness of the signal selection algorithm, we trained two classification systems, F^U and F^O using random forest (RF) [1]. F^U uses the 9 signals in X^U and F^O uses the original 20 signals in X^O. Both classifiers are trained with the same training data, Ω^O, and tested with the same test data, Ω^{test}, which contains the signals acquired from three different drivers.

The driver state is specified in two classes, if $A(t) > 30$, the driver alertness is "HIGH", otherwise, it is "LOW". Since the focus of this paper is on the time series signal selection, we used a simple method to extract signal features. All signals used in classification are transformed into the feature space that represents the deviation from signal mean. Using 10 trees in RF, F^O gives an error rate of 15.3%, while F^U, using less than half of the signals, produces an error rate of 12.4%. Fig. 2 shows the performances of both systems with different number of trees in RF. F^U has consistently out-performed F^O over all RF sizes.

In terms of computational time, F^O took an average of 0.4659 ms per training sample per tree, while F^U took only 0.2253 ms, on the computer with a platform of Intel i7 quad-core processor, 64 GB memory, Microsoft Windows 7, and Matlab 2009a.

4. Conclusion Remarks

In this paper, we have presented an innovative method for modeling causality and correlation in *MTS* signals, in order to select a subset of relevant signals. The proposed signal selection algorithm have been shown to be effective through a case study of driver state classification. The pattern classification system trained using the selected signals, which is only 45% of the original signal set, performed better than the system trained using the original set of the signals.

In conclusion, the proposed algorithm that utilizes Granger causality method and moving average correlation has generated a more efficient signal selection, resulting in more effective procedure for *MTS* pattern classification.

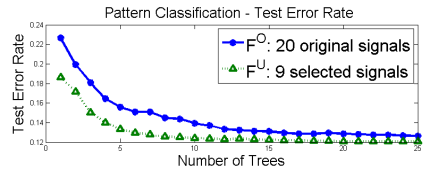


Figure 2. Comparison of system performances

References

[1] L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[2] B. Chakraborty. Feature selection and classification techniques for multivariate time series. In *Innovative Computing, Information and Control, 2007. ICICIC'07. Second International Conference on*, pages 42–42. IEEE, 2007.

[3] S. Cheema, T. Henne, U. Koeckemann, and E. Prassler. Applicability of feature selection on multivariate time series data for robotic discovery. In *Advanced Computer Theory and Engineering (ICACTE), 2010 3rd International Conference on*, volume 2, pages V2–592. IEEE, 2010.

[4] T. Fu. A review on time series data mining. *Engineering Applications of Artificial Intelligence*, 24(1):164–181, 2011.

[5] C. Granger. Some recent development in a concept of causality. *Journal of econometrics*, 39(1-2):199–211, 1988.

[6] L. Haugh. Checking the independence of two covariance-stationary time series: a univariate residual cross-correlation approach. *Journal of the American Statistical Association*, pages 378–385, 1976.

[7] J. Karvanen. A resampling test for the total independence of stationary time series: Application to the performance evaluation of ica algorithms. *Neural Processing Letters*, 22(3):311–324, 2005.

[8] S. Xu, R. Liu, D. Li, and Y. Murphey. A hybrid system ensemble based time series signal classification on driver alertness detection. In *Neural Networks (IJCNN), The 2011 International Joint Conference on*, pages 2093–2099. IEEE, 2011.