# Learning Time Series for Intelligent Monitoring *

**Stefanos Manganaris** **Doug Fisher**

*Dept. of Computer Science*
*Vanderbilt University*
*Box 1679, Station B*
*Nashville, TN 37235, U.S.A.*

Tel. (615)343-4111, Fax (615)343-8006
E-mail: `stefanos@vuse.vanderbilt.edu`

## KEY WORDS AND PHRASES

Bayesian learning, minimum description length, monitoring, time series data.

## ABSTRACT

We address the problem of classifying time series according to their morphological features in the time domain. In a supervised machine-learning framework, we induce a classification procedure from a set of preclassified examples. For each class, we infer a model that captures its morphological features, using Bayesian model induction and the minimum message length approach to assign priors. In the performance task, we classify a time series in one of the learned classes when there is enough evidence to support that decision. Time series with sufficiently novel features, belonging to classes not present in the training set, are recognized as such. We report results from experiments in a monitoring domain of interest to NASA.

## INTRODUCTION

Performance improvement in classification tasks has been a traditional area of machine learning. The objects to be classified are usually described by time-invariant attribute values. Our research is motivated by applications in temporal and sequential domains. In such domains, an object's properties often vary with time; objects are described by a *time series of values* for each attribute.

This paper focuses on learning to classify time series based on the morphological features of their behavior over time (i.e., the shape of their plots). We study univariate time series, where each object is described by one time-varying attribute. The term signature will be used synonymously with the term univariate time series.

## INDUCTION OF CLASS MODELS AND CLASSIFICATION

A set of preclassified signatures (the training examples) are presented to the learner simultaneously. Given that signatures in the same class share morphological characteristics, we design a learner that infers *class models*, represented by functions of time, that capture them. Functions in the space we consider can be decomposed into a set of polynomials and intervals, with one polynomial per interval. For example, Figure 1 shows a signature and the class model induced from it. We use a Bayesian model induction technique to find

*This research has been supported by a grant from NASA Ames (NAG 2–834).

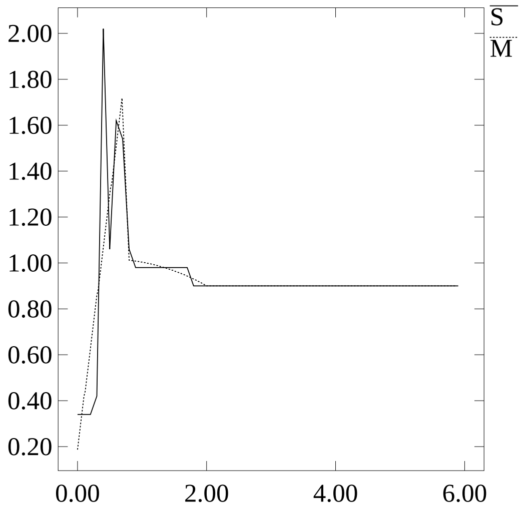


Figure 1: A signature (S) and the class model induced from it (M).

the function best supported by the training data [1]. For each class we search for the model $M$ with maximum posterior probability in light of prior information $I$ and training data $D$.

$$P(M|D,I) = P(M|I)\ \frac{P(D|M,I)}{P(D|I)} \qquad (1)$$

To assign priors, $P(M|I)$, we use the minimum message length approach [5, 6]. The negative logarithm of the prior probability of a model, $-\log_2 P(M|I)$, is equal to the theoretical minimum length of a message that describes $M$ in light of prior information $I$. Similar techniques have been used for surface reconstruction in computer vision [3], and for learning engineering models to support design [4], among other applications.

Class models are parameterized, thus the search for the best model extends in the space of parameters. We use the parameters in [3] and an additional precision parameter. Each class model has a partitioning of the time domain into a sequence of intervals. For a given interval we search through all possible families of parameterized models; we use polynomials of up to degree two, but, the method can be easily generalized. To facilitate probabilistic predictions, we assume a Gaussian noise model and independence of sampling errors. We also assume that the variance of the noise distribution is constant over an interval. For each interval we estimate the coefficients of the polynomial and the variance of the noise that maximize the posterior probability of the model.

After training, given a signature, $S$, and a set of class models, the goal is to find the model most likely to be correct for the signature in light of the prior knowledge. We treat this as a hypothesis testing problem: for each class, $C$, we compute the *evidence*, $e(C|D,I)$, that $S$ is an object of the class $C$ [2]:

$$e(C|D,I) = 10\log_{10}\left[\frac{P(C|D,I)}{P(\overline{C}|D,I)}\right] \qquad (2)$$

The probability that $S$ belongs in a class other than $C$, $P(\overline{C}|D,I)$, is computed from the posterior probabilities of all other classes and from the posterior probability of a special "novel" class. The likelihood of the "novel" class is set to zero when any of the known classes has a non-negligible likelihood. When all known classes have low likelihoods, its likelihood is computed so that it tends to one as the maximum likelihood among the known classes tends to zero. The prior of the "novel" class is set to an arbitrary low value. Under normal circumstances, the "novel" class plays no role in the computation of evidence, because of its very low posterior. Only when all known classes have low posterior probabilities, does the "novel" class become a viable alternative.

## A MONITORING APPLICATION

The Electrical Generation and Integrated Loading (EGIL) controllers at NASA monitor telemetry data from the Shuttle to detect various events that take place onboard. Typically, an event is the onset or termination of operation of an electrical device on a power bus. Each event has a signature with a set of distinguished morphological characteristics, based on which the controllers identify them. There are over two hundred different events of interest, making their accurate identification a challenging task.

Signatures are extracted from the telemetry stream whenever a change in one of the currents is detected that exceeds a preset threshold. All signatures have the same duration (6 sec. after the triggering change), and their baselines are normalized by subtracting a suitable DC value.

We have designed a set of experiments to demonstrate the feasibility of automating the classification of EGIL signatures using Calchas, a Bayesian induction system for time series data. Here we focus on the effect of training in classification performance. We use the percentage of correctly classified instances as our dependent measure of learning. In our experiments there are ten classes of signatures for ten different events; the average number of signatures per class is about 65. Our current implementation only handles univariate time series. There are many three-dimensional signatures in the EGIL domain; in these cases we ignore two of the phases.

In each run, we train Calchas on an equal number of randomly selected signatures from each class. We then evaluate its performance on the remaining signatures. We vary the amount of training by using different training set sizes. The results with training sizes of one and eight are summarized in the *confusion matrix* shown in Table 1. Each entry of the table shows the percentage of test signatures, in the class labeling the row, that were classified by Calchas to the class labeling the column. The top row for each class was obtained after training Calchas with one signature per class; the bottom row was obtained with training sizes of eight. All percentages are averaged over twenty runs; the standard deviations are shown. For example, with a training set of eight signatures, an average of 74% of the Wcs test signatures were correctly classified as Wcs, and 1% and 25% were incorrectly classified as Rcr and Novel, respectively. In general, the matrix diagonal indicates the percentage of correct classifications. Entries corresponding to Un1 and Un3 are for signatures whose actual class was unknown.

Table 1 indicates that increased training results in higher classification accuracies. A notable exception seems to be the Gal class, where training with eight signatures results in significantly lower accuracy than training with one signature. We suspect that Gal is an example of a disjunctive concept: there is more than one pattern of morphological features describing signatures in the class. Calchas is currently unable to handle disjunctive concepts; training on multiple patterns for a class results in a confused class model and thus lower classification accuracy.

Beyond the practical advantages of automatic vs. manual monitoring, a Bayesian learning approach offers the following technical advantages. It provides a principled way of discerning the distinguishing features of a signature from measurement noise; it mitigates the problem of overfitting. Calchas provides an estimate of the confidence in each classification. When more than one classification is supported by roughly the same evidence, we can recognize this fact and report it, as opposed to making an arbitrary classification. Similarly, we can report when no classification is supported with significant evidence. Signatures with sufficiently novel features, belonging to classes not present in the training set, are recognized as such and are classified as Novel; potentially costly classification mistakes are avoided.

Table 1: Classification of EGIL signatures (assumed univariate—see text).

| CLASS | | PHO | VAC | AWCS | H2O | CAB | PRP | WCS | TPS | RCR | GAL | NOVEL |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PHO | 1 | 40±29 | | | 1±4 | | | | 2±7 | | 57±29 | |
| | 8 | 96±5 | | | | | | | | | 4±5 | |
| VAC | 1 | | 68±32 | | | | | | | | | 32±32 |
| | 8 | | 93±2 | | | | | | | | | 7±2 |
| AWCS | 1 | | | 92±22 | 5±22 | | | | | | | 3±1 |
| | 8 | | | 96±2 | | | | | | | | 4±2 |
| H2O | 1 | 2±9 | | | 98±9 | | | | | | | |
| | 8 | | | | 100±0 | | | | | | | |
| CAB | 1 | | | | | 79±17 | | | | | | 22±17 |
| | 8 | | | | | 90±16 | | | | | | 10±16 |
| PRP | 1 | | | | | | 98±4 | | | | 2±4 | |
| | 8 | | | | | | 98±2 | | | | 2±2 | |
| WCS | 1 | | | | | | | 52±28 | | 1±0 | | 47±28 |
| | 8 | | | | | | | 74±4 | | 1±0 | | 25±4 |
| TPS | 1 | 7±14 | | | | | | | 76±17 | 3±5 | 15±11 | |
| | 8 | 8±7 | | | | | | | 85±8 | | | 7±7 |
| RCR | 1 | | | | | | 2±0 | | | 97±1 | | |
| | 8 | | | | | | 3±0 | | | 97±0 | | |
| GAL | 1 | 2±1 | | | | | | | | | 98±0 | |
| | 8 | 22±40 | | | | | | | | | 78±40 | |
| UN1 | 1 | 46±10 | | | 13±2 | | 12±2 | | 3±2 | 2±1 | 22±9 | 2±0 |
| | 8 | 55±4 | | | 12±1 | | 12±3 | | 1±1 | 3±1 | 15±7 | 2±0 |
| UN3 | 1 | 9±5 | | | 20±4 | | 30±4 | | 8±4 | 4±1 | 9±3 | 20±0 |
| | 8 | 18±2 | | | 15±1 | | 29±2 | | 11±2 | 4±1 | 3±2 | 20±0 |


## ACKNOWLEDGMENTS

We would like to thank Phil Laird, Robert Shelton, Ron Saul, Wayne Iba, and Deepak Kulkarni for providing the data we used in the experiments and information on the domain.


## REFERENCES


[1] Peter Cheeseman. On finding the most probable model. In J. Shrager and P. Langley, editors, *Computational Models of Discovery and Theory Formation*, chapter 3, pages 73–95. Morgan Kaufmann, 1990.

[2] E. T. Jaynes. Probability theory: The logic of science. Advance fragments of this book are made available by the author. Contact etj@wuphys.wustl.edu, 1993.

[3] Edwin P.D. Pednault. Some experiments in applying inductive inference principles to surface reconstruction. In *Proc. of the Eleventh International Joint Conf. on Artificial Intelligence*, pages 1603–1609, 1989.

[4] R. Bharat Rao and Stephen C-Y. Lu. Learning engineering models with the minimum description length principle. In *Proc. of the Tenth National Conf. on Artificial Intelligence*, pages 717–722, 1992.

[5] J. Rissanen. A universal prior for integers and estimation by minimum description length. *The Annals of Statistics*, 11(2):416–431, 1983.

[6] C. S. Wallace and P. R. Freeman. Estimation and inference by compact coding. *The Journal of the Royal Statistical Society*, 49(3):252–265, 1987.