

# MIXTURE OF DESIGNER EXPERTS FOR MULTI-REGIME DETECTION IN STREAMING DATA

*Evan Kriminger, José Príncipe\**

University of Florida  
Dept. of Electrical and Computer Engineering  
Gainesville, FL 32611-6200  
Email: ekriminger@ufl.edu

*Choudur Lakshminarayan*

HP Labs  
Palo Alto, California  
Email: Choudur.Lakshminarayan@hp.com

## ABSTRACT

Real-time streaming data takes on distinct visible patterns, known as regimes, as a result of changing external influences. Regimes corresponding to hazardous states, such as turbulent flow in oil pipelines or patients experiencing heart arrhythmias, must be identified quickly and accurately by on-line detection algorithms. In this paper, we propose a modification to the mixture of experts framework, which is traditionally used to model piecewise stationary time series. Our proposed modification allows experts to produce features specific to their designated regimes, rather than being limited to prediction error. This approach provides the flexibility to update the mixture modularly as new regimes emerge without the burden of retraining the entire mixture, as is typical in traditional classifiers. Our approach is tested on flow rate data from an oil and gas application, as well as detecting heart arrhythmias from electrocardiogram (ECG) signals. It outperforms traditional classification approaches both in terms of error rate and detector delay.

**Index Terms**— Detection, streaming data, mixture of experts

## 1. INTRODUCTION

In the problem of regime detection, or the related problem of anomaly detection [1], the goal of an algorithm is to label each sample of a time series as belonging to one of various regimes. When a data set includes several distinct regimes, it is difficult to capture it with a single model, or select a single set of discriminative features across regimes. As the algorithm has to detect different regimes, interference effects can lead to slow learning and poor generalizability [2]. Furthermore, in the on-line setting, traditional classifiers relying on joint distributions between classes must completely retrain, even if changes occur to only a single class.

In this setting, it is natural to decompose the problem into simpler sub-problems. The mixture of experts (ME) [2] is one

approach consisting of an ensemble of experts, each one being an adaptive model with parameters tuned specifically to a single regime. The experts are linked by a gating network which stochastically selects the expert which best models the input. The gate selects the expert based on hard competition over their single step prediction errors. The parameters of the expert that wins the competition are adjusted with the error. At the outset, the user need only have an approximate idea of the number of regimes present in the time series. The experts are initialized with small random parameters. Thus, a complicated time series can be modelled by distributing responsibility amongst an ensemble of experts.

The ME is useful for modelling data composed of distinct regimes, and it is our goal to modify it for use in the classification problem of regime detection. This modification requires associating labelled training data with each expert, rather than the unsupervised way in which experts train in the standard ME formulation. Furthermore, the expert and gating network must be redefined to work with features more general than single step prediction error.

In this paper, we propose a new framework for the mixture of experts. This paper is organized as follows: In Section 2 we present ME for the detection problem with a novel modification to allow greater freedom in the choice of experts. In Section 3 we provide an illustrative example of the modified ME on real-world oil pipeline data, by describing three experts used in our framework. In Section 4, we apply the modified ME approach to classification of electrocardiogram signals to detect heart arrhythmia in patients.

## 2. MODIFIED ME FOR REGIME DETECTION

In regime detection, the classifier has access to labelled training samples from each regime. Therefore, to modify the ME for classification, the experts are trained a priori. Each expert then has a known association with a particular regime. Then during testing, unlabelled samples are assigned to the regime represented by the expert which wins the competition. A second modification must be made, allowing for more general

\*Thanks to HP Labs for funding.

experts. For classifying regimes it may be difficult or unnecessary to model the time series. Rather, we wish to extract a characterizing feature specific to that regime. For example, in an oscillatory regime it may be best to work with a spectral feature, while in a highly specialized application the features may be designed with considerable expert knowledge.

To satisfy these requirements, we redefine the expert to be an entity which produces a measure of association between a given testing sample and its training set. In the case of the standard ME, the association measure is the single step prediction error, which characterizes how well the current sample fits the training data as captured in the expert’s model parameters. Other examples of association measures include nearest neighbor distances, correlations, and likelihood ratios. Competition amongst experts will then proceed over their association measures.

Note that we cannot directly compete over the association measures in a fair manner. Each expert,  $i$ , generates an association measure  $a_i$ , which may not even share the range of the association measures of other experts. For instance, if one expert produces a rank statistic (positive integer), while another produces a probability in  $[0, 1]$ , the latter will always win the competition. To overcome this issue, we build an empirical probability density from the association measures of the training samples for each regime. For regime  $i$ , this is  $f(a|i)$ . When a new sample arrives we compute the association measure for each class,  $a_i$ , and the classification decision is then given by the maximum likelihood class

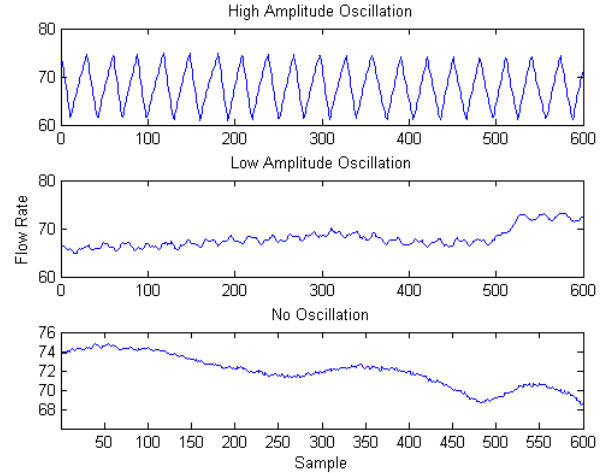
$$\operatorname{argmax}_i f(a_i|i).$$

The likelihood provides a measure of expert accuracy that is not dependent on the particular methods used by the expert.

The benefit of the modified ME framework is the decision making through competition. Competition across experts provides modularity, which is imperative in the streaming data setting. The experts can be designed separately, which allows new experts to be inserted into the mixture at any time without requiring changes in any of the other experts. Additionally, the decision system is not sensitive to the number of experts, nor their particular features. If a regime no longer appears, the competition is unaffected. Small likelihoods from all experts indicate the appearance of a new regime, signalling that a new expert should be implemented. In existing classification approaches, the features are selected a priori, hoping that they will be discriminatory for all regimes; present and future. Therefore, complete retraining is required when new regimes appear.

### 3. APPLICATION: OIL FLOW-RATE DATA FROM OFF-SHORE BORE-WELLS

We applied our modified ME framework to oil flow rates from sensors placed in a bore-well. The measurements are from a



**Fig. 1:** The high amplitude oscillation (top), low amplitude oscillation (middle), and no oscillation (bottom) regimes.

single well, recorded by a single sensor located at the surface of the sea, at 30 second intervals. This data is composed of many regimes, but here we will consider three of them (Figure 1). The first regime, characterized by high amplitude oscillation (HAO), consists of triangular oscillations with a period of 30 samples. The second characterized by low amplitude oscillation (LAO) consists of noisy oscillations at the same frequency superimposed on a normal time series. The third regime is simply a time series with no oscillation (NO).

Since the HAO oscillations are nonstationary, yet have very distinguishable structure, we chose to use a time embedding to capture the periodicity explicitly without the need for a model. The time embedding is a mapping,

$$n \mapsto (x_n, x_{n-\tau}, \dots, x_{n-[m-1]\tau}) = \mathbf{x}_n,$$

where  $n$  is the time,  $x_n$  is the time series value at  $n$ ,  $m$  is the embedding dimension, and  $\tau$  is the time delay. The justification for this embedding comes from Takens’ time delay embedding theorem [3], which states that the phase space of a high dimensional dynamical system can be reconstructed from the time embedded vectors obtained from a time series of observations from this system. The vectors  $\mathbf{x}_n$  preserve dynamical invariants of the dynamical system as long as  $\tau$  and  $m$  are chosen appropriately [4]. For this test  $m = 4$  and  $\tau = 3$ . The time embedding captures the triangular oscillations with only a few time lagged samples, which leads to a detector that can respond quickly to transitions between regimes. The training data for this regime was time embedded, forming a trajectory. For unlabelled streaming samples, we construct the appropriate time embedding vectors and compare a trajectory segment of these vectors of length  $T$  with the trajectory segments of the training set. This comparison is made with the correntropy function [5]. For the streaming length  $T$  trajec-

tory segment at time  $n$ ,  $\mathbf{Y}_n$ , and an arbitrary length  $T$  training set trajectory,  $\mathbf{X}$ , the correntropy function,  $\nu$ , is denoted by

$$\hat{\nu}(\mathbf{X}, \mathbf{Y}_n) = \frac{1}{T} \sum_{i=0}^{T-1} G_{\sigma} (\|\mathbf{X}(i) - \mathbf{Y}_n(i)\|)$$

where  $\mathbf{Y}_n(i)$  is the  $i^{\text{th}}$  element of the trajectory segment, and  $G_{\sigma}$  is the Gaussian kernel with bandwidth  $\sigma$ . If  $\mathcal{F}$  is the set of training trajectory segments, then the association measure we use for this expert is

$$a = 1 - \min_{\mathbf{X} \in \mathcal{F}} \hat{\nu}(\mathbf{X}, \mathbf{Y}_n).$$

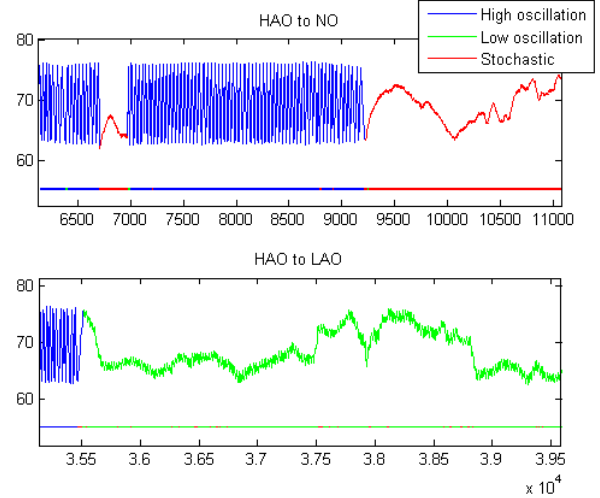
The NO regime is stationary only on relatively small time scales, and is very nicely modelled by a simple linear adaptive filter, with filter coefficients tuned by the least-mean-squares (LMS) learning rule

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \eta e(n) \mathbf{x}(n),$$

where  $\mathbf{w}_n$  is the length  $p$  vector of filter coefficients at time  $n$ ,  $\eta$  is the learning rate,  $e(n)$  is the prediction error at time  $n$ , and  $\mathbf{x}(n)$  is the vector  $[x(n), x(n-1), \dots, x(n-p+1)]^T$ . Since this is a predictive model, we use the magnitude of the one-step prediction error as an association measure. When this expert wins the competition, the error is used to tune the filter coefficients with LMS. This expert can therefore track gradual changes of the regime.

The LAO regime is composed of noisy oscillations of approximately the same frequency as the triangular oscillations, superimposed on a time series similar to the NO regime. We apply the short-time Fourier transform (STFT) to the streaming data, and use the power spectral value at the oscillation frequency as the association measure for this regime. We used a Hamming window that was 32 samples in length. The window slides through the time series each shifted by one time unit, producing one feature value for each streaming sample. The spectral component of the oscillation frequency is a good characterizing feature of the noisy oscillations, as we only know that LAO is composed of low amplitude oscillations of a known frequency. The time series is first bandpass filtered to remove trends and noise.

We tested the modified ME with the aforementioned experts on a test set composed of 26,709 samples ( $\tilde{24}$  hours) of real oil pipeline data, with 1000 training samples allocated to each regime. A segment of the test set and labels given by the modified ME detector are seen in Figure 2. We selected 12 regime transitions (4 transitions to each regime) and measured the detector delay and the error rate. The detector delay is defined as the number of samples that elapse after the change of a regime until the new regime is detected. The error rate is the ratio of detection errors to total samples, following this detector delay period. The total error rate and detector delay, averaged over the 12 transitions, is shown in Table 1.



**Fig. 2:** Two segments of the testing data, color-coded to illustrate the true regime label. The color bar below the time series represents the regime decision generated by the modified ME detector.

|             | Error Rate | Average Delay |
|-------------|------------|---------------|
| Modified ME | 3.22%      | 8.4           |
| STFT        | 3.37%      | 14.2          |
| State Space | 6.26%      | 3.4           |

**Table 1:** The performance the modified ME versus classification based on a spectral feature and competition amongst state space detectors. Average delay is in the units of samples.

Some statistics of the delay time for the ME detector in each regime are given in Table 2.

We compare the modified ME with a maximum likelihood classification approach using only the STFT spectral feature and with the standard mixture of experts, with each expert using the same time embedding feature as the HAO, as discussed earlier. The STFT coefficient is a discriminatory feature, which is reflected by its low error rate of 3.37%. However, as we consider a window of 32 past samples, its detector delay is high at 14.2 samples. The state space detectors consider fewer past data points and therefore have a shorter detection delay at 3.4 samples. This model is better suited to the dynamic oscillatory regime, and hence its error rate is higher at 6.26%. The modified ME combines the strengths of both methods and boasts the lowest error rate, 3.22%, as well as a short delay of 8.4 samples.

#### 4. APPLICATION: ECG BEAT DETECTION

We consider another important regime detection problem, the automatic classification of cardiac beat types from electrocardiogram (ECG) signals. There are many beat types that may be identified from the ECG, including normal beats and a vari-

| Modified ME Delay Stats |           |            |        |
|-------------------------|-----------|------------|--------|
| Regime                  | Tri. Osc. | Noisy Osc. | Stoch. |
| Mean                    | 14.5      | 3.5        | 7.5    |
| Min                     | 1.0       | 1.0        | 5.0    |
| Max                     | 31.0      | 6.0        | 11.0   |
| Std.                    | 14.0      | 2.9        | 2.6    |

**Table 2:** An evaluation of the delay times for the modified ME detector for transitions to each of the three regimes.

ety of irregularities such as *premature ventricular contraction* and *atrial premature contractions* as seen in Figure 3. The real-time detection of these abnormal beats is potentially life-saving in a clinical setting, as beat irregularities are used to diagnose heart arrhythmias, myocardial infarctions, ventricular hypertrophy, and other health problems. The ECG signal of a normal beat has a distinctive structure consisting of five successive deflections in amplitude; the so-called P, Q, R, S, and T waves. Attributes such as the time duration of the QRS complex and the R amplitude are useful defining characteristics of the different types of beats. Existing beat detection methods first extract these time and amplitude dependent features. Once the features are obtained, a variety of methods are applied to make a detection decision. Some techniques include decision trees, linear discriminant analysis, artificial neural networks, hidden Markov models, and template matching [6].

ECG beat detection is a fundamentally challenging problem. In an ECG, the morphology of a beat type will not remain fixed. Intra-patient, and inter-patient variability is quite pervasive and pose modelling challenges. Therefore, a detector must accommodate beats which differ from the training set. Secondly, each patient will have characteristic beat morphologies and a well-trained detector may fail when operating on a new subject. For these reasons, adaptive methods are desirable, as they can quickly learn to accommodate new patients and their unique beat morphologies.

Our modified ME approach satisfies the needs of an ECG classifier. To accommodate a new patient, annotated training data is added as it is made available. If this patient has a type of beat not previously seen, a small training set from this regime is obtained as in the case of the oil flow, and a new expert is added to the mixture.

A mixture of experts framework was used for ECG beat detection in [6], which provides grounds for comparison. In this approach, decision responsibility is split between global and local experts. Global experts are classifiers trained on a large training set composed of many patients. To adapt to a single patient’s unique characteristics, a local expert classifier is trained on a small segment of that patient’s annotated ECG data. A gating network dynamically weights the classification decisions of the global and local expert.

We use the well-known MIT/Beth Israel Hospital (BIH) Arrhythmia Database available on the PhysioBank archives

[7]. This database is composed of 2-channel ECG signals. For each patient, there is 30 minutes of ECG, with a sampling rate of 360 Hz. However, we downsampled the signal to 180 Hz. We use the upper channel signal, which was acquired through a modified limb lead (MLL) with electrodes attached to the chest. The MIT/BIH Arrhythmia Database provides ECG signals with each beat labelled by an expert cardiologist. There are 19 types of beat annotations, as well as 22 types of non-beat annotations, including rhythm annotations.

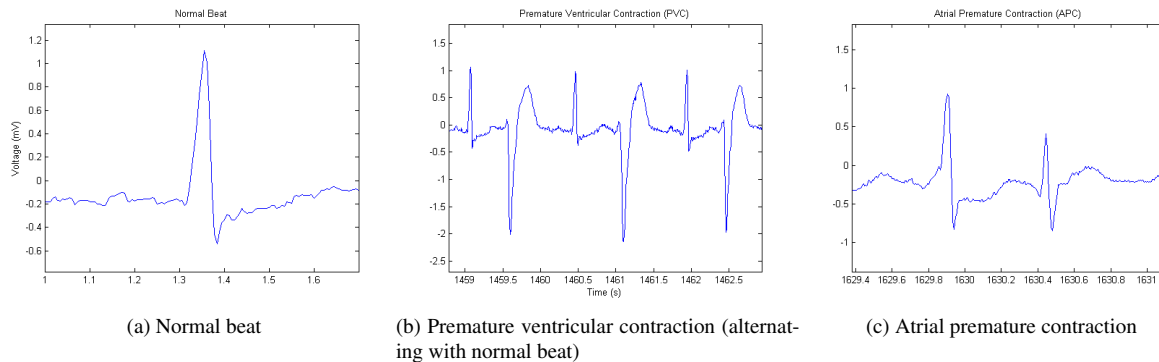
Testing protocol is carried out in accordance with American Association of Medical Instrumentation (AAMI) protocol [8] and follow the testing procedure of Hu et al. [6] for comparison. The goal is to place each beat in one of four more general categories, into which the beat annotations fall.

To employ the modified ME we must select the most natural feature for each known regime. In this case the signal structure of the QRS complex is the essential defining component of the beat of interest. Following Hu et al. we use the annotations to find the R wave peak (highest local amplitude point), and embed that R wave peak amplitude and the 6 adjacent samples to its left and 10 adjacent samples to its right into a 19 dimensional feature vector. We add a 20<sup>th</sup> dimension in the form of the RR interval, which is the time from the previous R wave to the current R wave. We subtract the 19 amplitude components of the feature vector by the first component, to align the ECG beats.

In our study, we consider 20 patients for training. The first 2.5 minutes of each patient’s ECG signal are used to build the training set. The feature vector is extracted from each beat in the training set, and the features are grouped by their annotation index. We ensure that no more than 200 vectors are stored for any one beat type, and randomly select 200 if more than that many exists in the training set. There are 16 beat types overall in this set of data and therefore we have 16 experts, each one storing up to 200 feature vectors. This is much less training data than that used by Hu et al., which included 13 full 30-minute ECG records, in addition to the 20, 2.5 minute segments we used, without the 200 sample limit we set.

Our association measure between a testing feature vector and the training set for a given beat type is the minimum weighted Euclidean distance between the testing vector and all of the training vectors of that beat type. We use leave-one-out on each training set to compute these Euclidean distances to build the empirical distributions. The expert that produces the maximum likelihood is selected as the winner, and its corresponding beat type is then mapped to one of the four AAMI groupings. We test our algorithm on the remaining 27.5 minutes of each patient’s ECG signal for 10 total patients.

The results are seen in Table 3. We compare to the results of Hu et al. The modified ME outperforms Hu et al. in 6 of 10 trials, but has a slightly less average success rate. Note that the modified ME is not tailored to ECG, and can perform on par with highly specified algorithms.



**Fig. 3:** ECG of three sample beat types.

| Detection Success Rate (%) |             |           |
|----------------------------|-------------|-----------|
| MIT/BIH Record #           | Modified ME | Hu et al. |
| 200                        | 87.3        | 81.0      |
| 201                        | 97.8        | 95.4      |
| 202                        | 99.8        | 71.7      |
| 203                        | 84.4        | 87.4      |
| 205                        | 98.7        | 97.1      |
| 207                        | 86.7        | 88.3      |
| 208                        | 66.3        | 90.5      |
| 209                        | 99.8        | 98.8      |
| 210                        | 96.4        | 93.4      |
| 213                        | 60.3        | 91.9      |
| avg.                       | 87.8        | 89.5      |

**Table 3:** Percent of successful detections, for the modified ME and for the mixture of experts in [6].

## 5. CONCLUSION

The ME is a useful framework for modelling time series with distinct regimes. However, the experts are limited to predictive models, making the approach unsuitable for many real world datasets which are noisy, nonlinear, and nonstationary. We extended the ME by allowing an experts to be from a larger class and other association measures. Competition is based on the likelihoods of the association measure. In the oil and gas example, the importance of being able to use a variety of experts is seen, as the oscillatory regimes vary from deterministic to stochastic regimes, requiring different approaches. Furthermore, the freedom of choice in experts allows quick detection of change points, with minimal detector delay. We also tested the modified ME on an ECG beat detection application. The flexibility in our method was demonstrated as the modified ME was able to contend with a specialized formulation of the mixture of experts for ECG, proposed by Hu et al. Our algorithm performed better in 6 of 10 trials, while also requiring much less training data.

## 6. REFERENCES

- [1] V. Chandola, A. Banerjee, and V. Kumar, “Anomaly detection: a survey,” *ACM Computing Surveys*, vol. 41, no. 3, July 2009.
- [2] Robert A. Jacobs, Michael I. Jordan, Steven J. Nowlan, and Geoffrey E. Hinton, “Adaptive mixtures of local experts,” *Neural Computation*, vol. 3, pp. 79–87, March 1991.
- [3] Floris Takens, “Detecting strange attractors in turbulence,” *Lecture Notes in Mathematics*, vol. 898, no. 1, pp. 366–381, 1981.
- [4] Henry D. I. Abarbanel, *Analysis of Observed Chaotic Data*, Springer, 1996.
- [5] Ignacio Santamaría, Puskal P. Pokharel, and José C. Príncipe, “Generalized correlation function: definition, properties, and application to blind equalization,” *IEEE Transactions on Signal Processing*, vol. 54, no. 6-1, pp. 2187–2197, 2006.
- [6] Yu Hen Hu, Surekha Palreddy, and Willis J. Tompkins, “A patient-adaptable ECG beat classifier using a mixture of experts approach,” *IEEE Trans. on Biomedical Engineering*, vol. 44, no. 9, pp. 891–900, 1997.
- [7] A. L. Goldberger et al., “PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals,” *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.
- [8] R. Mark and R. Wallen, “AAMI-recommended practice: testing and reporting performance results of ventricular arrhythmia detection algorithms,” *Tech. Rep. AAMI ECAR*, 1987.