An Ensemble of HMMs for Cognitive Fault Detection in Distributed Sensor Networks

Manuel Roveri, Francesco Trovò

An Ensemble of HMMs for Cognitive Fault Detection in Distributed Sensor Networks

Manuel Roveri and Francesco Trovò

Dipartimento di Elettronica, Informazione e Bioingegneria
Politecnico di Milano, Milano, Italy
{manuel.roveri,francesco1.trovo}@polimi.it

Abstract. Distributed sensor networks working in harsh environmental conditions can suffer from permanent or transient faults affecting the embedded electronics or the sensors. Fault Diagnosis Systems (FDSs) have been widely studied in the literature to detect, isolate, identify, and possibly accommodate faults. Recently introduced cognitive FDSs, which represents a novel generation of FDSs, are characterized by the capability to exploit temporal and spatial dependency in acquired datastreams to improve the fault diagnosis and by the ability to operate without requiring a priori information about the data-generating process or the possible faults. This paper suggests a novel approach for fault detection in cognitive FDSs based on an *ensemble of Hidden Markov Models*. A wide experimental campaign on both synthetic and real-world data coming from a rock-collapse forecasting system shows the advantages of the proposed solution.

1 Introduction

Distributed sensor networks represent an important and valuable technological solution to monitor and acquire data from an environment, a cyber physical-system or a critical infrastructure system (e.g., water, gas, electric or transport networks). Unfortunately, distributed sensor networks working in (possibly harsh) real-working conditions may suffer from permanent or transient faults, thermal drifts or ageing effects affecting both the embedded electronic boards and the sensors. Fault Diagnosis Systems (FDSs), which are able to detect, identify, isolate and possibly accommodate faults, have been widely studied in the related literature [1, 2]. These systems revealed to be particularly effective in several different application domains provided that a priori information about the system model or the possible faults is (at least partially) available.

Recently, cognitive FDSs have been introduced, representing a novel and promising approach for fault diagnosis [3–7]. The distinguishing features of these FDSs are the capability to work without a priori information about the process or the possible faults (which are learned directly from data) and the ability to exploit temporal and spatial dependencies in the acquired datastreams to improve the diagnosis of faults. Fault detection is obviously a key aspect in cognitive FDSs and, among the available techniques in the cognitive FDSs literature, we focus on

the change detection test (CDT) based on a Hidden Markov Model (HMM) suggested in [8]. This HMM-CDT revealed to be particularly effective in detecting faults affecting real-world monitoring applications (e.g., environmental monitoring [8] or water transport network [9]) in the scenario of networked intelligent embedded sensors or distributed sensor networks. Unfortunately, the detection abilities of this CDT are influenced by the initial conditions of the HMM training algorithm, possibly leading to suboptimal HMM parameters estimations [10].

In this paper, we propose an *ensemble-of-models* approach for cognitive fault detection based on HMMs, leading to a novel Ensemble HMM-based CDT (EHMM-CDT). More specifically, the proposed EHMM-CDT relies on a set of HMMs configured on the same training set, but with different randomly generated initial conditions for the HMM training algorithm. The ensemble approach aims at weakening the dependency of the HMM-CDT from the random initial conditions, hence providing better detection abilities. Interestingly, the possibility to combine different models to improve the generalization ability of a single model has been widely studied in the literature (mainly in the regression and classification scenarios) and effectively applied to different application fields [11–13]. Recently, ensembles of models have been successfully considered in time series prediction [14, 15] and on-line missing data reconstruction [16]. Moreover, an ensemble of HMMs within a Bayesian framework for parameter estimation has been derived in [17], while an algorithm for training HMM in the ensemble framework was presented in [18]. Remarkably, for the first time in the literature, in this paper we suggest an ensemble of HMMs for fault detection and we describe and evaluate two different aggregation mechanisms. In the experimental section, the proposed EHMM-CDT has been compared with the HMM-CDT: the advantages provided by the proposed solution have been evaluated both on synthetic and real datasets coming from a rock-collapse forecasting system.

The paper is organized as follows: Section 2 describes the problem formulation, while Section 3 summarizes the HMM-CDT suggested in [8]. Section 4 details the proposed ensemble approach for cognitive fault detection, while Section 5 shows the experimental results.

2 Problem Formulation

Let us consider a distributed sensor network acquiring scalar measurements - or datastreams - from a time invariant dynamic system whose model description is unavailable. Without loss of generality we focus on the relationship between two datastreams; the extension to multiple or multivariate datastreams is straightforward, e.g., see [6].

The unknown relationship $\mathcal{P}$ between the two streams of data is modeled by a Single-Input Single-Output (SISO) linear time invariant (LTI) predictive model belonging to a family $\mathcal{M}$ parametrized in $\theta \in \mathcal{D} \subset \mathbb{R}^p$, where $\mathcal{D}$ is a C^1 compact manifold in $\mathbb{R}^p$ and $p \in \mathbb{N}$ is the dimension of θ. For instance, SISO or Single-output/Multiple-input linear predictive models [19] and Reservoir Computing

Networks [20] are viable options for the choice of $\mathcal{M}$. In this paper, we consider linear SISO one-step-ahead predictive models:

$$\hat{y}(t|\theta) = f\left(t, \theta, u(t), \ldots, u(t-\tau_u), \ldots y(t-1), \ldots, y(t-\tau_y)\right), \quad \forall t \in \mathbb{N} \quad (1)$$

where $\hat{y}(t|\theta) \in \mathbb{R}$ is the prediction provided at time t, $f : \mathbb{N} \times \mathbb{R}^p \times \mathbb{R}^{\tau_u} \times \mathbb{R}^{\tau_y} \to \mathbb{R}$ is the approximating function in predictive form, $u(t) \in \mathbb{R}^m$ and $y(t) \in \mathbb{R}$ are the model input and output at time t, respectively, and τ_u and τ_y are the orders of the input and output, respectively.

Under mild assumption on $\mathcal{P}$, $\mathcal{M}$ and the estimation procedure of θ over a finite dataset $Z_N = \{(u(j), y(j))\}_{j=1}^N$, it holds that [19]:

$$\lim_{N \to \infty} \sqrt{N} \Sigma_N^{-\frac{1}{2}} (\hat{\theta} - \theta^o) \sim \mathcal{N}(0, I_p) \quad (2)$$

where $\hat{\theta}$ is the estimated parameter vector, θ^o is the parameter optimizing the structural risk in $\mathcal{M}$, Σ_N is a properly defined covariance matrix and I_p is the identity matrix of order p. Interestingly, Equation (2) assures that, given a sufficiently large N, the estimated parameter vector $\hat{\theta}$ follows a multivariate Gaussian distribution with mean θ^o and covariance matrix Σ_N. Moreover, this result remains valid even when $\mathcal{P} \notin \mathcal{M}$, i.e., a model bias $||\mathcal{M}(\theta^o) - \mathcal{P}|| \neq 0$ is present. This justifies the use of LTI models, even when the dynamic system under investigation $\mathcal{P}$ is non-linear.

3 The HMM-based Change Detection Test

The aim of this section is to summarize the key points of the HMM-CDT suggested in [8], representing the core element of the proposed EHMM-CDT. Under the formulation presented in Section 2, the use of a HMM ruled by a mixture of Gaussian (GMM) over a sequence of parameters $\hat{\theta}$s is the natural solution to model the functional relationship presented in Equation (1) in the parameter space. More formally, a HMM [21] can be defined as a tuple $\mathcal{H} = (S, A, \pi)$, where $S = \{S_1, \ldots, S_s\}, s \in \mathbb{N}$ is the indexed set of the states, each of which has an emission probability defined by a GMM, $A \in \mathbb{R}^{s \times s}, A = [a_{ij}], \sum_{i=1}^s a_{ij} = 1 \ \forall j \in \{1, \ldots, s\}$ is the transition matrix, i.e., a_{ij} is the transition probability from state i to state j and $\pi \in [0,1]^s$, with $\sum_{i=1}^s \pi_i = 1$, is the initial distribution probability over S.

The idea underlying the HMM-CDT is to analyse the statistical behaviour of $\hat{\theta}$s over time: when the statistical pattern of the estimated parameters does not follow what learned during an initial training phase, a change in the relationship is detected. Thus, the HMM-CDT relies on an initial parameter vector sequence $\Theta_L = (\hat{\theta}_1, \ldots, \hat{\theta}_L)$ estimated on a faulty-free training set Z_M, where $\hat{\theta}_j$ is estimated on a subsequence $Z_{j,N} = \{(x(h), y(h))\}_{h=j-N+1}^j$ of the training sequence, $j \in \{N, \ldots, M\}, L = M - N + 1$ (i.e., overlapping windows of length N). The parameter sequence Θ_L is used to train a HMM $\hat{\mathcal{H}}$, aiming at capturing the statistical behaviour of the estimated parameter vectors $\hat{\theta}_t$s in nominal

conditions (without any change/fault). Then, during the operational phase, the statistical affinity between the estimated parameter vectors $\hat{\theta}_t$s and $\hat{\mathcal{H}}$ is assessed by looking at the HMM loglikelihood: changes in the relationship expressed by Equation (1) are detected by inspecting the likelihood of $\hat{\theta}_t$ with respect to $\hat{\tilde{\mathcal{H}}}$. More specifically, given a sequence $\Theta_{t,k} = (\hat{\theta}_{t-k+1}, \ldots, \hat{\theta}_t)$, i.e., the last k consecutive parameter vectors at time t, a HMM $\hat{\mathcal{H}}$ provides an estimates on how well the sequence follows $\hat{\mathcal{H}}$ by means of the computation of the loglikelihood:

$$l(\hat{\mathcal{H}}, \Theta_{t,k}) = \log \Pr(\hat{\theta}_{t-k+1}|\hat{\mathcal{H}}) + \sum_{h=t-k+1}^{t-1} \log \Pr(\hat{\theta}_{h+1}|\hat{\theta}_h, \hat{\mathcal{H}}). \qquad (3)$$

If the relationship does not change over time, the loglikelihood $l(\hat{\mathcal{H}}, \Theta_{t,k})$ is comparable with the one computed during the training phase on a validation set $Z_O = \{(x(j), y(j))\}_{j=M+1}^O$. Otherwise, in case of variations in the relationship, $l(\hat{\mathcal{H}}, \Theta_{t,k})$ decreases, since $\Theta_{t,k}$ is no more compatible with the statistical model characterized by $\hat{\mathcal{H}}$. Further details about the HMM-CDT can be found in [8, 6].

4 Ensemble of HMM for Fault Detection

One of the key aspects of the HMM-CDT is the estimation of a HMM on Θ_L, aiming at characterizing the nominal state. The state-of-the art training algorithm for HMMs is the Baum-Welch (BW) algorithm [22], which aims at finding the maximum likelihood estimates of the HMM parameters given Θ_L. Nonetheless, BW algorithm does not provide any guarantee about the convergence to the global maximum of the likelihood function [10]. Since the BW algorithm requires a random initialization of the parameters, different HHMs can be obtained by repeating the training phase with different randomly initialized parameters on the same training sequence. Hence, a viable solution to weaken the effect of the initialization procedure on the HMM-CDT would be to repeat the BW algorithm and select the HMM guaranteeing the largest likelihood on a validation set. Unfortunately, as pointed out in [17], this solution may lead to overfit the training sequence Θ_L (and this is particular evident for reduced training sets), leading to false positive detections (false alarms) during the operational phase.

In this work, we suggest an ensemble approach for fault detection (EHMM-CDT) based on a set $E = \{\mathcal{H}_1, \ldots, \mathcal{H}_e\}, e \in \mathbb{N}$ of HMMs, where e is the ensemble cardinality. The main point of the proposed solution is that the HMMs $\mathcal{H}_i \in E$ are trained on the same training set Z_M with different initialization points of the BW algorithm. This ensemble approach allows to weaken the influence of the initial conditions of the HMM training algorithm, providing a better generalization ability and, consequently, better detection performance.

The proposed EHMM-CDT is detailed in Algorithm 1. To characterize the nominal state we rely on a training set Z_M, assumed to be change-free. We estimate the parameter vectors $\hat{\theta}_j$s on overlapping windows of N data $Z_{t,N}, t \in \{N, \ldots, M\}$. Without loss of generality, we assume a step of one sample between

---

Algorithm 1 EHMM-CDT algorithm

1: **Data:** Training set Z_M, Validation set Z_O, C, $\mathcal{A}$;
2: **Results:** Detection time T;
3: Estimate the sequence Θ_L from Z_M;
4: **for** $i \in \{1, \ldots, e\}$ **do**
5:     Estimate HMM $\mathcal{H}_i$ using Θ_L by considering a random initialization point;
6: **end for**
7: Built the ensemble $E = \{\mathcal{H}_1, \ldots, \mathcal{H}_e\}$;
8: Estimate the sequence $\Theta_{O,O-M+1}$ from Z_O;
9: **for** $h \in \{M + k, \ldots, O\}$ **do**
10:     **for** $i \in \{1, \ldots, e\}$ **do**
11:         Compute $l(\mathcal{H}_i, \Theta_{h,k})$ as in Equation (3);
12:     **end for**
13:     Compute $\mathcal{A}(h) = \mathcal{A}(l(\mathcal{H}_1, \Theta_{h,k}), \ldots, l(\mathcal{H}_e, \Theta_{h,k}))$;
14: **end for**
15: Compute $l_{\min}$ as in Equation (4);
16: **while** a new couple $(x(t), y(t))$ is available at time t **do**
17:     Estimate $\hat{\theta}_t$ by using $Z_{t,N}$;
18:     Built the sequence $\Theta_{t,k} = (\hat{\theta}_{t-k+1}, \ldots, \hat{\theta}_t)$;
19:     **for** $i \in \{1, \ldots, e\}$ **do**
20:         Compute $l(\mathcal{H}_i, \Theta_{t,k})$ as in Equation (3);
21:     **end for**
22:     Compute $\mathcal{A}(t) = \mathcal{A}(l(\mathcal{H}_1, \Theta_{t,k}), \ldots, l(\mathcal{H}_e, \Theta_{t,k}))$;
23:     **if** $\mathcal{A}(t) \leq l_{\min}$ **then**
24:         **return** $T \leftarrow t$;
25:     **end if**
26: **end while**
27: **return** $T \leftarrow \emptyset$;

---

two overlapping windows; larger steps could be considered as well (e.g., to reduce the computational complexity of the HMM training phase). The estimation procedure is performed by means of a suitable minimization technique, e.g., Least Square estimation on the empirical risk (see [19] for details), and let Θ_L be the sequence of estimated parameter vectors on Z_M, as shown in Section 3 (Line 3).

We build the ensemble $E = \{\mathcal{H}_1, \ldots, \mathcal{H}_e\}$ by repeating e times the training of a HMM on Θ_L, with random initial conditions for the BW algorithm (Line 5). Note that, since Z_M is assumed to be change-free, the ensemble E aims at modeling the statical behaviour of $\mathcal{P}$ in nominal conditions.

One of the key aspects of ensemble methods is the definition of the aggregation mechanism $\mathcal{A}$, that, in this case, specifies how to aggregate the loglikelihoods of the ensemble elements $\mathcal{H}_i \in E$ to provide the ensemble output. In this work, we considered two different aggregation mechanisms, i.e., $\mathcal{A} \in \{\mathcal{A}_{mean}, \mathcal{A}_{\min}\}$:

$$\mathcal{A}(t) = \mathcal{A}_{mean}(l(\mathcal{H}_1, \Theta_{t,k}), \dots, l(\mathcal{H}_e, \Theta_{t,k}))) = \sum_{i=1}^{e} \frac{l(\mathcal{H}_i, \Theta_{t,k})}{e}$$

$$\mathcal{A}(t) = \mathcal{A}_{\min}(l(\mathcal{H}_1, \Theta_{t,k}), \dots, l(\mathcal{H}_e, \Theta_{t,k}))) = \min_{i \in \{1,\dots,e\}} l(\mathcal{H}_i, \Theta_{t,k})$$

where $\mathcal{A}_{mean}$ computes the average value of the loglikelihoods, while $\mathcal{A}_{\min}$ takes into account their minimum. The last step of the EHMM-CDT training phase is the computation of the threshold $l_{\min}$ on a validation set Z_O(Line 15), which is computed as follows:

$$l_{\min} = \bar{l} - C \left[\bar{l} - \min_{h \in \{M+k,\dots,O\}} \mathcal{A}(h) \right] \tag{4}$$

where $\bar{l} = \frac{\sum_{h=M+k}^{O} \mathcal{A}(h)}{O}$ and $C > 1$ is a user-defined coefficient factor.

During the operational phase, when the aggregated loglikelihood decreases below $l_{\min}$, a change in the statical behaviour of $\mathcal{P}$ is detected by the EHMM-CDT. More in details, as soon as a new sample $(u(t), y(t))$, $t > M + O$ is available, the algorithm estimates a new parameter vector $\hat{\theta}_t$ on $Z_{t,N}$ (Line 17). The likelihoods for the ensemble of HMMs, i.e., $l(\mathcal{H}_i, \Theta_{k,t}) \forall \mathcal{H}_i \in E$, are computed (Line 20) and then aggregated, according to the aggregation mechanism $\mathcal{A}$ (Line 22). Afterwards, the aggregated likelihood $\mathcal{A}(t)$ is compared with the threshold $l_{\min}$ to asses if a change in $\mathcal{P}$ is detected (Line 23). In case of a change, an alarm is raised and the detection time $T = t$ is returned (Line 24), otherwise the algorithm keeps on monitoring data coming from $\mathcal{P}$.

While ensemble approaches are generally able to increase the generalization ability of single models, they are characterized by an increased computational complexity, that in this case scales linearly with the number of HMMs e. Two comments arise: 1) The most time consuming part of the EHMM-CDT refers to the HMMs training (the loglikelihood computation is much more lighter than training). Interestingly, the training phase is performed only once during the initial configuration of the cognitive FDS, while during the operational life only the likelihoods are computed. In addition, in scenarios where networked embedded systems are operating, the training of HMMs could be performed in a centralized high-powerful unit, leaving only the computation of likelihoods directly at the low-power distributed units of the network. 2) In scenarios where the fault detection ability is a relevant activity, the increase in the complexity induced by the ensemble approach is well compensated by the decrease in false and missed alarms and detection delays, as shown in the experimental section.

5   Experimental Results

To evaluate the performance of the proposed EHMM-CDT we considered both synthetic (Application D1) and real data (Application D2), coming from a rock collapse forecasting system deployed in Northern Italy. In both applications, we

considered a faulty free dataset divided into training (to create E), validation (to learn $l_{\min}$) and test sets (to evaluate the performance). In the last one we injected five different kinds of abrupt permanent fault at time $\bar{t}$:

- **A1-A3**: $y(t) = y(t) + \Delta_a \cdot a_{\max}, \forall t > \bar{t}, \Delta_a \in \{0.1, 0.2, 0.3\}$ (additive fault);
- **M** : $y(t) = y(t) \cdot (1 + \Delta_m), \forall t > \bar{t}, \Delta_m = 0.3$ (multiplicative fault);
- **S**: $y(t) = y(\bar{t}), \forall t > \bar{t}$ (stuck-at fault);

where $a_{\max} = \max_{j \in \{1,\ldots,M\}} y(j) - \min_{j \in \{1,\ldots,M\}} y(j)$.

As regards the model family $\mathcal{M}$, we considered SISO LTI AutoRegressive with eXogenous input (ARX) models, whose orders are chosen through a model selection procedure (i.e., minimizing the mean square one-step-ahead prediction error on a validation set). The HMMs are configured by considering batches of $N = 100$ samples, loglikelihood window length $k = 10$ and exploring $s \in \{3, \ldots, 6\}$ and the number of models in each GMM $\in \{1, 2, 4, 8, 16, 32\}$. The cardinality of the ensemble was set to $e = 30$. Finally, the parameter vectors estimation is performed with the least square method.

To evaluate the detection ability of what proposed the following figures of merit have been considered:

- false positive (FP) rate: fraction of the experiments where the method detected a change before it actually appears;
- false negative (FN) rate: fraction of the experiments where the method did not detected a change;
- detection delay (DD): the number of samples necessary to detect a change;

The proposed method is compared with the HMM-CDT, where a single HMM was considered (the one with largest loglikelihood on a validation set among those considered for the ensemble).

APP D1: Synthetic data. The data for the synthetic application have been generated from the following non-linear model:

$$y(t) = \sin(a^T \cdot (y(t-1), y(t-2)) + b \cdot x(t-1)) + \eta(t)$$

where $a \in [0,1]^2, b \in [0,1]$, $\eta(t) \sim \mathcal{N}(0, \sigma^2), \sigma \in \{0.01 a_{\max}, 0.05 a_{\max}\}$. Each experiment lasts 6125 samples (i.e., 3268 for training, 817 for validation and 2040 for testing): faults are injected at time $\bar{t} = 5105$. Results are averaged over 1000 runs.

Figure 1 shows the experimental results in the case of APP D1 and $\sigma = 0.05 a_{\max}$. Curves are obtained by considering values of C ranging from 1 to 5.1, where 5.1 is the largest value of C for which $FN \leq 0.1$ for all the faulty scenarios. Interestingly, in all the considered scenarios the ensemble approach is able to improve the detection ability (in terms of both FP and DD) of the single HMM-CDT. These curves show that the performance of EHMM-CDT cannot be achieved by the HMM-CDT by simply tuning the parameter C. In addition, in the faulty scenario **M** where FNs are present, the proposed EHMM-CDT behaves better than HMM-CDT even with respect to this figure of merit (see legend of Fig.1.d).

Table 1. Detection results for the single and ensemble HMM-CDT approaches from applications D1 (with different σs) and D2

| | | HMM-CDT | | | EHMM-CDT | | | | | |
| | | | | | $\mathcal{A}_{mean}$ | | | $\mathcal{A}_{\min}$ | | |
		FN	FP	DD	FN	FP	DD	FN	FP	DD
$\sigma = 0.01 a_{\max}$	**A1**	0.000	0.007	21.614	0.000	0.007	19.419	0.000	0.006	19.717
	A2	0.000	0.007	17.415	0.000	0.007	15.320	0.000	0.006	15.638
	A3	0.000	0.007	15.396	0.000	0.007	13.724	0.000	0.006	13.998
	M	0.000	0.007	37.057	0.000	0.007	35.355	0.000	0.006	35.794
	S	0.000	0.007	16.186	0.000	0.007	15.335	0.000	0.006	15.648
$\sigma = 0.05 a_{\max}$	**A1**	0.000	0.007	67.939	0.000	0.009	63.745	0.000	0.009	64.708
	A2	0.000	0.007	47.465	0.000	0.009	42.625	0.000	0.009	43.281
	A3	0.000	0.007	39.676	0.000	0.009	34.734	0.000	0.009	35.006
	M	0.077	0.007	185.989	0.057	0.009	161.401	0.055	0.009	166.641
	S	0.000	0.007	45.139	0.000	0.009	42.413	0.000	0.009	42.781
Rialba	**A1**	0.340	0.000	314.879	0.000	0.000	172.100	0.280	0.000	234.500
	A2	0.000	0.000	152.100	0.000	0.000	148.900	0.000	0.000	154.400
	A3	0.000	0.000	119.700	0.000	0.000	30.700	0.000	0.000	93.800
	M	0.040	0.000	163.250	0.000	0.000	153.200	0.000	0.000	162.100
	S	0.000	0.000	81.500	0.000	0.000	82.300	0.000	0.000	89.600

Experimental results presented in Table 5 refer to FN, FP and DD with fixed values of C. To ease the comparison we set C in the EHMM-CDT and HMM-CDT as the lowest values guaranteeing $FN = 0$ and $FP \leq 0.01$ in APP D1 with $\sigma = 0.01$ (i.e., $C = 4.814$ for HMM-CDT, $C = 4.629$ for EHMM-CDT with $\mathcal{A}_{mean}$ and $C = 4.917$ for EHMM-CDT with $\mathcal{A}_{\min}$). Remarkably, both EHMM-CDT aggregations provide lower DD than HMM-CDT. Interestingly, the *mean* aggregation mechanism generally outperforms the *minimum* one, since the mean is less influenced than minimum by outliers, which generally induce false alarms. Nonetheless, by comparing these results with those provided in Figure 1, we may also comment that the *mean* aggregation does not provide better results for each value of C: hence the *minimum* is a viable solution too.

APP D2: Rialba Data. We consider data coming from a real-world distributed sensor network for rock-collapse forecasting [23, 24]. This dataset is available at [25]. In particular, we analyzed two temperature datastreams composed by 5303 samples (3000 for training, 1000 for validation and 1303 for testing), coming from two network units. The sampling period is 5 minutes. We injected faults (**A1**-**A3**,**M**,**S**) at $\bar{t} = 4695$. Experimental results in Table 5 are obtained by averaging the results of 50 runs. Interestingly, the results on the real-world datasets are in line with those of the synthetic one: the ensemble approach provides lower FN and DD for most of the considered scenarios.
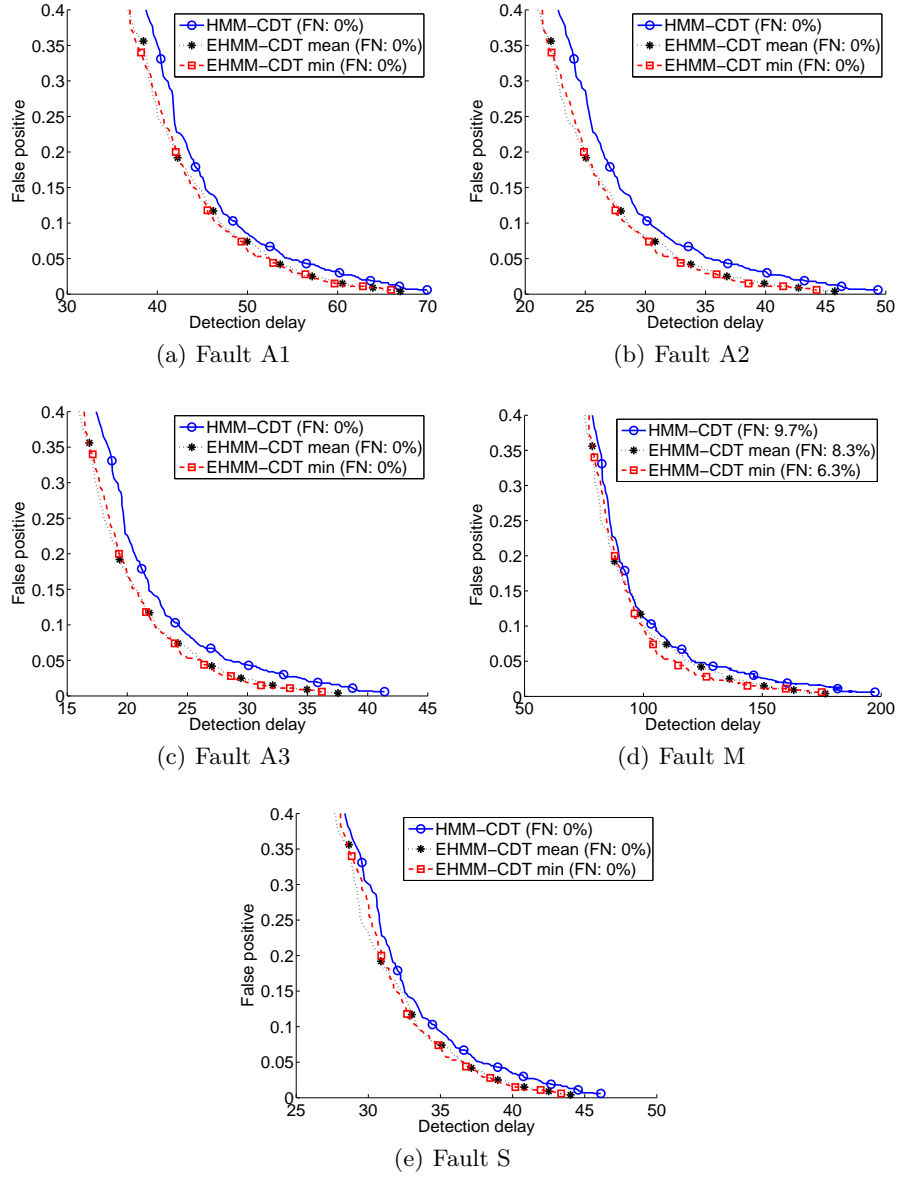
(a) Fault A1

(b) Fault A2

(c) Fault A3

(d) Fault M

(e) Fault S

Fig. 1. Experimental results showing the relationship between FP and DD in application D1 with $\sigma = 0.05$, where $C \in [1, 5.1]$

6   Conclusions

We presented an novel approach for cognitive fault detection based on an ensemble of HMMs. The distinguishing feature of the proposed solution is the ability to improve the generalization ability (in terms of detection performance) of a single HMM-CDT by considering an ensemble of HMMs trained on the same training set, but with different initialization points. The effectiveness of the proposed solution has been tested on both synthetic and real datasets coming from a real-world distributed sensor network for rock-collapse forecasting.

References

1. R. Isermann. *Fault-diagnosis systems.* Springer, 2006.
2. V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S.N. Kavuri. A review of process fault detection and diagnosis: Part i: Quantitative model-based methods. *Computers & chemical engineering*, 27(3):293–311, 2003.
3. M.A. Demetriou and M.M. Polycarpou. Incipient fault diagnosis of dynamical systems using online approximators. *Automatic Control, IEEE Transactions on*, 43(11):1612–1617, 1998.
4. J. Farrell, T. Berger, and B.D. Appleby. Using learning techniques to accommodate unanticipated faults. *Control Systems, IEEE*, 13(3):40–49, 1993.
5. A.B. Trunov and M.M. Polycarpou. Automated fault diagnosis in nonlinear multivariable systems using a learning methodology. *Neural Networks, IEEE Transactions on*, 11(1):91–101, 2000.
6. C. Alippi, S. Ntalampiras, and M. Roveri. A cognitive fault diagnosis system for distributed sensor networks. *Neural Networks and Learning Systems, IEEE Transactions on*, 24(8):1213–1226, 2013.
7. C. Alippi, M. Roveri, and F. Trovò. A "learning from models" cognitive fault diagnosis system. In *Artificial Neural Networks and Machine Learning–ICANN 2012*, pages 305–313. Springer, 2012.
8. C. Alippi, S. Ntalampiras, and M. Roveri. An hmm-based change detection method for intelligent embedded sensors. In *Neural Networks (IJCNN), The 2012 International Joint Conference on*, pages 1–7. IEEE, 2012.
9. C. Alippi, G. Boracchi, V. Puig, and M. Roveri. An ensemble approach to estimate the fault-time instant. In *Intelligent Control and Information Processing (ICICIP), 2013 Fourth International Conference on*, pages 836–841. IEEE, 2013.
10. A.P. Dempster et al. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal statistical Society*, 39(1):1–38, 1977.
11. Z.H. Zhou. *Ensemble methods: foundations and algorithms.* CRC Press, 2012.
12. A. Krogh and P. Sollich. Statistical mechanics of ensemble learning. *Physical Review E*, 55(1):811, 1997.
13. M.P. Perrone and L.N. Cooper. When networks disagree: Ensemble methods for hybrid neural networks. Technical report, DTIC Document, 1992.
14. J. Wichard and M. Ogorzalek. Time series prediction with ensemble models. In *Proceedings of IJCNN 2004*, pages 1625–1629. Citeseer, 2004.
15. G.P. Zhang. Time series forecasting using a hybrid arima and neural network model. *Neurocomputing*, 50:159–175, 2003.

16. C. Alippi, S. Ntalampiras, and M. Roveri. Model ensemble for an effective on-line reconstruction of missing data in sensor networks. In *Neural Networks (IJCNN), The 2013 International Joint Conference on*, pages 1–6. IEEE, 2013.

17. D.J.C. MacKay. Ensemble learning for hidden markov models. Technical report, Technical report, Cavendish Laboratory, University of Cambridge, 1997.

18. R.I.A. Davis and B.C. Lovell. Comparing and evaluating hmm ensemble training algorithms using train and test and condition number criteria. *Formal Pattern Analysis & Applications*, 6(4):327–335, 2004.

19. L. Ljung. *System identification*. Wiley Online Library, 1999.

20. B. Schrauwen, D. Verstraeten, and J. Van Campenhout. An overview of reservoir computing: theory, applications and implementations. In *European Symposium on Artificial Neural Networks*. Citeseer, 2007.

21. R.J. Elliott, L. Aggoun, and J.B. Moore. *Hidden Markov Models*. Springer, 1995.

22. L.E. Baum, T. Petrie, G. Soules, and N. Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *The annals of mathematical statistics*, pages 164–171, 1970.

23. C. Alippi, R. Camplani, C. Galperti, A. Marullo, and M. Roveri. An hybrid wireless-wired monitoring system for real-time rock collapse forecasting. In *Mobile Adhoc and Sensor Systems (MASS), 2010 IEEE 7th International Conference on*, pages 224–231. IEEE, 2010.

24. C. Alippi, R. Camplani, C. Galperti, A. Marullo, and M. Roveri. A high-frequency sampling monitoring system for environmental and structural applications. *ACM Transactions on Sensor Networks (TOSN)*, 9(4):41, 2013.

25. isense project website: `http://www.i-sense.org/`, mar 2014.