IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received July 8, 2018, accepted July 19, 2018, date of publication July 26, 2018, date of current version August 20, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2859922

K-PdM: KPI-Oriented Machinery Deterioration Estimation Framework for Predictive Maintenance Using Cluster-Based Hidden Markov Model

ZHENYU WU[1], HAO LUO[2], YUNONG YANG[2], PENG LV[2], XINNING ZHU[2], YANG JI[1,2], AND BIAN WU[3]

[1]Engineering Research Center of Information Network, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China
[2]Key Laboratory of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China
[3]Cancer Center, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430022, China

Corresponding author: Bian Wu (bian.wu@outlook.com)

ABSTRACT Explosive increase of industrial data collected from sensors has brought increasing attractions to the data-driven predictive maintenance for industrial machines in cyber-physical systems (CPSs). Since machinery faults are always caused by performance deterioration of components, learning the deteriorating mode from observed sensor data facilitates the prognostics of impeding faults and predicting the remaining useful life (RUL). In modern CPSs, several key performance indicators (KPIs) are monitored to detect the corresponding fine-grained deteriorating modes of industrial machines. However, the overall deterioration estimation and RUL prediction based on these KPIs with various patterns have been a great challenge, especially without labels of deteriorating index or uninterpretable of root causes. In this paper, we proposed *K-PdM*, a cluster-based hidden Markov model for the machinery deterioration estimation and RUL prediction based on multiple KPIs. The method uncovers the fine-grained deteriorating modes of machines through each unlabeled KPI data and learns a mapping between each deteriorating KPI index and RULs. Accordingly, an overall deterioration estimation and RUL prediction of machine are able to be achieved based on the combination of each KPI's deterioration estimation. Moreover, a set of interpretable semantic rules are setup to analyze the root cause of performance deterioration among KPIs. An experimental application is proposed to demonstrate its applicability based on the PHM08 data sets. The obtained results show their effectiveness to predict the RULs of machines.

INDEX TERMS Prognostics and health management, remaining life assessment, hidden Markov models (HMMs), time series analysis.

I. INTRODUCTION

Predictive maintenance becomes important and indispensable for machinery maintenance in modern Cyber-Physical Systems (CPSs) [1]–[3]. In real industrial scenarios, plenties of Key Performance Indicators (KPIs) of machines have been monitors by Internet of Things (IoT) infrastructure, and each KPI is time series data measuring one dimension of the life-cycle of machines. As Figure 1 shown, the *total_temperature*, *static_pressure* and *total_pressure* at HPC outlet collected from turbofan engine are three major KPIs measuring the HPC's working behavior from healthy to failure [4]. Mining the progression pattern of performance deterioration from monitored KPIs could facilitate assessing the health, predicting the RULs of machines and localizing the root-cause KPI of failure.

A rich body of literature exist on data-driven predictive maintenance methods, such as artificial neural network [5], support vector machine (SVM) [6], and Hidden Markov Model (HMM) [7]. All these methods rely on the assumption that the health index (HI) [8] and RULs have been labeled, and a direct mapping function is learnt from raw sensor data to represent features of HIs or RULs. However, in practice, some challenges exist that: 1) RULs are labeled as linear countdown time-to-failure curves [9], which is too straight-forward to represent actual health status; 2) actual faults occasionally occur because of periodic maintenance, which

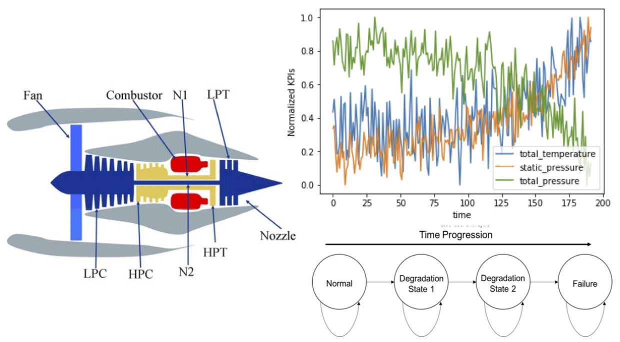

FIGURE 1. Run-to-failure process described by multiple KPIs of machine.

results in insufficient and inaccurate labels of failures. Thus, direct learning methods mapping from features of raw data to RUL are not effective; 3) the real deteriorating process could be usually separated into several phases (healthy state, sub-healthy state and failure state) which are unknown in previous, and each deteriorating phase follows different patterns. Thus, each phase should be learned individually. Consequently, a new method which could estimate the uncertain deteriorating phase from KPIs is necessary, and the method could also facilitate estimating an indirect health status to predict RUL.

Reference [10] has surveyed several indirect RUL prediction methods based on learning functional mapping between health index (HI) and RUL. The performance for RUL prediction depends on both mapping functions which are excellent research topics. The approaches for health index (degradation) estimation and modeling are of great importance for PHM. Since the operating conditions represent usage load profile as pointed out in publications [11], [12], their sequence is important. Therefore, the estimation of the health index can be improved by taking into account the sequence of operating conditions. References [13] and [14] uses a Encoder-Decoder model based on Recurrent Neural Networks (RNNs) to generate embeddings for multivariate time series subsequences, and the reconstruction errors are modeled as the HI to represent the health status of machine. These methods calculate HIs mainly based on domain knowledge or statistical errors between healthy and degraded samples, while the internal statistical transitions of performance deterioration based on physics-of-failure are rarely taken into consideration. Some papers use HMMs for deteriorating pattern learning and future state estimation. Reference [15] calculates remaining life time of turbofan engines based on features extracted by Artificial Neural Networks (ANN) and estimation by HMM. Reference [16] models the progression of deterioration into several consecutive states, and each state can be modeled as an HMM. The paper proposes a Multi-branch HMM (MB-HMM) model to recognize each state and estimate the RUL. However, these methods are based on assumptions that the number of deteriorating state are initialized by either expert's knowledge, clustering algorithms or identified by AIC [17], BIC [18] criteria, but not learnt from the data itself. It results in a fact that the transitional patterns are not able to be learned from observations simultaneously. Meanwhile, [19] proposes pattern-based HMM (pHMM) by using line segmentation, clustering and extended Viterbi algorithm of HMM to inference the high-level state from time series and predict the future value/trends. In our previous study [20], an unsupervised degradation estimation framework is proposed to discover high-level degradation states and the transitional patterns are learned to estimate machinery deterioration. Nevertheless, the method mainly focuses on fine-grained machinery deterioration estimation based on individual KPI, which are not able to estimate the overall health status based on multiple KPIs, and moreover, it did not consider in establishing a mapping between health status and RUL to facilitate overall machinery RUL prediction.

In this paper, we propose *K-PdM* framework which is a two-phase cluster-based Hidden Markov Model (cHMM) by leveraging clustering method, pHMM and MB-HMM in a uniform framework. The paper is an extension of our previous work and the contributions are summarized as below:

1) To estimate the performance deteriorating mode of machines based on unlabeled fine-grained KPI data, an extended Viterbi (eViterbi) algorithm based on *BestSymbol* function is proposed to learn the symbolic observation sequence and deteriorating states sequence simultaneously in a unsupervised way;
2) To estimate the overall deteriorating status and predict the RUL of machines, an HMM-based classifier is designed to identify the fine-grained deteriorating state of machines. Then, the mapping between the fine-grained health status based on each KPI and overall RUL is trained based on a linear regression model, and thus the overall RUL is able to be predicted;
3) To diagnose the root-cause KPI of the performance deterioration, a semantic rule-based reasoner is proposed to identify which KPI causes the overall performance deterioration;

The remainder of the paper is organized as follows: Section II gives the problem statement and notations. Section III introduces the K-PdM framework in details. In section IV, we use PHM08 challenge datasets to validate the feasibility of our model. Finally, we draw a concrete conclusion in section V.

II. PROBLEM STATEMENT

A. HMMS AND NOTATIONS

An HMM is a probabilistic model in which the state process are latent and can be only revealed through an observation process. Denoting h_t as the hidden state at time t and there are N hidden states $\mathbf{S} = \{S_1, S_2, \ldots, S_N\}$, where $h_t \in \mathbf{S}$. The observation sequence $\mathbf{O} = \{o_1, o_2, \ldots, o_T\}$, where T is the length of the sequence and o_t is the observation at time t. The elements of an HMM are defined as $\lambda = \{A, B, \pi\}$, where A refers to the transition probability from one state to another,

B refers to output probability of an observation generated by a certain state, and π refers to the initial probability of a certain state.

Given a sequence of observations $\mathbf{O}$, the optimal probability is the probability of HMM λ generating $\mathbf{O}$, along a state sequence $\mathbf{S} = \{s_1, s_2, \ldots s_N\}$, and it is computed as:

$$P(O, S|\lambda) = \delta_t(s_n) = b_{s_n}(o_j) \sum_{s_n \in S} \delta_{t-1}(s_{n-1}) a_{s_{n-1}, s_n} \tag{1}$$

and

$$\delta_1(s_1) = \pi_1 b_{s_1}(o_1) \tag{2}$$

B. PROBLEM FORMULATION

The core problem for the deterioration estimation and RUL prediction of machines are based on the assumption that the machines are working at different modes from health to failure, and the deterioration are caused by one or more KPIs' degradation. Denoting KPI_i, where $1 \leq i \leq n$, as the $i'th$ KPI time series to monitor the performance deterioration of machine. Each KPI_i indicates one dimension of fine-grained deterioration, and combination of KPIs reflects the overall HI and RULs of machine. Thus, the fundamental problems can be categorized as: (1) learning deteriorating mode of the system based on each KPI_i; (2) identification of current working state by combining different KPIs' deteriorating states; (3) prediction of remaining useful life of machines and root-cause analysis.

1) FINE-GRAINED DETERIORATING MODE LEARNING

Each KPI indicates the fine-grained deterioration, and the deteriorating process of the system on each KPI measurement could be modeled as an HMM. The discovery of deteriorating states could be modeled as finding the optimal hidden state sequence that generates the KPI observations. Additionally, this problem could be formulated as Viterbi decoding model of HMMs:

$$\delta_t(i) = \max_j(\delta_{t-1}(j) a_{ji}) b_i(o_t) \tag{3}$$

When the algorithm reaches the last time point n, we obtain all optimal probabilities: $\delta_n(i)$. By comparing all of them, and backtracking the largest one, this algorithm obtains the optimal state sequence.

Equation (1) would be adequate if the observation sequence are given. However, the observation sequence is unknown from given KPI time series, so it needs to be initialized and learned from time series itself. In our work, we formulate the problem as learning the observation and state sequence simultaneously. The formulation could be modeled as follows, according to Equation (1):

$$\delta_t(i) = \max_{d,j}(\delta_{t-d}(j) a_{ji}) b_i(C_t) \tag{4}$$

Similarly, in our algorithm, when $\delta_t(i)$ is computed, it implies a "new observation" is added to the observation sequence. However, here the observation C_t is a cluster of continuous KPI value starting at $t-d+1$ ending at t, instead of a single value. Since, the optimal observation sequence is unknown, d could not be determined beforehand. So $\delta_n(i)$ is computed by scanning all possible previous optimal probability and maximum result is chosen. Finally, we obtain the optimal observation sequence $\{C_1, C_2, \ldots, C_m\}$ and the corresponding state sequence $\{h_1, h_2, \ldots, h_m\}$.

To further learn the internal evolutionary pattern of each deteriorating states, every deteriorating states could be modeled as a left-right Gaussian HMM λ_i $(1 \leq i \leq K)$, where K is the number of deteriorating states and the output probability of each HMM follows Gaussian distributions. Baum-Welch [9] could be used to learn the parameters of λ_i

2) DETERIORATION ESTIMATION

Determination of the "current" fine-grained deteriorating states based on each KPI simply involves calculation of the log-likelihood probability for all the HMMs using O_i. Given the KPI observation sequence, current state with the largest log-likelihood probability according to λ_k is the estimated degradation state. The identification problem could be formulated as:

$$DS(\lambda_k^*) = \arg\max_{1\leq k\leq n} \log P(o_1, o_2, \ldots, o_t|\lambda_k) = \arg\max_{1\leq k\leq n} \log(\sum_{j=1}^{n} b_j(o_t) \sum_{i=1}^{n} \delta_{t-1}^{(k)}(i) * a_{i,j}^{(k))}) \tag{5}$$

Once each fined-grained deteriorating states have been determined on each KPI measurement, a combination of rules can be predefined to further identify the overall deteriorating state with its root-cause KPIs. As figure 2 shown, the rules divide the overall machinery deterioration into several levels based on each fine-grained KPI's deteriorating states, and each level represents the degree of overall deterioration.

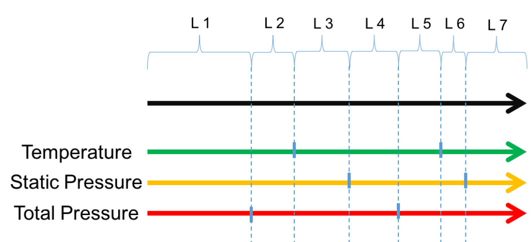

FIGURE 2. The overall machinery deterioration level division and identification based on each fin-grained KPI deteriorating states according to predefined rules.

3) RUL PREDICTION

The problem could be categorized as a regression learning and prediction tasks based on combination of fine-grained KPIs' deteriorating states. For each KPI measurement KPI_i $(1 \leq i \leq n)$, the fine-grained deteriorating states have been estimated by K HMM λ_i $(1 \leq i \leq K)$. The log-likelihood probability at time t is calculated for λ_i, and it is

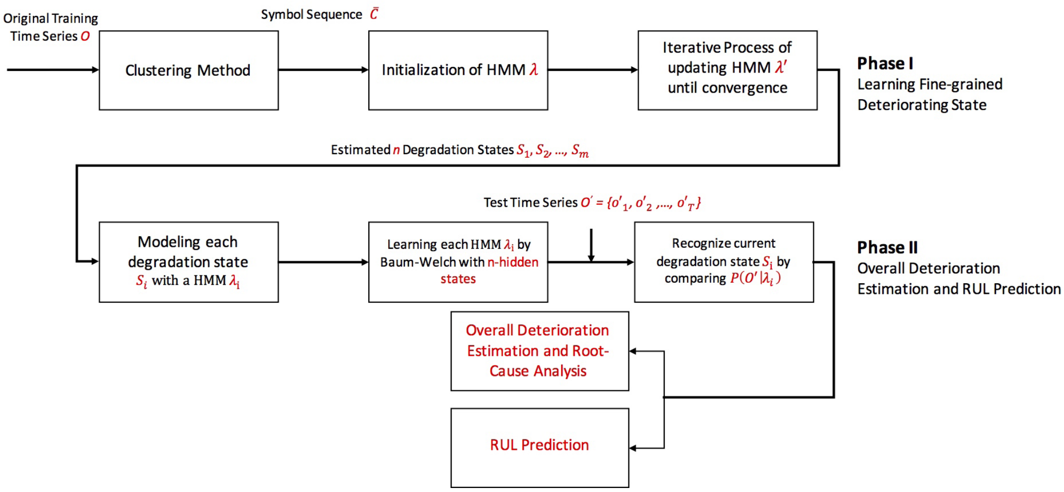

FIGURE 3. The K-PdM framework based on two-phased cHMM for machinery deterioration estimation and RUL prediction.

denoted as $L_t^{(i)}$ $(1 \leq i \leq K)$. Accordingly, the log-likelihood probabilities are able to represent the possibility of each fine-grained deteriorating states the machine perhaps stays at with given timestamp for each KPI. The log-likelihood probabilities for all the KPI compose a set of features to representing the overall HI and RULs, which can be modeled as a regression learning problem. The regression model could be formulated as:

$$RUL(t) = f(L^{(1)}_{KPI_1}(t), L^{(2)}_{KPI_1}(t), \dots, L^{(K)}_{KPI_1}(t), \dots, L^{(1)}_{KPI_n}(t), L^{(2)}_{KPI_n}(t), \dots, L^{(K')}_{KPI_n}(t)) \quad (6)$$

where $f(*)$ represents the regression function and $L^{(i)}_{KPI_j}(t)$ represents the log-likelihood probability from fine-grained deteriorating state i at time t for KPI_j.

III. SYSTEM MODEL

The proposed K-PdM is divided into two phases (shown in Figure 3): learning the fine-grained deteriorating mode and overall deterioration estimation along with RUL prediction.

A. PHASE I: LEARNING FINE-GRAINED DETERIORATING MODE

The fine-grained deteriorating mode can be modeled as a HMM according to each KPI. Specifically, clustering algorithms can be used to convert each KPI time series into symbolic sequence. Then, the HMM's observation sequence is initialized according to the symbolic sequence and then optimal state sequence is obtained by using extended Viterbi algorithm. The details will be introduced in following subsections.

1) INITIALIZE OBSERVATION SEQUENCE BY CLUSTERING

We regard the original KPI time series as $O = \{o_1, o_2, \dots, o_n\}$, $o_i \in O$ represents the tuple (t_i, v_i), where t_i is a timestamp and v_i is the KPI value at t_i. A clustering algorithm, such as K-means or spatial clustering method [21] can be used to cluster the original data into several clusters $C = \{c_1, c_2, \dots, c_N\}$, and each cluster could be represented as a symbol $c_i (1 \leq i \leq N)$. Thus, each data point is labeled with a symbol and the KPI sequence is transformed to a symbol sequence $S = \{s_1, s_2, \dots, s_T\}$, $s_t \in S$ represents a tuple (t, c_i), t is a timestamp and c_i is a symbol belongs to C.

We model the deteriorating process as an HMM, where the observation sequence is initialized according to the symbolic sequence S. Discovery deteriorating state sequence which generates the S is to find the optimal consecutive symbolic patterns which generates corresponding symbols. We firstly define the consecutive symbolic patterns as the hidden states. Since the consecutive symbolic patterns are unknown, we assume that there are N corresponding initial hidden states (the same number as clusters in C), which are denoted as $H = \{h_1, h_2, \dots, h_N\}$, to the symbol sequence S. Then, we should iteratively infer the most optimal consecutive symbolic patterns $H_{optimal} = \{h_1^{opt}, h_2^{opt}, \dots, h_{N'}^{opt}\}$ according to Equation 4.

For a HMM, the initial state transition probabilities, initial state and output probability B should be determined to inter the final parameters of HMM with finding the optimal consecutive symbolic patterns. We can use the maximum likelihood estimation method [22] to estimate the state transition probabilities $A = \{a_{ij}\}$. While, for the initial probabilities $\pi = \{\pi_i\}, 1 \leq i \leq N$, the estimate probability π_i is the frequency of h_i as initial state in all samples.

To initialize the output probability $B = \{b_i(s_t)\}$, where i represents hidden state h_i and symbol s_t belonging to symbol sequence S, we should firstly identify the distributions of state h_i generating symbol s_t. Each observation corresponding to one state can be regarded as a stable behavior with deviations, and the deviations are usually described as a Gaussian Distributions. Thus, we assume that the centroid of the symbol in cluster c_i follows one-dimensional Gaussian Distributions. Specifically, we define the symbols in cluster c_i as $s_k^{(i)}$, and the centroid of all the points in symbol $s_k^{(i)}$ is $centroid\_s_{ki} \sim N(\mu_{ki}, \sigma_{ki}^2)$. Kernel Density Estimation (KDE) [23] and Gaussian kernel [24] are used to estimate the output probability.

2) DISCOVERING FINE-GRAINED DETERIORATING STATES BY UPDATING HMM

It is intuitive and commonsense that same symbolic observations are more likely belonged to the deteriorating state, so it is assumed that symbols with similar distributions compose a new state based on initial symbolic observation and state sequences. Extended Viterbi algorithm (eViterbi) [19] based on *BestSymbol* function is used here to infer the deteriorating state sequence. The key part in eViterbi is to find the best consequent symbols generated by deteriorating state sequence with the optimal probability $\delta_t(i)$.

Algorithm 1 BestSymbol Algorithm

Require: S: Initialized observed symbolic sequence, t: time of current symbol, ε_r: threshold of average error
Ensure: Consequent Symbol sequence from $t-d+1$ to t

for $d = 2$ to t **do**
  $centroid_{avg} = centroid(s_t \cup s_{t-1} \cup \ldots \cup s_{t-d+1})$
  for $i = t-d+1$ to t **do**
    $ab\_err(s_i) = \left|centroid(s_i) - centroid_{avg}\right|$
  end for
  $mean\_ab\_err(clst) = \frac{1}{d}\sum_{i=t-d+1}^{t} ab\_err(s_i)$
  if $mean\_ab\_err(clst) > \varepsilon_r$ **then**
    BREAK
  end if
end for

BestSymbol function (shown in Algorithm 1) is to find the best symbolic sequence $\{s_{t-d+1}, s_{t-d+2}, \ldots, s_t\}$, whose mean average error is below the threshold ε_r. The $centroid_{avg}$ is the average centroid of all the points belonging to the symbolic sequence beginning at $t-d+1$ and ends at t. The $ab_err(s_i)$ measures the absolute error of symbol s_i with $centroid_{avg}$. And the $mean\_ab\_err(clst)$ is the average absolute errors for all the symbol belonging to the symbolic sequence which begins at $t-d+1$ and ends at t. If the $mean\_ab\_err(clst)$ is less than the threshold ε_r, it could be regarded as a best symbolic sequence.

In Algorithm 2, the eViterbi based on *BestSymbol* function is presented. Based on a dynamic iteration of eViterbi, the optimal observation and state sequence are able to be inferred simultaneously based on initial symbolic sequence. At each iteration, the new *'BestSymbol'* sequence is found corresponding to hidden states, thus the distribution of symbols belonging to each state will be changed and the output probability should be recalculated. It will converges until the symbolic sequence is unchanged, and the stable hidden states is denoted as $SS = \{h_1, h_2, \ldots, h_m\}$, where m is the number of hidden states.

Algorithm 2 eViterbi Inference Based on BestSymbol

Require: S_{init}: Initialized observed symbolic sequence, ε_r: threshold for *BestSymbol*
Ensure: deteriorating state sequence HS

1: Initialize $S_{temp} = S$
2: Initialize $\delta_1(i) = 0 (1 \le i \le K)$
3: **for** $t = 2$ to n **do**
4:   **for** $i = 1$ to K **do**
5:     $\delta_t(i) = 0$
6:     $S_{best}(t-d+1, t) = BestSymbol(S_{init}, t, \varepsilon_r)$
7:     $\delta_{temp} = \max\limits_{d,j}(\delta_{t-d}(j)a_{ji})b_i(S_{best})$
8:     **if** $\delta_{temp} > \delta_t(i)$ **then**
9:       $\delta_t(i) = \delta_{temp}$
10:       $prev_d(t) = t - d$
11:       $prev_h(t) = j$
12:     **end if**
13:   **end for**
14: **end for**
15: Obtain the new symbolic sequence S' by backtracking sequence of $prev_d$
16: **if** $S_{temp} \neq S'$ **then**
17:   $S_{temp} = S'$
18:   Goto Line 2
19: **else**
20:   Obtain the deteriorating state sequence HS by backtracking sequence of $prev_h$
21: **end if**

3) LEARNING FINE-GRAINED DETERIORATING MODE

To learn the internal deteriorating mode from a fine-grained KPI, we could model each fine-grained deteriorating state h_k as an HMM λ_k, where $1 \le k \le m$, and the HMM could be trained via Baum-Welch algorithm. The training observation O'_{train} can be divided into subsequences $\{\boldsymbol{o'_1}, \boldsymbol{o'_2}, \ldots, \boldsymbol{o'_m}\}$. For each $\boldsymbol{o'_i} = \{o_1^{(i)}, o_2^{(i)}, \ldots o_T^{(i)}\}$ $(1 \le k \le m)$, it is correspondent to h_k. It is assumed that there are n sub-states that generate $\boldsymbol{o'_i}$, and the sub-state spaces are denoted as $Sub\_SS^k = \{ss_1^k, ss_2^k, \ldots, ss_n^k\}$, where k represents h_k.

B. PHASE II: OVERALL DETERIORATION ESTIMATION AND RUL PREDICTION

Once the fine-grained deteriorating mode based on each KPI is learned, an overall deterioration is able to be estimated by combination of estimation on each KPI. Moreover, the health index and remaining useful life can be predicted based on the regression model. The details will be introduced in following subsections.

1) DETERIORATING STATE ESTIMATION

To estimate current fine-grained deteriorating state based on given KPI sequence, a comparable method based on calculation of log-likelihood is used according to section II. The given KPI data is denoted as O'_{test}, and it is segmented by sliding windows with window length l and sliding length s.

Thus, O'_{test} with length T is segmented into M windows: $O'_{test} = \{\boldsymbol{o}_1^{(test)}, \boldsymbol{o}_2^{(test)}, \dots \boldsymbol{o}_M^{(test)}\}$, where $M = \lfloor \frac{T-l}{s} + 1 \rfloor$. The fine-grained deteriorating state at time window t is estimated as:

$$h_k^t = \arg\max_{1 \le k \le m} (P(o_1^{'(t)}, o_2^{'(t)}, \dots, o_t^{'(t)} | \lambda_k)) \tag{7}$$

where $o_i^{'(t)}$ represents the observations in time window $\boldsymbol{o}_t^{(test)}$ and $1 \le t \le M$.

Corresponding to the estimated fine-grained deteriorating state h_k at time t, the log-likelihood at time t is calculated as $L_t(h_k) = \log \delta_t(h_k) = \log P(\boldsymbol{o}_t^{(test)}, q_t = h_k | \lambda_k)$. The definition could be used to represent the features of the system's health status, the calculation of $\delta_t(h_k)$ is given in Equation (4) and (5).

According to section II and our previous work [25], a set of rules can be defined to diagnose the overall deterioration based on a combination of fine-grained deteriorating states. For example, in CAMPSS datasets [4], three main KPIs are monitored to indicate HPC's performance, which are *total_temperature* at HPC outlet, *static_pressure* at HPC outlet and *total_pressure* at HPC outlet. It is assumed that each fine-grained KPI represents corresponding deterioration and has three states respectively, so the overall deterioration could be divided into 7 levels as Figure 4 shown, and each level indicates the severity of the HPC's deterioration. A rule-based reasoning can be executed to identify the overall deterioration according to each KPI's deteriorating state, and in addition, the root-cause KPI with the severity level is able to be judged.

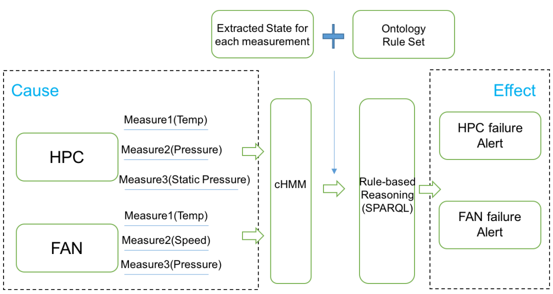

FIGURE 4. Root-cause analysis based on fine-grained deterioration estimation of each KPI.

2) LOG-LIKELIHOOD-BASED RUL PREDICTION

To predict RUL based on log-likelihood probability, a more flexible approach is to model the relationship between **RUL** and **L** through a regression model. While the regression model selection will depend on the properties of the (**L**, **RUL**) data, we will present here results based on linear polynomial regression models. Consistent with the principle of Occam's razor [26], we will prefer lower order models over higher order models, given reasonable accuracy.

The machinery performance are measured by several KPIs, and each KPI represents one dimension of features for RUL prediction. Therefore, data fusion method is able to be modeled to predict the RULs based on multiple KPI measurements. According to Equation (6), the log-likelihood probabilities generated by the fine-grained deteriorating states are the features of RUL for one KPI. Assuming that there are n KPIs, the linear regression function $f(L)$ is defined as:

$$f^{(\theta)}(\boldsymbol{L}) = \sum_{i=1}^{n} f_i^{(\theta_i)}(\boldsymbol{L_i}) \tag{8}$$

and

$$f_i^{(\theta_i)}(\boldsymbol{L_i}) = \sum_{k=0}^{n_i} \theta_i^{(k)} L_i^{(k)} = \boldsymbol{\theta_i^T} \boldsymbol{L_i} \tag{9}$$

where $\boldsymbol{L_i} = \{L_i^0, L_i^1, L_i^2, \dots, L_i^{n_i}\} (L_i^0 = 1)$ is a n_i-dimensional vector of log-likelihood probability from $KPI_i (1 \le i \le n)$ and $\boldsymbol{\theta_i} = \{\theta_i^0, \theta_i^1, \theta_i^2, \dots, \theta_i^{n_i}\}$ is the weight of each dimension.

IV. EXPERIMENTS AND EVALUATIONS

A. DATASETS AND EXPERIMENT SETUP

In our experiment, we use CMAPSS datasets [4] as the evaluation use case. The run-to-failure dataset consists of multiple multivariate time series for engine degradation simulation, which is simulated by C-MAPSS. We select three KPI time series named *total_temperature* at HPC outlet, *static_pressure* at HPC outlet and *total_pressure* at HPC outlet from dataset FD001 to evaluate our approach. The framework is implemented in Java and Python. The datasets and source code can be found on github.[1]

The data have been divided into training datasets and test datasets. The training datasets are used to learn the fine-grained deteriorating mode, and the test datasets are used to validate the performance of overall deterioration estimation and RUL prediction. Since there are not a uniform set of evaluating benchmarks to assess the entire proposed framework, three aspects of evaluations are setup to evaluate the framework step by step: 1)

1) The first one is evaluation on fine-grained deteriorating state discovery based on the first phase of cHMM. State preservation index (SPI) is used to evaluate the discovery effects of proposed methods based on spatial clustering and K-means algorithms. Then, the distributions of discovered deteriorating states are visualized and discussed;
2) The second one is evaluation on fine-grained/overall deterioration estimation and RUL prediction results based on the second phase of cHMM. Comparisons between proposed method and other state-of-arts methods will be discussed;
3) The third one is testify the feasibility and utility of overall deterioration estimation and root-cause analysis based on K-PdM framework.

[1] https://github.com/minelabwot/Semantic_State_Finder_Time_Series

B. RESULTS AND DISCUSSION

- **Evaluation on fine − grained deteriorating state discovery**

In this experiment, we use the *total_temperature* KPI of Engine 1 to evaluate fine-grained deteriorating state discovery. Spatial clustering and K-Means algorithms are used respectively as initialized clustering methods. In addition, according to Eberle *et al.* [21], **Statepreservationindex**(**SPI**) is used to measure the compactness among discovered deteriorating states.

By using spatial clustering algorithm to initialize symbolic observation sequence, we set the parameters as $grid\_size = 0.55$, $threshold = 4$ and $\varepsilon_r = 0.005$ according to [20]. The initialized symbolic observation sequence is correspondent to 5 initial deteriorating states, which is shown in Figure 5 (a). And after two iterations of HMM update, 3 final deteriorating state were discovered as Figure 5 (b) shown, which are labeled as symbolic number 0, 2 and 1 from healthy to failing state. While, by using K-means algorithm, where parameter K is set to 5 compared with spatial clustering, the symbolic observation sequence is initialized corresponding to 5 initial deteriorating states, which is shown in Figure 5 (c). And after And after three iterations of HMM update, 4 final deteriorating state were discovered as Figure 5 (d) shown, which are labeled as symbolic number 1, 2, 3 and 4 from healthy to failing state. Symbol 0 is labeled as a noisy state with a very short duration.

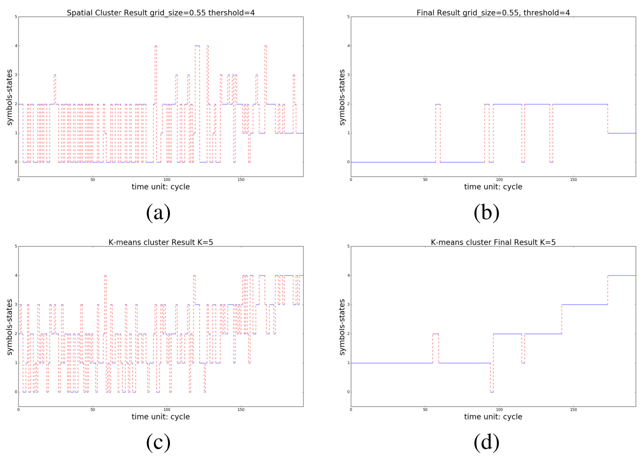

FIGURE 5. State discovery results based on clustering and cHMM methods. (a) 5 initial states (patterns) clustered by spatial clustering algorithm; (b) 3 final deteriorating states (symbolic number: 0-normal, 2-degraded and 1-faulty) discovered by cHMM (Spatial); (c) 5 initial states (patterns) clustered by K-means algorithm; (d) 4 final deteriorating states (symbolic number: 1-normal, 2-slightly degraded, 3-severely degraded and 4-faulty) discovered by cHMM (K-means).

As Table 1 shows, it can be found that (1) spatial cluster performs better than K-means on finding the initial symbolic observation sequence and states with more compactness; (2) HMM-based inference improves the accumulation and refinement of initial state dramatically; (3) The final *spi_value* of two methods are of little gap between each other (cHMM (Spatial) is a little better than cHMM (K-means)).

TABLE 1. spi_value comparison with and without HMM refinements.

	spi (cluster)	spi (HMM)
cHMM (Spatial)	0.97	7.33
cHMM (K-means)	0.46	7.27

To present the distributions of each cluster and final degradation-state, the probability density plots are shown in Figure 6. Kernel Density Estimation (KDE) is used to estimate the probability density, and Gaussian kernel is chosen as the kernel function. The distribution plots visualize that the state refinement of cHMM is effectively to discover the stable and compact deteriorating states, and moreover, the final results depend not too much on the type of clustering methods (no matter spatial clustering or K-means algorithm).

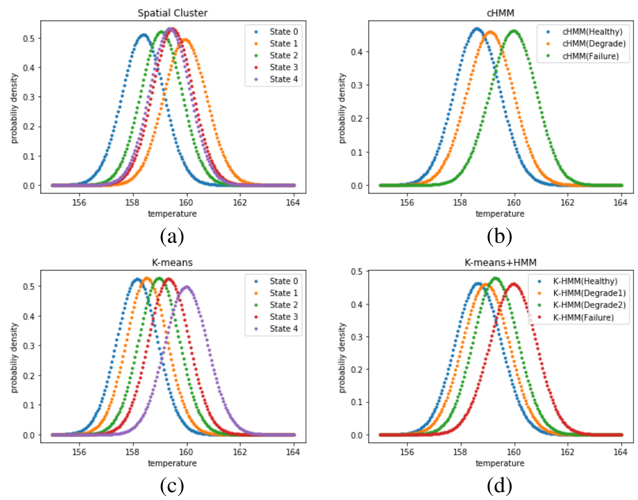

FIGURE 6. Probability density distribution of each cluster and deteriorating state. (a) Probability density distribution of each cluster by spatial clustering algorithm; (b) Probability density distribution of each deteriorating state by cHMM (Spatial); (c) Probability density distribution of each cluster by K-means algorithm; (d) Probability density distribution of each deteriorating state by cHMM (K-means).

- **Evaluation on deterioration estimation and RUL prediction**

Based on the cHMM, three fine-grained deteriorating states are discovered on *total_temperature*, *static_pressure* and *total_pressure* measurements of both engine 1 and 7 respectively. While for engine 6, two fine-grained deteriorating states are discovered on *total_temperature*, four fine-grained deteriorating states are discovered on *static_pressure* measurement and two fine-grained deteriorating states are discovered on *total_pressure* measurement. To train the sub-state sequences for each KPI, we initialize the hidden states as 3 and the time series are segmented into time windows as observation sequence. The window length is 10 and the step length is 5. The log-likelihood curve on training datasets have been shown in Figure 10 (a-c, g-i, m-o).

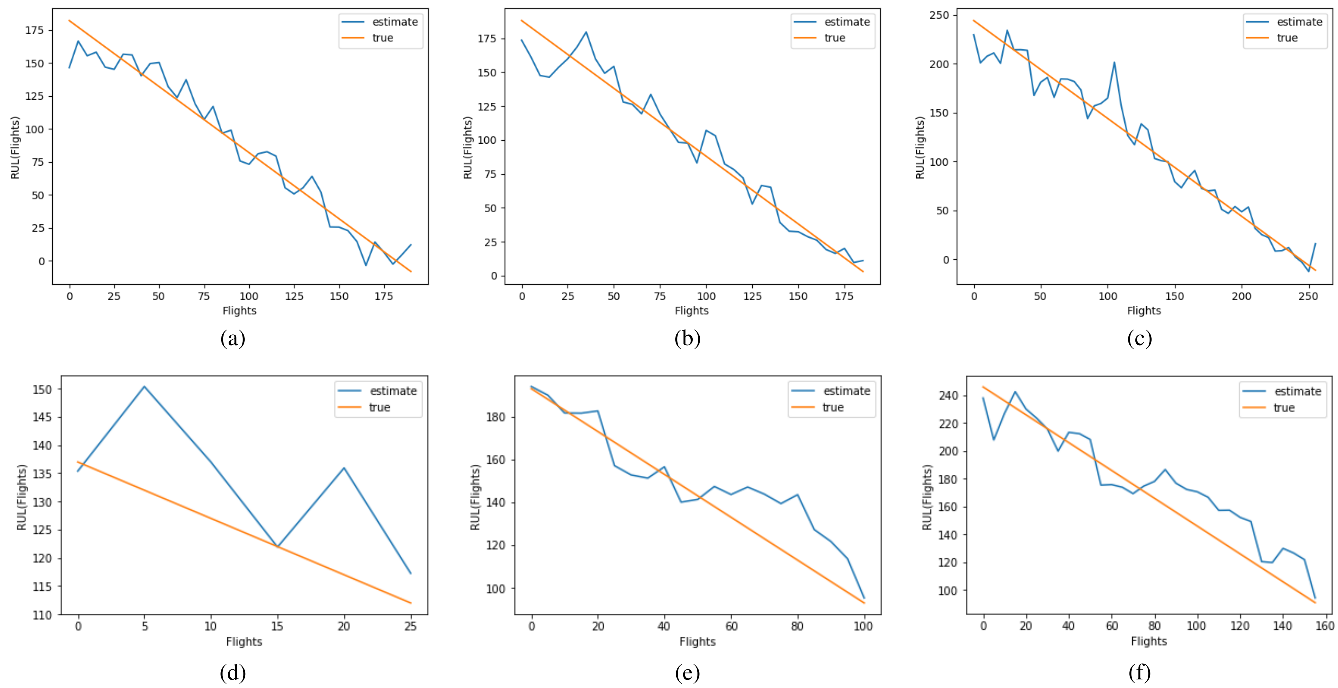

FIGURE 7. Log-likelihood trajectory on training data and test data of engine 1&6&7. (a)-(c): log-likelihood trajectory on training data of engine 1&6&7; (d)-(f): log-likelihood trajectory on test data of engine 1&6&7.

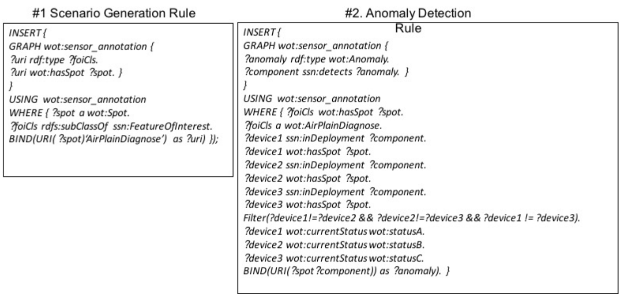

FIGURE 8. SPARQL syntax for rule-based failure detection.

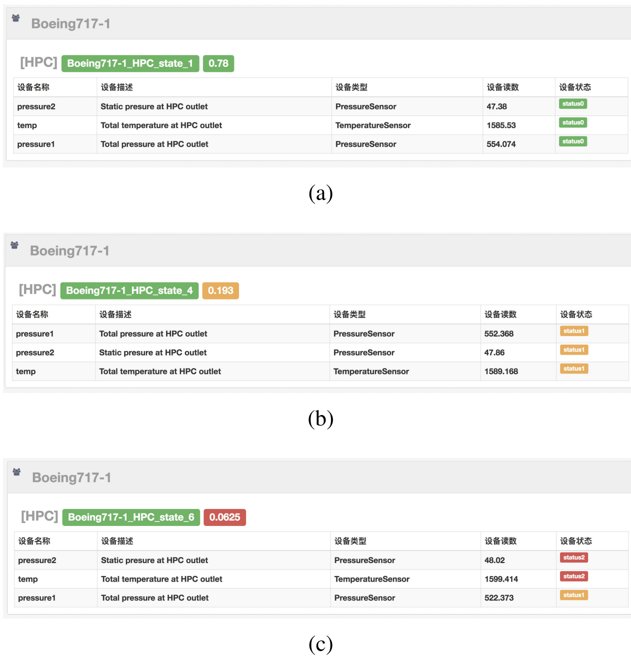

FIGURE 9. Proof-of-Concept: rule-based overall deterioration estimation and root-cause analysis.

Based on the trained model, the current fine-grained deteriorating state can be estimated based on cHMM corresponding to each KPI. The log-likelihood curves of test dataset for engine 1, 6 and 7 have been also presented in Figure 10 (d-f, j-l, p-r). The results show: (1) Engine 1's HPC temperature stays at state 1 (degraded) which is transferred from state 0 (normal) at recent time. Engine 6 stays at state 0 (normal). Engine 7 stays at state 0 (normal); (2) Engine 1's static pressure stays at state 0 (normal). Engine 6 stays at state 0 (normal). Engine 7 stays at state 0 (normal) and will transfer to state 1 (degraded) soon; (3) Engine 1's total pressure stays at state 0 (normal). Engine 6 stays at state 0 (normal). Engine 7 stays at 0 (normal) and transfer to state 1 (degraded).

Based on the fine-grained deteriorating state estimation, the overall RUL are able to be predicted. The linear regression models are trained for engine1, 6, and 7 respectively, and the RULs of corresponding engines are predicted as Table 2 shown. Since only the most recent RULs are given in the test datasets, the Absolute Error (AE) is used to evaluate the performance on each engine. It is shown that the estimated RUL on engine 1, 6 and 7 are very closed to the real RUL.

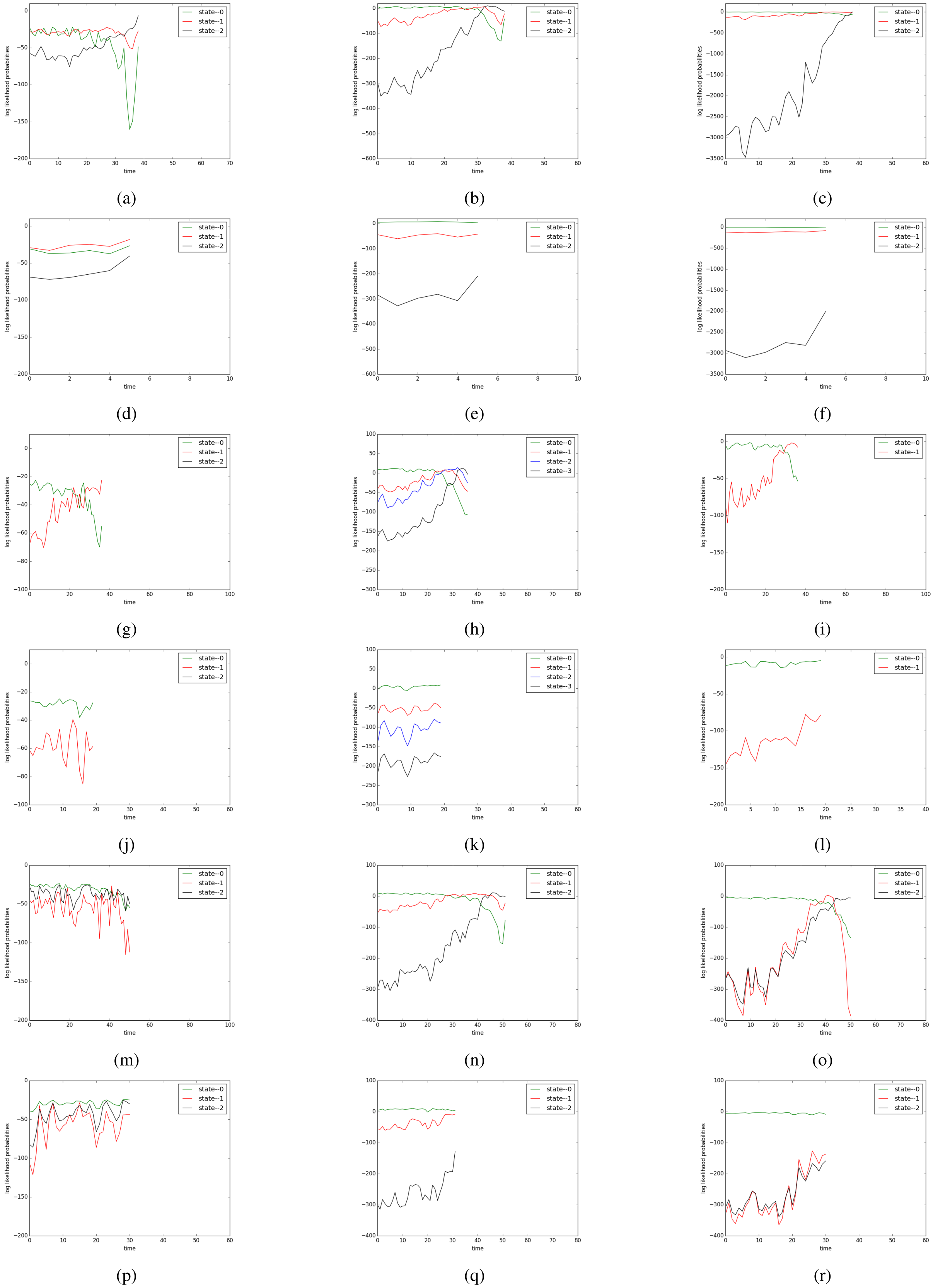

FIGURE 10. **Temperature, static pressure and total pressure's log-likelihood curve on training data and test data of engine 1&6&7. (a)-(c): curve on training data of engine 1; (d)-(f): curve on test data of engine 1; (g)-(i): curve on training data of engine 6; (j)-(l): curve on test data of engine 6; (m)-(o): curve on test data of engine 6; (g)-(i): curve on training data of engine 7; (p)-(r): curve on test data of engine 7.**

TABLE 2. Effects of KPI fusion on RUL prediction according to Absolute Error (AE) results. S1: total_temperature, S2: total_pressure and S3: static_pressure.

KPI Fusion	Engine 1 (True-112)		Engine 6 (True-93)		Engine 7 (True-91)	
	Estimated	AE	Estimated	AE	Estimated	AE
S1-2-3	**117.26**	**5.26**	**95.34**	**2.36**	**94.38**	**3.38**
S1-2	93.27	18.73	76.51	16.49	86.51	4.49
S1-3	103.65	8.35	89.29	3.71	105.09	14.09
S2-3	118.29	6.29	100.31	7.31	96.13	5.13
S1	61.24	12.50	90.22	2.78	203.05	112.05
S2	99.50	12.50	**91.27**	**1.73**	82.95	8.05
S3	105.17	6.83	98.38	5.38	110.03	19.03

Moreover, KPI fusion is able to gain the minimum AE on engine 1 and 7 by using all the three sensors. Although the minimum MAE is not performed by fusing three KPIs on engine 6, the KPI fusion method still gains a relative good AE result. The results illustrate that KPI fusion could improve the prediction accuracy, due to more useful features extracted from sensors. To illustrate performance of the polynomial regression function on predicting RULs, Figure 7 shows the log-likelihood trajectory on both training datasets and test datasets of engine 1, 6 and 7.

We also give comparisons between cHMM and other state-of-arts' methods [27] on RUL predictions. The actual RUL values for the last recorded cycles are given in the dataset, and the corresponding RUL labels for the previous life-time can be obtained accordingly. To evaluate the performance of RUL prediction, Root Mean Square Errors (RMSE) on engine 1, 6 and 9 are used. We use the average RMSE of engine 1, 6 and 9 as the overall performance of cHMM, and we compare it with other methods. As Table 3 shown, the results illustrate that cHMM performs better than most of other methods but not better than Deep LSTM on RUL predictions. The reason is perhaps that HMM-based method could probably not learn the long-term dependent patterns from observations well, while LSTM is good at it. However, deep learning methods are not good at translating the internal working patterns of how machines fail and not able to analyze the root-cause of the fault, while cHMM discovers the fine-grained deteriorating modes from each of the KPI measurements, which facilitates the overall estimation of the deteriorating states by each KPI measurements and evaluating which measurements cause the performance deterioration.

To estimate the overall deterioration and analyze the cause, a set of reasoning rules and ontological vocabularies are modeled into an expert system based on our previous work [28]. For example, the overall deteriorating degrees can be divided into 7 levels for Engine 1. When new KPI readings generated, the fine-grained deteriorating state of each KPI measurement will be estimated, and then the overall deteriorating levels with corresponding RUL will be diagnosed according to predefined rules. The SPARQL sentences of reasoning the fault are shown in Figure 8.

TABLE 3. The RMSE comparison between cHMM and other state-of-arts' methods.

Methods	RMSE@Average
cHMM	**14.28**
SVM	29.82
Random Forrest	17.91
Gradient Boosting	15.67
DCNN	18.45
Deep LSTM	12.81

Finally, the root-cause KPI of the deterioration will be traced with the deteriorating levels. The proof-of-concept demo is implemented and a snapshot is shown in Figure 9. The figure illustrates that the engine #1 is diagnosed in three different working conditions: 9(a) presents the engine is in relative healthy condition with health index value 0.78 (which means the remaining lifetime has 78% left), and the three measurements are all at its corresponding healthy state (status 0); 9(b) presents the engine is in median deteriorating condition with health index value 0.193 (which means the remaining lifetime has 19.3% left, very closed to its failure), and the three KPI measurements are all at its corresponding deteriorating state (status 1); 9(c) presents the engine is in severe deteriorating condition (nearly failure) with health index value 0.0625 (which means the remaining lifetime has 6.25% left), and the three KPI measurements are all at its corresponding late deteriorating state (temperature and static pressure at status 2, total pressure at status 1. When severe deteriorating condition happens, the failure alert will be triggered and the root-cause KPIs are able to be checked that *total_temperature* and *static_pressure* are abnormal.

V. CONCLUSION

In this paper, we presented $K - PdM$ based on cluster-based Hidden Markov Model (cHMM) for KPI-oriented predictive maintenance. $K - PdM$ is able to learn the deteriorating patterns from fine-grained KPI and extract the deteriorating states with transition from health to failure. Based on the learned pattern, overall RUL could be predicted according to the log-likelihood of given observations generated by fine-grained deteriorating states. Moreover, the root-cause diagnostic rules could be modeled according to extracted fine-grained deteriorating states and the root-cause KPI of overall performance deterioration is able to be reasoned. To validate the feasibility of cHMM, we conduct experiment based onPHM08 challenge data and implement a proof-of-concept predictive maintenance demonstration for turbofan engine. The experimental results illustrate that the model could discover the fine-grained deteriorating states with high compactness and RUL could be predicted within promising MAE and RMSE values compared with other methods. A proof-of-concept demo is implemented to show its application on turbofan engine's diagnostics and prognostics, which yields its feasibility and utility in real-world predictive maintenance applications.

REFERENCES

[1] T. Xia and L. Xi, "Manufacturing paradigm-oriented PHM methodologies for cyber-physical systems," *J. Intell. Manuf.*, vol. 9, pp. 1–14, Aug. 2017.

[2] J. Lee and B. Bagheri, "Cyber-physical systems in future maintenance," in *Proc. 9th WCEAM Res. Papers*, 2015, pp. 412–421.

[3] Z. Wu, W. Lin, and Y. Ji, "An integrated ensemble learning model for imbalanced fault diagnostics and prognostics," *IEEE Access*, vol. 6, pp. 8394–8402, 2018.

[4] A. Saxena and K. Goebel. (2008). "Turbofan Engine Degradation Simulation Data Set." NASA Ames Research Center, Moffett Field, CA, USA. [Online]. Available: http://ti.arc.nasa.gov/project/prognostic-data-repository

[5] F. J. Lin, Y. C. Hung, and M. T. Tsai, "Fault-tolerant control for six-phase PMSM drive system via intelligent complementary sliding-mode control using TSKFNN-AMF," *IEEE Trans. Ind. Electron.*, vol. 60, no. 12, pp. 5747–5762, Dec. 2013.

[6] M. Kang, J. Kim, and J. M. Kim, "An FPGA-based multicore system for real-time bearing fault diagnosis using ultrasampling rate AE signals," *IEEE Trans. Ind. Electron.*, vol. 62, no. 4, pp. 2319–2329, Apr. 2015.

[7] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.

[8] F. Yang, M. S. Habibullah, T. Zhang, Z. Xu, P. Lim, and S. Nadarajan, "Health index-based prognostics for remaining useful life predictions in electrical machines," *IEEE Trans. Ind. Electron.*, vol. 63, no. 4, pp. 2633–2644, Apr. 2016.

[9] F. O. Heimes, "Recurrent neural networks for remaining useful life estimation," in *Proc. Int. Conf. Prognostics Health Manage.*, Oct. 2008, pp. 1–6.

[10] E. Ramasso and A. Saxena, "Performance benchmarking and analysis of prognostic methods for CMAPSS datasets," *Int. J. Prognostics Health Manage.*, vol. 5, no. 2, pp. 1–15, 2014.

[11] M. Abbas, "System-level health assessment of complex engineered processes," Ph.D. dissertation, School Elect. Comput. Eng., Georgia Inst. Technol., Atlanta, GA, USA, 2010.

[12] H. Richter, "Engine models and simulation tools," in *Advanced Control of Turbofan Engines*. New York, NY, USA: Springer, 2012, pp. 19–33.

[13] P. Malhotra *et al.*, "Multi-sensor prognostics using an unsupervised health index based on LSTM encoder-decoder," in *Proc. 1st ACM SIGKDD Workshop ML PHM*, 2016, pp. 1–10.

[14] N. Gugulothu, T. V. Vishnu, P. Malhotra, L. Vig, P. Agarwal, and G. Shro, "Predicting remaining useful life using time series embeddings based on recurrent neural networks," in *Proc. ACM SIGKDD Workshop Mach. Learn. Prognostics Health Manage.*, 2017, pp. 1–10.

[15] A. Giantomassi, F. Ferracuti, A. Benini, G. Ippoliti, S. Longhi, and A. Petrucci, "Hidden Markov model for health estimation and prognosis of turbofan engines," in *Proc. ASME Int. Design Eng. Tech. Conf. Comput. Inf. Eng. Conf.*, 2011, pp. 681–689.

[16] P. Baruah and R. B. Chinnam, "HMMs for diagnostics and prognostics in machining processes," *Int. J. Prod. Res.*, vol. 43, no. 6, pp. 1275–1293, Mar. 2005.

[17] G. Prakash, S. Narasimhan, and M. D. Pandey, "Condition based maintenance of low speed rolling element bearings using hidden Markov model," *Int. J. Prognostics Health Manage.*, vol. 8, no. 6, pp. 1–16, 2017.

[18] T. T. Le, F. Chatelain, and C. Bérenguer, "Hidden Markov models for diagnostics and prognostics of systems under multiple deterioration modes," in *Proc. Eur. Saf. Rel. Conf.*, Wrocław, Poland, 2014, pp. 1197–1204.

[19] Wang, Peng, H. Wang, and W. Wang, "Finding semantics in time series," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2011, pp. 385–396.

[20] Z. Wu, H. Luo, Y. Yang, X. Zhu, and X. Qiu, "An unsupervised degradation estimation framework for diagnostics and prognostics in cyber-physical system," in *Proc. IEEE 4th World Forum Internet Things (WF-IoT)*, Feb. 2018, pp. 784–789.

[21] J. Eberle, T. K. Wijaya, and K. Aberer, "Online unsupervised state recognition in sensor data," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun.*, Mar. 2015, pp. 29–36.

[22] J. Pfanzagl and R. Hambker, *Parametric Statistical Theory*. Berlin, Germany: Walter de Gruyter, 1994, pp. 207–208.

[23] M. Rosenblatt, "Remarks on some nonparametric estimates of a density function," *Ann. Math. Stat.*, vol. 27, no. 3, pp. 832–837, 1956.

[24] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. London, U.K.: CRC Press, 1986, p. 48.

[25] Z. Wu, Y. Xu, Y. Yang, C. Zhang, X. Zhu, and Y. Ji, "Towards a semantic Web of Things: A hybrid semantic annotation, extraction, and reasoning framework for cyber-physical system," *Sensors*, vol. 17, p. 403, Feb. 2017.

[26] D. J. C. MacKay, *Information Theory, Inference and Learning Algorithms*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[27] X. Li, Q. Ding, and J.-Q. Sun, "Remaining useful life estimation in prognostics using deep convolution neural networks," *Rel. Eng. Syst. Saf.*, vol. 172, pp. 1–11, Apr. 2018.

[28] *SWoT Ontology*. Accessed: Mar. 13, 2018. [Online]. Available: https://github.com/minelabwot/SWoT

ZHENYU WU received the B.S. and Ph.D. degrees from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2008 and 2013, respectively. He is currently an Assistant Professor with the School of Information and Communication Engineering, BUPT. His research interests include Internet of Things, machine learning, prognostics and health management technology, and industrial big data analysis.

HAO LUO received the B.S. degree from the Chongqing University of Posts and Telecommunications, Chongqing, China, in 2015, and the M.S. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 2018. His research interests include machine learning, prognostics and health management technology, and industrial big data analysis.

YUNONG YANG received the B.S. and M.S. degrees from the Beijing University of Posts and Telecommunications, Beijing, China, in 2015 and 2018, respectively. His research interests include Semantic Internet of Things, machine learning, prognostics and health management technology, and industrial big data analysis.

PENG LV received the B.S. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 2016, where he is currently pursuing the M.S. degree with the School of Information and Communication Engineering. His research interests include prognostics and health management technology and industrial big data.

XINNING ZHU received the Ph.D. degree from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2010. She is currently an Associate Professor with the School of Information and Communication Engineering, BUPT. Her research interests include spatial temporal sequence mining, time-series analysis, and industrial big data.

YANG JI received the Ph.D. degree from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2002. He is currently a Full Professor with the School of Information and Communication Engineering, BUPT, and the Director of the Mobile Life and New Media Lab. His research interests include Internet of Things, machine learning, smart manufacturing, and smart education.

BIAN WU received the M.D. degree from the Tongji Medical College, Huazhong University of Science and Technology (HUST), in 2011. He is currently a Lecturer with the Cancer Center, Tongji Medical College, HUST. His research interests include bioinformatics, statistical methods on gene engineering, machine learning, and data-driven diagnostic methods.