Segmentation Improved Label Propagation for Semi-Supervised Anomaly Detection in Fused Magnesia Furnace Process

SHAOWEN LU\(^{\circledR}\), (Member, IEEE) 
State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang 110819, China 
e-mail: lusw@mail.neu.edu.cn

This work was supported in part by the National Natural Science Foundation of China under Grant 61833004 and Grant 61991404, and in part by the Youth Science and Technology Innovation Leader of Shenyang under Grant RC190443.

ABSTRACT 
This article introduces a semi-supervised solution to the practical task of identifying the semi-molten working condition for fused magnesium furnace. The classifier is trained under the semi-supervised learning framework using video data which are partially labeled, and it works on-line to classify the semi-molten working condition from the monitoring video of fused magnesium furnace. Firstly, a deep auto-encoder is used to extract features from the video. Next, the feature time series is segmented into subsequences through a bottom-up algorithm which maximizes the interior data correlation. Then, a graph based label propagation algorithm is employed to iteratively train a LSTM classifier using the subsequences as training samples and improve the accuracy of the LSTM classifier. The main contribution of this work is the novel segmentation algorithm for sequential data which can remarkably improve the classification accuracy. The advantages of the proposed method are demonstrated by experiments and comparative analysis based on industrial data.

INDEX TERMS 
Semi-supervised classification, long short-term memory, time series segmentation, fused magnesium furnace, anomaly detection, graph based label propagation.

I. INTRODUCTION 
Fused magnesia is an industrial material which can find applications in high end optical equipment, nuclear reactors and rocket nozzles [1]. It is produced at temperatures above the fusion point of magnesia oxide (2800°C) using electric arc furnaces (EAF). The industrial fusing process is referred as the fused magnesia furnace (FMF) process. The FMF process typically lasts 12 hours and then the melting block is being cooled down to obtain the magnesia crystal. Figure 1 shows the fused magnesia production process.

Semi-molten is one of the abnormal working condition occurring occasionally during FMF process. Under semi-molten, an overheated region is developing inside the molten pool whose temperature will eventually exceed the resistance limit of the furnace lining. The precise mechanism of semi-molten development is still unclear [2]. It is commonly believed that semi-molten is related to the impurities such as \(\text{SiO}_2\), \(\text{CaO}\) and \(\text{Fe}_2\text{O}_3\) in the magnesite powder which makes the electrical resistance of the molten pool being lower than the normal value. Therefore excessive the increased electrode current produces high intensity arc resulting in significant temperature increase in the furnace. It is vital to detect semi-molten in its early development to prevent burnt-through of the furnace wall. Missed detection or delayed detection of the semi-molten will result in leakage of high-temperature molten ore causing production halt and substantial economic loss. It may even endanger the safety of the operators at the scene.

In practice, early warning of semi-molten is largely relied on human operators by patrolling in the workshop. Operators...
estimate the working condition by inspecting the flame on the furnace mouth based on their experiences. This is because the characteristics of the flame such as its shape, brightness, color and the flashing patterns are closely correlated to the working condition. However, inexperience, unstable mental state of the operators may affect their judgment, which can lead to missed or false detection. Also, the on-site environment, such as intense light, high temperature, dust and noise, is human unfriendly, and continuous inspection involves high labor intensity. Therefore, the technology of automatic identification of semi-molten is very needed by FMF enterprises to ensure production safety and to free the operators from the burdensome patrolling and inspection tasks.

Currently, there are two approaches to the detection of semi-molten working condition of FMFs. The first approach identifies working condition based on the three-phase electrode currents of FMF. For example, Wu et al. [2] established a rule based technique to classify working condition using the features extracted from the tracking error of electrode current, the current change rate, and the current change time. Zhang [3] proposed a method of monitoring working condition based on subspace separation method. Indeed, the electrode current based semi-molten detection approach is intuitively sensible in that the local electrical resistance will remarkably change under the semi-molten condition. However, the patterns of electrode current are not sufficient to distinguish the semi-molten because other factors, such as refilling furnace, may also cause the similar fluctuation pattern of the current. The electrode current alone is not a sufficient source of features to distinguish semi-molten.

The other approach uses the machine vision based classifier trained with features extracted from the FMF monitoring video. It is inspired by the fact that experienced operators can tell the developing semi-molten state solely by “watching the flame”. This indicates that the image of the furnace mouth flame contains rich, ubiquitous features related to the working condition of FMF process. To this end, Guo et al. [4] established a semi-molten classifier using the stochastic configuration networks [5] trained using historical monitor video images. Lu et al. [6] established a RGB and infrared images based classifier using deep convolution network and the generative adversary networks (GAN) were employed to synthesize extra samples so as to overcome the problem of shortage of samples.

Compared with the electrode current approach, the machine vision based approach is much more promising because the visual patterns of furnace flame are sensitive and unique to the semi-molten working condition. However, the current machine vision based detection techniques are still far from practical because of low accuracy. There are two main challenges: First, the flame of a FMF is highly dynamic. The shape, brightness and color of the flame are always changing due to unmeasurable disturbances. It is not possible to accurately distinguish semi-molten according to a single or few images. One must also consider the temporal patterns of the video. Second, the historical video monitoring images can only be manually labeled by experienced operators. Labeling a large amount of video footage by human is very costly and not practical. Therefore, the problem has to be tackled with very limited number of labeled samples.

In view of the above problems, this article proposes a machine vision solution under the framework of semi-supervised learning. The semi-supervised learning (SSL) is a machine learning paradigm which utilizes a large number of unlabeled data with the assistance of a small number of labeled data [7]. Using sufficient unlabeled data, which are often easy to obtain, to improve the performance of classifier is the key to the success of many practical tasks. To address the dynamics in the FMF monitoring video, we consider using multivariate time series subsequences as samples in the classifier training. SSL classification of time series is especially challenging due to the dynamic characteristics of time series data [8], [9]. The subject has attracted increasing attentions driven by applications such as electroencephalography (EEG) pattern recognition, human dynamic behavior and facial expression recognition, automated trading etc. (see a review in [9]).

In this article, we introduce the development of the semi-molten working condition detection solution. We made three main contributions: 1) We developed an end-to-end solution to the practical task of abnormal condition detection in industrial settings. The proposed solution addresses the problem of insufficient labeled data by integrating semi-supervised learning approach with time-series segmentation method. And the proposed solution is tested using real industrial dataset. 2) We propose a dynamic multivariate segmentation method to adaptively partition the feature sequence extracted from video into homogeneous subsequences. Traditional studies on machine learning often assume that samples are well prepared. However, in practical tasks, the quality of samples is often the key to success. By segmenting the video sequence into homogeneous subsequences, it can make the classification performance. 3) We employ the graph based label propagation algorithm to train the condition classifier based on Long Short Term Memory (LSTM) model. When dealing with time series data, the common approach is to convert the sequential data to high dimensional static data and use the Euclidean distance measure. However when dealing with sequence data, such an approach loses the temporal information of the data. We propose a new measure of sequence distance by combining the dynamic predictability and the static reconstruction error in the construction of the affinity graph, which is the key to address the dynamics in FMF flame video.

The rest of the paper is organized as follows: Section 2 defines the FMF anomaly detection problem and briefly reviews the related works. Section 3 overviews the proposed method. Section 4 and section 5 elaborate the prepossessing and classifier training methods, which are the main...
The method used in this work is mostly related to the graph based label propagation algorithm, which is one of the prominent methods based on the cluster assumption [15]. The main idea of the algorithm is to construct a graph using observations as vertices and the similarity between observations represented as edges. The graph based label propagation smooths the spread of label because the nearest neighbors on the graph have the same prediction. For example, the authors of [16] proposes a probabilistic graphical model for label propagation in video sequences using an EM based algorithm. Wang and Tsotsos [17] extend the label propagation to be able to deal with multi-class problems.

Segmentation is to partition a given sequential data into subsequences which are internally homogeneous [18]. Proper segmentation of the sequential data is the key for semi-supersized time series classification [19] and it has been applied to anomaly detection [20], [21]. In semi-supervised learning, segmentation can serve as a data pre-treatment tool for representation reduction and feature extraction. Most of the basic segmentation approaches have a root on the idea of piecewise linear approximation (PLA) [22]. For example, the principal component analysis (PCA) based approach [23] and the dynamic PCA (DPCA) based approach [24]. For the clustering and labeling SSL approach, segmentation and classification are two closely related problems. For example, in [25] the authors cast the video segmentation problem as a semi-supervised learning problem.

III. PROPOSED METHOD OVERVIEW

A. DEFINITION

Let \( V = \{v_1, \ldots, v_n\} \) define the historical monitoring video of FMF process for training, where \( v_i \) represents the \( i \)th frame image and \( N \) is the total number of the images. Define the feature space mapping of image \( \phi : V \rightarrow F \), where \( F = \{x_1, \ldots, x_n\} \in \mathbb{R}^{m \times n} \) and \( x_i \) is the \( m \)-dimension feature vector extracted from \( v_i \). This forms a multivariate feature vector sequence. In this article, we use sequence as the atomic sample data for training. By performing segmentation on \( F \) we obtain the training set \( \mathcal{X} = \{X_1, \ldots, X_k\} \), where \( X_i \) is a subsequence of \( m \)-dimension feature vector. Note that the length of subsequences is not equal. The labeling matrix including \( l \) labels and the rest are unlabeled. It is defined as \( Y = \{y_1, \ldots, y_l, y_{l+1}, \ldots, y_k\} \) where \( Y_{i,j} \) is 1 if the corresponding sample \( x_i \) is of class \( j \) for \( i < l \).

B. THE PROPOSED METHOD

To train a classifier using partially labeled dataset \( \mathcal{X} \), our proposed method includes two parts: preprocessing and classifier training. The process of the proposed method is shown in FIGURE 3.

Preprocessing is to construct the dataset which includes multivariate time series subsequences as samples. First, the input video image sequence is compressed by the sparse auto-encoder (SAE) [26] to obtain the multivariate feature sequence \( F \). Then, we propose a multivariate time series

\[
X = \{x_1, \ldots, x_n\} \in \mathbb{R}^{m \times n}
\]

and the dynamic PCA (DPCA) based approach [24]. For the clustering and labeling SSL approach, segmentation and classification are two closely related problems. For example, in [25] the authors cast the video segmentation problem as a semi-supervised learning problem.
segmentation algorithm to segment the multivariate feature sequence into subsequences. The algorithm recursively merges neighborhood subsequences through a bottom-up procedure. The cost of merging is defined as the weighted combination of the dynamic latent model prediction accuracy and the principal component analysis reconstruction error.

The second part is to train a LSTM classifier for the detection of semi-molten working condition. The classifier is trained under the framework of graph-based label propagation. First, we construct a graph whose nodes are data points and edges are similarities between points. Next, the network parameters of the LSTM model is trained using the initial available labeled samples. Then, the unlabeled samples are assigned with pseudo label which is calculated as the combination of the initial label information and those received from its neighbors weighted by the corresponding edges. And then the network is retrained using both labeled and pseudo labeled datasets.

IV. SEQUENCE SEGMENTATION BASED ON LINEAR DYNAMIC PREDICTABILITY
A. SEQUENCE SEGMENTATION PROBLEM
It has been noted that human operators learn to recognize the semi-molten working condition by watching the dynamic pattern of the furnace flame. And human experts are more comfortable to label the working condition based on video rather than single images. To develop a machine detector which mimics the human cognitive process and learning, we will use video sequences as the basic training samples. This requires a segmentation operation on the multivariate feature time series extracted from the video for training.

Give a sequence data \( T = \{ x_i \in \mathbb{R}^m | 1 \leq i \leq N \} \), the segmentation problem can be defined as constructing \( q \) non-overlapping subsequences

\[
S^q_T = \{ s_i(a_i, b_i)|1 \leq i \leq c, a_i < b_i \}
\]  

The objective of segmentation is to make the subsequence homogeneous according to given criteria, which is expressed as a cost function \( \text{cost}(\cdot) \) which maps the criteria to a scalar.

The global objective of a given segmentation scheme is

\[
G = q^{-1} \sum_{i=1}^{q} \text{cost}(s_i)
\]  

Since the subsequences are used as the training samples for a higher level classifier model, it is desirable to keep the structure of subsequence as simple as possible. Therefore, we seek a segmentation scheme that each subsequence can be linearly expressed. Considering that the multi-variariate feature sequence shows both cross-correlation and auto-correlation, we use the dynamic latent variable (DLV) modeling method to approximate the multivariate sequence. The segmentation objective is then to minimize the overall averaged approximation error.

B. MULTIVARIATE TIME SERIES SEGMENTATION BASED ON DiPCA
1) PRELIMINARY
For a multivariate sequence the latent variable obtained by PCA can be expressed as linear combination of the sample variable

\[
t_k = x_k^T w
\]  

where \( w \) is weight vector. The dynamics in latent variables can further be expressed as a linear combination of the past \( s \) variables as

\[
t_k = \beta_1 t_{k-1} + \beta_2 t_{k-2} + \cdots + \beta_s t_{k-s} + r_k
\]  

Dong and Qin [27] proposed the dynamic inner principal component analysis (DiPCA) algorithm which can simultaneously learn \( w \) and \( \beta \). The objective function has the following form

\[
\max_{w, \beta} \sum_{i=1}^{T} Z_i (\beta \otimes w)
\]

\[
s.t. \| w \| = 1, \| \beta \| = 1
\]

where

\[
X_i = [x_i, x_{i+1}, \cdots, x_{N+i-1}]^T \quad \text{for } i = 1, \cdots, s + 1
\]

\[
Z_i = [X_1, X_2, \cdots, X_s]
\]  

The DiPCA algorithm uses an iterative process to optimize the above objective function and the details can be found in [27].

2) SEGMENTATION USING DYNAMIC PREDICTABILITY
In [28], we propose a multivariate time series segmentation algorithm. The algorithm realizes the "bottom-up" approach [29]. It starts with the sequence partitioned into the maximum allowed number of subsequences. Then, in each iteration, it merges one of adjacent subsequences which maximizes the objective function. The iteration repeats until the termination condition is satisfied. The key of the segmentation algorithm is the definition of merging cost. The main idea of the algorithm is to learn a DLV model of each subsequence using the DiPCA algorithm and then the merging
cost is defined as the mutual predictability of the subsequence models.

For a new data \( x_{t+1} \), form the data matrix \( X_{t+1} \) as in eq. (6). The estimated latent score is calculated as

\[
\hat{T}_{s+1} = \sum_{i=1}^{k} T_i \Theta_{s+1-i}
\]  

(7)

where \( \Theta \) is estimated using the least squares method. The prediction accuracy (PA) of a next score can be measured by cosine similarity between \( T_{s+1} \) and the prediction \( \hat{T}_{s+1} \).

\[
PA = \frac{T_{s+1} \cdot \hat{T}_{s+1}}{\|T_{s+1}\| \|\hat{T}_{s+1}\|}
\]  

(8)

Eq. (8) can be used to calculate the merging cost of a pair of two adjacent subsequence \( S_i \) and \( S_{i+1} \) as

\[
S_D(i, i + 1) = \frac{\text{\( T_{s+1}^{i+1} \cdot \hat{T}_{s+1}^{i+1} \)}}{\|T_{s+1}^{i+1}\| \|\hat{T}_{s+1}^{i+1}\|}
\]  

(9)

3) AN EXTENDED SEGMENTATION ALGORITHM

One limitation of the segmentation algorithm proposed in [28] is that it only considers the dynamic predictability. The cost function defined using eq.(9) uses the dynamic latent variables which are mostly relevant to the dynamic auto-correlation. In view that the features extracted from the furnace flame image are highly correlated, it is also necessary to take into consideration the cross-correlation structure.

The authors of [27] suggested that DiPCA can be used as a dynamic whitening filter. This indicates that the prediction error can be considered as the residual after filter off the autocorrelation in the original data. For \( X_{t+1} \) its dynamic prediction using a learned DiPCA model is

\[
\hat{X}_{s+1} = \hat{T}_{s+1} P^T
\]  

(10)

Therefore, we can extract the static latent variables by applying PCA on the prediction error and obtain the following

\[
E_{s+1} = X_{t+1} - \hat{T}_{s+1} P^T = T_P P_T^T + E_r
\]  

(11)

Using the similar idea of eq. (9), we can define the merging cost of the static part as

\[
S_S(i, i + 1) = \frac{E_{s+1}^{i+1} \cdot \hat{E}_{s+1}^{i+1}}{\|E_{s+1}^{i+1}\| \|\hat{E}_{s+1}^{i+1}\|}
\]  

(12)

where \( \hat{E}_{s+1}^{i+1} \) is the reconstruction using \( T_P P_T^T \).

Next, we introduce a coefficient \( \lambda \in [0, 1] \) to control the weight of the dynamic part and the static part, and the final merging cost is:

\[
S(i, i + 1) = \lambda S_S(i, i + 1) + (1 - \lambda) S_D(i, i + 1)
\]  

(13)

The segmentation algorithm is listed in the following.

### Algorithm 1 Multivariate Sequence Segmentation

1. Partition \( T \) into \( k \) equal length subsequences
2. while \( k > K_{target} \) do
3. for \( i = 1 \) to \( k - 1 \) do
4. Calculate merging cost \( S(i, i + 1) \) using eq. (13)
5. end for
6. Merge the two adjacent subsequences which having the largest merging cost
7. \( k \leftarrow k - 1 \)
8. end while
9. for each subsequence whose length > \( \kappa \) do
10. Equally divide the subsequence to have length \( \kappa \)
11. Discard the reminder if its length is smaller than \( \kappa_{min} \)
12. end for

V. LABEL PROPAGATION BASED SEMI-SUPERVISED LEARNING

In this section, we develop a semi-supervised learning strategy under the framework of label propagation. The central idea is to assign pseudo labels to unlabeled samples and train the classifier using all labeled and pseudo labeled data. The overall learning process alternatively carries out the label propagation step and the training step [15].

A. LABEL PROPAGATION

The label propagation operation is performed using the graph based method introduced in [<zhou2003>]. Given the training sample set \( X \), first we construct the affinity matrix \( W \) whose diagonal elements are zero and the \( w_{ij} \) for \( i \neq j \) which represents the similarity of sample \( x_i \) and \( x_j \). Then, we calculate the normalized graph Laplacian \( L = D^{-1/2}WD^{-1/2} \) where \( D \) is the diagonal degree matrix of \( W \). Let \( \hat{Y} \) be the label matrix estimated using the current classifier network trained using the available labeled training samples. The label spreading procedure proposed in [15] uses the following iteration

\[
\hat{Y}^{(t+1)} = \alpha L \hat{Y}^{(t)} + (1 - \alpha) \hat{Y}^{(0)}
\]  

(14)

and the above iteration equation converges to

\[
\hat{Z} = (I - \alpha L)^{-1} \hat{Y}^{(0)}
\]  

(15)

Instead of solving eq. (15), we solve the following linear system using conjugate gradient as it has been shown [30] to be more efficient.

\[
(I - \alpha L) \hat{Z} = \hat{Y}^{(0)}
\]  

(16)

Note that in the context of this article, each training sample is multivariate sequence. It is not appropriate to use Euclidean distance to define the similarity. In this work, we use the idea developed in the segmentation section to construct the affinity matrix. The merging cost function of eq. (13) is actually the accuracy of using the model trained with the first sequence to predict the other sequence. By its definition it can serve as the
similarity measure of two sequences. To make it symmetric, we define the following similarity measure
\[
\tilde{S}(i, i + 1) = \frac{S(i, i + 1) + S(i + 1, i)}{2}
\]  
(17)

The affinity matrix is defined as
\[
w_{ij} = \begin{cases} 
\tilde{S}(i, j), & \text{if } i \neq j \text{ and } x_i \in \text{NN}_k(x_j) \\
0, & \text{otherwise} 
\end{cases}
\]  
(18)

where \(\text{NN}_k\) is the set of \(k\) nearest neighbors in \(X\) which is introduced to make the affinity matrix sparse [31].

B. CLASSIFIER TRAINING

The network model consists of a sequence input layer, an LSTM layer, a fully connected layer and a softmax layer. We use the weighted form of loss function for the semi-supervised training process and it will be briefly introduced here. The loss function for the classifier training is defined as

\[
L(X, Y_L, \hat{Y}; \theta) = \sum_{l=1}^{L} L_{s}(f_{\theta}(x_l), y_l) + \sum_{i=1}^{n} w_{i} L_{s}(f_{\theta}(x_i), \hat{y}_i)
\]  
(19)

where \(L_{s}\) denotes the supervised loss function of network \(f_{\theta}\) with parameter \(\theta\) using labeled sample set \(Y\) and pseudo labeled sample set \(\hat{Y}\). The weight \(w_i\) of the training sample \(x_i\) measures the uncertainty of the pseudo label using entropy [31]

\[
w_i = 1 - \frac{\mathbb{E}[-\log(\text{Pr}(\hat{z}_i))] }{\log(c)}
\]  
(20)

The main training process includes two steps which are repeated for \(T\) epochs.

Algorithm 2 Semi-Supervised Training

1: Initialize the training sample set \(X\), the label matrix \(Y_L\) and the pseudo label matrix \(\hat{Y}_U^{(0)}\)
2: for \(t := 1 : T\) do
3: \hspace{1em} Train the network \(f_{\theta}\) using \(X\) and \(Y_L \cup \hat{Y}_U^{(t-1)}\) by minimizing eq. (19)
4: \hspace{1em} Perform the label propagation process to get \(\hat{Y}_U^{(t)}\) by solving eq. (16);
5: end for

It is worth noting that when calculating \(\tilde{S}\) it doesn’t need to learn the model parameters for each sequence as it has been done in the segmentation step. It only needs to cache the learned model of eq. (13) and use those to calculate the prediction accuracy. This can significantly reduce the computation demand in the label propagation step.

VI. EVALUATION WITH INDUSTRIAL DATA

A. DATASET PREPARATION

The data used for evaluating the proposed method were collected from the fused magnesia processing plant of Xinfazhan Ltd. of China. The original data are obtained with the industrial video surveillance camera system of RGB video with resolution of 1.3 million pixels (1280*1024) and frame rate of 25 frames per second. There are in total 22 hours of video selected for training and evaluation. We invited experienced operators of the FMF process plant to label all the video. There were 7 clips identified as semi-molten and the operators marked the beginning and the end of those occurrences.

Next, we extract the flame region near the furnace mouth making the ROI of dimension 1024*512*3 because the visual information of the semi-molten condition of the FMF concentrates on the flame area. Those images are further resized to a dimension of 128*64*3 and then fed into the SAE which consists of 6 pooling layers and each pooling layer connects to a convolution layer to smooth the image. Each frame of the encoded ROI image is then unrolled to a 64 dimensional feature vector.

The segmentation algorithm is used to segment the feature vector sequence extracted by SAE into subsequences. The segmentation operation is performed to the sequence clipped by a moving window of size 2000 and the segmentation algorithm terminates when the smallest subsequence is smaller than 250, which corresponds to clip length of 10 seconds.

FIGURE 4 shows an example of the segmented subsequences. Each of the stacked subfigures represents a feature and only some of the features are displayed here. The red vertical lines on the time axis represent division points found by the segmentation algorithm. Note that for subsequences whose length are longer than 250, we equally divide those into pieces with 250 length each and discard the fragments with length smaller than 50. Those are the green vertical lines shown in the figure. The short subsequence started at 648 and ended at 686 is discarded because it is shorter than the predefined threshold 50. As a result, for the example shown in the figure, we get 9 valid subsequences. After the segmentation operation, the whole training set includes 4320 labeled sequences and among those 330 samples are labeled as the semi-molten working condition.

B. EVALUATION RESULTS AND COMPARISON

For the network model we use a single layer LSTM controller, where the step length is 125, the input dimension is 64,
the number of the hidden layer units is 32 and the output dimension is 2. The gradient descent optimizer is used in training LSTM at a learning rate of 0.01. We choose and the training is carried out 300 epochs. We start training with 90% of the training samples being unlabeled. FIGURE 5 shows the prediction accuracy on the unlabeled samples during the training process sampled with every 30 epochs. The overall prediction accuracy is above 80% after 150 epochs and it reaches 85.8% after 300 epochs of training. The trained classifier is tested using the test dataset with 1296 samples and the confusion matrix is shown in Table 1.

Since the main idea of this article is to use dynamic segmentation to build the training set on the label propagation algorithm, we will evaluate the effects of introducing segmentation by comparison studies. In the following, the proposed method will be abbreviated as DynSeqLP (Dynamic Segmentation enhanced Label Propagation).

We implemented 5 other methods to compare with the proposed method. The first is the fully supervised method, which uses only labeled samples without label propagation, and it is served as the baseline. The rest methods are divided into 2 groups. The first group includes some of the most commonly used semi-supervised learning methods such as S3VM, self-training, co-training and entropy regularization (ER) methods [10]. Note that the methods in this group train the classifiers using the encoded feature vector directly without performing segmentation. The second group of methods train the LSTM classifiers using sequences as the training samples. It includes the method proposed in this article and a fixed length segmentation based label propagation method. The later one constructs the training samples by directly dividing the feature sequence into subsequences of equal length, where denotes the minimal length of the subsequence.

The later one constructs the training samples by directly dividing the feature sequence into subsequences of equal length, where denotes the minimal length of the subsequence. All the results of classification accuracy are listed in Table 2.

C. DISCUSSIONS
1) EXPLANATION OF THE RESULTS
We can see from Table 1 that the proposed method significantly outperform the other methods for comparison. There are sever reasons contributing to the improvements. First, the “patterns” in the furnace flame related to the semi-molten of FMF process are identified over continuous video. The poor performance of the algorithms in the first group is due to those are trained based on single image which is unable to learn the temporal structure in the video sequences. In the proposed method, the segmentation operation can be considered as a pre-clustering of the images subsequences having close temporal structure. By doing so, the granularity of the following label propagation operation is increased from image to video so that the LSTM model can effectively extract the temporal “patterns” presented in the video, and thus improve the classification accuracy.

Second, the significant improvements compared to the FixSeqLP method is due to the segmentation operation which ensures that each subsequence contains single working condition. Without using the dynamic segmentation approach, the samples constructed by arbitrarily dividing the sequence may result that one subsequence includes multiple working conditions. This will cause the training of LSTM network difficult to converge and having poor accuracy.

Finally, we will emphasize that the performance improvement observed is largely due to the adaptive segmentation method proposed in this article. The methods in group I for comparison have a common limitation that they strongly rely on the assumption that the unlabeled samples can be well separated by clusters. When using feature of single image, the highly noisy nature of the FMF flame image will cause it difficult to cluster the samples, which decreases the classification performance of those methods. However, using uniformly constructed sequences as training samples will not notably alleviate the problem because it may introduce mixture of conditions in one sequence as illustrated by the FixSeqLP method.

2) PARAMETERS
The parameters of the proposed approach including those related to the label propagation and the segmentation algorithm. The discussions of the former can be found in [15], [31] and we only discuss the later part. There are two parameters for segmentation: the minimal length of subsequence and the dynamic coefficient .

If we take a close look at Table 1 it can be seen that when the value of is between 250 and 300 its effects on the algorithm performance is not very noticeable. Its effects only becomes significant when is above 300. A particular visual pattern of the furnace flame of FMF process mostly lasts about 7-15 seconds. If the minimal length of segmentation is too large, it is difficult to keep the homogeneity in the subsequence and it will make consequent classifier training more difficult. In the experiment, we found that when , which corresponds to a video clip of 10 seconds, the algorithm achieves the best overall performance.

The dynamic coefficient provides a control of balancing between the contributions of cross-correlation and auto-correlation in the segmentation. The Former tries to make a
TABLE 2. Comparison of Classification Accuracy With Related Methods.

| Group   | 97% unlabeled | 95% unlabeled | 90% unlabeled | 80% unlabeled | 60% unlabeled |
|---------|---------------|---------------|---------------|---------------|---------------|
| Supervised | 0.323         | 0.3643        | 0.432         | 0.489         | 0.571         |
| Group I  |               |               |               |               |               |
| S3VM     | 0.792         | 0.706         | 0.802         | 0.848         | 0.892         |
| Self-training | 0.701     | 0.723         | 0.733         | 0.757         | 0.829         |
| Co-training | 0.686     | 0.708         | 0.731         | 0.764         | 0.836         |
| ER       | 0.735         | 0.747         | 0.770         | 0.803         | 0.873         |
| Group II |               |               |               |               |               |
| FixSeqLP |               |               |               |               |               |
| κ = 250 | 0.706         | 0.719         | 0.730         | 0.773         | 0.814         |
| κ = 300 | 0.689         | 0.691         | 0.702         | 0.779         | 0.790         |
| κ = 350 | 0.702         | 0.715         | 0.712         | 0.754         | 0.769         |
| κ = 400 | 0.673         | 0.709         | 0.692         | 0.722         | 0.763         |
| DynSeqLP |               |               |               |               |               |
| κ = 250 | 0.806         | 0.807         | 0.850         | 0.889         | 0.929         |
| κ = 300 | 0.811         | 0.807         | 0.842         | 0.877         | 0.905         |
| κ = 350 | 0.792         | 0.811         | 0.834         | 0.833         | 0.912         |
| κ = 400 | 0.720         | 0.743         | 0.757         | 0.790         | 0.835         |

linear approximation of each subsequence using the linear combination of the extracted latent variables. While the later locates those segmentation points where the dynamics pattern change. In the experiment, we choose $\lambda = 0.5$ so as to give equal weights to the two aspects.

VII. SUMMARY

In this article, we analyzed the problem of identifying the semi-molten working condition of FMF process. To tackle the problem of insufficient labeled samples, we introduce a semi-supervised classifier training strategy. First, the features of images are extracted by a convolution auto-encoder. Then, based on the data features, a dynamic multivariate segmentation method is proposed to segment the multivariate time series into subsequences which are served as samples for classifier training. The graph based label propagation is used to iterative spread the labeling information and train a LSTM classifier. Our method was compared with S3VM, co-training, self-training, entropy minimization algorithms and the label propagation method without using dynamic segmentation, showing significantly better accuracy than the other approaches. The proposed method was tested using industrial video data and showed promising results. The algorithm can be used to solve similar problems in a variety of video pattern identification applications.

One limitation of the proposed method is that the segmentation algorithm is based on linear approximation, which is fast and suit for online applications. However, this approach corresponds to the assumption that the temporal pattern of the same working condition can be linearly expressed and it might not be accurate in some cases. For more complex applications, it might be necessary to consider using nonlinear approximation methods to more accurately partition the feature sequence.

REFERENCES

[1] M. A. Shand, The Chemistry and Technology of Magnesia. Hoboken, NJ, USA: Wiley, 2006.
[2] Z. Wu, Y. Wu, T. Chai, and J. Sun, “Data-driven abnormal condition identification and self-healing control system for fused magnesium furnace,” IEEE Trans. Ind. Electron., vol. 62, no. 3, pp. 1703–1715, Mar. 2015.
[3] Y. Zhang and S. Li, “Modeling and monitoring of nonlinear multi-mode processes,” Control Eng. Pract., vol. 22, pp. 194–204, Jan. 2014.
[4] Z. Guo, K. Wang, S. Lu, and Z. Wu, “Image based semi-molten condition diagnosis system of fused magnesium furnace,” Control Eng. China, vol. 27, pp. 219–225, Feb. 2020.
[5] D. Wang and M. Li, “Stochastic configuration networks: Fundamentals and algorithms,” IEEE Trans. Cybern., vol. 47, no. 10, pp. 3466–3479, Oct. 2017.
[6] S. Lu and H. Gao, “Deep learning based fusion of RGB and infrared images for the detection of abnormal condition of fused magnesium furnace,” in Proc. IEEE 15th Int. Conf. Control Autom. (ICCA), Jul. 2019.
[7] X. Zhu and A. B. Goldberg, Introduction to Semi-Supervised Learning. Morgan & Claypool: San Rafael, CA, USA, 2009.
[8] W. Li and E. Keogh, “Semi-supervised time series classification,” in Proc. 12th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD). New York, NY, USA: Association for Computing Machinery, 2006, pp. 478–483.
[9] M. González, C. Bergmeir, I. Triguero, Y. Rodríguez, and J. M. Benítez, “Self-labeling techniques for semi-supervised time series classification: An empirical study,” Knowl. Inf. Syst., vol. 55, no. 2, pp. 493–528, May 2018.
[10] O. Chapelle, B. Schölkopf, and A. Zien, Eds., Semi-Supervised Learning. Cambridge, MA, USA: MIT Press, 2006.
[11] K. Murusy and K. Bura, “Success: A new approach for semi-supervised classification of time-series,” in Proc. Int. Conf. Artif. Intell. Soft Comput., 2013, pp. 437–447.
[12] Y. Chen, B. Hu, E. Keogh, and G. E. A. P. A. Batista, “DTW-D: Time series semi-supervised learning from a single example,” in Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD). New York, NY, USA: Association for Computing Machinery, 2013, pp. 383–391.
[13] H. A. Dau, N. Begum, and E. Keogh, “Semi-supervision dramatically improves time series clustering under dynamic time warping,” in Proc. 25th ACM Int. Conf. Knowl. Manage., Oct. 2016, pp. 999–1008.
[14] L. de Carvalho Pagliosa and R. F. de Mello, “Semi-supervised time series classification on positive and unlabeled problems using cross-recurrence quantification analysis,” Pattern Recognit., vol. 80, pp. 53–63, Aug. 2018.
[15] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schölkopf, “Learning with local and global consistency,” in Proc. Adv. Neural Inf. Process. Syst. 20, vol. 52, pp. 75–84, Apr. 2016.
[16] V. Badriarayanan, F. Galasso, and R. Cipolla, “Label propagation in video sequences,” in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., Jun. 2010, pp. 3265–3272.
[17] B. Wang and J. Tsotsos, “Dynamic label propagation for semi-supervised multi-class multi-label classification,” Pattern Recognit., vol. 52, pp. 75–84, Apr. 2016.
[18] K. T. Vasko and H. T. Toivonen, “Estimating the number of segments in time series data using permutation tests,” in Proc. IEEE Int. Conf. Data Mining, Dec. 2002, pp. 466–473.
[19] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York, NY, USA: Springer, 2009.

[20] Y. Su, Y. Zhao, C. Niu, R. Liu, W. Sun, and D. Pei, “Robust anomaly detection for multivariate time series through stochastic recurrent neural network,” in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2019, pp. 2828–2837.

[21] H. Ren, M. Liu, X. Liao, L. Liang, Z. Ye, and Z. Li, “Anomaly detection in time series based on interval sets,” *IEEE Trans. Electr. Electron. Eng.*, vol. 13, no. 5, pp. 757–762, May 2018.

[22] E. Keogh, S. Chu, D. Hart, and M. Pazzani, “An online algorithm for segmenting time series,” in *Proc. IEEE Int. Conf. Data Mining*, Nov. 2001, pp. 289–296.

[23] S. Spiegel, J. Gaehler, A. Lommatzsch, E. De Luca, and S. Albayrak, “Pattern recognition and classification for multivariate time series,” in *Proc. 5th Int. Workshop Knowl. Discovery From Sensor Data - SensorKDD*, 2011, pp. 34–42.

[24] L. Dobos and J. Abonyi, “On-line detection of homogeneous operation ranges by dynamic principal component analysis based time-series segmentation,” *Chem. Eng. Sci.*, vol. 75, pp. 96–105, Jun. 2012.

[25] A. Fathi, M. F. Balcan, X. Ren, and J. M. Rehg, “Combining self training and active learning for video segmentation,” in *Proc. Brit. Mach. Vis. Conf.*, 2011, pp. 1–11.

[26] M. A. Ranzato, Y. L. Boureau, and Y. LeCun, “Sparse feature learning for deep belief networks,” in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2007, pp. 1185–1192.

[27] Y. Dong and S. J. Qin, “A novel dynamic PCA algorithm for dynamic data modeling and process monitoring,” *J. Process Control*, vol. 67, pp. 1–11, Jul. 2018.

[28] S. Lu and S. Huang, “Segmentation of multivariate industrial time series data based on dynamic latent variable predictability,” *IEEE Access*, vol. 8, pp. 112092–112103, 2020.

[29] E. Keogh, S. Chu, D. Hart, and M. Pazzani, “Segmenting time series: A survey and novel approach,” in *Data Mining in Time Series Databases*, vol. 57. Singapore: World Scientific, Mar. 2004.

[30] S. Chandra and I. Kokkinos, “Fast, exact and multi-scale inference for semantic image segmentation with deep Gaussian crfs,” in *Computer Vision—ECCV*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham, Switzerland: Springer, 2016, pp. 402–418.

[31] A. Iscen, G. Tolias, Y. Avrithis, and O. Cham, “Label propagation for deep semi-supervised learning,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 5070–5079.

**SHAOWEN LU** (Member, IEEE) received the B.S. and M.S. degrees in computer science from Northeastern University, China, and the Ph.D. degree in electronic engineering from the Queen Mary University of London, U.K. He is currently a Professor with the State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University. His current research interests include industrial process modeling and simulation, and high performance industrial machine learning algorithms, especially focusing on industrial time series analysis, working with both data driven and model driven methods.