Process monitoring based on symbolic episode representation and hidden Markov models - A moving window approach

Nima Sammaknejad and Biao Huang*

***Abstract*— In this paper a new procedure for classification of normal and abnormal operating conditions of a process when multiple observation sequences are available is introduced. Signals are converted to discrete observations using the method of triangular representation. For overall classification of the process, the combinatorial method is used to train hidden Markov models when multiple observation sequences are available. The proposed methodology has the advantage of considering time varying window sizes for the input data. Therefore, a search algorithm to find the optimal window size at each time step is proposed. Application of the method is tested for overall monitoring of the continuous stirred tank reactor (CSTR) system. Results of overall classification are compared to the previous approach based on back propagation neural networks (BPNN) and show better performance in detecting normal and abnormal operating conditions.**

Keywords: Process Diagnosis, Hidden Markov models

I. INTRODUCTION

Process monitoring and fault diagnosis are two important steps in order to achieve safe operations. Three main approaches exist for process monitoring. The knowledge based approach which is based on qualitative models, the model based approach which is based on analytical models and the data driven framework which is appropriate for large multivariate processes [1]. Statistical approaches are among the most popular ones for the data-driven process monitoring of the multivariate processes. Simoglou et al. proposed process monitoring based on a state space model to capture the system dynamics where the parameters of the model were obtained based on canonical variate analysis (CVA) and partial least squares (PLS) [2]. Haiqing et al. improved fault detection based on the principal component analysis (PCA) by considering the relations between T^2 and squared prediction error statistics and their effect on the system parameters [3]. More details in this area is available in [4]. Pattern recognition techniques, which are based on the extraction of main features of the signals from plant data, are more popular among other knowledge based approaches [5].

Based on the method of triangular representation, initially introduced by Bakhshi et al. [6], Wong et al. (1998) developed a new strategy to detect abnormal trends from important qualitative information of a signal [7]. Later, Wong et al. (2001) introduced two different methodologies for overall classification of the process [8]. In the first scheme, observations from individual variables are classified based on separate hidden Markov models. Then, a symbolic string, called the sequence of events, is produced from this classification. A final HMM step is applied for the overall classification afterwards. In the second scheme, the probabilities generated from classification of individual variables are input to a back propagation neural network (BPNN) to decide on the overall classification. They show that the BPNN approach provides a better classification while increasing the computational time.

One of the drawbacks of their proposed methodology is the consideration of only individual variable effect and interactions between different inputs are not considered. Furthermore, time varying property of the process cannot be considered in the original approach. Finally, high computational time and a large number of components are required for overall classification. For example, at least four components (three HMMs and one BPNN) are required for the decision making having only three inputs available.

In this paper we propose to use a multivariate approach (Li et al. (2000)) to build the hidden Markov model for multiple observations [9]. Consequently, only one component is required for the overall decision making and interactions between different inputs are automatically considered. Furthermore, unlike the BPNN approach, time varying window sizes of the input data can be used to deal with time varying property of the process. Therefore, a search algorithm to find the optimal window size at each time step is proposed. Application of the new methodology is studied on the CSTR system developed by Henson et al. (1990) [10]. The new approach shows fewer false alarms in detection of abnormal behaviour of the process.

The remainder of this paper is organized as follows:
In section 2, the proposed methodology will be briefly introduced and compared to the previous approach. In section 3, the data pre-processing steps are summarized. Section 4 explains our proposed methodology in detail and compares it with the BPNN approach. Sections 5 and 6 include the case study, results and conclusion.

II. PROCESS MONITORING APPROACHES

Wong et al. (2001) introduce a methodology for classification of abnormal process operations when multiple observation sequences are available in [8]. First, they use wavelet analysis to remove the high frequency noise of the

Biao Huang is with Faculty of Chemical and Materials Engineering, University of Alberta, Edmonton, Alberta, Canada T6G 2V4 bhuang@ualberta.ca

Nima Sammaknejad is PhD candidate in the Department of Chemical and Materials Engineering, University of Alberta, Edmonton, Alberta, Canada T6G 2V4 sammakne@ualberta.ca

signals. Second, continuous time observations are discretized to a sequence of numbers based on the method of triangular representation and using appropriate fuzzy rules and membership functions. In the overall classification step, first, each variable is classified based on its corresponding HMM. These probabilities are then input to a BPNN for overall decision making.

In the proposed methodology of this paper, the data preprocessing steps are similar to Wong et al. (2001) which include wavelet analysis, fuzzification and triangular representation. However, in the critical modelling step, we adopt the multivariate hidden Markov modelling procedure [9], which results in significant performance improvement. Since the new methodology is flexible in adapting variable window sizes for the input data, a procedure to optimize the window size is also proposed. In comparison to the BPNN, the new approach helps in reducing the number of false alarms and the required computational time for overall decision making. The proposed fault detection method of this paper is presented in figure 1.

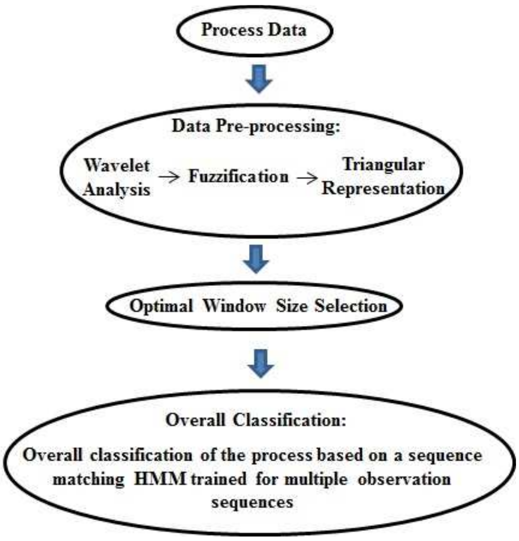

Fig. 1. The proposed methodology of this paper for overall classification of a process

The algorithm consists of the following steps: 1. Data pre-processing, 2. Window size selection and 3. Overall classification. These three steps will be presented respectively in the following sections.

III. DATA PRE-PROCESSING

A. Wavelet Analysis

Since typical process data is noisy, the first step in classification of a signal is data filtering. Several approaches exist to filter a signal, e.g., moving average, Gaussian filter, Fourier transform or wavelet analysis. Wavelet analysis uses a time-scale region and has excellent time-frequency properties. Furthermore, it gives a multi-scale description of trends and features which enables us to analyse the data more efficiently. Therefore, wavelet analysis is also used in this paper. More information in this area can be found in [11].

B. Feature Extraction

The method of triangular representation by Bakhshi et al. (1994) converts a continuous time signal into a sequence of symbolic discrete observations which capture the most important qualitative and quantitative information contained in the signal [6]. The symbolic form of observations is appropriate as the input of a classifying system such as hidden Markov models. This method is based on the idea that at extrema and inflection points, the first and second derivatives are zero respectively. They named any part of the signal with a constant sign of first and second order derivatives as an episode. These episodes are generated after filtering the signal and contain an extremum with a neighboured inflection point which makes a triangle. Seven types of triangles, e.g. A, B, C, D, E, F and G, can be defined as illustrated in figure 2.

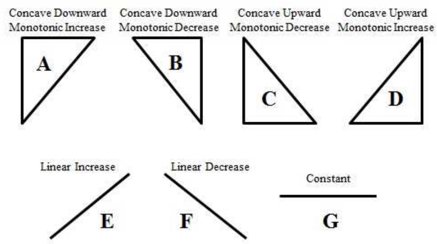

Fig. 2. Seven types of triangles for triangular representation of a signal [7]

In addition to the qualitative information of the episode, semi-quantitative information such as duration and magnitude is also used for classification. The duration is defined as the time interval between two end points of an episode. The magnitude is the vertical difference between those two end points. Using fuzzy classification, three kinds of magnitudes (large, medium and small) and three kinds of durations (long, middle and short) will be generated. Figure 3 shows the results of classification of an arbitrary type triangle, A, to nine sub-types, using fuzzy classification. In fuzzy classification, elements of a fuzzy set have a degree of membership in the unit interval [0,1], while the membership of elements in a classical set is only defined by binary terms 0 and 1 [12].

Straight lines in figure 2 are limited to very smooth trends and can be approximated with triangles of the same type. Therefore, the total possible number of discrete observations is equal to $4 \times 9 = 36$.

IV. DATA CLASSIFICATION FOR MULTIPLE OBSERVATION SEQUENCES

A. BPNN Approach

Wong et al. (2001) introduce the BPNN approach as a classification method when multiple observation sequences are available [8]. In this method, each variable is classified

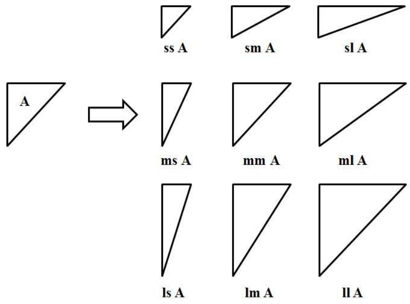

Fig. 3. Nine sub-types of triangle "A" using fuzzy classification (Wong et al. (1998)) [7]

separately based on its corresponding HMM. Probabilities generated from classification of each individual variable are then input to a BPNN for the overall classification.

B. Multivariate HMM modeling for Multiple Observation Sequences

Since process variables are correlated, the sequence resulted from the triangular representation must also be correlated. It is natural to adopt multivariate modeling approach. Li et al. (2000) has laid theoretical foundation on multivariate HMM modelling, which is adopted here [9].

Consider the following set of observation sequences

$$O = \{O^{(1)}, O^{(2)}, \dots, O^{(K)}\} \tag{1}$$

where

$$O^k = o_1^{(k)} o_2^{(k)} \dots o_{T_k}^{(k)},\ 1 \le k \le K \tag{2}$$

are individual observation sequences.

To calculate the probability of the observation sequence given the model, the following expression is always true from the chain rule:

$$\begin{aligned} P(O|\lambda) &= P(O^{(1)}|\lambda)P(O^{(2)}|O^{(1)},\lambda)\dots P(O^{(K)}|O^{(K-1)},\dots,O^{(1)},\lambda) \\ P(O|\lambda) &= P(O^{(2)}|\lambda)P(O^{(3)}|O^{(2)},\lambda)\dots P(O^{(1)}|O^{(K)},\dots,O^{(2)},\lambda) \\ &\dots \\ P(O|\lambda) &= P(O^{(K)}|\lambda)P(O^{(1)}|O^{(K)},\lambda)\dots P(O^{(K-1)}|O^{(K)},O^{(K-2)},\dots,O^{(1)},\lambda) \end{aligned} \tag{3}$$

where λ is the model. Therefore, the probability of the multiple observations given the model can be expressed as:

$$P(O|\lambda) = \sum_{k=1}^{K} \omega_k P(O^{(k)}|\lambda) \tag{4}$$

where

$$\begin{aligned} \omega_1 &= \frac{1}{K} P(O^{(2)}|O^{(1)},\lambda)\dots P(O^{(K)}|O^{(K-1)},\dots,O^{(1)},\lambda) \\ \omega_2 &= \frac{1}{K} P(O^{(3)}|O^{(2)},\lambda)\dots P(O^{(1)}|O^{(K)},\dots,O^{(2)},\lambda) \\ &\dots \\ \omega_K &= \frac{1}{K} P(O^{(1)}|O^{(K)},\lambda)\dots P(O^{(K-1)}|O^{(K)},O^{(K-2)},\dots,O^{(1)},\lambda) \end{aligned} \tag{5}$$

are weights. Considering the serial independence assumption within the individual observation sequences, the formulation can be simplified to:

$$P(O|\lambda) = \prod_{k=1}^{K} P(O^{(k)}|\lambda) \tag{6}$$

and the combinatorial weights become:

$$\omega_k = \frac{1}{K} \frac{P(O|\lambda)}{P(O^{(k)}|\lambda)},\ 1 \le k \le K \tag{7}$$

Implementing the EM algorithm with calculation of the auxiliary function at the E-step and maximization over the auxiliary variable at the M step, the re-estimation formulas to train HMMs for multiple observation sequences can be obtained as follows:
State transition probability is

$$\bar{a}_{mn} = \frac{\sum_{k=1}^{K} \sum_{t=1}^{T_k-1} \xi_t^{(k)}(m,n)}{\sum_{k=1}^{K} \sum_{t=1}^{T_k-1} \gamma_t^{(k)}(n)},\ 1 \le m \le N, 1 \le n \le N \tag{8}$$

where $\xi_t^{(k)}(m,n)$ is the probability of the new observation being in state m at time t and in state n at time $t+1$ for the k^{th} observation sequence and $\gamma_t^{(k)}(n)$ is the probability of the new observation being in state n at time t for the k^{th} observation sequence. Further information regarding the calculation of $\xi_t^{(k)}(m,n)$ and $\gamma_t^{(k)}(n)$ can be found in [9].
Symbol emission probability is

$$\bar{b}_n(m) = \frac{\sum_{k=1}^{K} \sum_{t=1}^{T_k} \delta(o_t^{(k)}, v_m) \gamma_t^{(k)}(n)}{\sum_{k=1}^{K} \sum_{t=1}^{T_k-1} \gamma_t^{(k)}(n)},\ 1 \le m \le N, 1 \le n \le N \tag{9}$$

where $\delta(o_t^k, v_m) = 1\ if\ o_t = v_m\ and\ 0\ otherwise.$
Initial state probability is

$$\bar{\pi}_n = \frac{1}{K} \sum_{k=1}^{K} \gamma_1^{(k)}(n),\ 1 \le n \le N \tag{10}$$

V. Optimal Window Size Selection

Although BPNNs are limited to fixed window sizes, time-varying input dimensions can be considered at each time step using hidden Markov models [7]. Therefore, a search algorithm to find the optimal window size is proposed in this section. Unlike the previous approaches [13], this algorithm is looking for a fixed episode of more informative observations in the window.

Selection of large window sizes has the drawback of remaining in transition zones for large time intervals where no decision can be made on the operating condition of the system. The minimum number of observations (N_{min}) required to thoroughly explain the operating condition might differ according to the level of noise removal, on-line sampling rate, etc. Although N_{min} contains a window of most recent observations, making the final decision only based on N_{min} may cause many false alarms and affect critical decisions. One solution to this problem is searching for N_{min} number of observations in a window of most recent data (N_W). Intuitively, using the proposed methodology in figure

5, we are looking for a small window of observations which maximizes the difference of the likelihood given each model. Therefore, assuming $e_{max} = N_W - N_{min}$,

$$e_{opt} = argmax_{e\in[0,e_{max}]}\{(P(O|\lambda_{Normal}) - P(O|\lambda_{Abnormal}))^2|_{O=O(\tau+e:\tau+e+N_{min})}\} \quad (11)$$

where $\tau = t - N_W$ and $O = O(\tau + e : \tau + e + N_{min})$ is the optimal episode of observations in the window. $P(O|\lambda)$ is calculated from the forward-backward algorithm [14].

Using the search algorithm in (11), we are looking for an episode of N_{min} observations which best classifies the normal and abnormal operations in the window of N_W observations. A schematic of the proposed algorithm is illustrated in figure 4.

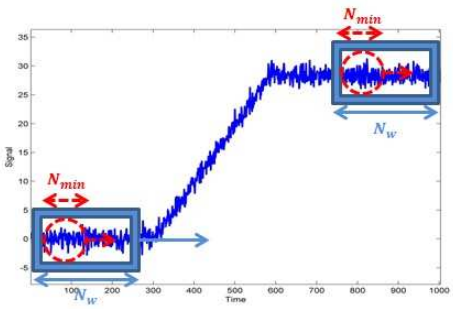

Fig. 4. Optimal window size selection

A summary of the proposed algorithm (4.*B* and 5) is presented in figure 5.

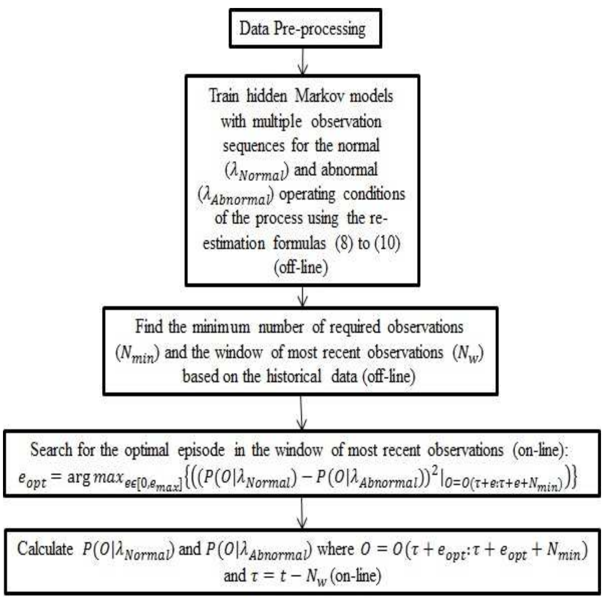

Fig. 5. Summary of the proposed methodology of this paper

VI. CASE STUDY

A. Two CSTRs in Series

Application of the proposed method is studied on the CSTR system of Henson et al. [10]. The irreversible exothermic first order reaction $A \rightarrow B$ occurs in two CSTR reactors in series. The product of the first reactor is the feed to the second reactor and a parallel flow (q_C) is used as the coolant. A schematic view of the system is shown in figure 6.

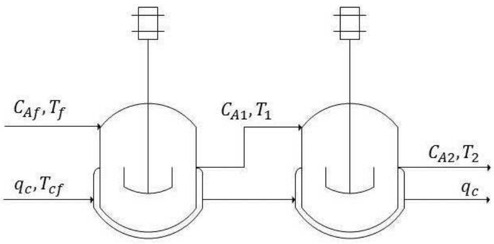

Fig. 6. Two CSTRs in Series [10]

B. Normal and Abnormal Operating Conditions

The feed flow rate q_f is used as the disturbance to the process. The outputs of the process (C_{A2} and T_2) are selected as the indicators of the abnormality in the process. The abnormal behavior in the process outputs is due to the sudden changes in the feed flow-rate (a pulse disturbance with the amplitude $15(L/min)$ and period of five sample times in time steps between $610-1210$) which cause overshoots in process outputs. The system starts at the initial value of C_{A2} equal to $0.05(\frac{mol}{L})$ and the set-point is selected as $C_{A2} = 0.075(\frac{mol}{L})$. A PI controller with parameters given as $\tau_I = 0.25(min)$ and $K_C = 350(\frac{L^2}{mol.min})$ is implemented as explained by Henson et al. (1990). It is also assumed that q_f always has a white noise disturbance with variance $0.5(\frac{L}{min})$. Normal and abnormal process operations are shown in figure 7. The disturbance is illustrated in figure 8.

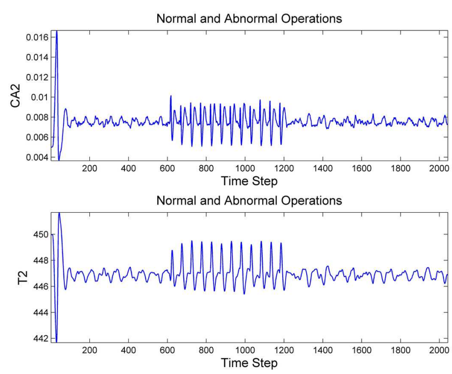

Fig. 7. Normal and abnormal process operations and the indicator variables

Other model parameters used in simulation are presented in table I.

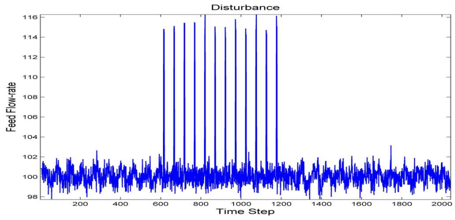

Fig. 8. Disturbance to the process

TABLE I
SIMULATION PARAMETERS [10]

$q = 100(L/min)$	$k_0 = 7.2 \times 10^{10} min^{-1}$
$C_{Af} = 1(mol/L)$	$E/R = 1 \times 10^4 K$
$T_f = 350K$	$-\triangle H = 4.78 \times 10^4 J/mol$
$T_{Cf} = 350K$	$\rho, \rho_C = 1000g/L$
$V_1, V_2 = 100L$	$C_P, C_{PC} = 0.239J/g.K$
$UA_1, UA_2 = 1.67 \times 10^5 J/min.K$	

VII. RESULTS AND DISCUSSION

A. Triangular Representation

A combination of normal and abnormal operating regions after reaching the desired set-point is presented in figure 9.

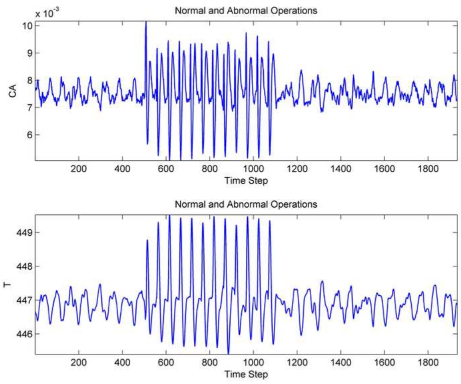

Fig. 9. Normal and abnormal operations after reaching the desired set-point

After removing the high frequency noise in two levels using wavelet analysis and normalizing the data, the minimum, maximum and inflection points of the signals are calculated. Then, using appropriate fuzzy membership functions and rules for durations and magnitudes of the signals, they are converted to discrete observations. Peaks of the membership functions are selected as minimum, maximum and average of the magnitudes and durations. Standard deviation of the membership functions is a linear function of the standard deviation of durations and magnitudes. Discretized observations of the output concentration and temperature signals are presented in figures 10 and 11.

The abnormal region is the area between two direct red lines. As it is clear from these two figures, larger size triangles are generated when the process enters the abnormal region

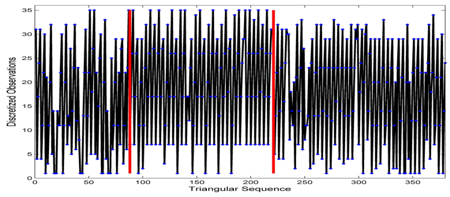

Fig. 10. Discritized observations for the output concentration

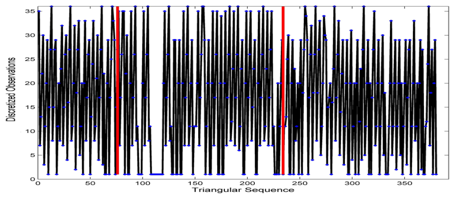

Fig. 11. Discretized observations for the output temperature

B. Large Window of Input Data (non-adaptive window size)

Large window of input data for process classification provides similar results between BPNN and the proposed method of this paper. Figures 12 and 13 compare the results of the two approaches considering 10 last observations as the input to the classification algorithms. A total number of 772 discrete observations, including 579 observations for normal and 193 for abnormal regions, are used to train the models of the normal and abnormal operating conditions. In each data set, 2/3 of the data is used for training and 1/3 is used for validation. The number of HMM hidden states, which is usually selected as the average number of symbols in a sequence of the training set, is equal to 8 in this study. The number of neurons in the hidden layer of the neural network is selected equal to be equal to 6 in the BPNN approach.

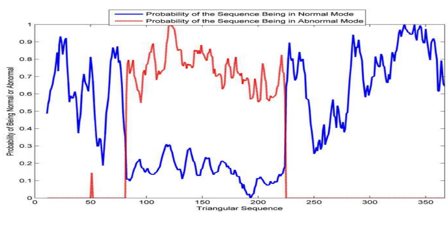

Fig. 12. Normalized probability of the observation sequences belonging to normal and abnormal regions using HMMs with fixed window of data ($N_W = N_{min} = 10$)

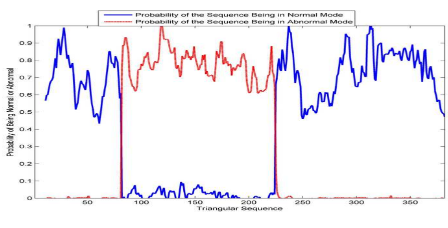

Fig. 13. Normalized probability of the observation sequences belonging to normal and abnormal regions using the BPNN approach ($N_W = N_{min} = 10$)

As presented in figures 12 and 13, with a large window of the input data, both approaches provide similar results.

C. Small Window of Input Data (adaptive window size)

As explained in section 5, using large window sizes, a long time will be required for any classification algorithm to capture the most recent behavior of the process due to a large amount of old data in the window. Consequently, the classification algorithm remains in transition zones where no decision can be made on the operating condition of the system. Furthermore, by training a hidden Markov model for multiple observation sequences directly, interactions between different inputs are considered. Figures 14 to 16 present the results of the overall classification with 5 observations as the input of the classification system.

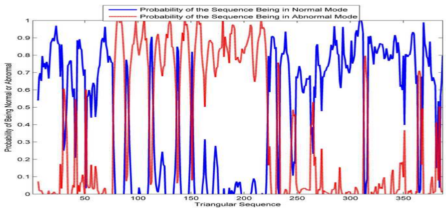

Fig. 14. Normalized probability of the observation sequences belonging to normal and abnormal regions using the BPNN approach ($N_W = N_{min} = 5$)

As it is clear from figure 14, a large number of false alarms appear when the window size is reduced to half using the BPNN approach.

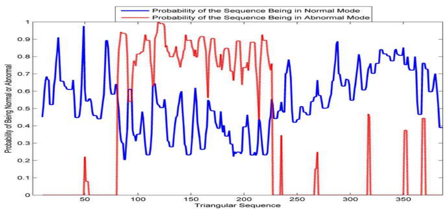

Fig. 15. Normalized probability of the observation sequences belonging to normal and abnormal regions using the proposed methodology of this paper ($N_W = 9, N_{min} = 5$)

Figure 15 shows the result of overall classification based on the proposed method. The number of false alarms is reduced. However, the likelihood ratio is decreased. In other words, the overall decision making is improved while the individual effect of each variable is reduced. The number of shifts to find the optimal window (e_{opt}) is presented in figure 15. This number varies between 1 and $e_{max} = N_W - N_{min} = 9 - 5 = 4$ and indicates the $O_{opt} = O(\tau + e_{opt} : \tau + e_{opt} + N_{min})$ sequence of observations which are selected for overall classification.

VIII. CONCLUSIONS

In this paper an improved method for multivariate symbolic based classification of the process is introduced and compared to the previous approaches. Results demonstrate that the new approach shows fewer false alarms. However, the final decision will be made based on smaller likelihood ratios. Following the proposed procedure, interactions between inputs will be automatically considered and abnormalities due to the effect of each single variable (which could often be an outlier or noise) will be avoided. Furthermore, in comparison to the BPNN approach, only one component (one hidden Markov model trained for the overall classification) is required for decision making and computation time is greatly reduced. Finally, in the new procedure, the optimal window of input data can be found using the introduced search algorithm. Therefore, the final decision is based on the more informative observations in the window and the old information will not affect the overall decision making.

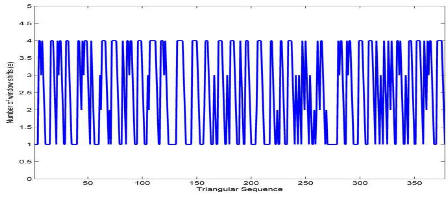

Fig. 16. Number of shifts to find the optimal window

REFERENCES

[1] S. Verron, J. Li, T. Tiplica. Fault detection and isolation of faults in a multivariate process with Bayesian network, Journal of Process Control, 20, 902-911, 2010.

[2] A. Simoglou, E. B. Martin, A. J. Morris. Statistical performance monitoring of dynamic multivariate processes using state space modelling, Computers and Chemical Engineering, 26, 909-920, 2002.

[3] H. Wang, Z. Song, P. Li. Fault detection behavior and performance analysis of principal component analysis based process monitoring methods, Industrial Engineering and Chemistry Research, 41, 2455-2464, 2002.

[4] S. J. Zhao, J. Zhang, Y. M. Xu. Performance monitoring of processes with multiple operating modes through multiple PLS models, Journal of Process Control, 16, 763-772, 2006.

[5] M. F. S. V. D Angelo, R. M. Palhares, R. H. C. Takahashi, R. H. Loschi. Fuzzy/ Bayesian change point detection approach to incipient fault detection, IET Control Theory Appl, 5, 539-551, 2011.

[6] B. R. Bakhshi, G. Stephanopoulos. Representation of process trends - Part III, Multi-scale extraction of trends from process data, Computers chem. Eng, 18, 267-302, 1994.

[7] J. C. Wong, K. A. McDonald, A. Palazoglu. Classification of process trends based on fuzzified representation and hidden Markov models, Journal of Process Control, 8, 395-408, 1998.

[8] J. C. Wong, K. A. McDonald, A. Palazoglu. Classification of abnormal plant operation using multiple process variable trends, Journal of Process Control, 11, 409-418, 2001.

[9] X. Li, M. Parizeau, R. Plamondon. Training hidden Markov models with multiple observations - A combinatorial method, IEEE Transactions on Pattern Analysis and Machine Intelligence, 22, 371-377, 2000.

[10] M. A. Henson, D. E. Seborg. Input-output linearization of general nonlinear processes, AIChE, 36, 11, 1990.

[11] S. G. Mallat. A theory for multi-resolution signal decomposition: the wavelet representation, IEEE Transactions on Pattern Analysis and Machine Intelligence, 11, 674-693, 1989.

[12] L. A. Zadeh, Fuzzy sets, Information and Control, 8, 338-353, 1965.

[13] W. Sun, A. Palazoglu, J. A. Romangoli. Detecting abnormal process trends by wavelet-domain hidden Markov models, AIChE, 49, 1, 2003.

[14] L. R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition, Proceedings of the IEEE, 77, 2, 1989.