

Fault Diagnosis Using Dynamic Time Warping

Rajshekhhar², Ankur Gupta², A.N. Samanta², B.D. Kulkarni^{1,*},
and V.K. Jayaraman^{1,*}

¹ Chemical Engineering Division, National Chemical Laboratory, Pune-411008, India
{bd.kulkarni,vk.jayaraman}@ncl.res.in

² Department of Chemical Engineering, Indian Institute of Technology,
Kharagpur-721302, India

Abstract. Owing to the superiority of Dynamic Time Warping as a similarity measure of time series, it can become an effective tool for fault diagnosis in chemical process plants. However, direct application of Dynamic Time Warping can be computationally inefficient, given the complexity involved. In this work we have tackled this problem by employing a warping window constraint and a Lower Bounding measure. A novel methodology for online fault diagnosis with Dynamic Time Warping has been suggested and its performance has been investigated using two simulated case studies.

1 Introduction

The process deviation from the normal operating range leads to deterioration in product quality and can be source of potential hazard. The control of such deviations comes under abnormal event management (AEM) in chemical process industry. The first step in AEM consists of timely detection and diagnosis of fault, so that it can lead to situation assessment and planning of supervisory decisions to bring the process back to a normal and safe operating state. However due to the size and complexity involved in the modern process plants, traditional method of complete reliance on human operators has become insufficient and unreliable. The advent of computer based control strategies and its success in process control domain has lead to several automated fault diagnosis methodologies.

Currently available fault diagnosis techniques can be classified into three broad categories: quantitative model based, qualitative model based and process history based approaches. In this work, a novel process history based approach for fault detection has been proposed. It employs the concept of Dynamic time warping (DTW) for the similarity measurement. Direct application of DTW leads to poor computational efficiency of the methodology. This problem has been rectified in this work by using window warping constraint in DTW with the application of lower bounding technique. We demonstrate the efficiency of our proposed methodology by performing online fault diagnosis on two simulated case studies.

* Corresponding author.

2 Dynamic Time Warping

Consider two multivariate time series Q and C , of length n and m respectively,

$$Q = q_1, q_2, q_3, \dots, q_i, \dots, q_n \quad (1)$$

$$C = c_1, c_2, c_3, \dots, c_j, \dots, c_m \quad (2)$$

such that, $q_i, c_j \in \mathbb{R}^p$. Since the DTW measure is symmetric with respect to the order of the two time series, without any loss of generality we can assume that $n \geq m$ in our work. To align these two sequences using DTW, we construct a $n \times m$ matrix where the (i^{th}, j^{th}) element of the matrix corresponds to the squared distance,

$$d(i, j) = \sum_{r=1}^{r=p} (q_{i,r} - c_{j,r})^2 \quad (3)$$

In order to find the best match between these two sequences, one finds a path through the matrix that minimizes the total cumulative distance between them. Such a path will be called a warping path. A warping path W is a contiguous set of matrix elements that characterizes a mapping between Q and C . The k^{th} element of W is defined as $W(k) = (i, j)_k$. The time-normalized distance [1] between the two time series is defined over the path as,

$$DTW(Q, C) = \min_W \left[\sqrt{\frac{\sum_{k=1}^{k=K} d(W(k)) \cdot \phi(k)}{\sum_{k=1}^{k=K} \phi(k)}} \right] \quad (4)$$

Where, $\phi(k)$ is the non-negative weighting co-efficient and K is the length of the warping path W , which satisfies the condition,

$$\max(m, n) \leq K \leq m + n - 1 \quad (5)$$

The normalization is done to compensate for K , the number of steps in the warping path W , which can be different for different cases. The symmetric normalization, owing to its established superiority [1] has been used for purpose, given as,

$$\phi(k) = (i(k) - i(k-1)) + (j(k) - j(k-1)) \quad (6)$$

The warping path however, is constrained to the following conditions [2]:

Boundary conditions: The path must start at $W(1) = (1, 1)$ and end at $W(K) = (n, m)$.

Continuity and monotonic condition: From a point (i, j) in the matrix, the next point must be either $(i, j+1)$, $(i+1, j+1)$ or $(i+1, j)$.

Warping Window condition: In case of equal length time series, the path as we know intuitively should not wander too far from the diagonal. For the case

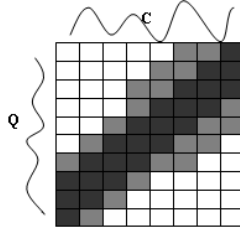


Fig. 1. A sample warping window band. Note the two regions in the band, in black and grey.

of unequal length time-series, a unique diagonal does not exist, thus we define a region analogous to the diagonal, which is bounded by the two lines whose equations are given by (7).

$$i = j \ \& \ i = j + n - m \text{ where } 1 \leq j \leq m \text{ for } n \geq m \quad (7)$$

Figure 1 shows an example of such a region for the case of $n \geq m$. This region, analogous to the diagonal has been shown in darker (black) shades in the figure. It should be noted that this region reduces to a diagonal, in case the two time series are of equal length. In order to restrict the warping path from straying away from the diagonal (or defined-region for unequal length time-series), we limit the distance by which it can wander away from the diagonal/defined-region. This limits the warping path to a warping-window or a band, which allows the warping path a distance (say) R_i to wander directly above and to the right of the diagonal/defined-region. This definition of band makes it symmetric with respect of the diagonal/defined-region. This is a generalization of the warping window originally proposed by [2].

We now mathematically define the band as,

$$j - R_i \leq i \leq j + R_i + (n - m) \text{ for } n \geq m \quad (8)$$

Where, $R_i = d$ such that $0 \leq d \leq \max(m)$, and $1 \leq i \leq n$. $\max(m)$ is the length of the longest time-series in the data available, and R_i is the permissible range of warping above and to the right of the region defined in (7). In the case where the time series are univariate and of equal length i.e. $n = m$, the band reduces to the R-K band as defined in [2].

In general, the above definition of band allows the formulation of any arbitrary shaped band. However, we are employing the bands for online fault diagnosis purpose and hence the temporal variation in bandwidth should be avoided. Hence, for current work, R_i has been considered independent of i . In order to solve the optimization problem given by (4), we employ the dynamic programming technique as done in [1,2]. For this cumulative distance matrix is calculated [1,3],

$$D(i, j) = \min \begin{cases} D(i-1, j) + d(i, j) \\ D(i-1, j-1) + 2 \cdot d(i, j) \\ D(i, j-1) + d(i, j) \end{cases} \quad (9)$$

This takes care of the continuity and monotonic conditions along with the path normalization. This recursive formula for dynamic programming enables us to find the warping path W , along which the values of $d(i, j)$ can be summed up, and the final value of DTW distance can be found out as given by (4). However, when the warping window constraint is employed, (9) admits only those points which lie inside the warping window or the band. Thus, only those points are eligible for consideration in (9) which satisfy (8). We can see that application of warping window constraint drastically reduces the number of paths that are needed to be considered, thus speeding up the DTW calculation. For a detailed introduction to DTW, lower bounding measures, warping windows and their application to classification we refer the readers to [1,2,3].

3 Lower Bounding Measures

Computation of DTW distances is computationally very expensive, consequently a classification algorithm based on DTW distance as a similarity measure is bound to be computationally inefficient. This problem has been circumvented by using a fast lower bounding measure, which saves unnecessary calculations by pruning off the time-series which cannot be nearest to the given time-series [3]. The lower bounding measure is obviously computationally cheaper than the actual DTW calculation, but gives a fairly tight approximation of the DTW distance. The original method has already been successfully applied to the case of time series of equal length [2]. In this work, we present a modification of the lower bounding measure which can be applied to time series of unequal lengths, which we will call as LB_UMV (stands for Lower-Bound Unequal Multivariate).

3.1 Lower-Bounding Measure (LB_UMV)

Let us consider the two time-series Q and C defined in (1) and (2). Using the global constraints on the warping path given by (8), we construct an envelope using the warping window across Q bounded by two time-series U and L given as,

$$u_{j,r} = \max(q_{j-R,r} : q_{j+R+(n-m),r}) \quad (10)$$

$$l_{j,r} = \min(q_{j-R,r} : q_{j+R+(n-m),r}) \quad (11)$$

Using the above definitions of U and L , we define LB_UMV as,

$$LB_UMV(Q, C) = \sqrt{\frac{1}{(m+n-1)} \sum_{j=1}^{j=m} \sum_{r=1}^{r=p} \begin{cases} (c_{j,r} - u_{j,r})^2 & \text{if } c_{j,r} > u_{j,r} \\ (l_{j,r} - c_{j,r})^2 & \text{if } c_{j,r} < l_{j,r} \\ 0 & \text{otherwise} \end{cases}} \quad (12)$$

It is important to note that the lower bounding measure defined in (12) for the two time series will always be lower than the corresponding DTW distance. A mathematical proof of this has been provided in [3] for the case of equal length time series. The same can be proved for the current definition of band by following a similar approach.

4 Learning Multiple Bands for Classification Using Heuristics

A lower value DTW distance implies that the given time series are similar to each other, while a large value implies dissimilarity between the two. In a classification task, one usually compares a test sample with the samples from the reference dataset, and then assigns a label to the test sample of the class to which it resembles the most. In this work, both the test and reference samples are time series and we utilize the concept of DTW distance to define the similarity between them. Each class in the dataset will have a warping window or a band of its own, which will be utilized when the test time-series is compared to a reference time-series of that class. Our aim is to automatically find these bands using the reference dataset. Similar to approach employed by [2], we pose this problem as a search problem and make use of the generic heuristic techniques to obtain bands. The search can be performed as either forward or backward hill-climbing search algorithm, with the aim of maximizing a heuristic function. An intuitive selection for the heuristic function is the accuracy of classification, which is defined as the percentage of correct classifications obtained by using a given set of bands. A backward search begins with maximum bandwidth, above and to the right of the region defined in (7); the corresponding accuracy is evaluated. The bandwidth of these bands is reduced and the accuracy is re-evaluated. The change is accepted if an improvement in accuracy is registered. The reduction in bandwidth is continued until a reduction in accuracy is observed. This is done until we reach zero bandwidth.

5 Online Fault Diagnosis

We follow the methodology described in Section 4 to obtain a set of trained bands (one for each class) by using the reference dataset, which is then used to compute DTW distances wherever necessary. A process plant continuously provides us with a multivariate time-series in the form of measured variables at some time intervals. We utilize the DTW concept for the purpose of online fault diagnosis by using a window wise pattern matching approach. We select a window from the current sequence of measured variables, which is used as a test sample to compare with windows (of nearly equal length) from the reference dataset. This allows us to predict the nature of the current sequence of variables to be faulty or not as well as the type of fault. The selection of the current and reference windows has been illustrated in Figure 2.

The approach consists of three major steps:

1. *Selection of Current Window:* A window is selected from the current batch of measured variables obtained from process plant, which is then matched against similar windows from the reference batches. There are three parameters involved in this step: Initial point (IP), Window Width (WW) and Window Shift (WS). We start selecting windows from an Initial Point (IP).

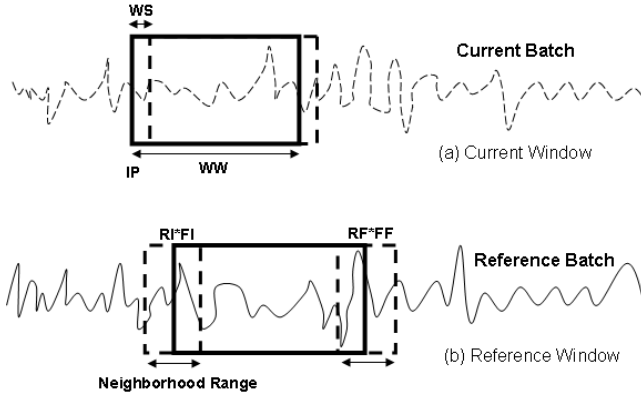


Fig. 2. (a) Selection of Current Window, the current window shifts by WS . (b) Selection of Reference Window, the reference window can be of various sizes due to variation in the neighborhood on both sides.

Initial Point decides the reliability of initial transient part in identifying the fault type. Window Width is the size of the current window and Window Shift is the number of time points by which the window shifts. Every current window acts as a test sample which is then compared with windows from the reference dataset in order to determine the fault.

2. *Selection of Reference Window:* The main parameter involved in this step is the window location in a particular reference time series. We will identify this window with initial point RI and final point RF . However, in order to counter the effect of different or uneven sampling time, these parameters will be updated by two other parameters, FI and FF respectively. A more effective independence from the sampling variation can be obtained by allowing the RI and RF to deviate in the neighborhood. This can be achieved by introducing two parameters, number of neighboring points (N) by which the RI and RF can deviate and the time (n) by which these points are apart. In a particular batch of the reference dataset, we consider all the possible windows as a match against the current window and nearest window is selected as the reference window. This will also update the FI and FF parameters.
3. *Fault Diagnosis:* After the current window and corresponding reference windows are selected, we rank the reference windows in increasing order of their distances from the current windows, such that the nearest window is ranked first. The current batch is expected to be of the same class to which its nearest neighbor belongs to. However, instead we follow a more robust probability based approach to identify the class of the current batch as described subsequently. If the assigned class belongs to a fault type, the process is likely to be faulty. But if it belongs to normal class, the process is considered to be operating normally and we modify the current window by shifting it by WS , and repeat the classification procedure. However, it is possible that

a current window is falsely assigned a faulty class. In order to suppress such false alarms, another parameter viz. Fault Count (FC) has been introduced. This parameter requires the current batch to be assigned a faulty class for a given number of times (given by FC) before it is declared faulty. It should be noted that the reference dataset is first used to generate trained bands and then later to compare with current windows, which are obtained from the test dataset.

Similarity Measure: To make the similarity search more robust, a probability based method has been used. The probability value will indicate the chances of current time series to be a faulty or not. The similarity measurement has been performed by finding 50 best matched reference windows and based on this set a probability function was defined as,

$$P_i = \frac{\sum_{\substack{j=1 \\ \text{s.t. } j \in \text{class } i}}^{50} \frac{1}{r_j \times d_j}}{\sum_{j=1}^{50} \frac{1}{r_j \times d_j}} \quad (13)$$

Where, P_i is the probability of the current batch to belong to class i (i.e. particular fault or normal type), r_j and d_j are rank and distance measure with current time series of j^{th} reference time series respectively.

6 Simulated Case Studies

6.1 Case Study 1: Batch Acetone-Butanol Fermentation

This case study employs the mathematical model of batch acetone-butanol fermentation, originally proposed by Votruba et al. [4]. This model has been further explored by Singhal [5]. He introduced different operating parameters consisting of one normal operation and four abnormal operations.

Generation of historical database: Each abnormal operating condition was characterized by an abnormal value of a cell physiology parameter. The magnitude of the abnormal parameter varied randomly from batch to batch. The duration of each batch was 30 h and sampling interval was randomly varied from 9 to 12 minutes. The operating conditions and parameter ranges can be found in [5]. Each of the five operating conditions was simulated 60 times to provide a historical database of 300 batches. Gaussian measurement noise was added to the measured variables so that the signal-to-noise ratio for a normal batch run was equal to ten. These batch profiles were divided into two subsets; a set of 250 profiles (50 profiles per class) was taken as the reference dataset and another set of 50 profiles (10 profiles per class) was used for the online diagnosis. The reference dataset was first used to generate trained bands for each class, and then later to compare with the current window in online fault diagnosis. The test dataset is used to generate the current windows, and the proposed methodology

was investigated. In order to exploit the temporal nature of the batch profiles, we considered each profile in the reference dataset to contain a transient, steady state and an intermediate part. Thus, each profile was split into three equal time-series, and a new dataset was created which contained 750 time-series.

Diagnosis Result: The various parameters defined above were tuned with an objective of achieving maximum diagnosis efficiency. The chosen values of these parameters are shown in Table 1.

When the fault count parameter was set at 1, the classification accuracy is 100% for all the classes except the normal class. This is because many of the test batch profiles were declared faulty due to false alarms. These were controlled when fault count was increased to 2. Furthermore, we determine the reaction time of the proposed methodology in detecting a fault. The detection times for the four faults lie between 985 minutes and 1308 minutes. We compare our results with the LLE-SVDD approach, has been found to be superior to the PCA approach for the given case study, as given by Kumar et al. [8]. We can see that both the methods are able to perform perfect diagnosis efficiency; however LLE-SVDD approach detects a fault at an earlier stage than the proposed method. The LLE-SVDD approach however performs only fault detection, while our proposed methodology performed fault diagnosis. Thus, our proposed approach is advantageous in cases where it is imperative to diagnose a fault.

Table 1. Fault Diagnosis Results for a particular set of parameters

Parameters	Parameter Values	Fault Type	Diagnosis Efficiency(%)	
			Fault Count=1	Fault Count=2
Initial Point	51	Normal	80	100
Window Width	50	Fault 1	100	100
Window Shift	10	Fault 2	100	100
N	2	Fault 3	100	100
n	2	Fault 4	100	100

6.2 CSTR Batch Profiles

In the second case study, we have performed fault diagnosis of a jacketed CSTR in which an irreversible, exothermic, first order reaction ($A \rightarrow B$) is taking place using the proposed methodology. The system is equipped with three control loops, controlling the outlet temperature, the reactor holdup and the outlet concentration. A detailed description of mathematical model has been provided by Luyben [6]. This model has been further explored by Venkatasubramanian et al. [7] in their work on application of neural networks for fault diagnosis. They have introduced different ranges of operating parameters resulting in one normal operation and six abnormal operations. The variables that cause malfunctions are inlet flowrate, inlet temperature and inlet concentration of the reactant. Deviations beyond $\pm 5\%$ of the normal values of these variables were considered as faults, while values within $\pm 5\%$ were considered normal.

Generation of the historical database: The normal operations were simulated by varying any of the input variables in the neighborhood of the perfectly normal operation. This variation was kept in the range of 2.0%. The magnitudes of the input variable were varied randomly in each simulation. The duration of each simulation was 4 hours and sampling interval was randomly varied from 2 minutes. Each of the seven operating conditions was simulated 50 times to provide a historical database of 350 batches. Gaussian measurement noise was added to the measured variables so that the signal-to-noise ratio for the CSTR profile was approximately equal to ten. These batch profiles were divided into two subsets; a set of 175 profiles (25 profiles per class) was taken as the reference dataset and another set of 175 profiles (25 profiles per class) was used for the online diagnosis.

Diagnosis Result: Various parameters defined above were tuned with an objective of achieving maximum diagnosis efficiency, as shown in Table 2.

Table 2. Fault Diagnosis Results for a particular set of parameters

Parameters	Parameter Values	Fault Type	Diagnosis Efficiency(%) Fault Count=2
Initial Point	11	Normal	68
Window Width	50	Fault 1	100
Window Shift	10	Fault 2	88
N	2	Fault 3	80
n	2	Fault 4	96
		Fault 5	92
		Fault 6	80

We can see that the overall classification accuracy is 86.2857% for the above case. Very high diagnosis efficiency is obtained for Fault1, Fault4 and Fault5 operations, while the diagnosis efficiency for the case of Normal is quite less. We compare the results of the proposed algorithm the SVM-based classification technique, where one-against-all strategy is employed to classify the test profiles. An overall classification accuracy of 80% is achieved by using this approach, which is significantly less than the classification accuracy of 86.2857% obtained by the DTW approach. Thus, our method performs better in cases where fault diagnosis is required.

7 Result

In this work we have proposed a novel methodology for the purpose of online fault diagnosis employing the concept of Dynamic Time Warping as a superior distance measure between two time series. Warping window constraints and lower bounding measures have been utilized for the case of unequal length time-series, which substantially reduce the computational expense required. The proposed

methodology is capable of handling time-series of unequal lengths with different sampling times. This allows the application of our methodology to time-series with missing time points. A moving window approach has been demonstrated for the purpose of fault diagnosis in two simulated case studies. In both the cases, a simulated reference dataset is used to train bands for different classes using a heuristic hill-climbing approach, which were then employed for similarity measurement. The diagnosis efficiency is heavily dependent on the parameter values, however for appropriate set of parameters, the methodology is found to be satisfactorily efficient in detecting the faults in the process and can predict the type of the fault also.

Acknowledgements

We gratefully acknowledge financial assistance provided by Department of Science and Technology, New Delhi, India.

References

1. Sakoe, H., Chiba, S.: Dynamic-Programming Algorithm Optimization for Spoken Word Recognition. *IEEE Trans Acoust Speech Signal Process* 26, 43–49 (1978)
2. Ratanamahatana, C.A., Keogh, E.: Making Time-series Classification More Accurate Using Learned Constraints. In: Jonker, W., Petković, M. (eds.) *SDM 2004*. LNCS, vol. 3178, pp. 11–22. Springer, Heidelberg (2004)
3. Rath, T.M., Manmatha, R.: Lower-Bounding of Dynamic Time Warping Distances for Multivariate Time Series. Technical Report MM-40, Center for Intelligent Information Retrieval, University of Massachusetts Amherst (2002)
4. Votruba, J., Volesky, B., Yerushalmi, L.: Mathematical model of a batch acetone-butanol fermentation. *Biotechnol. Bioeng.* 28, 247–255 (1986)
5. Singhal, A.: Pattern-matching in multivariate time-series data. Ph.D. dissertation, Univ. of California, Santa Barbara (2002)
6. Luyben William, L.: Process modeling, simulation and control for Chemical Engineers. McGraw Hill, New York (1973)
7. Venkatasubramanian, V., Vaidyanathan, R., Yamamoto, Y.: Process fault detection and diagnosis using Neural Networks-I. Steady-state processes. *Computers Chem. Engg.* 14(7), 699–712 (1990)
8. Kumar, R., Jade, A.M., Jayaraman, V.K., Kulkarni, B.D.: A Hybrid Methodology For On-Line Process Monitoring. *IJCRE* 2(A14) (2004)