Pattern recognition based on episodes and DTW. Application to diagnosis of a level control system.

Joan Colomer, Joaquim Meléndez, Fco. Ignacio Gamero

Grup eXiT, UdG (IIiA i LEA-SICA) Av. Lluis Santalo s/n Girona E-17071 {quimmel, colomer, gamero}@eia.udg.es

Abstract

This work is oriented towards situation assessment of dynamic processes. Process conditions and abnormalities can be detected from patterns of measured process variables. Then, a correct representation and classification of these patterns allows identifying a particular class of operating situation. Nevertheless, different patterns belonging to the same class of situations could have different time duration or magnitudes. In this paper a modification of Dynamic Time warping (DTW) algorithm is presented in order to compare and classify patterns by means of a measure of similarity. The main improvement introduced in this algorithm is the use of qualitative representation of process variables by means of episodes

Introduction

Interpretation of measured process signals is an important task in Situation Assessment of dynamic processes, namely for Fault Detection and Diagnosis. For this reason, it is necessary to have tools for dealing with signal coming from processes. Qualitative representations are proposed to represent trends of signals (tendencies, oscillation degrees, alarms, degree of transient states...) needed in supervision, especially in fault detection and diagnosis. According to knowledge about process and its behaviour several techniques could be used with this aim.

One of these techniques is the representation of signals by means of episodes. In this case, series of episodes are used to describe patterns that identify particular classes of operating situation. Then the problem is to obtain a classification mechanism of these patterns in order to identify the state of the process. In this paper a description of the tools used in this pattern recognition methodology is shown.

The paper is organized as follows. In the following section, similarity methods applied to time series are discussed. Then Dynamic Time Warping (DTW) is introduced and basic concepts related to episodes are presented. Finally, a new approach of DTW is proposed and tested in a diagnosis application example.

Comparing Time Series.

There are numerous studies that have been carried to compare time series of data in several applications. Next, some models of distance-similarity are observed.

Agrawal et al. (1995b) present a shape definition language (SDL) for retrieving objects contained in the histories based on shapes. SDL allows converting original data in a qualitative description of its evolution that allows a comparison between sequences.

In Agrawal et al. (1995a) another model of similarity is introduced, it is based on the notion that two time sequences are said to be similar if they have enough nonoverlapping time-ordered pairs of subsequences that are similar. Given this similarity model, fast search techniques are used for discovering all similar sequences in a set of sequences by creating a indexable data structure. Other indexing methods to locate subsequences within a collection of sequences are presented by Faloutsos et al (1994) or Chan and Fu (1999) where a Haar wavelet transformation is used for the time series indexing problem.

A new representation, adopted by Keogh and Pazzani (1998), consists of piecewise linear segments to represent shape and a weight vector containing the relative importance of each individual linear segment, allowing the user to define a variety of similarity measures. (Keogh & Pazzani 2000) introduce a dimensionality reduction technique that supports an indexing algorithm.

A useful measure of similarity for strings is the length of a longest common subsequence (LLCS), based on the edit distance required in passing from one string to another one. Paterson and Dancík, (1994) carry out a revision of some existing solutions.

In (Konstantinov and Yoshida 1992) the qualitative shape of a signal is represented by the combination of strings. Hence, two temporal shapes are considered qualitatively equivalent if their *qshapes* coincide. A real time analysing procedure extracts *qshapes* over a predefined time interval and compares them with those of an expandable shape library that stores all interesting behaviours.

A methodology for pattern recognition based on episodes is described in (Bakshi and Stephanopoulos 1994b). Each pattern is represented by a string of primitives, also identified by means of a pattern grammar. The string that captures all the features necessary for classification is determined by matching the distinct syntactic descriptions, which represent similar events in these trends. Pattern matching facilitates extraction of qualitative and quantitative features used for solving the classification problem resolved by means of decision trees.

Dynamic Time Warping.

Most of algorithms that operate with time series of data use the Euclidean distance or some variation. However, Euclidean distance could produce an incorrect measure of similarity because it is very sensitive to small distortions in the time axis.

A method that tries to solve this inconvenience is Dynamic Time Warping (DTW), this technique uses dynamic programming (Sakoe and Chiba, 1978; Silverman, 1990) to align time series with a given template so that the total distance measure in minimised (Fig. 1). DTW has been widely used in word recognition to compensate the temporal distortions related to different speeds of speech. Next, a brief notion of DTW is described.

Given two time series X and Y, of length m and n respectively

$$X=x1,x2,...,x_{i},...,x_{m}$$
 $Y=y1,y2,...,y_{i},...,y_{n}$ (1)

To align the two sequences, DTW will find a sequence W of k points on a m-by-n matrix where every element (i,j) of the matrix contains the distance d(xi,yj) between the points xi and yj. The path W is a contiguous set of matrix elements that minimise the distance between the two sequences.

$$W=w1,w2,...,w_{k} max(m,n)\bullet k\bullet m+n$$
(2)

$$\mathbf{w}_{\mathbf{k}} = [\mathbf{i}_{\mathbf{k}}, \mathbf{j}_{\mathbf{k}}] \qquad (3)$$

where ik and jk denote the time index of trajectories X and Y respectively. In order to find the best path W, some constraints on the matching process are considered, main ones are:

- Constraints at the endpoints of the path, w1=[1,1] and wk=[m,n]
- Continuity constraints, matching paths cannot go backwards in time, this is achieved forcing ik+1≥ik and jk+1≥jk.

The path is extracted by evaluating the cumulative distance D(i,j) as the sum of the local distance d(xi,yj) in the current cell and the minimum of the cumulative distances in the previous cells. This can be expressed as:

D(i,j)=d(xi,yj)+min[D(i-1,j-1),D(i-1,j),D(i,j-1)] (1)



Fig. 1 Two signals with similar shapes. a) The Euclidean distance produce a pessimistic result of similarity since the signals are not aligned in time b) DTW find an alignment that allows a correct measure of similarity.

Several modifications of this technique have been introduced in order to apply the method in several situations. In (Keogh and Pazzani 1999) a modification of DTW is introduced to operate on a higher level of data abstraction through a piecewise linear representation. (Keogh and Pazzani 2001) consider a higher level feature of shape considering the first derivative of the sequences. Caiani et al. (1998) adapt the DTW approach to the analysis of the left ventricular volume signal for an optimal temporal alignment between pairs of cardiac cycles. (Vullings et al. 1998) implement a piecewise linear approximation and segment the signal into separate heartbeats. DTW also is used in (Kassidas et al. 1998) to synchronise batch process trajectories in order to reconcile timing differences among them.

Episodes Based Representations

Representations by means of episodes provide a good tool for situation assessment. On the one hand, uncertainty, incompleteness and heterogeneity of process data make the qualitative reasoning a good tool. On the other hand, reasoning not only with instantaneous information, but with historic behaviour of processes is necessary. Moreover, since a great deal of process data is available for the supervisory systems, to abstract and use only the most significant information is required. The representation of signals by means of episodes provides an adequate response to these necessities.

The general concept of **episode** was introduced in the field of qualitative reasoning by Williams (1986), who defined an episode as a set of two elements: a time interval, named **temporal extent** and a **qualitative context**, providing the temporal extension with significance. This definition allows defining an episode as explicitly as the qualitative context.

A general formalism for the representations of signals by means of episodes, the Qualitative Representation of Process Trends, can be found in (Cheung and Stephanopoulos 1990). This formal approach introduces the concept of trend as a sequence of episodes characterised by the signs of the first and the second derivative. It has a practical extension in the triangular and trapezoidal representations. Stephanopoulos et al. (1997) have applied the above representation to the analysis of industrial fermentation data. The methodology consists of three components developed by Bakshi and Stephanopoulos (1994a,b). First, a wavelet signal decomposition that acts as a noise removing filter; second, the triangular representation of smoothened process signals, and finally a search algorithm that makes use of decision trees and Shannon's entropy comparisons for the identification of certain classes of process outcomes. The methodology has been implemented as a part of a broader system referred to as dbminer[©] aimed at fermentation database mining, diagnosis and control.

Janusz and Venkatasubramanian (1991) proposed a qualitative description (TDL) of signals consisting of primitives, episodes, trends and profiles. Primitives are based on the sign of first and second derivatives (positive, zero or negative). Thus, nine basic types compose the set of primitives. The trend of a signal consists of a series of episodes, and a profile is obtained by adding quantitative information. Drawbacks due to noise and discontinuities are rectified by an error correcting code (ECC) acting as a postprocessor (Rengaswamy 1995). Later, Rengaswamy and Venkatasubramanian (1995) refined the language using a syntactic pattern recognition approach where a fixed-size neural network was used to identify the primitives. Vedam and Venkatasubramanian (1997) proposed an adaptive trend identification algorithm based on wavelet theory. Then, the identified primitives are used as input to a knowledge base to perform fault diagnosis. This system, called W-ASTRA, is demonstrated on a fluidised catalytic cracking unit. An improvement is developed by Rengaswamy et al. (2001) and utilised in (Dash et al, 2001), where a new procedure that identifies piecewise unimodals represented as quadratic segments is used to identify qualitative shapes of trends.

The formalism described in (Meléndez and Colomer, 2001) extend previous formalisms to both qualitative and numerical context in order to be more general. It means that allows building episodes according to any feature extracted from variables. According to this formalism, a new representation allows to describe signal trends

depending on the second derivative, that can be computed by means of a band-limited FIR differentiator (Colomer and Meléndez, 2001) in order to avoid noise amplification. The qualified first derivative at the beginning and end of each episode is used in order to obtain a more significant representation. Then, a set of 13 types of episodes is obtained (Fig. 2).



Fig. 3 Example of a qualitative representation

A major benefit of this set of episodes for supervisory tasks is that discontinuities and stability periods (usual in fault situations and in normal situations respectively) are explicitly represented by means of 5 types of episodes (] [-]].

Combining DTW and Episodes based Representations

In previous section, DTW has been shown as a good method to determine the similarity between two sequences of episodes due to its capacity to align sequences with different longitudes. As disadvantages, it is a computationally expensive algorithm and it could fail in the alignment by trying to solve the variability in the Y-axis by warping the X-axis. In this section, a modification of the DTW algorithm that allows solving these inconveniences is introduced.

The proposed solution consists on apply DTW not in original time series but in its episodes based representations. The representation of a sequence as episodes reduces the calculation time by decreasing the amount of manipulated data. Likewise, the qualitative character that defines an episode avoids the problem of the variability in the Y-axis. Therefore DTW can be used to align episodes to obtain a global distance.

The only problem is to define a local distance between episodes. In this sense, a chart of distances has been defined where the 13 types of episodes described in the previous section are related. Distances are based on the qualitative state and auxiliary characteristics that define the different types of episodes (Table 1). However, these local distances could be subject to the criterion of the user, so one could give more importance to some episodes concerning another obtaining a different global distance and preserving the essential features of the process signal. This way, a new approach (EpDTW) of the DTW algorithm is created using episodes as a higher level representation of the signal.

	7	Γ	(\cap		\backslash	_	/	L	U)		L
1	0	.72	.85	.7	.62	.67	.75	.9	.8	.87	.95	1	.67
Γ	.72	0	.7	.62	.77	.82	.75	.75	.87	.95	.87	.67	1
(.85	.7	0	.52	.8	.85	.6	.27	.9	.82	.65	.8	.87
\cap	.7	.62	.52	0	.45	.6	.6	.6	.82	.9	.82	.87	.95
	.62	.77	.8	.45	0	.27	.6	.85	.65	.82	.9	.95	.87
Ń	.67	.82	.85	.6	.27	0	.55	.8	.27	.6	.85	.9	.75
_	.75	.75	.6	.6	.6	.55	0	.55	.6	.6	.6	.75	.75
/	.9	.75	.27	.6	.85	.8	.55	0	.85	.6	.27	.67	.82
U	.8	.87	.9	.82	.65	.27	.6	.85	0	.4	.8	.85	.7
Ú	.87	.95	.82	.9	.82	.6	.6	.6	.4	0	.45	.7	.62
J	.95	.87	.65	.82	.9	.85	.6	.27	.8	.45	0	.57	.77
j	1	.67	.8	.87	.95	.9	.75	.67	.85	.7	.57	0	.72
L	.67	1	.87	.95	.87	.75	.75	.82	.7	.62	.77	.72	0

Table 1 Local distance between episodes.

It is necessary to keep in mind that compared sequences could have different duration. This fact complicates the generalisation of the proposed technique, in the next example the length of the analysed sequences is different although not too dissimilar.

A Diagnosis Application

As application example, the proposed approach has been used in a laboratory plant for diagnosis purposes. In this plant (See Fig. 4), level in tank A is controlled by means of a PID controller by pumping water from a reservoir (tank B).



Fig. 4 Laboratory Plant

Three valves (V1,V2 and V3) can be handled in order to simulate obstructions and leakages. Then several behaviours are possible by appropriate combination of opening and closing valves. Table 2 represents these situations. Additionally, system dynamics can by slightly modified by filling or emptying the reservoir with external water. Then, input and output of external water are also interesting situations to detect. The experiments have been developed under the assumption that two situations can not be overlapped. Thus, changes in the configuration of valves are only performed when process is in steady state. Monitored signals are the level in tank A and the control signal (pump).

	Description	V1	V2	V3
1	Normal Behaviour	OPEN	OPEN	CLOSED
2	Obstruction in pump or input pipe	CLOSED	OPEN	CLOSED
3	Obstruction in the output pipe	OPEN	CLOSED	CLOSED
4	Leak in input pipe or pump	OPEN	OPEN	OPEN

Table 2

The monitoring system will be able to detect such situations and diagnose about the origin of misbehaviours according to the behaviour of measured signals described by sequences of episodes. The monitoring system acquires data periodically and represents them as sequences of episodes according to the previous description. These sequences are compared by means of EpDTW with other well-known patterns with the purpose of detecting and diagnose the situation.

Execution Example

The example described in this section corresponds to the manipulation of the three valves as described in Table 2. First the valves are manipulated in order to simulate the failure and later are manipulated again in order to return to the normal operation. Three reference patterns (R1,R2 and R3) have been obtained to represent each abnormal situation, each one (Fig. 5-7) is composed by the two monitored signals (level in tank A and control) and its representation in episodes.



Fig. 5 Reference pattern R1: Level and control signals for situation 2



Fig. 6 Reference pattern R2: Level and control signals for situation 3



Fig. 7 Reference pattern R3: Level and control signals for situation 4

Then, three test patterns T1, T2, and T3 (Fig. 8-10) corresponding to the same situations but with different set points are used in order to compare them with the reference patterns and diagnose the situation.



Fig. 8 Test pattern T1: Level and control signals for situation 2



Fig. 9 Test pattern T2: Level and control signals for situation 3



Fig. 10 Test pattern T3: Level and control signals for situation 4

First, the level and control signals of each pattern have been compared with a classical DTW algorithm after normalising the signals. The obtained results are shown in the Table 3 and Table 4. Then, the sequences of the test patterns are compared by means of EpDTW with the wellknown sequences of the reference patterns. In the Table 5 and Table 6 the results of the comparison for the level and control signals respectively are presented. In both cases the obtained values are a normalised distance, so 0 represent a complete equality.

Level	R1	R2	R3
T1	0.054	0.521	0.42
T2	0.489	0.045	0.197
T3	0.494	0.217	0.074

Table 3 Results of the comparison for the level signals via DTW

Control	R1	R2	R3
T1	0.04	0.53	0.13
T2	0.64	0.208	0.635
T3	0.11	0.61	0.075

Table 4 Results of the comparison for the control signals via DTW

Level	R1	R2	R3
T1	0.045	0.335	0.31
T2	0.391	0.03	0.462
T3	0.383	0.425	0.068

Table 5 Results of the comparison for the level signals via EpDTW

Control	R1	R2	R3
T1	0.039	0.382	0.078
T2	0.382	0.039	0.46
T3	0.034	0.368	0

Table 6 Results of the comparison for the control signals via EpDTW

Finally, the mean value of the distances for the two signals is obtained in order to have a global distance between each situation and the patterns in both cases (DTW and EpDTW), the results from this similarity assessment are presented in Table 7 and Table 8.

Global	R1	R2	R3
T1	0.047	0.525	0.275
T2	0.564	0.126	0.416
Т3	0.302	0.413	0.0745

Table 7 Results from similarity assessment via DTW

Global	R1	R2	R3
T1	0.042	0.3585	0.194
T2	0.3865	0.0345	0.461
T3	0.2085	0.3965	0.034

Table 8 Results from similarity assessment via EpDTW

It can be shown that it is easier to distinguish between correct and incorrect diagnostics by using Table 8 than Table 7. Another point to take into account is the computing time. In these examples, the maximum time of execution by using the DTW algorithm in a AMD K2 computer has been 5.34 seconds, while with EpDTW it has been 0.3 seg.

Conclusions

This work shows the possibilities of combining qualitative representations of signals by means of episodes and Dynamic Time Warping for pattern recognition in the field of diagnosis.

Since different patterns belonging to the same class of situations could have different time duration or magnitudes, a modification of DTW algorithm is presented in order to compare and classify similar patterns by means of a measure of similarity. So, the advantage of the temporal alignment produced by DTW is added to the advantage of the representation in episodes, that considers the behaviour of the signals.

The utility of the representation is shown in the diagnosis of a level control system where the correct identification of operating situations is obtained through the comparison of the current pattern with well-known reference patterns.

Acknowledgements

This work is partially supported by the Spanish government within the projects TAP99-1079-C03-03 "Estructuras inteligentes. Desarrollo de un sistema integrado de monitorización y control: aplicación a los puentes atirantados" and DPI2001-2198, "Supervisión Experta de Calidad del Servicio Eléctrico" partially charged to FEDER funds, and by the CTP project ITT 2000-18 "Monitorització Intel·ligent Preventiva d'EDAR" supported by the Generalitat de Catalunya.

References

- Agrawal, R., Psaila, G., Wimmers, E.L., Zaït, M., (1995a). Querying shapes of histories. *Proc. of 21st VLDB Conf.*, Switzerland, pp. 502-514
- Agrawal, R., Lin, K.I., Sawhney, H.S., Shim, K., (1995b). Fast similarity search in the presence of noise, scaling, and translation in time series database. *Proc. of 21st VLDB Conference*, Switzerland, 1995, pp. 490-501.
- Bakshi, B. R., Stephanopoulos, G. (1994a). Representation of process trends part III. Multiscale extraction of trends from process data. *Comp. Chem. Eng.* 18, 267-302.
- Bakshi, B. R., Stephanopoulos, G. (1994b). Representation of process trends part IV. Induction of real-time patterns from operating data for diagnosis and supervisory control. *Comp. Chem. Eng.* 18, 303-332.
 Caiani, E.G., et al. (1998) Warped-average template
- Caiani, E.G., et al. (1998) Warped-average template technique to track on a cycle-by-cycle basis the cardiac filling phases on left ventricular volume. *IEEE Computers in Cardiology. Vol.* 25, 73-76.
- Chan, K. and Fu, W. (1999). Efficient time series matching by wavelets. *Proc. of the 15th IEEE Int. Conf. on Data Engineering.*
- Cheung, J. T., Stephanopoulos G., (1990). Representation of Process Trends, parts I and II. *Computers Chemical Engineering* 14, 495-540.
- Colomer J, Meléndez J, (2001). A family of FIR differentiators based on a polynomial least squares estimation. *Proc. Of the European Control Conference*, 2001, pp: 2802-2807.
- Dash, S., Rengaswamy, R., Venkatasubramanian, V. (2001). A Novel Interval-Halving Algorithm for Process Trend Identification", 4th IFAC Workshop on On-Line Fault Detection & Supervision in the CPI, Seoul, Korea, 2001.
- Faloutsos, C., Ranganathan, M., Manolopoulos, Y. (1994) Fast subsequence matching in time series databases. *Proceedings of the 1994 ACM SIGMOD Int. Conf. on Management of Data*, 1994, pp. 419–429.
- Janusz, M. and Venkatasubramanian, V., (1991). Automatic generation of qualitative description of process trends for fault detection and diagnosis". *Engineering Applications of Artificial Intelligence 4*, 329-339.
- Kassidas, A.; MacGregor, J.F.; Taylor, P.A., (1998), "Synchronization of Batch Trajectories Using Dynamic Time Warping", *AIChE Journal* 44, No. 4, pp. 864–875.
- Keogh and Pazzani (2001). Derivative Dynamic Time Warping. First SIAM International Conference on Data Mining (SDM'2001) 2001, Chicago, USA.

- Keogh and Pazzani (2000). A simple dimensionality reduction technique for fast similarity search in large time series databases. *Fourth Pacific- Asia Conference* on Knowledge Discovery and Data Mining. Kyoto, Japan.
- Keogh and Pazzani (1999) Scaling up dynamic time warping to massive datasets. 3rd European Conference on Principles and Practice of Knowledge Discovery in Databases. Prague.
- Keogh and Pazzani (1998). An enhanced representation of time series which allows fast and accurate classification, clustering and relevance feedback. *Proceedings of the Fourth conference on Knowledge Discovery in Databases and Data Mining*, New York. pp 239-241, AAAI Press.
- Konstantinov, K. and Yoshida T., (1992). Real-Time Qualitative Analysis of the Temporal Shapes of (Bio) Process Variables. *AIChE Journal 38 (11)*, 1703-1715.
- Meléndez, J.; Colomer, J., (2001) Episodes representation for supervision. Application to diagnosis of a level control system. *Workshop on Principles of Diagnosis DX'01*, Sansicario (Italy)
- Paterson, M.S. and Dancik, V., (1994) Longest common subsequences. Proceedings of 19th Int. Symposium Mathematical Foundations of Computer Science, Lecture Notes in Computer Science 841, 1994, pp. 127-142.
- Rengaswamy, R.,Hagglund, T.,Venkatasubramanian, V. (2001). A qualitative shape analysis forma-lism for monitoring control loop performance. *Engng. Applic. Artif. Intell.*, 14(1):23–33.
- Rengaswamy, R. (1995). A framework for integrating process monitoring, diagnosis and supervisory control. *Doctoral Thesis, Purdue University.*
- Rengaswamy R., Venkatasubramanian V. (1995) A syntactic pattern-recognition approach for process monitoring and fault diagnosis. *Engng Applic. Artif. Intell.*, 8(1):35–51.
- Sakoe, H., Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans. Acoustics, Speech, and Signal Proc., Vol. ASSP-26(1)*, 43-49
- Silverman, H.F., Morgan, D.P. (1990). The application of dynamic programming to connected speech recognition. *IEEE ASSP Magazine.*, 7, 6-25
- Stephanopoulos, Gregory; Locher, Georg; Duff, Michael J.; Kamimura, Roy; Stephanopoulos, Georg, (1997), Fermentation Database Mining by Pattern Recognition, *Biotechnology and Bioengineering, Vol. 53, No. 5*, pp. 443–452.
- Vedam, H.; Venkatasubramanian, V. (1997). A Wavelet Theory-Based Adaptive Trend Analysis System for Process Monitoring and Diagnosis, *Proc. of the American Control Conference, Piscataway/NJ, USA*, pp. 309–313.
- Vullings, H.J.L.M., Verhaegen, M.H.G., Verbruggen, H.B. (1998) Proceedings of the 20th Int. Conf. Of the IEEE Engineering in Medicine and Biology Society, 20, (1), 163-166
- Williams, B.C., (1986). Doing Time: Putting qualitative reasoning on firmer ground. *Proc. of AAAI-86, National Conf. on Artif. Intel.*, 105-112.