PERIODICA POLYTECHNICA SER. CIVIL ENG. VOL. 37, NO. 1, PP. 43–55 (1993)

APPLICATION OF CLUSTER ANALYSIS TO EVENT RECOGNITION

Béla PALÁNCZ, Mária JORDÁN MAGYAR and János VÁGÁSI

Faculty of Civil Engineering,
Laboratory of Informatics
Technical University of Budapest
H–1521 Budapest, Hungary

Received: August 25, 1992

Abstract

A new method based on cluster analysis was applied to failure recognition of a distillation process. Two coupled indicator functions were introduced for identifying malfunction transients. Different distance definitions were analysed and the Wards method was found the most effective one in this case.

Keywords: cluster analysis, failure recognition, dynamical modelling, distillation process.

Introduction

To ensure safety operations in chemical and nuclear technologies, early recognition and diagnostic of unexpected fast transients and malfunctions are required. Many different methods have been developed to solve this problem: heuristic rules for operators in emergency situations, automatic alarm systems, algorithms employing Kalman filters, Pontrjagin maximum principle known from optimal control theory, noise diagnostic and the nowadays fashionable on-line expert systems were recommended, [HIMMELBLAU (1978), MAZETIS (1984)].

The idea of the proposed method presented in this paper is based on the cluster analysis. The clustering algorithm using hierarchical procedure is employed to classify unknown transients. The dendrogram, as a result of the classification, shows clearly which specified malfunction transient is in the same class with the transient to be identified on line.

In this way we have a precise information about the type of malfunction event.

Formulation of the Problem

Let us consider a nonlinear multivariable dynamic system with I measurable variables. The trajectories of K different malfunctions, $x_{ik}(t)$ are known and stored in a databank for $i := 1, 2, \ldots, I$ and $k = 1, 2, \ldots, K$.

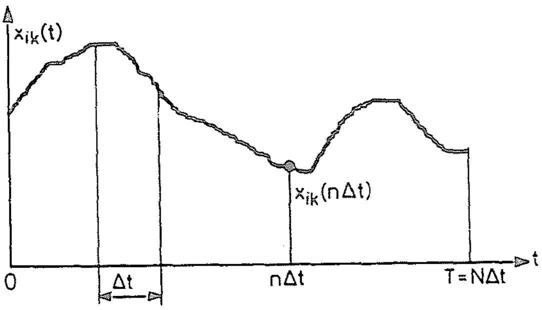

Fig. 1. Illustration of the notations

The following notations are used for these trajectories, see *Fig. 1*, where

Δt – time step size for measurement
T – total observation time
N – total number of the points of measurement during observation
n – number of the measurement up to the time $t \leq T, n = 1 \dots N$

The problem is to identify the on-line measured, unknown transient, $i = 1, 2 \dots I$ on the basis of the earlier specified and stored malfuction trajectories. We suppose that the malfunction event occurred at $t = 0$.

Application of Cluster Analysis

This application procedure can be carried out in three steps.

Preparation of the Data

The first step is the preparation of the data being at our disposal. The so-called data matrix can be constructed on the basis of the stored and the on-line measured data of dimension $(K+1)^*(n^*I)$.This data matrix can be expressed by a vector of dimension $(K+1)$. Each element of this vector is a row vector of dimension n^*I.

$$\mathbb{X}(t) = \begin{bmatrix} \mathbf{x}_1^*(t) \\ \mathbf{x}_2^*(t) \\ \vdots \\ \mathbf{x}_k^*(t) \\ \vdots \\ \mathbf{x}_K^*(t) \\ \mathbf{x}_D^*(t) \end{bmatrix}, \tag{1}$$

where $\mathbf{x}_k^*$ – the transposed of the event vector of the k-th event – is the k-th row of the data matrix, namely:

$$\mathbf{x}_k(t) = \begin{bmatrix} x_{1k}(\Delta t) \\ x_{2k}(\Delta t) \\ \vdots \\ x_{Ik}(\Delta t) \\ \vdots \\ x_{1k}(n\Delta t) \\ x_{2k}(n\Delta t) \\ \vdots \\ x_{Ik}(n\Delta t) \end{bmatrix} \tag{2}$$

and the last row represents the on-line measured, unknown transient, $x_D^*(t)$.

The distance matrix can be computed from the data matrix:

$\mathbb{X}(t) =$

$$\begin{bmatrix} x_{11}(\Delta t) & x_{21}(\Delta t) & . & x_{I1}(\Delta t) & \dots & x_{11}(n\Delta t) & x_{21}(n\Delta t) & . & x_{I1}(n\Delta t) \\ x_{12}(\Delta t) & x_{22}(\Delta t) & . & x_{I2}(\Delta t) & \dots & x_{12}(n\Delta t) & x_{22}(n\Delta t) & . & x_{I2}(n\Delta t) \\ \dots & \dots & . & \dots & \dots & \dots & \dots & . & \dots \\ x_{1k}(\Delta t) & x_{2k}(\Delta t) & . & x_{Ik}(\Delta t) & \dots & x_{1k}(n\Delta t) & x_{2k}(n\Delta t) & . & x_{Ik}(n\Delta t) \\ \dots & \dots & . & \dots & \dots & \dots & \dots & . & \dots \\ x_{1K}(\Delta t) & x_{2K}(\Delta t) & . & x_{IK}(\Delta t) & \dots & x_{1K}(n\Delta t) & x_{2K}(n\Delta t) & . & x_{IK}(n\Delta t) \\ x_{1D}(\Delta t) & x_{2D}(\Delta t) & . & x_{ID}(\Delta t) & \dots & x_{1D}(n\Delta t) & x_{2D}(n\Delta t) & . & x_{ID}(n\Delta t) \end{bmatrix}. \tag{3}$$

$$D(t) = \begin{bmatrix} d_{1,1} & \dots & \dots & d_{1,K} & d_{1,D} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & d_{r,s} & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ d_{K,1} & \dots & \dots & d_{K,K} & d_{K,D} \\ d_{D,1} & \dots & \dots & d_{D,K} & d_{D,D} \end{bmatrix}, \tag{4}$$

where $d_{r,s}$ is the distance between the vectors $\mathbf{x}_r(t)$ and $\mathbf{x}_s(t)$ represented by two different rows in $\mathbb{X}(t)$.

Here, the Mahalanobis-type distance is employed: (ANDERBERG, 1973)

$$d_{r,s}(t) = [\mathbf{x}_r(t) - \mathbf{x}_s(t)]^* S^{-1} [\mathbf{x}_r(t) - \mathbf{x}_s(t)], \tag{5}$$

where S is the covariance matrix of dimension $p = n * I$

$$S_{ij} = (\hat{\mathbb{Q}} - \mathbb{Q}_i)(\hat{\mathbb{Q}} - \mathbb{Q}_j) \ i,j = 1 \dots p, \tag{6}$$

where Q_i and Q_j are the i-th and j-th elements of the vector

$$\mathbb{Q} = \mathbb{X}_r - \mathbb{X}_s \tag{7}$$

and

$$\hat{\mathbb{Q}} = \frac{1}{p} \sum_{k=1}^{p} \mathbb{Q}_k. \tag{8}$$

Classification Procedure

In order to identify the unknown transient, we consider the earlier stored transients together with the unknown one, and carry out the hierarchical classification on this set of vectors. The dendrogram as the result of this procedure represents the hierarchical structure of the classes. Now the unknown transient can be assigned to the malfunction event which is belonging to that of the stored transient being in the same class whith it.

The main steps of the clustering algorithm are the following:

- Let us find the minimal value of the distance matrix $\mathbb{D}$, namely for the pair of (a,b) vector;

$$d_{a,b} = \min(d_{r,s}); \tag{9}$$

- Link these two elements into one (a, b) element;
- Compute the distance matrix on the basis of the new elements;
- If the number of the linkages are less than K, go back to the first step.

To compute the new elements of the distance matrix the following recursive formula can be used (HENVIRON et al., 1988):

$$d_{r,ab} = \alpha_a d_{r,a} + \alpha_b d_{r,b} + \beta d_{a,b} + \gamma |d_{r,a} - d_{r,b}. \tag{10}$$

Table 1
Constants of the recursive formula (10) in the case of different clustering methods
$(m = m_a + m_b)$

Method	α_a	α_b	β	γ
a. Single linkage	1/2	1/2	0	−1/2
b. Complete linkage	1/2	1/2	0	1/2
c. Average linkage (weighted)	1/2	1/2	0	0
d. Average linkage (unweighted)	$\frac{m_r}{m}$	$\frac{m_b}{m}$	0	0
e. Median	1/2	1/2	−1/4	0
f. Centroid linkage	$\frac{m_a}{m}$	$\frac{m_b}{m}$	$\frac{-m_a m_b}{m^2}$	0
g. Wards method	$\frac{m_r + m_a}{m + m_r}$	$\frac{m_r + m_b}{m + m_r}$	$\frac{-m_r}{m + m_r}$	
h. Flexible strategy	α	α	1–2α	0

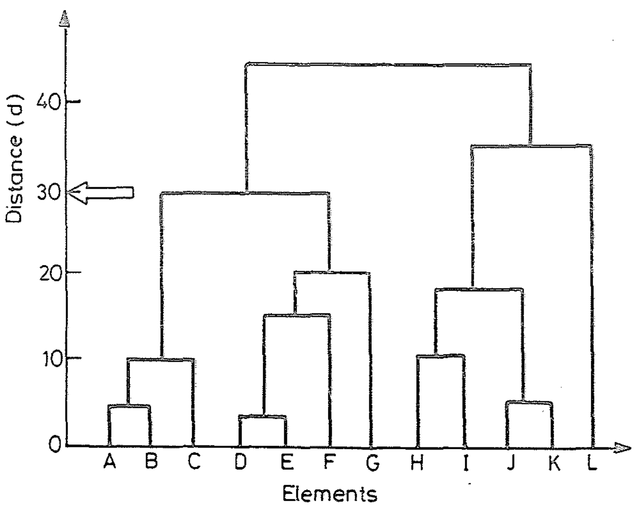

Fig. 2. Hierarchical structure of the classes demonstrated by a dendrogram

The constants for different clustering methods can be seen in *Table 1*.

Generally, elements a, b and r can result from earlier linkage of other single elements, therefore, let us consider m_a, m_b and m_r, the numbers of these single elements.

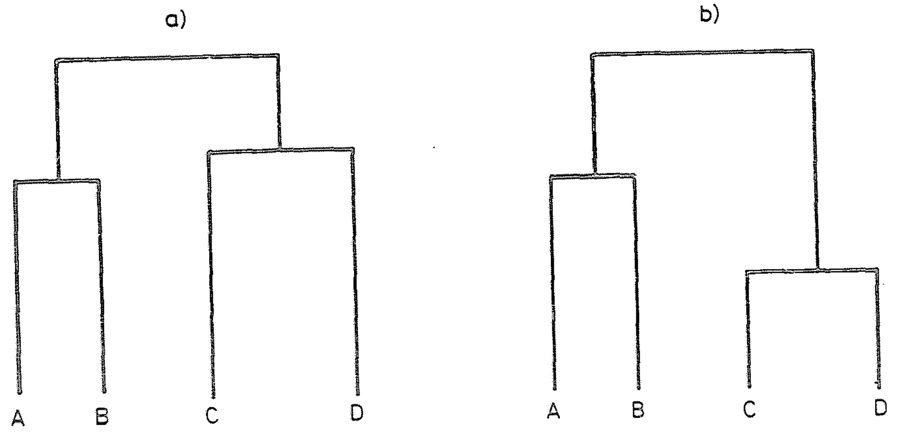

Fig. 3. Conditions can be separated

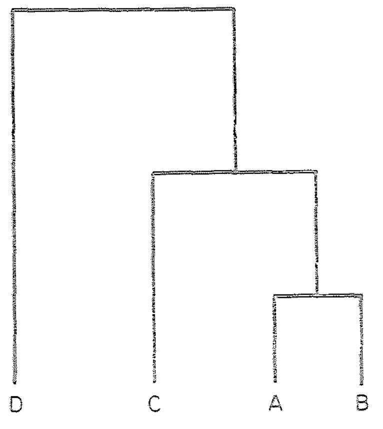

Fig. 4. Conditions are not separable

The level of the linkages, the distances among single elements being in the same class as well as the hierarchical structure of the classes can be demonstrated by the dendrogram, see *Fig. 2*.

This dendrogram illustrates that at level $d = 10$ elements (1, 2, 3) are in the same class, while at level $d = 30$ elements (1, 2, 3, 4, 5, 6, 7) belong to the same class.

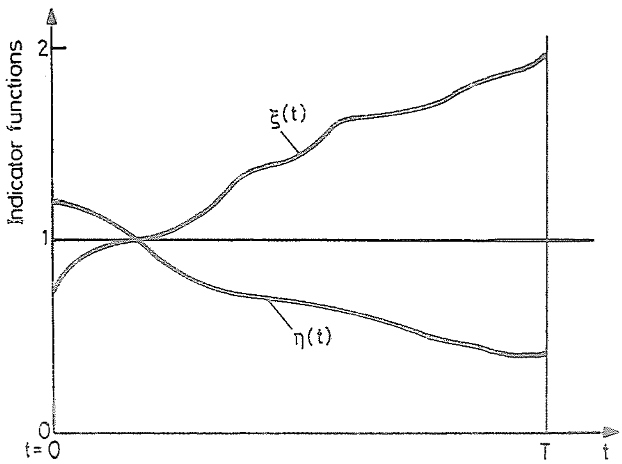

Fig. 5. The indicator functions in the case of a successful recognition

Identification of the Transient

Transient A can be assigned to a malfunction event if the following conditions for the transient B belonging to this malfunction are true:

$$d_{A,B} = \min_{1 \leq j \geq K} d_{A,j} \qquad j = 1 \dots K \tag{11}$$

and

$$d_{l_1,l_2} > d_{A,B},$$

$$l_1, l_2 = 1 \dots K, \qquad l_1 \neq l_2. \tag{12}$$

It can happen that the two conditions are not satisfied at the same time, see *Fig. 3*.

On the other hand, when these two conditions cannot be separated, both of them are true, see *Fig. 4*.

Conditions (11) and (12) also can be expressed in the form of dimensionless variables, namely

$$\eta(T) < 1 \quad \text{(first condition)} \tag{13}$$

and

$$\xi(T) > 1 \quad \text{(second condition)}, \tag{14}$$

where

$$\eta(t) \hat{=} \min_j \frac{d_{A,B}(t)}{d_{A,j}(t)}, \qquad j = 1 \dots K \tag{15}$$

and

$$\xi(t) \hat{=} \min_{l_1 \neq l_2} \frac{d_{l_1,l_2}(t)}{d_{A,B}(t)}, \qquad l_1, l_2 = 1 \dots K. \tag{16}$$

Let us consider $\eta(t)$ and $\xi(t)$ as indicator functions.

Table 2
Possible events and measured variables
(see also *Fig. 6*)

Malfunction Events
1. Malfunction of temperature controller
2. Plate plugging
3. Malfunction of reboiler
Measured variables
1. Distillate temperature
2. Bottom temperature
3. Distillate rate
4. Bottom rate
5. Temperature of feeding plate

Fig. 5 shows how the values of these functions change in time during the recognition procedure. This figure represents a successful identification process.

Application Example

In the following, the application of the method to continuous distillation process is presented. The scheme of the distillation column can be seen in *Fig. 6*. The dynamic model used for simulation of malfunction was adopted from the literature (ROBINSON, 1975)). The malfunction events and the measured variables are in *Table 2*.

The malfunction of the temperature controller was simulated by increasing instantly its setpoint by 10 °C.

Plate plugging was imitated by decreasing the liquid downflow rate of the 4-th (feeding) plate from 1 kmol/h to 0.1 kmol/h, and the malfunction of the reboiler was taken into consideration by decreasing its vapor production by 50 per cent.

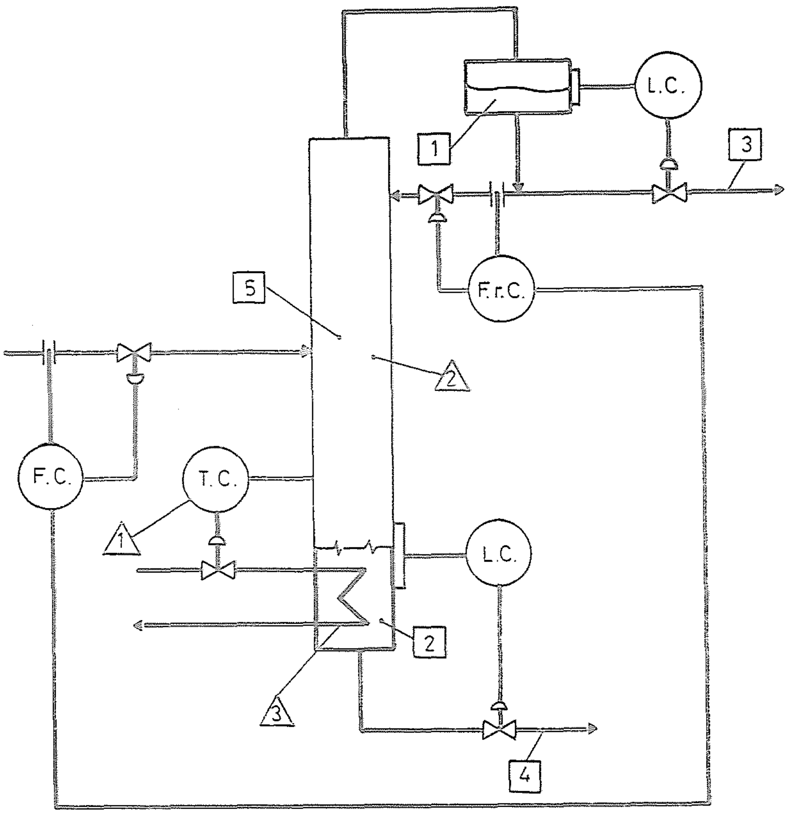

Fig. 6. The scheme of the distillation column

As illustration of the simulation, the change of distillate rate in time caused by the three different malfunction events can be seen in *Fig. 7*. Similar transients are available for the other four variables.

They were computed by numerical simulation. In our case $I = 5$ and $K = 3$. The observation time is $T = 0.08$ hour and the total number of observation is N=20.

In real cases, the measured transients deviate from the ideal (simulated) ones because of modelling error, measurement noise and other unknown effects and changes.

Fig. 8 shows the result of the recognition procedure in the case of 1% noise level. To carry out the classification, the average linkage (weighted) method was used. It can be seen according to the indicator functions that the procedure worked successfully.

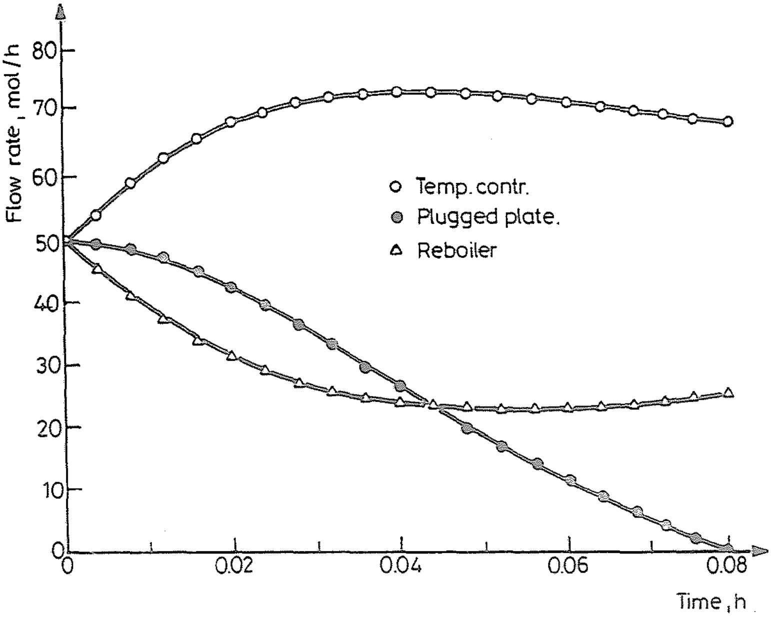

Fig. 7. Distillate rate in the case of different events

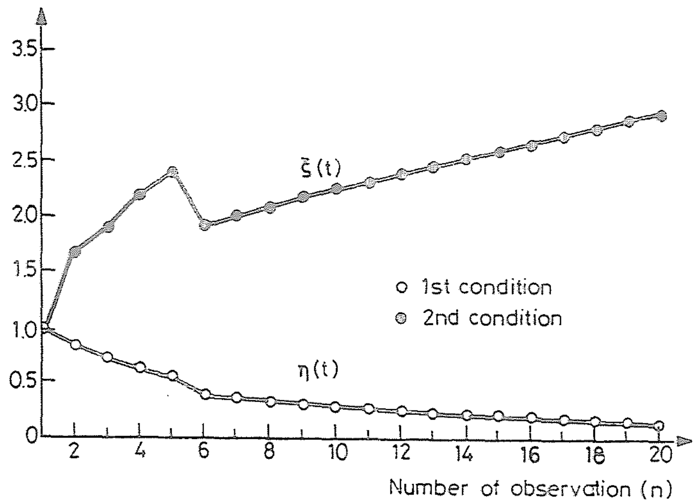

Fig. 8. Event recognition (noise 1%). Malfunction of temperature controller

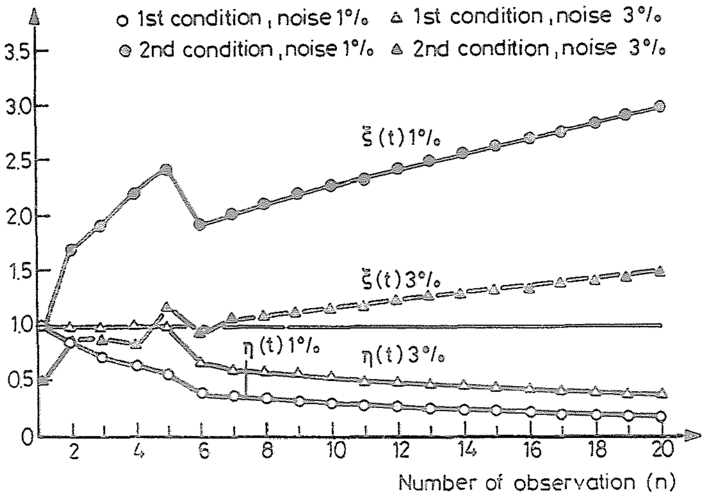

Fig. 9. Event recognition. Malfunction of temperature controller in the case of different noise levels

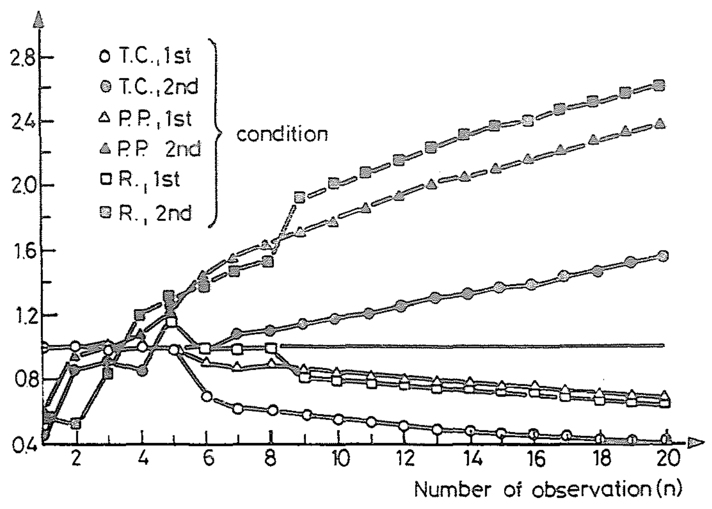

Fig. 10. Event recognition (noise 3%) in the case of different malfunctions

In *Fig. 9* one can see the effect of the increasing noise level. In the case of higher noise level the efficiency of the procedure is deteriorated, the separation of $\xi(t)$ and $\eta(t)$ is not so characteristic.

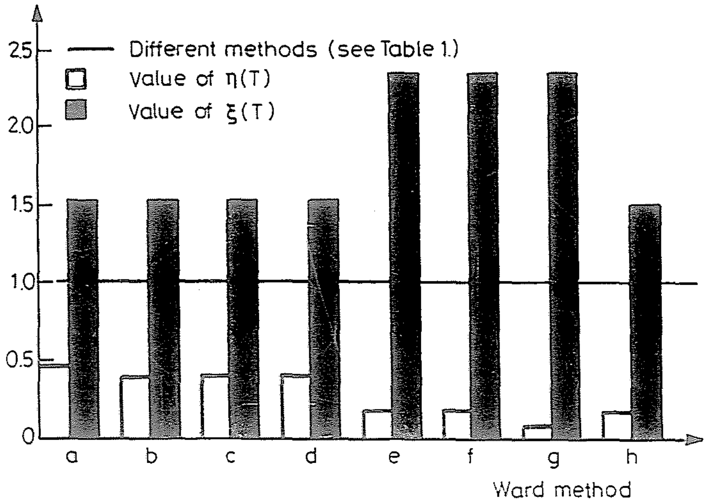

Fig. 11. Results of different strategies. Values of the indicator function at $t = T$

The next figure shows the effectivity of the recognition procedure in the case of different malfunction events. The two conditions are demonstrated in *Fig. 10*.

According to *Fig. 11*, where the values of the indicator function valves can be seen at the end of the observation time $t = T$, the most efficient classification method is the Wards method ($\xi(t) - \eta(t)$ difference is the greatest).

Conclusion

A new method based on cluster analysis for event recognition in dynamical systems was proposed. Two coupled conditions represented by two indicator functions were introduced and employed to indicate malfunction events.

As illustration a continuous distillation process was analysed.

Different distance definitions were studied and the Wards method proved to be the best with parameter $0.5 < \alpha < 1$.

Numerical results show that measurement noise can deteriorate the effectivity of the recognition significantly.

The method has two important advantages:

- it can be applied even in the case of more events than measured state variables;

- malfunction can be detected progressively even earlier than $t = T$, which is important in the case of on-line application.

Acknowledgement

This work was partly supported by the National Research Found (OTKA T 4445).

References

ANDERBERG, M. R. (1973): Cluster Analysis for Applications, Academia Press, New York.

HENVIRON, G. et al. (1988): Beispiele zur Datenanalyse mit BASIC-Programmen, VEB Deutscher Verlag der Wissenschaften, Berlin.

HIMMELBLAU, D. M. (1978): Fault Detection and Diagnosis in Chemical and Petrochemical Processes, Elsevier, Amsterdam.

MAZETIS, G. R. (1984): Malfunction Diagnostic Analysis of Nuclear Power Plant System, Ph. D. The Catholic Univ. of Am., Washington.

ROBINSON, E. R. (1975): Time Dependent Chemical Processes, Applied Science Publisher, London, pp. 165–178.