

# Technology Extraction for Future Generations from Process Time Series Data Reflecting Expert Operator Skills

Setsuya Kurahashi  
University of Tsukuba, GSSM  
3-29-1 Otsuka Bunkyo-ku  
Tokyo, Japan

kurahashi@gssm.otsuka.tsukuba.ac.jp

Takao Terano  
Tokyo Institute of Technology  
4259 Nagatsuda-Cho Midori-ku  
Yokohama Japan

terano@dis.titech.ac.jp

## ABSTRACT

This paper proposes a novel method to develop a process response model from continuous time-series data. The main contribution of the research is to establish a method to mine a set of meaningful control rules from Learning Classifier System using the Minimum Description Length criteria. The proposed method has been applied to an actual process of a biochemical plant and has shown the validity and the effectiveness.

## Categories and Subject Descriptors

I.2.6 [Artificial Intelligence]: LearningKnowledge acquisition

## General Terms

Algorithms

## Keywords

Learning Classifier Systems, Minimum Description Length, Process Control, Data mining

## 1. INTRODUCTION

Although there are so many theoretical researches on Learning Classifier Systems (LCSs), however, very few applications have been reported in the literature. This paper is concerned with a practical application of LCSs in order to extract of plant operation knowledge from actual operation data of a biochemical plant. So far, Many kinds of automatic control systems have been established in such plants as biochemical plants. Operator confirmation and manual procedures are essential for a wide variety of products used in small quantities requiring stringent quality control, such as advanced materials. Transfer functions like the delay time function have built up a process model by describing an individual response process. However, process circumstances

might change significantly, according to variations of infused material stuff or operating conditions. Thus, automated acquisition or data mining of processes from actual daily data is desirable to manage these changes. In this research, we propose a heuristic search method for plant operation rules, which could provide guidance on human operators, building up a process response model from a large amount of time series data. The basic principles of the model are 1) to maximize the correlation coefficient among time series data, and 2) to apply LCSs with Minimum Description Length (MDL) criteria [10]. The paper also describes results from applying the proposed method to actual operation data for a biotechnical plant.

## 2. RESEARCH OBJECTIVE

### 2.1 A target plant

We are concerning a biotechnical plant with a distillation tower as in Fig.1. In the distillation tower, low-pressure treatment performs constituent separation after the basic ingredient is infused into the tower.

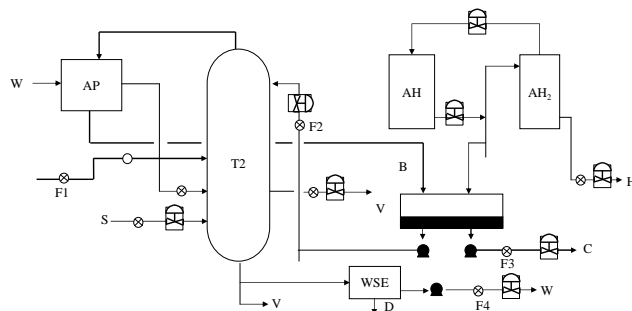


Figure 1: Outline of the biotechnical plant

### 2.2 Problem Description

The purpose of this research is to extract significant information from such time series data that appear to be complicated. The following phases in Fig.2 show how to build up a model in order to analyze the process data. (1) "Reciprocal correlation analysis": It selects two sets of normalized process data and searches for the time difference that indicates the biggest correlation by gradually shifting the time. (2) "Process response model": Like the response

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

GECCO'06, July 8–12, 2006, Seattle, Washington, USA.  
Copyright 2006 ACM 1-59593-186-4/06/0007 ...\$5.00.

model, it describes the relation among the process data from the shifted time and the correlation coefficient. (3) "Extraction of control rules": The process extracts the control rule by executing the classifier system that handle noted process data as a process response model class. (4) "Extraction of a workflow": The classifier system that counts occurring time of operational events and process data is executed so as to find the work flow.

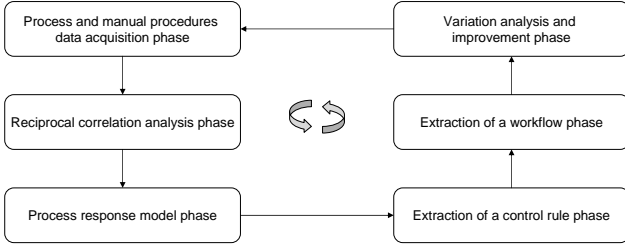


Figure 2: Analyzing phases

### 3. PRINCIPLES OF LSC WITH MDL CRITERIA

Operation rules of such plants require simple and clear description in order for operators to recognize the target process conditions. We use MDL criteria to minimize the complexity of the model and data [13] [9]. MDL criteria are shown as follows: Here,  $m_1$  and  $m_0$  are set to  $m = m_1 + m_0$ , in the data row  $y^m$  with the length  $m$ , as each occurrence number of  $y = 1$  and  $y = 0$ . Also,  $c_i$  indicates 0 in case each condition of the former part has a wild card #, or 1 for other cases. And  $t_i$  indicates division number of the process data in each condition. At that time, description length of data and model are as follows.

$$dataLength = mH\left(\frac{m_1}{m}\right) + \frac{1}{2} \log\left(\frac{m\pi}{2}\right) + o(1),$$

$$modelLength = \sum_{i=1}^k c_i(1 + \log t_i),$$

$$\text{Here, } H(x) = -p \log(x) - (1-x) \log(1-x)$$

Although MDL principle generates a simple and assured model, this does not always mean the model is easy to understand. Thus, we apply the improvement rate of association rules in data mining literature [2]. The following formula expresses the improvement rate: Here,  $P(r_i)$  expresses the rate that the latter part  $r_i$  appears without condition, and  $P(r_i|\mathbf{p})$  expresses the rate that the latter part  $r_i$  appears with the condition of the former part,  $\mathbf{p}$ .

$$improvement = \frac{P(r_i|\mathbf{p})}{P(r_i)}.$$

The description length in MDL principles is calculated as follows: It is known that probability distribution  $P(\cdot)$ , on the assembly of data row  $y^m = y_1 \dots y_m$  with the length  $m$ , exists. Also, the length  $L(y^m)$  of binary code string  $\phi(y^m)$  can be expressed as

$$L(y^m) = -\log P(y^m).$$

Expressing the occurrence rate in marketing basket analysis through a logarithm with the same description length

as the description length of MDL principles, it is possible that the improvement rate is the differential of the information amount between before refining and after refining with certain conditions. As a result, we cope with improved information amount and the description length in MDL principles at the same time. So, MDL criteria are expanded, in order to maximize the differential of the description length (model length + data length) obtained for classification by the learning classifier system. The following method shows the calculation of the learned classifier weight. Here,  $dataLength_f$  and  $modelLength_f$  express the initial description length of data and model,  $dataLength_l$  and  $modelLength_l$  express the final description length of data and model. The weight of classifier is calculated by

$$\begin{aligned} DL_{first} &= dataLength_f + modelLength_f, \\ DL_{last} &= dataLength_l + modelLength_l, \\ Weight &= DL_{first} - DL_{last}, \end{aligned}$$

When the knowledge not noticed without the rule is able to be discovered, and when the rule is simple, it is an unpredictable rule. The expansion proposed here allows for detailed evaluation of the simple rule that can reveal a valuable fact with copious amounts of information, though it is unnoticeable. Evaluating all classifiers hit in the former part, counting the result of its classifier can allow for calculation of the estimated value of a classification error. MDL criteria are used to get the weight, selecting the smallest dataLengths in the classification of results. Then the learning classifier system is implemented to maximize the weight as a fitness function. As a learning classifier system, an original system based on the Pittsburg approach is introduced[11]. This system corresponds to a lot of events, using the learning method it estimates the event distribution by random sampling. Each individual consists of a condition part (as *disjunctive normal form*) and of a conclusion. Fig.3 shows the concept of the learning classifier system in this model. First of all, random generated rules, set as classifiers, classify

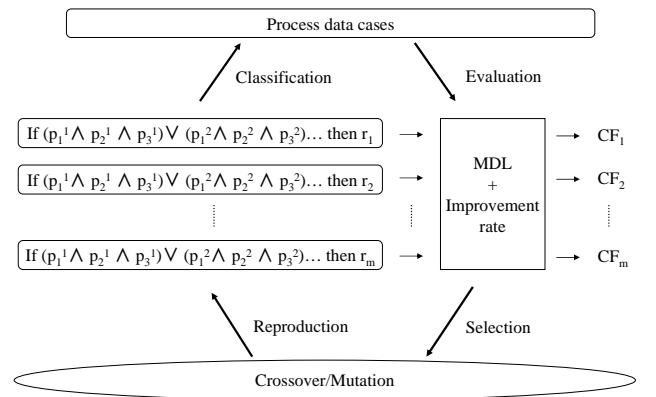


Figure 3: Learning classifier system

the process data. MDL criteria and the improvement rate evaluate these rules and classification results, and set the result to  $CF_i$ . To each classifier, a new classifier is generated, conducting tournament selection based on the obtained  $CF_i$ , crossing and mutation. The number of classifiers : 200, data : 9 items \* 300min, crossover probability : 0.7, mutation

probability : 0.5%. Each numeric data was divided every 25% after it normalized. It was also classified into down-trend, stability and uptrend respectively. When  $\mathbf{p}$  is set as a conditional expression of the former part and  $\mathbf{r}$  is set as a conditional expression of the latter part, the structure of classifier becomes

$$\begin{aligned}\mathbf{p} &= (p_1^1 \wedge p_2^1 \dots \wedge p_k^1) \vee (p_1^2 \wedge p_2^2 \dots \wedge p_k^2) \dots, \\ \mathbf{r} &= r_1, r_2, \dots, r_n.\end{aligned}$$

The result  $r_i$  shows all the kinds of result that the target event would obtain, and counts the number of hits in all  $r_i$  that hit in the former part. This gives the estimated value of reliability for the latter part event, in accord with the agreed-upon event of the former part.

## 4. EXPERIMENTS

### 4.1 The response model

In the case of the continuation process, a correlation is recognized among the data. The following operation produces a reciprocal correlation coefficient of time series data: (1) Select time series data  $\mathbf{x}$  and  $\mathbf{y}$  that of couple process tags. When  $k$  is set as time shift amount of each tag and  $\bar{x}$ ,  $\bar{y}$  are averages, the next formula calculates the next reciprocal correlation coefficient of  $k$ ,  $r_{xy}(k)$ .

$$r_{xy}(k) = \frac{\sum_{t=k+1}^T (x_{t-k} - \bar{x})(y_t - \bar{y})}{\sqrt{\sum_{t=k+1}^T (x_{t-k} - \bar{x})^2} \sqrt{\sum_{t=1}^T (y_t - \bar{y})^2}}$$

(2) Obtain  $k$  that shows maximum correlation.

$$\max_k r_{xy}(k).$$

This operation which consists of all data, can produce a maximum correlation coefficient and shift time table. Table.1 shows a part of it. The above operation can lead to the

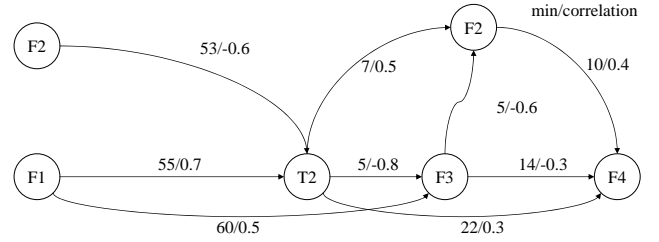
**Table 1: Maximum correlation coefficient/Shift time**

Cor/Time	F4	F2	F3	T2
F4	1.00/0	0.41/-10	-0.32/-14	0.32/-22
F2	0.41/10	1.00/0	-0.57/-5	-0.63/53
F3	-0.32/14	-0.57/5	1.00/0	-0.80/-5
T2	0.32/22	-0.63/-53	-0.80/5	1.00/0

building of the process response model, extracting tags with high correlation from a quantity of time series tag data. Fig.4 shows an example. It shows the structure of process response, by means of time series correlation among process tags and the information of time shift.

### 4.2 Heuristic search for operation rules

In the actual operation, it is significant to discover a control point that makes the final quality stable. The learning classifier system, based on MDL criteria and the improvement rate, searches the control rule targeting tag data with high correlation obtained by the process response model. Fig.5 shows an example of the classifier obtained. At this moment, the improvement rate is 3.1, and MDL value is 32.9 bit. The next example shows the classifier in the case that



**Figure 4: Process response model**

considers the improvement rate. In this case, the improvement rate is 6.6, and MDL value is 54.8 bit. The former becomes a simpler model though, the result is close to commonplace with a low improvement rate. In the case that gives consideration to the improvement rate, in addition to MDL value, an unpredictable rule is more easily revealed. As a result of the interview to the person in charge of the

MDL:

75% < F2 and 75% < F3 then 50% < T2

"F2 flow is 75% or more, and F3 flow is 75%

or more, then T2 temperature becomes 50% or more."

MDL+Improvement:

25% < F3 <= 50% and 75% < F4 and F3 is down  
then 75% < T2

"F3 flow is from 25% to 50%, and F4 flow is 75%  
or more, and F3 flow is decreasing, T2 temperature  
becomes 75% or more."

**Figure 5: control rule**

operation, T2(temperature of the tower) provides an important control point that greatly affects product constituent quality in this biotechnical plant, and it is too difficult to control the temperature. In order to control it more accurately, the classification rule that operators find hard to notice becomes precious information. In the case of Fig.5, F2 flow is expected to be related to T2 temperature, but it is not noticeable that F3 flow transition rate away from T2, or F4 flow are also connected to T2. Furthermore, in Fig.6 too, comparing with F1 ingredient flow which directly infused into the distillation tower and F2 return flow, operators found it hard to notice that F3 flow away from T2 is related to T2 temperature. As mentioned above, the application of Minimum Description Length criteria considering the improvement rate can result in such unpredictable information.

MDL:

(50% < F1 <= 75% and 25% < F4 <= 50%

and F4 is up) or 75% < F2 then T2 <= 50%

MDL+Improvement:

(75% < F2 and 50% < F3 <= 75% and F4 is up)

or 75% < F2 then T2 <= 25%

**Figure 6: Condition rule 2**

## 5. COMPARISON WITH THE CONVENTIONAL METHODS

There has been a lot of research in data mining to obtain useful information from a large amount of data [1][12][4][6]. Also MDL principle has been applied in many fields, for example, Genetic Programming and LCSs as *Bloat* control and the generalization pressure [8][3][5]. Several indicators are proposed in order to measure the interestingness of the extracted knowledge: *J*-Measure, *i*-Measure, *I*-Measure or *IShannon*-Measure for example[7]. However, these indicators do not directly represent the length of the model nor classified data. Moreover, they do not give a human operator specific operation rules.

A simple application of decision tree methods also produces very large trees. For example, our results on actual process data using C4.5 have generated a huge tree with 87 nodes before pruning, 43 nodes even after pruning. They are hard for human operators to understand the meanings. They need a simple model to represent authentic information. Table.2 shows results of the learning classifier system with decision tree C4.5, the learning classifier system on the basis of MDL criteria and MDL + improvement rate criteria.

**Table 2: Decition tree C4.5, LCS on MDL**

	Node	MDL	Model Length	Errors
C4.5 before pruning	87		297.2bit	12.0%
C4.5 after pruning	59		191.6bit	14.3%
MDL criteria		121.3	9.8bit	6.7%
MDL+Imprv rate		109.4	9.8bit	7.0%

The classification error rate indicates a lower value in MDL criteria or MDL + improvement rate criteria, than decision tree 4.5 before and after pruning. And in the rule description length, the MDL + improvement rate criterion classifier considerably decreases compared to C4.5.

## 6. EXTRACTING WORKFLOWS FROM HUMAN OPERATORS

Using the proposed method, we also generate workflow processes for operations. First, we collect event data such as switching and value setting operations with time stamps. Then, using the proposed methods, we search for rules during given time intervals. Finally, we sort the acquired rules with the time key, then we get the corresponding workflows. Fig.7 shows the result of discovering rules with learning classifiers. The workflow states: If switch 1(blower) is turned on, the state on surface of the product will confirm "sticky" in 15 minutes, and switch 2(chiller) is turned on, dry temperature T2 will become 25% or less in 20 minutes. Although the workflow in the figure is very small, however, process in the workflow occurs certainly on the condition.

The method enables even human expert operators to support to discover implicit plant operation knowledge from both operation manuals data and process data. Such knowledge is useful to transfer experts' special skills to naive operators.

## 7. CONCLUSION

SW1=ON/0 and Visual exam=sticky/15 and SW2=ON/15  
then T2<=25%/20

**Figure 7: Work flow**

This paper has proposed a novel method to extract plant operation knowledge from time series data using LCSs with MDL principle. The method has generated useful but simple operation knowledge with high reliability. The effectiveness of the proposed method has demonstrated using actual plant data. We believe the proposed method is one of the good practise of LCS applications. Although the accuracy of the rules was about 93%, they are valuable because they have revealed important operation check points, which are very difficult to find even for human experts. These knowledge will be of use for the future technology transfer of the plant operation.

## 8. REFERENCES

- [1] P. Adriaans and D. Zantinge. *Data Mining*. Addison-Wesley, 1996.
- [2] R. Agrawal, H. Mannila, R.Srikant, H. Toivonen, and A. I. Verkamo. *Fast Discovery of Assosiation Rules*. AAAI Press and The MIT Press, 1996.
- [3] J. Bacardit and J. M. Garrell. Bloat control and generalization pressure using the minimum description length principle for a pittsburgh approach learning classifier system. In *Sixth International Workshop on Learning Classifier Systems (IW LCS-2003)*, July 2003.
- [4] D. J. Berndt and J. Clifford. *Finding Patterns in Time Series: A Dynamic Programming Approach*. AAAI Press and The MIT Press, 1996.
- [5] M. V. Butz, M. Pelican, X. Llorca, and D. E. Goldberg. Extracted global structure makes local building block processing effective in xcs. *IlliGAL Report*, 2005011, February 2005.
- [6] M. L. Hetland. Evolutionary rule mining in time series database. *Machine Learning*, 58:107–125, 2005.
- [7] R. J. Hilderman and H. J. Hamilton. *Knowledge Discovery and Measures of Interest*. Kluwer Academic Publishers, 2001.
- [8] H. Iba, H. de Garis, and T. Sato. Genetic programming using a minimum description length principle. In K. E. Kinnear, Jr., editor, *Advances in Genetic Programming*, pages 265–284. MIT Press, 1994.
- [9] M. Mehta, J. Rissanen, and R. Agrawal. MDL-based decision tree pruning. In *Proceedings of the First International Conference on Knowledge Discovery and Data Mining (KDD'95)*, pages 216–221, 1995.
- [10] J. Rissanen. Modeling by shortest data description. *Automatica*, 14:465–471, 1978.
- [11] S. Smith. Flexible learning of problem solving heuristics through adaptive search. In *Proceedings 8th International Joint Conference on Artificial Intelligence*, August 1983.
- [12] S. M. Weiss and N. Indurkha. *Predictice Data Mining, A Practical Guide*. Morgan Kaufmann Publishers, Inc, 1998.
- [13] K. Yamanishi. A learning criterion for stochastic rules. *Machine Learnibg*, 8:165–203, 1992.