Proceedings 2002 IEEE International Conference On Control Applications, Glasgow, Scotland, pp 687-688.

ANALYSIS OF PLANT-WIDE DISTURBANCES USING CONDITIONAL ENTROPY

J. Paul Dobson and Nina F. Thornhill

Department of Electronic and Electrical Engineering, Torrington Place, University College London, London WC1E 7JE, UK (*n.thornhill@ee.ucl.ac.uk*).

ABSTRACT

Plant-wide disturbances in chemical plants have an impact on product quality and running costs, thus there is a need for automated diagnosis of the root cause. This short paper reports work in progress and preliminary results from a novel signal processing application using conditional entropy. It analyses measurements from normal process operations in a chemical plant to determine cause and effect relationships. The method suggests which measurement is the root cause of a plant-wide disturbance and, additionally, maps the physical structure of the plant.

Key words: Chemical industry; conditional entropy; control loop performance; fault diagnosis; non-linearity; plant-wide disturbance; process control; signal processing.

1. INTRODUCTION

This paper explores the potential of conditional entropy calculations for the diagnosis of root causes of plant-wide disturbances. Conditional entropy has the potential to determine cause-and-effect because it can show that measurement *A* influences measurement *B* more than *B* influences *A*. A key motivation is to provide a diagnosis while the plant is still running so that the maintenance effort may be optimally directed, thus reducing the costs of lost production during a maintenance shutdown.

Preliminary results correctly found the measurement closest to the root cause of a plant-wide disturbance in an industrial data set. In addition, the method may be used to map the physical structure of the plant from the data alone. The significance of such a map will be that the analysis is able to explain the propagation of a plant-wide disturbance from its root cause. It is noted that a cause and effect map of the plant also has the potential to identify suitable pairings of manipulated and controlled variables in the design of a multivariable control systems.

2. METHODS

2.1 Conditional entropy

For univariate random variables *X* and *Y* having values x_i and y_j with associated probabilities:

$$P\left(X = x_i\right) = p_i \quad \text{and} \quad P\left(Y = y_j\right) = q_j$$

the conditional entropy of *X* with respect to *Y* is:

$$H\left(X|Y\right) = -\sum_j \sum_i r_{ij} \log\left(\frac{r_{ij}}{q_j}\right)$$

where $r_{ij} = P\left(X = x_i, Y = y_j\right)$

$H\left(Y|X\right)$ is defined similarly with q_j replaced by p_i. Conditional entropy is a quantitative measure of the remaining uncertainty of *X* given information about the state of *Y*. In the case of *Y* providing significant information about *X*, for example if there is a linear relationship between their probability distributions, then $H\left(X|Y\right)$ tends to zero. Conditional entropy is asymmetric because generally $H\left(X|Y\right) \neq H\left(Y|X\right)$.

2.2 Practical implementation

For a practical implementation the continuous time domain signal is quantized using quantizer levels $c_1, c_2, ..., c_\ell$. The entropy calculation then proceeds using the probabilities that values fall into a given quantizer interval. Thus:

$$p_i = P\left(c_i \leq X < c_{i+1}\right),\ \left(i = 1, 2, ...\ell - 1\right)$$

$$q_j = P\left(c_j \leq Y < c_{j+1}\right),\ \left(j = 1, 2, ...\ell - 1\right)$$

$$r_{ij} = P\left(c_i \leq X < c_{i+1},\ c_j \leq Y < c_{j+1}\right)$$

For a dynamic time series the history of the signal over a number of samples is of interest. Consecutive samples from the time signals are grouped to form words of finite length *L*, and new random variables *W* and *V* are derived from *X* and *Y* respectively. *W* and *V* are defined such that the words w_i and v_j are composed from a current instance of x or y together with those *L-1* samples previous in time. For example $w_i = \{1, 3, 4\}$ indicates a short sequence such that the first instance of x fell into quantizer interval 1, the next into interval 3 and the last into interval 4. The probabilities of the w_i are determined by enumeration of all possible words and counting their occurrences. Similar comments apply to *V*. Conditional entropy calculations then proceed on *W* and *V* rather than on *X* and *Y*.

2.3 Refinery data set

Data from a refinery unit were kindly provided by a SE Asian refinery. Previous work (Thornhill *et.al,* 2001) investigated a plant-wide oscillation in the data set. Time trends of the oscillatory tags are shown in Fig 1. The first challenge for the conditional entropy analysis is to identify the root cause of the oscillation (known to be in a recycle loop that includes Tags 33 and 34). A further challenge is to map the structure of the plant from an analysis of the data.

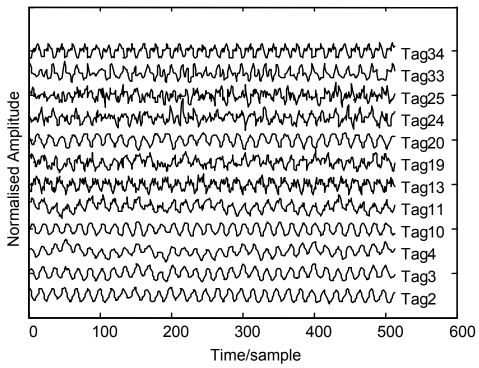

Fig 1. Time trends from refinery data

3. RESULTS

3.1 Root cause analysis

Table 1 shows conditional entropy calculations for all pairs of the oscillatory variables. The conditional entropy of Tag *A* given information from Tag *B* is to be found in the column headed *B* and row headed *A*. The column sums give an indication of the remaining uncertainty of the other signals given knowledge of the signal to which the column refers (Rouncefield, 1998). A candidate for the root cause of the plant-wide oscillation is that signal which gives the most information about the other oscillating signals.

Tag 34 is a candidate for the root cause because the conditional entropies in its column have the smallest sum, meaning that tag 34 gives the most information about the other signals. This is the same conclusion as was previously reported using non-linear time series analysis (Thornhill *et.al*., 2001). Tag 34 is the flow in a recycle, which is why the disturbance propagates widely in the plant.

3.2 Plant structure analysis

The case when $H\left(W\middle|V\right) << H\left(V\middle|W\right)$ means that V derived from tag Y contains more information about W (derived from X) than vice versa. In terms of cause and effect, Y influences X. Relationships for each tag, for instance 34, may be inferred from the calculations. Fig 2 shows which tags influence and are influenced by Tag 34. The numerical values are $\left|H\left(W\middle|V\right) - H\left(V\middle|W\right)\right|$. The cause-and-effect relationships analysis from Table 1 shows:

- Tag 24 influences 34;
- Tag 34 influences 2, among others;
- Tag 2 influences Tag 24.

The feedback implicit in the relationship matches the known plant structure because tag 34 is in a recycle. The fact that 34 was found to be the root cause suggests that another unmeasured influence comes into the recycle loop at tag 34. This is most likely to be a non-linearity caused by a faulty actuator.

4. CONCLUSIONS AND FURTHER WORK

The paper has presented preliminary results from work in progress on the use of conditional entropy for the analysis of operating data from chemical plants. Application to refinery data has shown that the method has the potential to infer cause-and-effect relationships from plant data and for diagnosis of the root cause diagnosis of a plant-wide disturbance. Several implementation issues remain to be addressed in the next phase of the work. These include:

- Automated selection of the quantization regime;
- Optimization of the word length;
- Determination of a statistical threshold for the decision $H\left(W\middle|V\right) << H\left(V\middle|W\right)$;
- Automation of cause and effect network mapping.

5. REFERENCES

Rouncefield, M., 1998, *Statistics for experimenters*, Wiley

Thornhill, N.F., Shah, S.L., and Huang, B., 2001, Detection of distributed oscillations and root-cause diagnosis, *CHEMFAS4, Korea, 7-8 June*, 167-172

Table 1: Conditional entropy results and totals

tag	2	3	4	10	11	13	19	20	24	25	33	34
2	0.00	0.21	0.29	0.19	0.22	0.22	0.25	0.23	0.42	0.31	0.33	0.09
3	0.15	0.00	0.18	0.23	0.50	0.29	0.25	0.27	0.31	0.45	0.20	0.15
4	0.42	0.41	0.00	0.32	0.25	0.21	0.32	0.29	0.29	0.37	0.36	0.07
10	0.23	0.37	0.29	0.00	0.39	0.17	0.37	0.30	0.21	0.49	0.40	0.11
11	0.19	0.37	0.24	0.33	0.00	0.26	0.23	0.34	0.48	0.46	0.39	0.27
13	0.34	0.32	0.22	0.34	0.42	0.00	0.40	0.38	0.30	0.76	0.58	0.01
19	0.21	0.18	0.26	0.28	0.35	0.25	0.00	0.40	0.22	0.37	0.33	0.15
20	0.37	0.13	0.28	0.16	0.57	0.32	0.15	0.00	0.38	0.52	0.48	0.18
24	0.16	0.36	0.35	0.27	0.63	0.27	0.46	0.42	0.00	0.29	0.41	0.56
25	0.59	0.42	0.44	0.36	0.55	0.51	0.33	0.52	0.29	0.00	0.58	0.31
33	0.33	0.22	0.20	0.46	0.51	0.20	0.25	0.36	0.27	0.31	0.00	0.20
34	0.37	0.26	0.36	0.32	0.39	0.28	0.31	0.32	0.25	0.29	0.31	0.00
sum	**3.36**	**3.25**	**3.41**	**3.26**	**5.29**	**3.08**	**3.32**	**3.82**	**3.42**	**4.62**	**4.17**	**2.10**

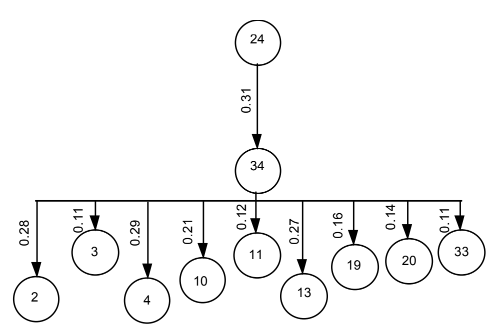

Fig 2. Network map showing the influences measurements have on one another