From time series measurements to rules of causality

K J Raivio
Department of Information and Computer Science, Aalto University School of Science,
P.O.Box 15400, FIN-00076 Aalto, Finland
E-mail: kimmo.raivio@tkk.fi

Abstract. Data analysis procedure is described which can be used to extract events and rules of causality from time series measurements of industrial processes. The proposed procedure incorporates well-known data analysis methods, that have not been widely used in condition monitoring systems. Here, it will be demonstrated how the algorithms can be utilised in mobile work machine condition monitoring. The analysis process starts with selection of representative measurements which describe the operation of machine reasonable well. Here, variables are selected using a clustering method, in order to find groups of variables. One measurement from each group is selected for later analysis. Next, data streams are segmented to find areas on which the operation of machine continues without any abrupt changes. The segments or combinations of them are analysed to be able to extract sequences of operational states or change points. Event sequences are further analysed to extract association rules for the events. The extracted rules contain information about occurrence probability of certain event sequence. This facilitates, e.g., identification of fault precursors. Probabilities computed from measurements using this procedure can be used to adjust expert knowledge based fault-trees.

1. Introduction
The machines used in the process industry and on mobile work sites get all the time more complicated and more expensive. Thus, new advanced condition based maintenance systems are developed to be able to monitor and analyse the machines. New sensor technologies are also utilised to be able to collect the measurements.

Huge amounts of data need to be processed effectively. Thus, new advanced data analysis methods are needed to be able to extract the essential information out of the data. The operators and the staff working in the research and development in those companies would like to know what kind of events occur in the machines. They also would like to know what is the operational state of the process or machine in a format which is easy to understand.

Advanced adaptive data analysis methods can be used to extract event sequences and load profiles which can be further analysed with other tools. Such tools can perform state recognition, work cycle modeling and anomaly detection based on probability distributions of events. In state or event detection, the data is fed through several detectors which are specialised to detect certain events. These events can be stored in log files or they can be analysed to find causal dependencies. Event sequences can also be used as inputs to work task recognition methods [1; 2]. In these Hidden Markov Model based methods, the events are expected to occur with certain probability during each task of work cycle. In this paper, a method, which is able to build rules of causality from time series measurements will be described. At first, a variable selection method for monitoring purposes is described. Then, two segmentation based event detection
methods will be described. Sequences are combined and association rules are computed from the event sequences. The main objective is to demonstrate in practice how causality analysis of time series measurements can be performed using as small amount of expert knowledge from the process as possible. The method is an alternative to traditional fault tree based analysis and can also be used to update the fault trees. Association rules can be utilised, for example, in setting alarms.

The contents of this article is divided in sections as follows. The analysis methods are described in section 2 and the results are shown in section 3. Finally, the conclusions are drawn in section 4.

2. Methods

The first step of the method is the variable selection (see section 2.1). Variables are selected for monitoring purposes i.e. such variables are proposed which are as dissimilar with each other as possible to be able to include as much information as possible from the process into a small set of variables.

When the variables have been selected, they are preprocessed and the time series are segmented. Segments are classified as such or consequent segments are classified together. Classified segments or segment groups are interpreted as event sequences which can be combined and analysed together to find association rules for events. Consequent segments can be used to detect change points from data [3]. The method described in figure 1 has two different event detectors analysing the same data streams. The right branch detects steady states, while the left branch detects change point events.

In general, several event detectors should be used to analyse the same data (see figure 2). They should differ from each other in respect of variable selection, type of event detected or something else. Type of event can be continuous state described by one segment, edge described by one segment of one variable and limited by a threshold or a change point described by difference between successive segment features. Typically, segment features are slopes given by linear regression or interpolation. Also, other definitions for events can be used.
The methods presented in this paper are demonstrated using time series measurements collected from test drives of medium-sized mobile machine [4] engineered at the Department of Intelligent Hydraulics and Automation at Tampere University of Technology. Machine has a length of 6.5 m, width of 2.0 m and it weighs about 4000kg. There were 79 drives and the length of drive varied from a couple of seconds to 1min 45s. The ground type was either asphalt or gravel. The machine was driven by a person on board or remotely using laptop. In some cases, an additional load of 1000kg was also used.

2.1. Variable selection

Variable selection method described in this paper is based on clustering. For later analysis that variable from each cluster is proposed which is the nearest of the prototype vector. About same kind of analysis can be performed using correlations. However, clustering gives the groups and an estimate how many variables should be selected. Clusters of variables can be visualised using principal component analysis [5] to map the result into two dimensions. The clustering result of all studied variables can be seen in figure 3.

The result varies somewhat depending on time. When average linkage hierarchical clustering method [6] is used, the number of clusters can be estimated from a dendrogram (see figure 4). The dendrograms drawn from the hierarchical clustering of variables vary only slightly depending on the time window of samples used in the analysis. If the dynamics of the process is more changing the results can vary a lot more. So, test should be performed using several selections of time window. Distances between horizontal levels in the dendrogram are defined by distances between clusters. The leaves of the dendrogram correspond to clusters of one variable.

The names of variables corresponding to the most interesting variable indexes can be found from table 1. For example, 'Diesel rotational speed reference' represents the most separate variable group. From figure 4 we can see that this group is found even in case of two clusters.
Figure 3. Variables are mapped to 2D using the eigenvectors corresponding to two largest eigenvalues given by principal component analysis. Variables closest to cluster prototype vectors have been shifted by a small amount to avoid overlapping names and they have been typed with red font.

Figure 4. Dendrogram of clustered variables. Variable indexes are on horizontal axis and cluster distances can be computed from the differences between the values on vertical axis.
Table 1. Selected variables.

| Variable group          | No | Variable name        | Range               | Unit |
|-------------------------|----|----------------------|---------------------|------|
| Driving behaviour       | 13 | Frame angle          | 193,1…117,1        | °    |
|                         | 19 | Steering reference   | -16384…16383       | -    |
| Hydrostatic transmission| 2  | Diesel load          | 0…100              | %    |
|                         | 7  | Diesel rotational speed reference | 0…2200 | rpm |
|                         | 20 | HST pump measured angle | -1000…1000 | °    |

2.2. Preprocessing

Now, the variables in the table 1 are selected for further analysis. The variables are divided manually into two groups, which are variables describing driving behaviour and variables related to hydrostatic transmission.

In the preprocessing stage, outliers and noise are removed. Here, moving average filtering is used to filter out noise and downsample the time series so that the time step between successive samples is one second. Variables are also scaled to zero mean and unit variance in order to guarantee that all variables have equal influence on segmentation result. Removal of noise by moving average filtering also removes short segments which are not so desirable. Additionally, filtering also speeds up the computation.

2.3. Segmentation of time series

Preprocessed time series measurements are segmented using sliding window method [7] in order to find continuous operational states. Here, continuous means a state with almost constant measurements or with about fixed increasing or decreasing trend. Linear regression with cumulative squared error function is used. It is also possible to use linear interpolation instead of regression and maximum of squared error instead of cumulative squared error. In the used method, the value of a cost function is computed over the variables of a group and samples are added to the segment until the value of a cost function exceeds the threshold. Then, a new segment is started and costs are computed from the beginning of the new segment. All the variables in a group define together the edges of segment when they have a common cost function. In figure 5, the segments of one test drive are shown.

2.4. Clustering and classification of segments

The segments found as described in the previous section are characterised by the slope and the offset of the regression line and the length of segment. These can be used as features when the segments are clustered and classified to have a label for each segment. Of course, segments can be classified using predefined models or some part of the data can be used to define models and the other segments can be classified using these models.

Number of clusters can be defined using dendrograms as in variable selection procedure (see section 2.1) or a cluster validity index can be used. Here, the average linkage clustering is performed using segment slopes as features. Several number of clusters were tested and the one giving minimum value of Davies-Bouldin validation index is selected [8]. In figure 6, the cluster prototype vectors are shown. For driving behaviour related measurements 20 classes has been defined. From the models it can be seen that for class 20 of driving behaviour related measurements both measurements (Steering reference and Frame angle) increase with the same rate during the segment. Frame angle is the measured angle between the front and the rear part.
2.5. Postprocessing segment classes
Labels of consequent segments can be interpreted as operational states or events. In case transitions from states to themselves are undesirable, successive segments with same labels can be concatenated. Of course, this operation changes probability distributions of states and transition probabilities between them.

Segments can also be further processed to find more interesting events. It is possible to filter out non-interesting segment classes. Such a class can be, for example, a class with zero or almost zero slope features. This kind of class represents a stable state without any changes in measured signal levels. If segments have been built using only one variable, the slope coefficient of regression line can be used in edge detection. Successive segments can also be analysed together to find more complicated events like change points.

2.6. Change point detection
If the distance between slope coefficients of two consecutive segments is larger than some predefined threshold, a change point is detected [3]. Euclidean distance between successive segments is computed using the selected features. In this article, distance of 0.5 is used as a threshold.
2.7. Clustering and classification of change points

Change points are classified to be able to label them. Feature vectors are either clustered or earlier defined model vectors are used in classification. In case new models are defined, number of the models have to be defined. In this paper, the number of change point classes has been fixed to 8 for both variable sets and average linkage hierarchical clustering is used. Prototype vectors of change point classes in case of driving behaviour related features are shown in figure 7. One change point class per variable group has been filtered out because the change was smaller than selected threshold.

Change points build up an event sequence of their own. In a sequence, each event has a class label or a type, and a time stamp.

2.8. Combining sequences

Event sequences based on different variable sets and either on steady operational states or change points are combined for further analysis. Thus, the new sequence consists of events found using two kind of detectors. Total amount of event types is a sum over all event types generated by the detector pool. Exact time stamps of events are needed when the sequences are combined.

2.9. Association rules

Combined event sequence is analysed using Apriori algorithm [9; 10] to find association rules, i.e. rules of causality, for the events. The algorithm finds frequent subsequences from an event sequence by counting the occurrences of subsequences in a fixed length window. The window is moved over the sequence from the beginning to the end. The computational efficiency of the algorithm is based on the fact that all subsequences of a candidate frequent episode must also be frequent. An episode is frequent if the probability of an episode is above a predefined threshold. Also, a window length and a maximum length of episodes to be searched have to be defined. The total amount of windows have to be counted to be able to compute the probabilities. The algorithm starts with episodes of length one. Length of episodes is increased by one every round. At the second round, it is possible to search for serial episodes, in which the order of events is taken into account or parallel episodes in which the order is not important. In parallel case, it is just checked whether the events occur in the same window. If longer episodes are searched, it is possible to mix serial and parallel episodes if desired.

Frequencies of episodes are converted to conditional probabilities using
Figure 8. Segmented time series measurements. Segments have been coloured using their classes. Segment classes and change point classes are shown under the curves.

$$P(B|A) = \frac{P(A, B)}{P(A)}.$$  

Here, event $A$ precedes event $B$ and $P(A)$ is the probability of event $A$. The idea is to find association rules in which event $B$ will follow event $A$ with reasonable high conditional probability compared to how often event $A$ occurs. The rule is more important if the events occur more frequently.

3. Results

Segmented driving behaviour related measurements of one test drive are shown in figure 8. There are 20 classes or events due to driving behaviour related measurements and 12 due to hydrostatic transmission related measurements. In addition, there are 8 change point classes or events per measurement group. Thus, there are 32 events based on segment classes and 16 events based on change points, which makes total 48 event types. Event 27 of the original events is filtered out, because all features building this event are below threshold 0.1 and thus the event is considered non-interesting.

Also, two event types given by change point detection are filtered out, because the change in them is not large enough to be considered as change point. If the distance between pair of segments in change point models is smaller than 0.5, the model is not considered to be a change point. There exists one such change point class for both measurement groups. Event labels are also shown below the segment wise coloured curves in figure 8. The first label row is for segment based events and the second one for change point related events.

The probabilities of events are shown in figure 9. Event 27 is included in the figure, but its probability is zero. Change point classes which are filtered out are removed from the probability distributions and further analysis. Event 6 is the most probable and when checked from figure 6, it can be seen that it is characterised by almost constant levels of signals. Thus, it is possible to adjust the threshold and filter events of this type out if desired.

The most important association rules found by Apriori algorithm and their conditional probabilities are shown in table 2. Events occurring less than 4 times are filtered out. From the
Figure 9. Event probabilities.

Table 2. Association rules and their conditional probabilities.

| A   | B   | $P(A, B)$ | $P(A)$  | $n(A)$ | $P(B|A)$ |
|-----|-----|-----------|---------|--------|----------|
| 32  | 42  | 0.02336   | 0.03271 | 21     | 0.7143   |
| 13  | 35  | 0.00779   | 0.01090 | 7      | 0.7143   |
| 25  | 44  | 0.02336   | 0.03583 | 23     | 0.6522   |
| 18  | 34  | 0.01090   | 0.01713 | 11     | 0.6364   |
| 14  | 37  | 0.00779   | 0.01246 | 8      | 0.6250   |
| 23  | 44  | 0.02492   | 0.04206 | 27     | 0.5926   |
| 37  | 6   | 0.00779   | 0.01402 | 9      | 0.5556   |
| 44  | 6   | 0.04206   | 0.08100 | 52     | 0.5192   |

table we can see, for example, that event 32 occurs 21 times and that event 42 follows event 32 with over 71% probability.

4. Conclusions
A method which is able to extract rules of causality from time series data has been presented. The main objective has been to demonstrate how the knowledge can be extracted using general purpose data analysis methods and a minimum amount of manual work. However, measurements have to be selected and detector types and some parameters have to be defined. Similar method can be used in the analysis of any industrial process or work machine.

In the method, time series measurements are segmented, segments are classified as themselves or a combination of successive segments is classified to build up an event sequence. Multiple types of event detectors, can be utilised in building up an event sequence. Even log files can be processed for that. Rules of causality are searched from the combined sequence, in order to give a more compact description of the process.
Described data analysis procedure provides the operator new useful information from the process. That is, because many conventional analysis procedures rely on tests based on predefined constants and fault-tree structures and probabilities which might not be valid any more. The parameters can also be outdated due to change of machine operating mode, environment or ageing. It is also possible that the parameters have never been exactly correct, because they are originally just estimates. Fault-trees also do not necessary cover all types of faults.

Acknowledgements
The author would like to thank the Tolkku project of FIMECC EFFIMA program and Adaptive Informatics Research Centre for funding. The latter is a centre of excellence of the Academy of Finland. The co-operation with colleagues from VTT Technical Research Centre of Finland and Tampere University of Technology is also acknowledged.

References
[1] Aulanko S 2009 Work cycle modeling of Ship-to-Shore container cranes Master’s thesis Helsinki University of Technology Espoo, Finland
[2] Palmroth L 2011 Performance Monitoring and Operator Assistance Systems in Mobile Machines Ph.D. thesis Tampere University of Technology Tampere, Finland
[3] Bhattacharya P 1994 Institute of Mathematical Statistics Lecture Notes – Monograph Series 23 28–56
[4] Backas J, Ahopelto M, Huova M, Vuohijoki A, Karhu O, Ghabcheloo R and Huhtala K 2011 The Twelfth Scandinavian International Conference on Fluid Power (Tampere, Finland) pp 161–176
[5] Haykin S 2009 Neural Networks and Learning Machines (Upper Saddle River, New Jersey, USA: Pearson Education)
[6] Everitt B 1993 Cluster Analysis (London, UK: Arnold)
[7] Keogh E, Chu S, Hart D and Pazzani M 2001 IEEE International Conference on Data Mining (San Jose, California, USA) pp 289–296
[8] Davies D L and Bouldin D W 1979 IEEE Transactions on Pattern Analysis and Machine Intelligence 1 224–227
[9] Agrawal R, Mannila H, Srikant R, Toivonen H and Verkamo A I 1996 Advances in Knowledge Discovery and Data Mining (AAAI/MIT Press) pp 307–328
[10] Mannila H, Toivonen H and Verkamo A 1997 Discovery of frequent episodes in event sequences Tech. Rep. C15 University of Helsinki, Department of Computer Science Helsinki, Finland