December 2004

Ontology trend analysis of dynamic signals

David A. Stirling
University of Wollongong, stirling@uow.edu.au

P. Zulli
Bluescope Steel, Australia

Ontology trend analysis of dynamic signals

Abstract

This paper describes a novel approach to analysing trends of a performance signal indicator from an industrial metallurgical reactor over a number of years of operation. Using a minimum message length algorithm, a detailed ontology of the signal behaviours or modalities was established. An abstraction of these yielded a number of related super states that in turn provided an insightful correspondence for the domain experts. Further detailed identification of the likely composition and causal influences contributing to each mode was subsequently induced with supervised learning.

Keywords

expert systems, learning (artificial intelligence), metallurgical industries, ontologies (artificial intelligence)

Disciplines

Physical Sciences and Mathematics

Publication Details

This paper originally appeared as: Stirling, D & Zulli, P, Ontology trend analysis of dynamic signals, Proceedings of the 2004 Intelligent Sensors, Sensor Networks and Information Processing Conference, 14-17 December 2004, 445-449. Copyright IEEE 2004.

Ontology Trend Analysis of Dynamic Signals

D. Stirling [1], P. Zulli [2]
[1]*School of Electrical, Computer and Telecommunications Engineering, University of Wollongong*
NSW 2522, stirling@elec.uow.edu.au
[2] *Steel Research, BlueScope Steel*
P.O Box 202, Port Kembla, NSW 2505, paul.zulli@bluescopesteel.com

Abstract

This paper describes a novel approach to analysing trends of a performance signal indicator from an industrial metallurgical reactor over a number of years of operation. Using a minimum message length algorithm, a detailed ontology of the signal behaviours or modalities was established. An abstraction of these yielded a number of related super states that in turn provided an insightful correspondence for the domain experts. Further detailed identification of the likely composition and causal influences contributing to each mode was subsequently induced with supervised learning.

1. INTRODUCTION

Ordinarily the learning of ontologies is usually associated with discovering taxonomic relationships from such fields as knowledge acquisition and sharing, knowledge representation or distributed knowledge-based systems [1-3]. Others have developed methods to symbolise and classify various modalities in trends or time series data [4,5]. However, in this paper we explore the use of an ontology-based analysis of a single numeric performance indicator over a number of years.

The performance indicator in question, known as the Stave Heat Load (SHL) is used to monitor the aggregate heat loss to the cooling elements lining the outer surface of the metallurgical reactor. In reality, measurement is a composite of several sensor measurements sampled simultaneously. The SHL is critically important to both the long- and short-operation of the process. Long-term, the operating life (campaign) of the reactor may be maximised by reducing both the magnitude and the variability of SHL. Short-term, the fuel consumption (and product quality) may be improved by maintaining the heat loads as low as possible. The fuel consumption is a reflection of the type and amount of feed materials consumed over time, as well as the various control strategies that are utilised.

Skilled operators and other domain experts often judge the impact of any change in type, rate or distribution of feed material, or the associated control strategy, by the magnitude and trend variations of SHL on a daily basis. There are possibly over 200 parameters that could influence the dynamic reaction process and consequently alter the SHL levels. Because of this multiplicity of possible drivers, operational changes are carefully regulated, monitored and managed by operational personnel. An according body of knowledge has consequently been amassed through this process. Within this body of knowledge are well-understood modes of process behaviours that are often used to successfully characterise the dynamic processes that may be occurring. The driving influences in such cases can be adjusted so as to correct undesirable trends or promote favourable ones. This scenario, however, does not work for all cases. On occasions a seemingly familiar modal trend may not respond to remedial actions as it had in the past—in fact it may have worsened. Clearly new or unfamiliar behaviours (process modalities) might exist, manifesting as familiar SHL trends, and yet have different underlying drivers (process influences). A major objective for this work was to understand the driving influences behind the unexpected trends in SHL. This was achieved by learning an extensive ontology of the process behaviours and subsequently employing supervised learning to identify driving influences in each.

In Section 2 a comprehensive base level range of modalities (clusters or behaviour classes) is discussed. In Section 3, these are further abstracted into a higher level of more discernible super-modes. Section 4 outlines the use of supervised learning to identify key drivers for these.

2. USING UNSUPERVISED CLUSTERING WITH MML

Unsupervised clustering is based on the premise that there are several underlying classes that are hidden or embedded within a data set. The objective of such processes is to identify an optimal model representation of these intrinsic classes, by separating the data into multiple clusters or subgroups. The partitioning of data into candidate subgroups is usually subject to some objective function like a probabilistic model distribution, e.g. Gaussian. From any arbitrary set of data, several possible models or segmentations might exist with a plausible range of clusters. Accordingly an appropriate evaluation method, such as Minimum Message Length encoding, is used on each in order to identify the best model. This methodology is also known as mixture modelling or intrinsic classification [6,7]. Algorithms such as “SNOB” or “AutoClass”, which employ MML metrics have been found to reliably form clusters that are statistically distinct (unique) within the multi-variate parameter space of the data. The specific software used in this work was ACPro i.e. a well known data mining program used for some of the NASA Mars Lander studies in 1998 [8].

The day-averaged SHL measurements, as previously indicated, are in fact a composite average of 16 individual heat loads from areas covering the majority of the reactor surface. Thus in order to examine this problem more comprehensively, the total heat load data was also disassociated into its 16 constituent heat loads being four sectional heights and four circumferential zones (each quarter of the reactor). These are illustrated in Figure 1 (a) and (b) as normalised trends over a number of years of operation.

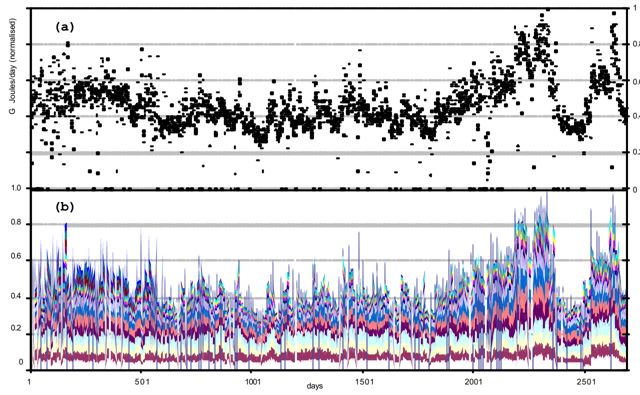

Figure. 1: Normalised SHL trends over 2688 campaign days (a) total daily average (b) stacked profile of sixteen individual zones.

One can observe in Figure 1 (a) that after approximately 500 to 1800 days into the campaign, the initial elevated SHL levels have been reduced to lower average values. Subsequently the SHL trend appears to rise and fall as if some unknown process behaviours or uncontrolled modes have been activated. The relative proportions of the sixteen zones, seen in Figure 1(b), provide no obvious visual insight as to why. However, by employing unsupervised MML clustering to the stream of these, a Gaussian mixture model of 38 classes (or potential modes) was established—as represented in Figure 2.

Each cluster in this model possesses a unique profile of means and standard deviations across all sixteen subcomponents of the SHL. As seen in Figure 1 (a) there are numerous isolated point values, many at very low levels of SHL (several of these appear in the data as zero). These data were associated with a number of likely contexts. The most obvious is the various maintenance shutdowns carried out on the furnace for short periods (typically 1-2 days), where the feed and fuel levels would decrease leading to a proportional decrease in the generated SHL. In other more dramatic situations (decrease to zero) may have arisen from either database failures (missing values) or sensory equipment failures (zero or erroneous signals). In most data mining work, such events would be dealt with in some data cleaning procedure, however for the purposes of capturing a comprehensive set of operational modes these were not conditioned or removed. Many of the cluster profiles identified in the 38 state model illustrated in

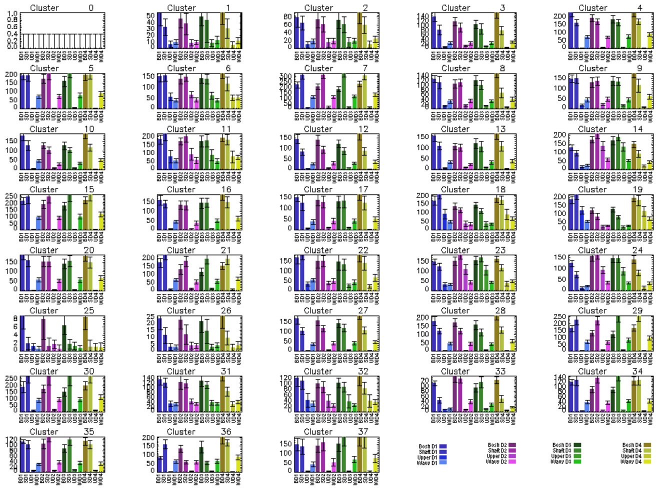

Figure. 2: Gaussian mixture model of 38 clusters: each cluster represents the 16 member SHL subcomponents as a histogram of mean values of each together with specific standard deviations (error bars).

Figure 2 were familiar or understandable to the domain experts, whilst several others in contrast were difficult to reconcile to their prior understanding.

In an attempt to understand these in some overall context, the clusters were visualised through a number of alternative perspectives, such as that in Figure 3. Here, each cluster in the model is represented by its combined total SHL value, and further the set of clusters has been ordered according to their abundance.

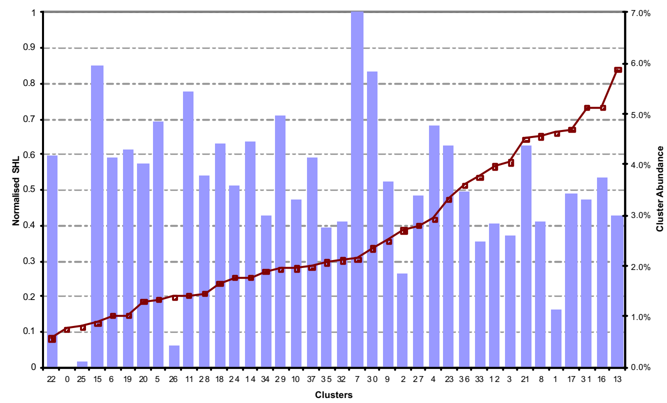

Figure. 3: Normalised SHL trends over 2688 campaign days (a) total daily average (b) stacked profile of sixteen individual zones.

Based on these views one could develop a coarse, yet effective ontology, a *zero-ordered-ontology*, with each of the 38 raw states of Figure 2, converging to the target concept, here SHL. Having discovered this array of modalities, a greater understanding of their interrelationships was required.

As mentioned many of the subtle differences between these states had no previously recognised basis in the perception of the experts. Many modes have very similar SHL, yet resolve to significantly different patterns in their subcomponents such as cluster 21 and 23. In order to address this further, attention was turned to the identification of an abstract layer between the target and base levels.

3. ABSTRACTION LAYER

Concerns arising from a review of the array base level clusters lead to various approaches with which to analyse their similarities or differences. For example the precision or measurement accuracy used for the original data in the unsupervised clustering process may lead to a situation of *over-training,* were the variance inherent in all of the SHL elements may unduly bias both the number and structure of modes in the mixture.

A useful feature of ACPro (also in SNOB [7]) is the ability to deem attributes of a cluster irrelevant. This is determined when an attribute's characteristic within a cluster does not differ significantly from the original population [8]. As each of the sub-elements here are considered significant for the daily SHL index, the upper bound for an element's accuracy is readily resolved when it is found to be missing (irrelevant) in a resultant mixture model.

Using a suitable metric the range of specific interrelationships was subsequently explored through link analysis. As each cluster of the model represents a mixture of 16 normal probability density functions (pdf) the Kullback-Leibler (KL) distance metric was determined between each pair. Ordinarily this is used to measure the distance between a statistical model and the true distribution. In order to accommodate all interstate comparisons, in this instance, only a 37×37 matrix (triangular, assuming symmetry between pairs) of KL distances is required from the base level 38 cluster model. These distances were examined again through a number of visual techniques, the most insightful of which was the

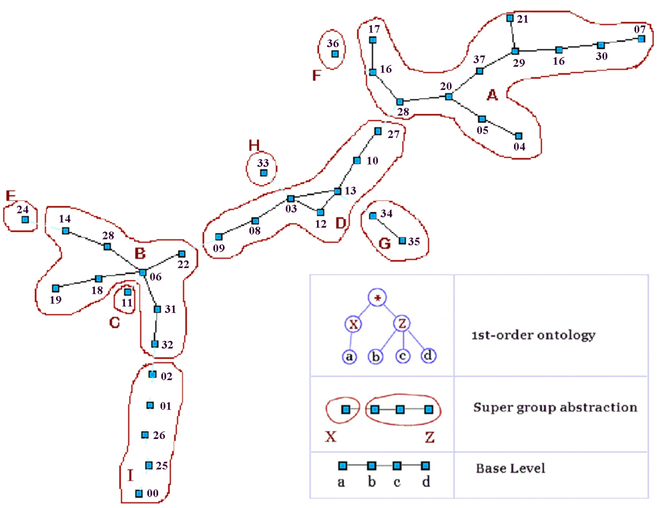

Figure. 4: Multidimensional scaling: KL-distances are mapped as cumulative link lengths in the graph between any pair of clusters; Super group abstractions are formed through removal of links exceeding a dissimilarity threshold.

Knowledge Network Organizing Tool (KNOT) [9] which is essentially a multi-dimensional scaling algorithm (MDS). In this specific case, the 38 dimensional model is viewed through a convenient projection as a two-dimensional graph, as illustrated in Figure 4.

The graphical representation in Figure 4 shows an ordered projection of the proximity (or similarity) between each pair of clusters or nodes of the graph. These are suitably arranged through an MDS projection such that non-adjacent nodes represent larger KL-distances, i.e. the cumulative link distance between any pair of nodes is proportional (qualitatively) to their KL-distance. As also illustrated in example of Figure 4, the four base level nodes, `a-d`, are arranged to reflect their cumulative dissimilarity distances such that: `a-b < a-c <a-d`.

Having determined the fully connected associations of the mixture model, a threshold prohibiting the visualisation of any link that exceeds it was progressively reduced until a number of segmentations occurred, as seen in Figure 4. This reveals a number of super groups, labelled A to I.

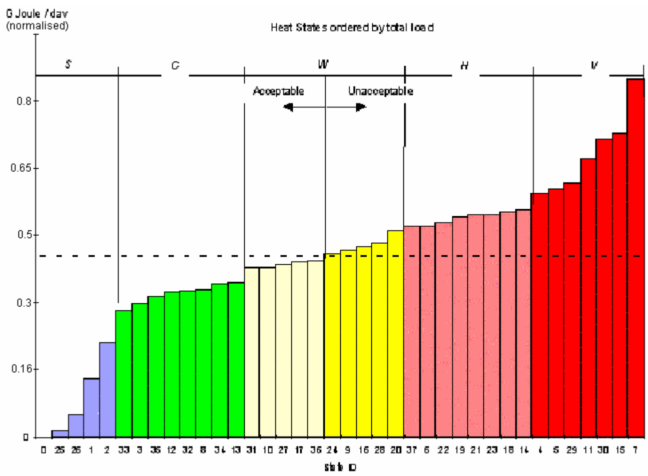

Figure. 5: Perato of base level clusters with domain expert classifications S, C, W, H, and V are "*Start-up, Cool, Warm, High and Very-high*" respectively.

As previously mentioned the motivation for developing an abstraction level was to find a better correspondence between the prior knowledge and understanding of the domain experts. A strategy used toward this is seen in Figure 5, which illustrates an expert's characterisation of the same set of base level modes ordered with increasing SHL. These classifications together with the derived SHL modes are compared and contrasted in Table 1.

TABLE 1: ABSTRACTION & BASE LEVEL ALIGNMENT

Base-level	Super-Group	Expert
4, 5, 7, 11, 15, 16, 17, 20, 21, 28, 29, 30, 37	A, C	Very High
6, 14, 18, 19, 22, 23, 31, 32	B	High
3, 8, 9, 10, 12, 13, 18, 27, 36	D, E, F	Warm
33, 34, 35	G, H	Cool
0, 1, 2, 25, 26	I	Shut Down Start

Importantly, the "High" to "Very High" heat states depicted in Figure 5, and Table 1, map to the classes A, B and C, whereas all of the "Cool" and a number of "Warm" states map to classes D, E, F, G and H, and all of the transition states map to class I.

This now provides a new basis for selecting a focused and more aligned set of targets representing SHL, through which supervised learning can be readily applied. These appear as rules or patterns, defined as significantly repetitive patterns over the elements and relationships of target data, the feed materials, operations, production and so on, that are predictive of each class of the SHL. Qualitatively, however, not all of the 9 groups found (i.e. A to I) were associated with unique SHL values, for example it was discovered that there were two distinct, but separate, classes (modes) for very high total heat loads. This presupposed that there were at least two different drivers, or at least combinations of drivers behind these. It was also conjectured that a change from one class to another might involve a critical transition, such that it acts as a trigger of heat load change in the new state, or subsequent sequence of states.

4. MODAL DRIVERS

Once the association alignment for this abstraction level was calibrated or accepted, a significant amount of other data detailing the daily variations across some 140 process attributes were combined with the abstraction level "class" data. Subsequent supervised learning was used to find patterns that promoted each Super group.

Each of nine super-groups (A to I) where further targeted to identify their associated driving influences. These were derived through a series of machine-learning trials where several decision trees and rule sets where derived. Generally, a number of iterations of each target where made to evaluate, in part the robustness of suggested drivers, and to test for any confounding effects in the data, where a proposed candidate influence was in fact a surrogate for another.

Using a supervised learning algorithm (C5, [10]) a set of some 18 rules were induced as likely predictors for each of the super-groups. These were found to correctly classify 80.5% of all SHL instances based on 10-way cross validation. The distribution of the number of rules per class is related to the complexity of each class. A number of example rules are detailed in Figure 6. Here, one can appreciate that although abstracted classes A and B both have elevated SHL levels, they appear to stem from diverse influences. For example, Rule 1 implicates various quantities and types of feed materials, whereas Rule 8 infers a specific charging or distribution pattern used to enter the feed materials (spatially) into the furnace. In contrast, Rule 9 predicting the desirable class D of SHL levels has a more complex combination of both feed materials and charging parameters.

The comparative performance of the predictive rules for all classes is detailed in Table 2. The number of predictive rules for each class is also included.

```
Rule 1: (cover 128)
        a45_FePcnt_Sntr <= 67
        a53_FePcnt_A_PELLETS > 13.2
        a77_Availability > 87.3
        -> class A  [0.992]

Rule 8: (cover 534)
        a15_CkChute_9 > 1
        a45_FePcnt_Sntr > 67
        a77_Availability > 74
        -> class B  [0.787]

Rule 9: (cover 59)
        a11_FeChute_3 <= 2
        a17_CkChute_7 <= 3.5
        a45_FePcnt_Sntr > 59.8
        a45_FePcnt_Sntr <= 67
        a48_FePcnt_SMALL_Sntr <= 3.5
        a54_FePcnt_B_PELLETS <= 10.3
        a77_Availability > 87.3
        -> class D  [0.885]
```

Figure. 6: Example predictive rules for abstract layer classes A, B and D. The cover parameter indicates the extent or generality of each rule, being the number of instances (days) predicted from the SHL data.

A further aim of this stage was to identify triggers, or changes of context mentioned above (feed, operation or production), that may drive the changes in the SHL trends. Initial work in this area suggested that such triggers might in fact be a preceding series of conditions. Due to the limited time available— this was not completed.

TABLE 2: COMPARATIVE MATRIX OF SHL PREDICTION RULES

Prediction of super group modes									True modes	no. rules
A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*		
755			132		3	9	4	1	A	7
2	**421**		32					7	B	1
1	45		3				1		C	1
129	37		**683**	1	1	14	2	27	D	5
27									E	1
1					**22**				F	1
3			3		5	**82**		3	G	2
2			8			1	**40**		H	1
4		**5**		1				**109**	I	5

Although the various heat states, and their aggregated classes or super groups, were originally derived from clustering the 16 sectional point loads, it is important to remember these ultimately infer various underlying process contexts. These would most likely arise through variations in all or any of the feed materials, burden distribution, or operational parameters. In as much as the data and time allowed, estimations of the most plausible drivers were made.

From the initial work, the region of class A appears to be strongly related to the amount of sinter and pellets used, as well as specific pellet types. Both class B and class A are predominantly associated with undesirable heat loads, whereas class D (and E-H) relate to more acceptable heat loading. Several misclassifications are seen as the off-diagonal elements are observed in Table 2. These arise from a degree of pruning in the supervised learning algorithm in order to cope with noise in the data and also obtain a compact, discernable rule set. Relaxing the amount of pruning will reduce the misclassifications, but also increase the number of rules—ongoing work is planned to optimise this.

Further, by classifying each daily instance of the SHL, an alternative perspective of the day-to-day dynamic behaviour was obtained. Here, for illustration only, strongly repetitive transitions, above a minimum threshold (6 daily transitions) are considered to obtain the dynamic ontology seen in Figure 7. Apart from the various established relationships between process modes (A to I), the historical pattern of such transitions in the operating campaign is also overlaid as varying arc widths. Here we see that there is a large number of transitions between the desired mode, D, and the high SHL mode of A. Interestingly, the desired mode of D also has the highest number of entry-exit paths (5) suggesting that there are more constraints that need to be met to ensure it is maintained—in other words, there are more ways to upset the balance and end up in an undesirable mode. Whereas once having entered an alternative (undesirable) mode, such as A, there are fewer exit conditions.

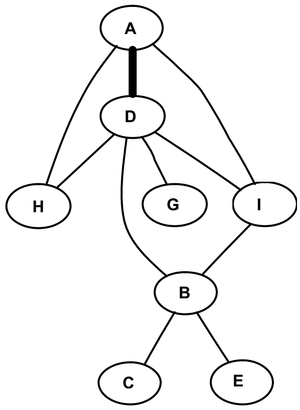

Figure. 7: Dynamic ontology of reactor process behaviours, where for each bidirectional link the width reflects the number of historical transitions.

CONCLUSIONS

Various machine learning procedures, and data mining techniques, were adopted to provide an objective means of identifying valid and useful structure or patterns in this large, complex, multi-factorial and highly coupled data set. The technique involves visualisation, transformation and multi-dimensional scaling of data, as well as decision tree analysis and rule set generation.

Through a number of data mining trials on the historical data supplied from blast furnace operations, a number of key variables together with several structural, or contextual relationships between them, were uncovered. These, in turn demonstrated a significantly high accuracy in predicting nine defined bands (or modes) of stave heat loads ranging from low to very high.

ACKNOWLEDGEMENT

The authors wish to thank BlueScope Steel for permission to publish this work. The countless discussions with Catherine Welsh and other members of the Ironmaking Technology and Development group are gratefully appreciated.

REFERENCES

[1] H. Suryanto and P. Compton, Discovery of Ontologies from Knowledge Bases.
[2] Chandrasekaran, J. Josephson, V. Benjamins, What Are Ontologies, and Why Do We Need Them?, IEEE Intelligent Systems, Jan-Feb. 1999.
[3] P. Cimiano, A. Hotho, S. Staab. Comparing conceptual, partitional and agglomerative clustering for learning taxonomies from text. In Proc of ECAI-2004, Valencia, August 2004.
[4] S. Sripada, E. Reiter, I. Davy, J. Hunter, J. Yu, Modelling the Task of Summarising Time Series Data Using KA Techniques. In: Macintosh, A., Moulton, M. and Preece, A. (ed) Proceeding of ES-2001, 2001 pp 183—196.
[5] D. Calvelo a, M.-C. Chambrin, D. Pomorski, Towards symbolization using data-driven extraction of local trends for ICU monitoring, Artificial Intelligence in Medicine 19 (2000) 203–223.
[6] C. Wallace, "Intrinsic Classification of Spatially Correlated Data", The Computer Journal, Vol. 41, No. 8, 1968.
[7] C. Wallace, D. Dowe, "Intrinsic classification by MML – the Snob program", Proc. 7th Australian Joint Conf. on Artificial Intelligence, World Scientific Publishing Co., Armidale, Australia,1994.
[8] J. Oliver, T. Roush, P. Gazis, W. Buntine, R. Baxter, S. Waterhouse, Analysing Rock Samples for the Mars Lander, In: Knowledge Discovery and Data Mining, pp 299-303, AAAI press, 1998.
[9] Interlink, Knowledge Network Organising Tool KNOT, http://www.interlinkinc.net/KNOT.html, 24-August, 2004.
[10] RULEQUEST RESEARCH (1997-2004), See5 and C5.0, http://www.rulequest.com/, 30-May 2004