FUTURE DIRECTIONS IN NETWORKED INSTRUMENTATION

J.C. Tapson

UNIVERSITY OF CAPE TOWN
Future Directions in Networked Instrumentation

Author: JC Tapson

Author’s Affiliation: Department of Electrical Engineering
University of Cape Town
UCT Private Bag
Rondebosch 7701

Introduction

Industrial automation systems generally consist of a compromise between theoretically desirable features and available hardware. A classic example of this is the use of Ethernet as a network technology; Ethernet does not provide real-time communication and the CDMA/CD protocol is not best suited to many control applications, but nonetheless it has become widely used in industrial applications because of its low cost, high bandwidth, and easy configuration.

In this paper, we will address the future of industrial automation as it will be shaped by current problems and emerging technologies. Current problems which are encountered in most applications include:

- The high cost of system installation, particularly "soft costs" such as communications and power infrastructure
- The configuration burden, particularly in SCADA systems
- The poor return in information from existing infrastructure

Future technologies which will shape the way we do automation include:

- Wireless self-configuring networks
- The web model
- Intelligent software

System Installation Costs

The installation of an automation system currently includes costs for infrastructure that may well exceed the cost of the instrumentation and control elements. All communication and power signals are carried by cable which has to be laid, trunked, conduit, trenched, scaffolded, and eventually connected. In some environments, particularly in hazardous areas, the cost of carrying a single communication cable a distance of 50m can exceed R 100 000. In First World countries, there are added complications introduced by labour unionisation (a big problem in the USA) and occupational safety concerns (a big problem in the UK).

The Configuration Burden

One of the consequences of versatility is the requirement for configuration. We will never get away from this entirely, because at the highest level a system specification is a configuration description, but as present there is a great deal of industrial
hardware that is not sufficiently self-configuring. The configuration level requires highly skilled human intervention and is also an area in which errors can be introduced to the system.

Return on Infrastructure

In theory, a typical SCADA system should be able to provide a vast amount of non-core information to the users. In practice, most SCADA systems seem to barely manage to perform their core function, and the SCADA support teams seem to spend their whole effort in maintaining this fragile grip on the process. There is an enormous gap between the potential of a SCADA system and the delivery experienced by the operators.

Wireless Self-Configuring Networks

The Bluetooth radio system is the first of many such systems that we will see that offer self-configuration. This is a quality that we can use extensively in the control environment to build networks that are non-cabled and to some degree self-connecting.

The Web Model

The World Wide Web has given us a model whereby information is formatted in a universally acceptable fashion, and is delivered by means of a client-server model. While this horrifies many control engineers with concerns about security and reliability, the model offers a great deal of hope for solution of the above problems.

Intelligent Software

There is a vast body of academic knowledge on so-called soft computing, such as fuzzy logic, neural networks, genetic algorithms, and stochastic optimisation techniques. The penetration of this into industry has been negligible. Partly because the strength of these techniques cannot be brought out except with the assistance of the above enabling technologies, we may see some change to this in the future.

A Vision of the Future

Our control engineer moves round the plant, placing sensors on various pieces of the process machinery. As they are powered up, they connect wireless with the SCADA system and announce their configuration specification. The SCADA system is expecting them; it begins a seamless integration of the new data. The general manager notices his process output improve by 1% - he wonders why...
This paper describes some of the experiences within De Beers in the use of data mining and soft sensors.

Keywords: Expert systems, Data Mining, Optimization, Soft Sensors

1. INTRODUCTION

De Beers has been actively pursuing the goal of increased automation, and process optimisation for some years now. Early work was based on the automation of diamond sorting machines, where the need for hands-off processing of concentrate and high security resulted in a high level of mechanisation and automation. As a result, the level of instrumentation and control on many of the modern diamond processing plants has resulted in an infrastructure that lends itself to the further automatic control and optimisation of the process. Initial work was based on a rule-based (expert system) approach, which resulted in effective plant balancing across the entire process, and resulted in significant improvements in throughput, plant utilisation and process efficiency.

The rule-based approach was deemed to be appropriate due to the difficulties in modelling either the variability of the feedstock or the process units to a degree that could be used with confidence in a conventional control application. This was further complicated by the fact that the control problem of interest was for the optimisation of the entire process, and not just selected unit processes.

Nonetheless, the rule-based approach is limited to the extent of the knowledge that is available on the process at hand. The process knowledge in turn is determined by the experience of people available (typically on site), and the amount of data (sensors) that is available to reason with. It is thus obvious that the next step towards further improvement of the knowledge base is to

a) measure further key performance indicators and/or improve the accuracy and reliability of current sensors, and

b) provide the tools to turn this data into useful information, based on which the relevant operation knowledge and experience can be built.

Figure 1 illustrates the problem diagrammatically. The frustrating aspect of the situation is that the tools for the capture and archiving of data from plant is readily commercially available and installed on most modern plants. Yet, we are generally not deriving the benefit from the availability of this data store. This is because the relevant tools are not available to enable the responsible people to extract, visualise and analyse the data in such a way that it becomes useful information. Many
may claim that the tools are available, but we maintain that they are not sufficiently integrated into the process engineer’s toolkit that he can perform his required task effectively.

It is for this reason that we have been tracking the field of data mining and visualisation techniques for several years. Of late, we have identified three potentials for some of the techniques:

1) for the data mining and “modelling” of process to increase operational knowledge base, so that the process can be further optimised using a rule- (or model-) based approach. (the cloud in figure 1).

2) For the generation of soft sensors which can be used to either continuously estimate key performance indicators which are generally not available continuously.

3) To compare the softsensor with actual measurement, to detect deviations and to alarm for maintenance prior to critical failure. Alternatively to replace critical measurements in the event of severe sensor degradation or failure, until maintenance is convenient.

The following paper will describe some of the experiences in this journey.

2. DATA MINING

Data mining is defined as “the process of discovering meaningful new correlation, patterns and trends between variables of data sets by sifting through large amounts of historical data stored in repositories, using pattern recognition technologies as well as statistical and mathematical techniques”\(^1\). It is, therefore, a “knowledge discovery process of extracting previously unknown, actionable information”\(^2\).

There is a plethora of Data Mining packages available commercially. Many of these have their roots in providers of statistical analysis packages, while many of the newer ones are based on neural network techniques.

Either way, the three requirements of a data mining package are:

1) Data handling: this consists of import and export from various data formats, as well as connectivity to data sources such as data bases, spreadsheets and the like.

2) Data pre-processing: this consists of utilities for the filtering, validation and editing of data as well as the creation of new (calculated) variables to calculate more useful data e.g. calculate kWh/t from kW and t/h

3) Data visualisation: this is linked very closely to pre-processing where visualisation techniques are required to see if data is “reasonable”.

4) Modelling: this is the critical step, in which the relationship between key variables is established and quantified. Data visualisation techniques to visualise results of the modelling step are typically also required.

5) Deployment – the ability to take the output of the data mining exercise and to deploy or automate it.

Some of the packages that we evaluated as part of our early technology scan were:

- Integrated Solutions Limited’s Clementine 5.0
- Pavilion Suite of Data mining and real-time OLAP software for the process industries;
- SPSS data-mining product suite family;
- Silicon Graphics’ MineSet™ 2.5
- Pilot Software’s Dicovery Server Suite
- Crusader System’s ModelGen
- IBM’s Intelligent Miner for data;
- WizSoft’s WizWhy and WizRule
- SDI’s e – The evolutionary algorithm

One of the major requirements of our process was that the bulk of our data has a major time-based component. That is, the data mining systems must be able to handle both time series and very large data sets, rather than just transactional data. This requirement eliminated a large number of potential solutions, since the bulk of data mining packages are designed for transactional databases such as would be used in the banking or retail industries.

The initial survey indicated that there are a multitude of potential solutions (and eager vendors), ranging from $100s to $10 000s, so it is important that the correct solution is found for the problem at hand.

The next step was thus to define a problem on which some of these methods could be tested.

3. CRUSHER MODELLING

The first problem which we attempted was to be able to infer ore type and/or material properties from crusher performance data on-line. Intuitively it was felt that this should be a tractable problem, since the primary crushing product is scrubbed and screened...
into 4 size fractions (oversize, coarse, fine and slimes/grits). The distribution of material can change markedly with ore-type, so it was felt that by monitoring crusher parameters such as throughput, power, crusher gaps and size distributions a good model for ore-type should be achievable.

Initial attempts at dealing with the data were hampered by the distributed nature of the data. Crusher data was readily available from SCADA but data on ore type or source of ore being treated by the plant was less readily available and had to be retrieved from separate databases.

Some of the following problems had to be dealt with:

- Alignment of data from different data sources and different sampling rates.
- Elimination of data where faulty sensors were evident.
- Checking for saturation of measurements.
- Calculating inferred variables.
- Eliminating outliers.
- Filtering of data to reduce measurement noise.

The steps above are very reliant on knowledge of the process as well as good common sense, and are virtually impossible to automate. We found Insights from Pavilion particularly good at doing the required pre-processing and visualisation on very large time series datasets. It also has excellent facilities for merging datasets with different time bases.

The next step was to use Principal Component Analysis to determine the most relevant variables and to eliminate redundant variables. The primary Principal Components were then plotted against one another to determine the existence of clusters of data (see Figure 2).

Although the clusters illustrated were relatively clear, subsequent classification on a verification data set did not consistently identify the correct ore types and sources. In fact, the trend observed is that in order to ensure that crushers can deal effectively with the variation in feedstock, they tend to be designed to be relatively insensitive to feed type. Hence the changes in ore type do not result in significant changes in crusher operating parameters.

4. SOFT SENSOR

After having tried the “traditional approach”, it was decided to try the heuristic and neural network based methodology, and also to tackle a less ambitious problem.

The CSENSE toolkit from Crusader Systems was used to create a softsensor model of weightometers on one of the De Beers mines. Using OPC, the system allows easy access to real-time data from the mine SCADA, and the CSENSE toolkit allows the development of on-line models with relative ease. Variables used to infer ore feedrate were

- power draw of processing equipment such as scrubbers, sizing and dewatering screens and pumps,
- ore particle size from the online particle size analyser, and
- other weightometers.

The objective was to

1) create a soft sensor for each weightometer that can validate the on-line measurement or alarm a deviation. In the case of a weightometer failure, the soft sensor can then be used to continue production until the next convenient maintenance opportunity.

2) Create a backbone for more reliable on-line mass balance in order to improve metallurgical accounting.

A total of 31 weightometers were modelled. Most of these have correlation with the actual measurement of $R^2 > 0.90$, resulting in the calculation of the throughput in tons/h within a daily error of < 4%, and with confidence of > 90% on a moving real-time trend sampled at 2 second intervals.

Figure 3 illustrates some of the results achieved.
5. CONCLUSION

Although the use of data mining, including both traditional statistical as well as neural network (black-box) methods is inherently appealing and has great potential, the data mining practitioner is faced with a multitude of tools, and requires a thorough understanding of the nature of the data, the result to be achieved, and the pros and cons of the various methods that could be applied, in order to effectively "mine" this data.

In our experience, in excess of 85% of effort is spent in the process of acquiring, cleaning and pre-processing the data, and only the remainder is spent on modelling and analysis of results. This is not an area for dabbler's or the faint-hearted!

6. BIBLIOGRAPHY

12. Ginsberg DW, Whiten WJ, “Cluster Analysis for Mineral Processing Applications”, Transactions...
of the Institute of Mining and Metallurgy, 100, 1991


27. Medich L, Ball Mill Circuit Model, internal publication, Pavilion Technology, June, 1997

28. META Group Application Development Strategies: “Data Mining for Data Warehouses: Uncovering Hidden Patterns”, 7/13/95


ACKNOWLEDGEMENTS
This paper is published by permission of the General Manager, Technical Support Services - Technology, De Beers Consolidated Mines Ltd.