

Practical Applications of Data Mining in Plant Monitoring and Diagnostics

Scott M. Strachan, Stephen D.J. McArthur, *Member, IEEE*, Bruce Stephen

Abstract—Using available expert knowledge in conjunction with a structured process of data mining, characteristics observed in captured condition monitoring data, representing characteristics of plant operation may be understood, explained and quantified. Knowledge and understanding of satisfactory and unsatisfactory plant condition can be gained and made explicit from the analysis of data observations and subsequently used to form the basis of condition assessment and diagnostic rules/models implemented in decision support systems supporting plant maintenance. This paper proposes a data mining method for the analysis of condition monitoring data, and demonstrates this method in its discovery of useful knowledge from trip coil data captured from a population of in-service distribution circuit breakers and empirical UHF data captured from laboratory experiments simulating partial discharge defects typically found in HV transformers. This discovered knowledge then forms the basis of two separate decision support systems for the condition assessment/defect classification of these respective plant items.

Index Terms—Circuit breakers, transformers, monitoring, trip coil current signature, UHF, partial discharge, data mining, condition-based maintenance, decision support system

I. INTRODUCTION

Condition monitoring technology enables the periodic or continuous acquisition and measurement of data characterizing the ‘vital signs’ of operational plant. Typically, domain experts are then required to interpret this data to assess plant health and/or predict incipient failures, and where possible diagnose the root cause of any plant failure.

Novel condition monitoring sensor, data acquisition and communications technologies, designed to detect and capture key plant parameters in larger volumes, at a faster rate, using less bandwidth [1] presents plant experts with a rich vein of data. Experts have the onerous task of effectively managing and interpreting large volumes of data for an indication of the plant health. Overwhelming plant experts with large volumes of unfamiliar plant data for the extraction of meaningful diagnostic information is clearly something to be avoided.

In addition, as utilities continue to downsize, outsource, lose and retire staff, they run the risk of losing the specialist knowledge and experience required for plant condition assessment and defect diagnosis. All this precipitates the need for new methods of analyzing and interpreting condition monitoring data to alleviate the burden placed on what is widely regarded as a diminishing population of plant experts. In particular, where new plant parameters are made available

from novel condition monitoring technologies, a means of actively driving diagnostic knowledge out of this data is required.

In conjunction with plant experts, the causal relationships between data observations and known plant behavior may be determined, modeled or encoded in rules, enabling meaningful diagnostic information to be interpreted directly from condition monitoring data captured in future. Knowledge implicit in condition monitoring data may be extracted and made explicit using a structured method of data mining tailored to meet the specific goal of classification common to plant condition monitoring. This ‘discovered knowledge’ may then be embedded within a decision support system assisting plant experts (and non-experts) in the future condition assessment and defect diagnosis of plant items.

The level of expert interaction in the data mining process and the choice of techniques used will depend on the level of domain knowledge available, the level of data interpretation knowledge available and the a priori knowledge of the data classes used in classifier training.

This paper proposes a structured method of analyzing plant data for the development of suitable plant condition classifier models implemented in practical condition monitoring decision support systems.

II. BUILDING PLANT CONDITION CLASSIFIERS USING A STRUCTURED METHOD OF DATA MINING AND KNOWLEDGE ELICITATION

Data mining is often cited as “*the non-trivial extraction of implicit, previously unknown, and potentially useful information from data.*” [2]. Data mining may therefore be considered a knowledge extraction process aimed at “making sense” of large data sets through the discovery of hidden and often unsuspected facts, patterns and correlations implicit in the data. This is achieved by analyzing large volumes of data through a combination of advanced processing techniques, including machine learning, statistical methods and visualization tools to make observations on the data. Using these techniques, the observations gleaned from the data provide an insight into the behavior of the system/domain responsible for bearing the data. These observations can be unified with knowledge and understanding of the system/domain elicited from a domain expert (Figure 1).

Domain knowledge can be used to drive the data mining process (e.g. extraction of salient features from data).

Similarly, the data mining observations may be used to drive the knowledge elicitation involving the domain expert (e.g. to classify data clusters representing undefined faults).

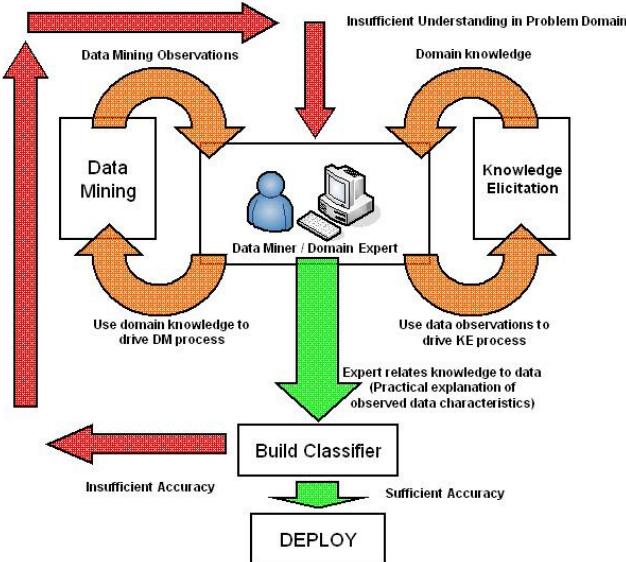


Figure 1 Unification of data mining observations and expert domain knowledge

Domain experts can then explain characteristics of the system behavior represented in these data observations and suitable rules or models may be created for the classification of future system behavior and/or condition. The discovered knowledge relating to plant condition assessment and defect diagnosis can then be used to develop a decision support tool offering maintenance staff advice in the condition assessment of distribution circuit breakers following routine trip testing.

In the condition monitoring domain *a priori* knowledge of the plant condition associated with acquired condition monitoring data may or may not be readily available (e.g. data representing healthy/unhealthy conditions may be labeled or unlabelled). When data-class associations are known, classifier models may be built using supervised learning techniques and incorporated directly into a decision support system (e.g. using labeled data to train an artificial neural network to identify known partial discharge defect sources in HV transformers). Where data-class associations are unknown, a more complex data mining process is required to extract hidden associations between the data characteristics and those of the plant behavior. Figure 2 illustrates the data mining objectives (defined by CRISP-DM) [3] involved in developing a classifier model where the data-class associations are unknown prior to developing a classifier model.

The objectives at each stage of this data mining process are described as follows:

- Segmentation – separation of data into interesting subgroups representing concepts (e.g. clusters representing varying plant condition/behavior).
- Concept Description – describing the thresholds defining different plant condition and behavior (i.e. quantitative description of the identified clusters).

- Expert interpretation – using domain knowledge, where it exists, to verify and explain the observations made and concepts described from the segmented data.
- Classification – building classifier models or deriving rules to correctly assign conditions to previously unseen plant data.

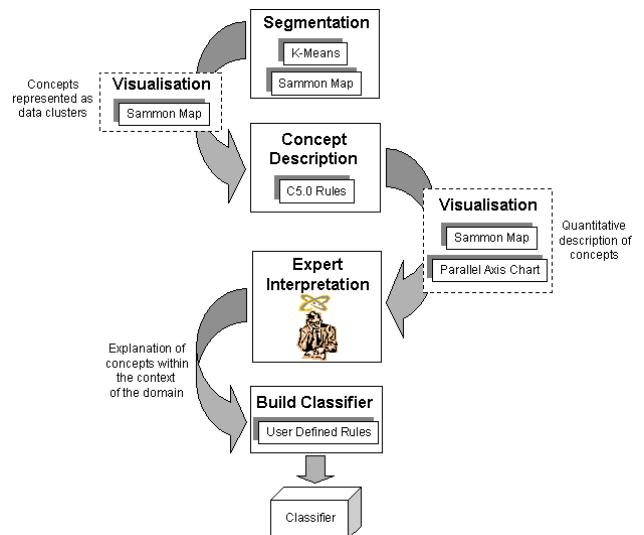


Figure 2 Data mining process for trip signature data analysis and classifier development

The visualization stages of the process outlined in Figure 2 are concerned with improving the '*understandability*' of patterns observed in the data for the consumption of data analysts and experts [2]. Visualization of the outcomes from each stage of the data mining process, using suitable techniques, make it easier to describe and convey the concepts and observations identified in the data, and so facilitates the discovery of meaningful diagnostic knowledge in conjunction with experienced domain experts.

Figure 2 also illustrates the techniques applied at each stage of the data mining process applied to the trip signature data.

III. BUILDING A DSS FOR CIRCUIT BREAKER CONDITION ASSESSMENT USING A STRUCTURED METHOD OF DATA MINING

Measurement of a breaker's trip signature provides a useful insight into the operation of the circuit breaker and its condition. In addition to its primary function of controlling the breaker trip operation, the trip coil performs a secondary role as a non-invasive condition monitoring sensor, capturing the trip coil current (signature) from the 'all-important' first trip for subsequent breaker condition assessment [4]. Previously, the decision determining whether a breaker should undergo maintenance or not was based on a simple heuristic – "IF trip time > 100ms THEN maintain breaker", irrespective of model type and manufacturer. This approach ignored the nuances associated with different circuit breaker designs and the condition defined by the timings of the various stages of breaker operation.

Domain knowledge relating to the circuit breaker operation

and the idiosyncrasies associated with various manufacturers and models is available from utility switchgear experts. The data mining process seeks to bridge the gap between the experts knowledge of the application domain (or plant behavior) and the understanding and interpretation of the data captured using novel circuit breaker condition monitoring technology.

Through a combination of knowledge elicitation and data mining activity, expert understanding of the breaker operation and behavior could be correlated with observations made from the outcomes of the data analysis conducted.

Using a structured method of data mining, distinct clusters of trip signature shapes could be identified and quantified for various models and manufacturers encompassing the entire asset base of distribution circuit breakers. These clusters notionally represent different circuit breaker conditions. Using the practical knowledge of experienced switchgear experts to explain the data observations derived from the data mining process, condition assessment and ultimately diagnostic rules could be derived and implemented in a rule-based system.

A prototype was developed which presents, in a visually appreciable format to the data analyst and maintenance expert, the outcomes of the data analysis conducted on a population of captured trip signatures. This allows the expert to define the operational thresholds characterizing the varying degrees of breaker condition observed in the data. This has also led to tentative steps being taken in the elicitation of basic diagnostic knowledge.

A.