Operational Event Detection: A supervised Machine learning method to detect events from multivariate time-series data

Author: Prasanth Gopalakrishna

Abstract: Extracting insights from high density data in real-time from Oil & Gas drilling rigs and turning those insights into decisions that can be executed in a closed-loop system is an exciting proposition. Towards this process of automating drilling, 'Operational Events' provide context for the interpretation of streaming IoT data from the rig. Accurate detection, in real-time of Operational Events becomes a crucial first step.

Discussed below are some approaches utilizing Artificial Intelligence techniques to detect Operational Events against high velocity, high density, streaming data. Random Forest based multivariate time-series classification is one such approach. Supervised learning using this technique has yielded an extremely accurate model for use with real-time streaming data.

Keywords: Random Forest, Multivariate Time Series classification, Supervised learning, Artificial Intelligence.

1. Introduction

Multivariate time series data are ubiquitous and broadly available in many fields including finance, medicine, Oil & Gas industry and other business domains. A time series T is a series of ordered observations made sequentially through time. We denote the observations by:

$$x_i(t);[i=1,\dots,n;\ t=1,\dots m] \text{ where:}$$

- i is the index of the different measurements made at each time point t,
- n is the number of variables being observed, and
- m is the number of observations made

If the time series has only one variable (n = 1) then this time series is referred to as univariate, if it has two variables or more (n > 1) then it is referred to as multivariate. Data from a drilling rig is an example of multivariate time-series data; where many mechanical parameters such as torque, hook load and block position, are continuously measured by surface sensors and stored in real-time in databases.

Fig. 1 shows drilling multivariate time series consisting of eight channels or variables.

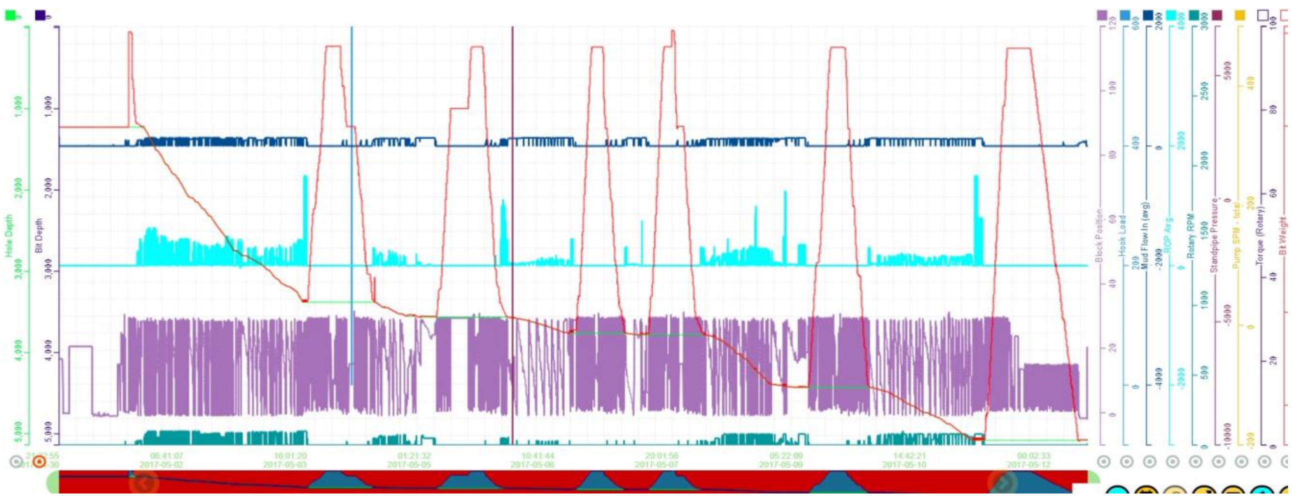

Fig 1. A multivariate time series of drilling data. This time series consists of eight variables representing eight mechanical parameters measured at the surface

Multivariate time series classification is a supervised learning problem aimed at labelling multivariate series of variable length.

2. **Approaches**

A number of approaches including but not limited to MCPD, FCN and Decision-tree were attempted, and the feature engineering designs are beyond the scope of this document.

Conclusion and Future Work

The following conclusion can be drawn from the final approach discussed in this paper

- Reducing dimensionality of the time-series data is in and of itself an objective
- The reduced representation can be used as alternative to the time series without losing any important characteristics or patterns existing in the original time series data.
- Having more intuitive domain inspired features help aid the model improve its predictive power,

As Future scope, to further improve the performance of the model to satisfactory level, esp in a few classes where predictive power is low, below methods can be adopted,

- Visually analyse plots to design features for resolving confusion between classes.
- For classes with low support(instances), devise a mechanism to assign class weight based on Bayesian logic.

References:

1. Zahra Ebrahimzadeh, Samantha Kleinberg.: Multi-Scale Change Point Detection in Multivariate Time Series
2. Fazle Karim, Somshubra Majumdar, Houshang Darabi, Samuel Harford.: Multivariate LSTM-FCNs for Time Series Classification.
3. Ricardo Chavarriaga, Hesam Sagha, Alberto Calatroni, et al. The Opportunity Challenge: A Benchmark Database for On-body Sensor-based Activity Recognition. Pattern Recognition Letters, 34(15):2033–2042, 2013.
4. M. Lichman, UCI machine learning repository (2013). URL http://archive.ics.uci.edu/ml
5. Bilal Esmael, Arghad Arnaout, Rudolf K. Fruhwirth and Gerhard Thonhauser.:Multivariate Time Series Classification by Combining Trend-Based and Value-Based Approximations
6. Lin, J., Keogh, E., Lonardi, S. & Chiu, B.: A Symbolic Representation of Time Series, with Implications for Streaming Algorithms. In proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery. San Diego, CA. June 13 (2003).
7. [Cleveland et al. 1990] Cleveland, R. B.; Cleveland, W. S.; McRae, J. E.; and Terpenning, I. 1990. STL: A seasonal trend decomposition procedure based on loess. Journal of Official Statistics 6(1):3–73