STATISTICAL METHODS FOR PREDICTIVE MODELING USING SENSOR DATA

ORGANIZER: BIANCA ZADROZNY, IBM RESEARCH, BRAZIL

TALK 1: BUS TRAVEL TIME PREDICTIONS USING ADDITIVE MODELS

SPEAKER: MATTHIAS KORMAKSSON, IBM RESEARCH, BRAZIL

Many factors can affect the predictability of public bus services such as traffic, weather and local events. Other aspects, such as day of week or hour of day, may influence bus travel times as well, either directly or in conjunction with other variables. However, the exact nature of such relationships between travel times and predictor variables is, in most situations, not known. In this paper we develop a framework that allows for flexible modeling of bus travel times through the use of Additive Models. In particular, we model travel times as a sum of linear as well as nonlinear terms that are modeled as smooth functions of predictor variables. The proposed class of models provides a principled statistical framework that is highly flexible in terms of model building. The experimental results demonstrate uniformly superior performance of our best model as compared to previous prediction methods when applied to a very large GPS data set obtained from buses operating in the city of Rio de Janeiro.

TALK 2: A DATA DRIVEN METHOD FOR SWEET SPOT IDENTIFICATION IN SHALE PLAYS USING WELL LOG DATA

SPEAKER: BIANCA ZADROZNY, IBM RESEARCH, BRAZIL

In recent years, interest in shale plays has grown substantially due to horizontal drilling and hydraulic fracturing techniques. Special interest is in shale plays previously exhausted with vertical wells that are believed to still have high potential if developed with horizontal wells. However, with drilling costs at an all time high, choosing the right locations for new wells is a crucial issue. Therefore, identifying so called “sweet spots” with high potential for oil and gas is of great importance for oil companies worldwide. Well log data from millions of wells drilled using conventional techniques since the industry’s inception is available and generally not used. We propose a data analytical solution that 1) automatically extracts simple features from complex and high-dimensional well log curves arising from vertical wells using functional Principal Component Analysis (fPCA), and 2) builds models that predict sweet spots in shale plays by correlating extracted features with production data from horizontal wells. Our solution builds predictive models for production directly using previous production data and petrophysical well logs alone, thus circumventing time consuming and expensive geological analysis. To the best of our knowledge this is the first time fPCA is applied to well log curves in the context of oil and gas exploration.