Smart Grid Data Analytics for Decision Support

Prakash Ranganathan, Department of Electrical Engineering, University of North Dakota, Grand Forks, ND, USA
Prakash.Ranganathan@engr.und.edu, 701-777-4431
Kendall Nygard, North Dakota State University, Fargo, ND, USA
Department of Computer Science

October 5 2011, EPEC 2011, Winnipeg, Canada
Blackouts and ‘human error’

On Sep 9 2011, San Diego and south Orange counties experienced the most widespread power outage in their history. A short-circuit at a substation in North Gila, Arizona, set off a series of failures that led to the massive power outage that left millions of people in California, Arizona and Mexico in the dark. The Utilities took 12 hours to restore the power to nearly everyone during this unprecedented event.

CAUSE: The power outage began Thursday afternoon. An Arizona Public Service Co. worker was handling a capacitor at a substation in North Gila when the short-circuit occurred. That knocked out a major transmission line that carries power to California. Low-voltage readings sent the nuclear plant into shutdown, and the blackout was on.
Objective: Enhance Situational Awareness (SA) – System Operators

Despite upgrades and advances in technology, power system operators must assimilate overwhelming amounts of data to keep the electric utility grid operating. Studies of recent blackouts have demonstrated the need to enhance the operator’s ability to understand the state of the system and anticipate possible problems.

PSO’s can better understand Situation Awareness (SA) through mining these sensor data and arrive at a good estimate and control mechanisms to undertake necessary switching actions.

This paper focuses on identifying the variables of interest that are important in the electric grid embedded in distributed real time data engines which will help decision support process for system operators. So, here we employ Decision Trees to classify grid data and provide PSO better global vision of the data they manage.
Multi-agent oriented approach (work in progress)

Figure 1a. Communication Flows and Reporting Responsibilities in the Simulation

The right sided Figure 1b. illustrates the multiplicity of the relationships among physical system devices with intelligent computational agents.
Decision framework for PSO

Data Sources
1, 2, ..., n
(Sensors, Relays, PMUs etc.,)

Data preprocessing, Extraction and Transformation

(Applicability of J48/MSP Decision Tree models to classify SG data)

System Operators interpret data in meaningful way and/or activate the processes in the Smart Grid.
(All levels: Generation, Transmission, Distribution)
Data Source

- NYISO
- Unit commitment

Table 1: NYISO Data Set from 28 April to 3 May 2011 for NYC

Method: M5 Model – [Quinlan’93]

- Numerical prediction (regression) methods, that we found to be practically unknown to practitioners is so-called M5 model tree of Quinlan. It is based on ideas of a popular classification method, a decision tree that follows the principle of recursive partitioning of input space using entropy-based measures, and finally assigning class labels to resulting subsets.

- The advantages of M5 model trees (Solomatine & Dulal ; Solomatine ) are that they are more accurate than regression trees, more understandable than, for example, ANNs, easy to use and to train, robust when dealing with missing data, can handle large number of attributes and high dimensions. Decision trees are used for predicting or explaining outputs from observations. In such a tree, each node is a leaf indicating a class or an internal decision node that specifies some test to be carried out. If the output values conform to intervals, then the decision trees are called regression trees, whereas if they do correspond to a nominal or ordinal scale they are called classification trees. There are many tree-building algorithms such as C4.5 (Quinlan, 1993) which determine which attributes best classifies the remaining data, and then the tree is constructed iteratively. The main advantage of decision trees is their immediate conversion to rules that can be easily interpreted by decision-makers.

- For numeric prediction in data mining, it is common to use regression trees or model trees. The M5 algorithm builds trees whose leaves are associated to multivariate linear models and the nodes of the tree are chosen over the attribute that maximizes the expected error reduction as a function of the standard deviation of output parameter.
M5 (Contd.,)

- M5 combines a conventional decision tree with the possibility of linear regression functions at the nodes. First, a decision-tree induction algorithm is used to build a tree, but instead of maximizing the information gain at each inner node, a splitting criterion is used that minimizes the intra-subset variation in the class values down each branch. The splitting procedure in M5 stops if the class values of all instances that reach a node vary very slightly, or only a few instances remain.

- Second, the tree is pruned back from each leaf. When pruning an inner node is turned into a leaf with a regression plane.

- Third, to avoid sharp discontinuities between the subtrees a smoothing procedure is applied that combines the leaf model prediction with each node along the path back to the root, smoothing it at each of these nodes by combining it with the value predicted by the linear model for that node.

Techniques devised by Breiman et al. [BRE84] for their CART system are adapted in order to deal with enumerated attributes and missing values. All enumerated attributes are turned into binary variables so that all splits in M5 are binary. As to missing values, M5 uses a technique called “surrogate splitting” that finds another attribute to split on in place of the original one and uses it instead. During training, M5 uses as surrogate attribute the class value in the belief that this is the attribute most likely to be correlated with the one used for splitting. When the splitting procedure ends all missing values are replaced by the average values of the corresponding attributes of the training examples reaching the leaves. During testing an unknown attribute value is replaced by the average value of that attribute for all training instances that reach the node, with the effect of choosing always the most populous subnode.
M5 Model

- Input space $X_1 \times X_2$ is split into regions; separate regression models can be built for each of the regions.
WEKA
(Data mining tool)
Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes.

Link: http://www.cs.waikato.ac.nz/ml/weka/
Weekly Power Demand for NYC

![Weekly Power demand for NYC on 24 hour period](image)
A Sample NYISO data run on WEKA

\[ \text{demand} \leq 4568 \]
\[ | \text{demand} \leq 4256: \text{t2} (4.0) \]
\[ | \text{demand} > 4256: \text{t1} (9.0/3.0) \]
\[ \text{demand} > 4568: \text{to} (8.0/1.0) \]

Number of Leaves : 3

Instances: 21
Attributes: 3
  time
demand
date

Table:

<table>
<thead>
<tr>
<th>Time (in seconds)</th>
<th>Demand (in Mega Watts of power)</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>to</td>
<td>4808</td>
<td>ap 27</td>
</tr>
<tr>
<td>to</td>
<td>4789</td>
<td>ap 28</td>
</tr>
<tr>
<td>to</td>
<td>4772</td>
<td>ap 29</td>
</tr>
<tr>
<td>to</td>
<td>4620</td>
<td>ap 30</td>
</tr>
<tr>
<td>to</td>
<td>4572</td>
<td>1-May</td>
</tr>
<tr>
<td>to</td>
<td>4592</td>
<td>2-May</td>
</tr>
<tr>
<td>to</td>
<td>4714</td>
<td>average</td>
</tr>
<tr>
<td>t1</td>
<td>4568 x</td>
<td>ap 27</td>
</tr>
<tr>
<td>t1</td>
<td>4585</td>
<td>ap 28</td>
</tr>
<tr>
<td>t1</td>
<td>4525 x</td>
<td>ap 29</td>
</tr>
<tr>
<td>t1</td>
<td>4408 x</td>
<td>ap 30</td>
</tr>
<tr>
<td>t1</td>
<td>4296 x</td>
<td>1-May</td>
</tr>
<tr>
<td>t1</td>
<td>4329 x</td>
<td>2-May</td>
</tr>
<tr>
<td>t1</td>
<td>4469 x</td>
<td>average</td>
</tr>
<tr>
<td>t2</td>
<td>4409 x</td>
<td>ap 27</td>
</tr>
<tr>
<td>t2</td>
<td>4402 x</td>
<td>ap 28</td>
</tr>
<tr>
<td>t2</td>
<td>4323 x</td>
<td>ap 29</td>
</tr>
<tr>
<td>t2</td>
<td>4256 $</td>
<td>ap 30</td>
</tr>
<tr>
<td>t2</td>
<td>4123 $</td>
<td>1-May</td>
</tr>
<tr>
<td>t2</td>
<td>4145 $</td>
<td>2-May</td>
</tr>
<tr>
<td>t2</td>
<td>4247 $</td>
<td>average</td>
</tr>
</tbody>
</table>

The first number is the total number of instances (weight of instances) reaching the leaf. The second number is the number (weight) of those instances that are misclassified. If your data has missing attribute values then you will end up with fractional instances at the leaves. When splitting on an attribute where some of the training instances have missing values, M5/J48 will divide a training instance with a missing value for the split attribute up into fractional parts proportional to the frequencies of the observed non-missing values. This is discussed in the Witten & Frank Data Mining book as well as Ross Quinlan's original publications on C4.5.
A LM decision tree with Reserve, Time, Wind data
RESULTS

In statistics as well as in data mining, with linear regression models used in this work, the goodness of fit of a model is usually measured by the correlation and by the root mean squared error.

<table>
<thead>
<tr>
<th>Method</th>
<th>RMSE</th>
<th>Correlation coefficient</th>
<th>Relative Absolute error</th>
<th>Data set</th>
</tr>
</thead>
<tbody>
<tr>
<td>J48</td>
<td>.3037</td>
<td>-</td>
<td>51.3%</td>
<td>NYISO</td>
</tr>
<tr>
<td>M5</td>
<td>54.4</td>
<td>.968</td>
<td>14.5%</td>
<td>Unit commitment</td>
</tr>
<tr>
<td>Decision Stump</td>
<td>125.19</td>
<td>.816</td>
<td>62.79%</td>
<td>Unit commitment</td>
</tr>
<tr>
<td>M5</td>
<td>17.6</td>
<td>.99</td>
<td>8.68</td>
<td>NYISO</td>
</tr>
<tr>
<td>Decision Stump</td>
<td>.3407</td>
<td>-</td>
<td>64%</td>
<td>NYISO</td>
</tr>
</tbody>
</table>
Work in Progress

- Real-time decision tree need to be implemented and represented through GUI
- Attention needed in decision making mechanisms using grid data where system variables of the grid is visible to system operators to take remedial actions.
- The number of attributes is reduced mainly using expert knowledge although the data mining algorithms can help us to identify the most relevant attributes in relation to the output parameter, that is, the attribute that wants to be estimated.
- The number of instances or samples in the dataset is reduced by selecting those that contribute to a better accuracy of the estimates.
References


