Large Scale Predictive Modeling for Micro-Simulation of 3G Air Interface Load

Dejan Radosavljevik – T-Mobile NL & Leiden University
Peter van der Putten – Leiden University
Data Mining Research

Over emphasis on technology, under emphasis on the user

“Everyone talks about rock these days; the problem is they forget about the roll.”- Keith Richards
“Everyone talks about rock these days; the problem is they forget about the roll.” - Keith Richards

What they thought would happen

What actually happened

USAID Nile Delta Project forgot about the roll too 😒

How to Make Mobile Network Analysts Who (Want to) Know Nothing About Data Mining Use It To Run Their Own Simulations of Mobile Network Service Quality Across Thousands of Cells And Trust The Results
This research
Addresses both the rock and the roll

The rock:
- Simple standard data mining algorithm
  - Multivariate Linear Regression + Wrapper
- Fully Automated Data Mining process, no human interaction
- Uses tools that are either Open Source or part of regular Telco infrastructure- Low Cost
- Massive number of models in a short amount of time
- High impact business problem, not addressed with Data Mining before

The roll:
- Emphasis on Deployment
- Real business world applied study
- Embedded and Deployed Data Analysis System
- Data mining for non-data miners
- Simulation framework decoupled from data mining
New DM use within telecommunications (not marketing)
- First time a mobile operator uses DM for Network Upgrades
- Network is a crucial resource, major proportion of CAPEX
- Overloaded cells are detrimental to customer experience
- Predict congestion/overload of cells in a given timeframe, for each cell separately
- From reactive to proactive
- Targeted investments
Load Approximation and Prediction

Using Linear Regression to gain acceptance

- One model per cell
  - Load pockets*
  - Interest in cells that do not behave normally

- Why Linear Regression
  - Focus on the high end of the load curve
    - Remove zero instances
  - Fast- Speed is crucial
  - Low variance -> less over-fitting
  - Easy to implement and explain
  - Transparent, no ‘black box’
  - Easy to export to Excel

How does it all work-The Rock?

Data and Structure

- 9 input parameters, 4 output parameters
- Hourly measures for six weeks
- More than 20,000 cells
- About 80,000 models in less than a week
  - When run in parallel, in just 3 days
How does it all work - The Rock?

Data and Structure

- Query the data for the cell
- Filter out zero instances output
- Use a wrapper for feature selection
- Create, validate and store the model in Excel
Original Use Case
How we thought the model will be used

Budget for Network upgrades using forecasts of future load

1. Evaluate Current Load And The Respective Input Parameters
2. Scale Input Parameters Based on Regular Growth
3. Feed the Scaled Input Parameters into the Regression Load Formulas
4. If the Value is Above a Threshold, Mark the Cell for Upgrade
Original Use Case
How we thought the model will be used

Budget for Network upgrades using forecasts of future load

1. Evaluate Current Load And The Respective Input Parameters
2. Scale Input Parameters Based on Regular Growth
3. Feed the Scaled Input Parameters into the Regression Load Formulas
4. If the Value is Above a Threshold, Mark the Cell for Upgrade
Unforeseen Use Cases

We created a monster

Scaling the input parameters can be based on anything

Several other use cases developed:

- Business case evaluation for adding an MVNO like Lycamobile
- Business case evaluation for M2M
- Prioritization of picking locations for 4G
- Calculating network improvements necessary when opening **shops**-> implies more customers
- Calculating network improvements necessary for specific geo-marketing campaigns
Deployment- This is how we roll
The least analyzed step of the data mining process and key to our success

Addressing Product Based Causes of Innovation Failure* by decoupling data mining from simulation

- Relative Advantage
  - A huge number of models and allows simulation
- Compatibility
  - Use Current Infrastructure
- Complexity
  - Data Mining reduced to Excel 😊
- Observability
  - It’s a formula in Excel
- Trialability
  - Change an input value Excel

Deployment- This is how we roll
The least analyzed step of the data mining process and key to our success

Large Scale Predictive Modeling for Micro-Simulation of 3G Air Interface Load, KDD 2014, NYC

Addressing Product Based Causes of Innovation Failure*

• Relative Advantage
• A huge number of models and allows simulation
• Compatibility
• Use Current Infrastructure
• Complexity
• Data Mining reduced to Excel 😊
• Observability
• It’s a formula in Excel
• Trialability
• Change an input value Excel

Deployment- This is how we roll
The least analyzed step of the data mining process and key to our success

Complexity
- Data Mining reduced to Excel 😊

Observability
- It’s a formula in Excel

Trialability
- Change an input value Excel

4 other country operators are using it too!
Thank you for your attention! Questions?