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

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

What they thought would happen

What actually happened

USAID Nile Delta Project forgot about the roll too ☹

*Rogers,Everett M. (1995),Diffusion of Innovations 4th ed.New York, N.Y.: The Free Press.

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:

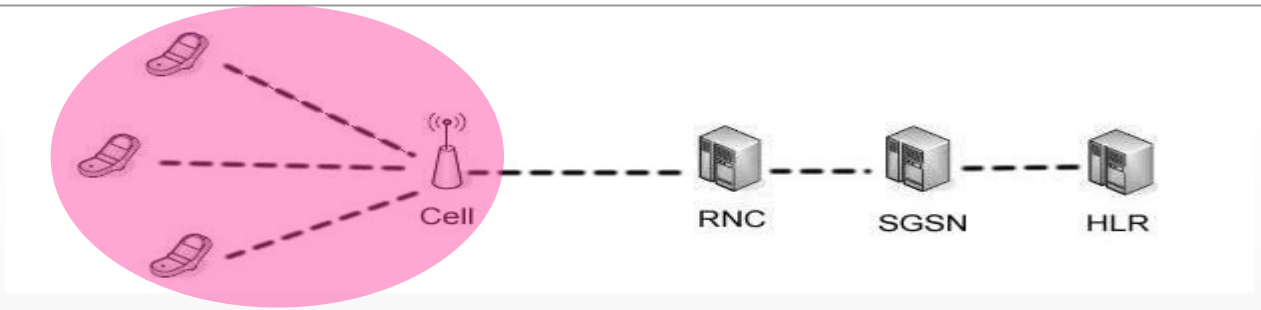
- Simple standard data mining algorithm
  - Multivariate Linear Regression + Wrapper
- Fully Automated Data Mining process, no human interaction
- Uses tools that are either Open Source or part of regular Telco infrastructure- Low Cost
- Massive number of models in a short amount of time
- High impact 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

3G Air Interface Load

What's the problem?



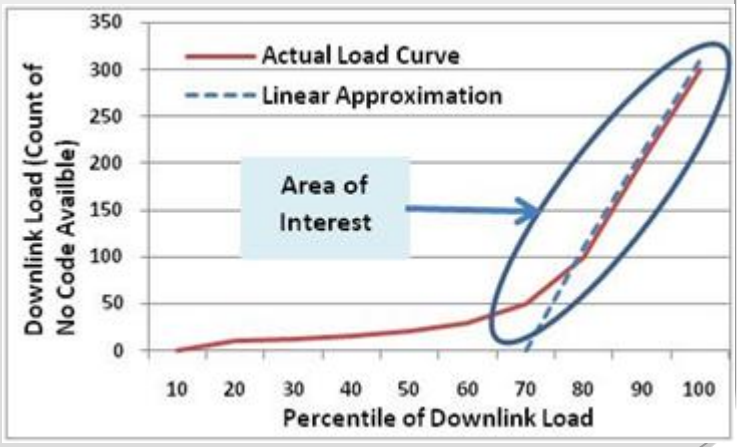
New DM use within telecommunications (not marketing)

- 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



Feinberg, E.A., Genethliou, D.: Load Forecasting. In: Chow, J.W., Wu, F.F., Momoh, J., (eds.) Applied Mathematics for restructured electric power systems, pp. 269-285, Springer, Heidelberg (2005)

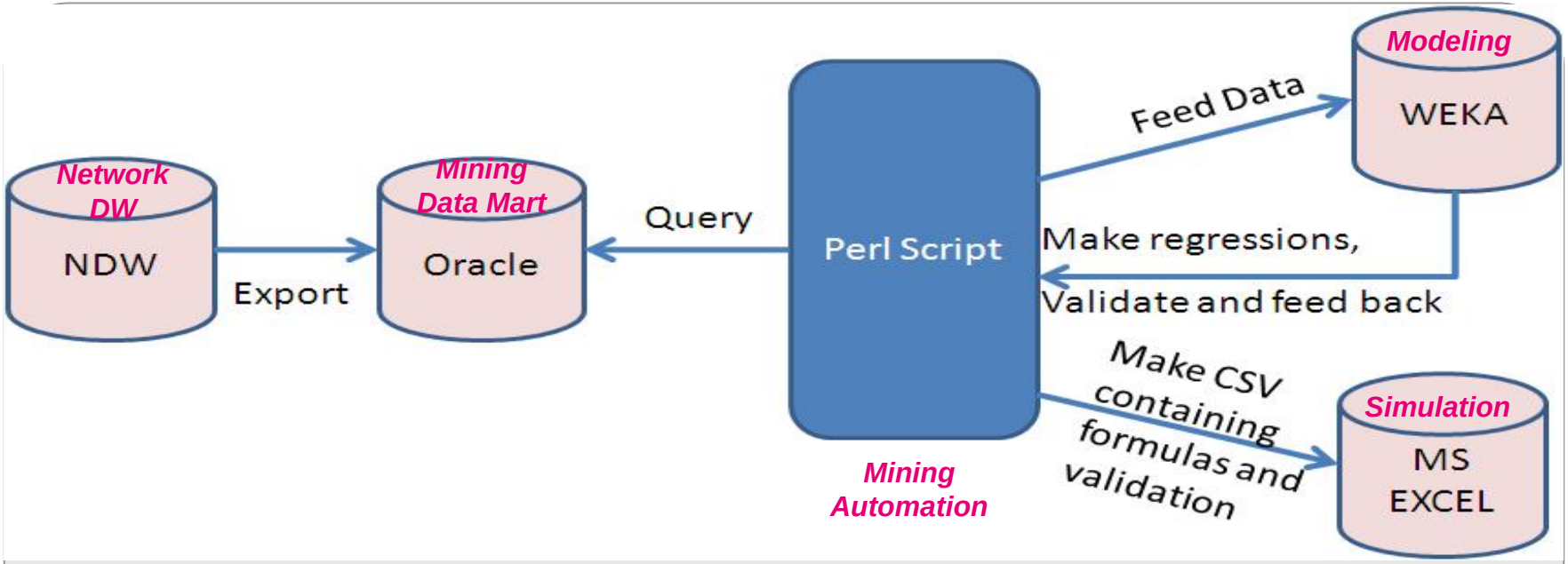
How does it all work-The Rock?

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

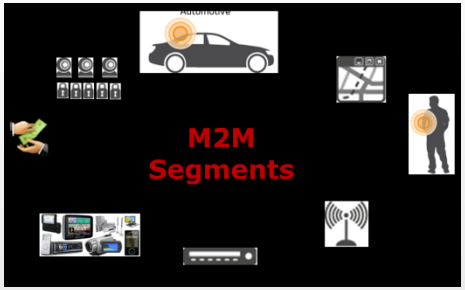
1. Evaluate Current Load And The Respective Input Parameters
2. Scale Input Parameters Based on Regular Growth
3. Feed the Scaled Input Parameters into the Regression Load Formulas
4. If the Value is Above a Threshold, Mark the Cell for Upgrade

Unforeseen Use Cases

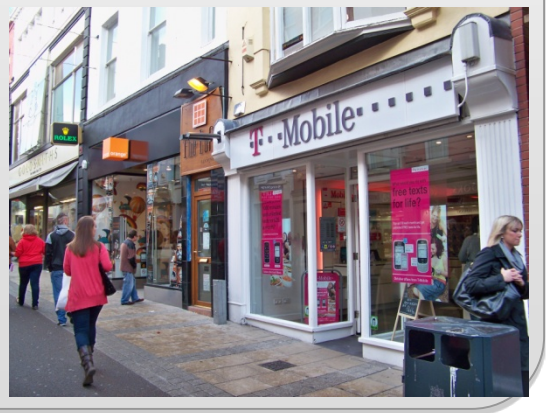
We created a monster

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



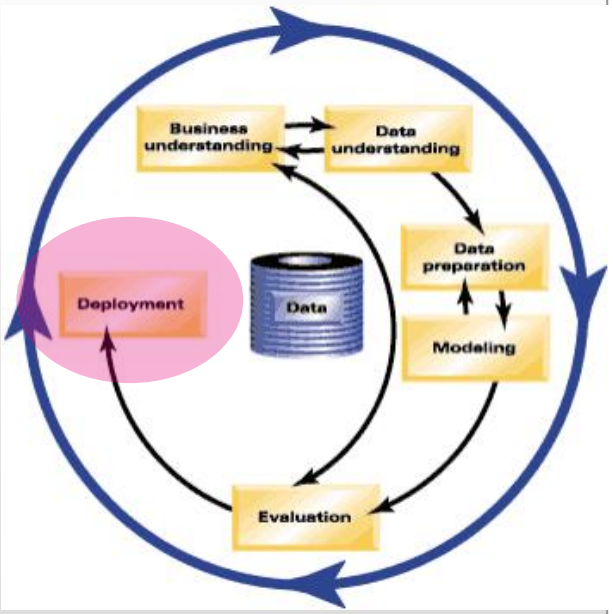
Deployment- This is how we roll

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

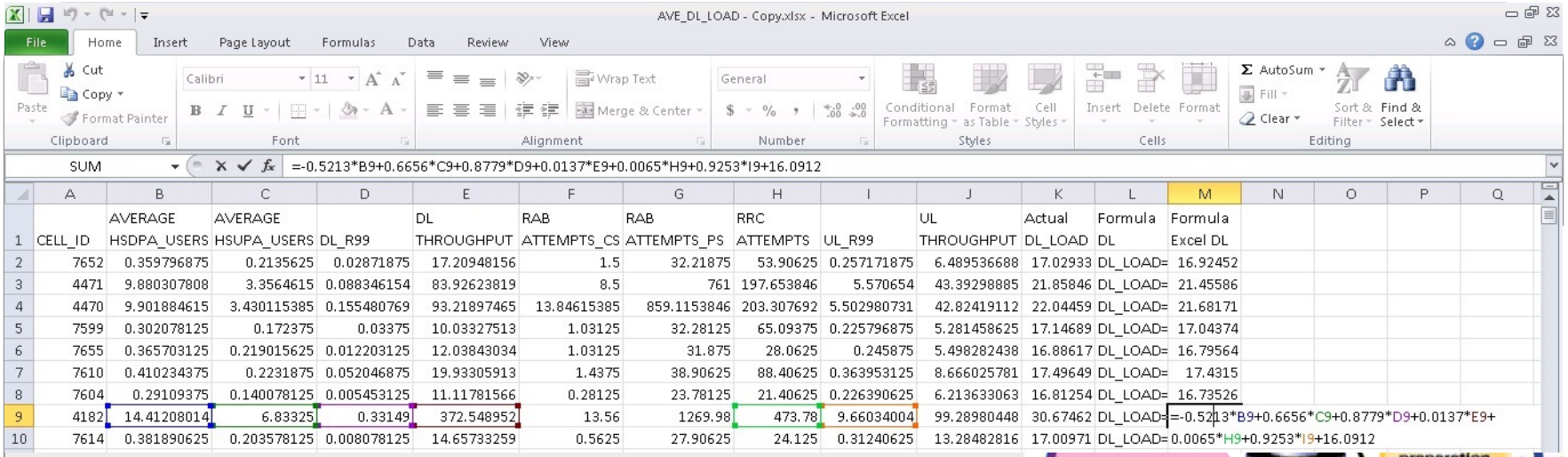
- Relative Advantage
  - A huge number of models and allows simulation
- Compatibility
  - Use Current Infrastructure
- Complexity
  - Data Mining reduced to Excel ☺
- Observability
  - It's a formula in Excel
- Trialability
  - Change an input value Excel



*Rogers,Everett M. (1995),Diffusion of Innovations 4th ed.New York, N.Y.: The Free Press.

Deployment- This is how we roll

The least analyzed step of the data mining process and key to our success



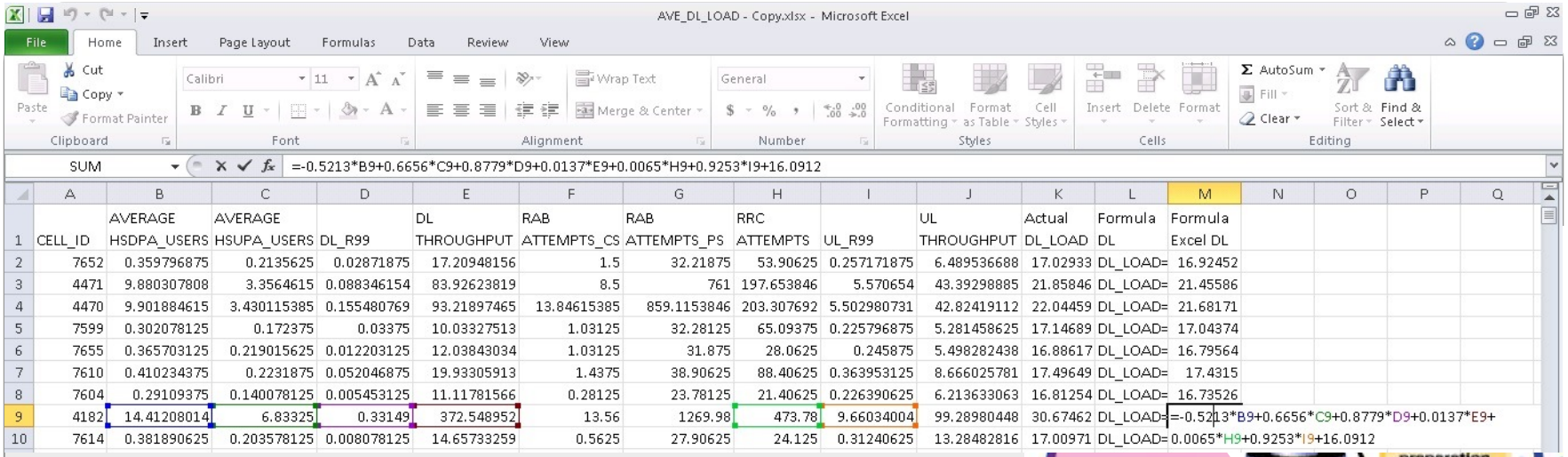
	A	B	C	D	E	F	G	H	I	J	K	L	M
1	CELL_ID	AVERAGE HSDPA_USERS	AVERAGE HSUPA_USERS	DL_R99	DL THROUGHPUT	RAB ATTEMPTS_CS	RAB ATTEMPTS_PS	RRC ATTEMPTS	UL_R99	UL THROUGHPUT	Actual DL_LOAD	Formula DL	Formula Excel DL
2	7652	0.359796875	0.2135625	0.02871875	17.20948156	1.5	32.21875	53.90625	0.257171875	6.489536688	17.02933	DL_LOAD=	16.92452
3	4471	9.880307808	3.3564615	0.088346154	83.92623819	8.5	761	197.653846	5.570654	43.39298885	21.85846	DL_LOAD=	21.45586
4	4470	9.901884615	3.430115385	0.155480769	93.21897465	13.84615385	859.1153846	203.307692	5.502980731	42.82419112	22.04459	DL_LOAD=	21.68171
5	7599	0.302078125	0.172375	0.03375	10.03327513	1.03125	32.28125	65.09375	0.225796875	5.281458625	17.14689	DL_LOAD=	17.04374
6	7655	0.365703125	0.219015625	0.012203125	12.03843034	1.03125	31.875	28.0625	0.245875	5.498282438	16.88617	DL_LOAD=	16.79564
7	7610	0.410234375	0.2231875	0.052046875	19.93305913	1.4375	38.90625	88.40625	0.363953125	8.666025781	17.49649	DL_LOAD=	17.4315
8	7604	0.29109375	0.140078125	0.005453125	11.11781566	0.28125	23.78125	21.40625	0.226390625	6.213633063	16.81254	DL_LOAD=	16.73526
9	4182	14.41208014	6.83325	0.33149	372.548952	13.56	1269.98	473.78	9.66034004	99.28980448	30.67462	DL_LOAD=	=-0.5213*B9+0.6656*C9+0.8779*D9+0.0137*E9+
10	7614	0.381890625	0.203578125	0.008078125	14.65733259	0.5625	27.90625	24.125	0.31240625	13.28482816	17.00971	DL_LOAD=	0.0065*H9+0.9253*I9+16.0912

- Complexity
  - Data Mining reduced to Excel ☺
- Observability
  - It's a formula in Excel
- Trialability
  - Change an input value Excel

*Rogers,Everett M. (1995),Diffusion of Innovations 4th ed.New York, N.Y.: The Free Press.

Deployment- This is how we roll

The least analyzed step of the data mining process and key to our success



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	CELL_ID	AVERAGE HSDPA_USERS	AVERAGE HSUPA_USERS	DL_R99	DL THROUGHPUT	RAB ATTEMPTS_CS	RAB ATTEMPTS_PS	RRC ATTEMPTS	UL_R99	UL THROUGHPUT	Actual DL_LOAD	Formula DL	Formula Excel DL
2	7652	0.359796875	0.2135625	0.02871875	17.20948156	1.5	32.21875	53.90625	0.257171875	6.489536688	17.02933	DL_LOAD=	16.92452
3	4471	9.880307808	3.3564615	0.088346154	83.92623819	8.5	761	197.653846	5.570654	43.39298885	21.85846	DL_LOAD=	21.45586
4	4470	9.901884615	3.430115385	0.155480769	93.21897465	13.84615385	859.1153846	203.307692	5.502980731	42.82419112	22.04459	DL_LOAD=	21.68171
5	7599	0.302078125	0.172375	0.03375	10.03327513	1.03125	32.28125	65.09375	0.225796875	5.281458625	17.14689	DL_LOAD=	17.04374
6	7655	0.365703125	0.219015625	0.012203125	12.03843034	1.03125	31.875	28.0625	0.245875	5.498282438	16.88617	DL_LOAD=	16.79564
7	7610	0.410234375	0.2231875	0.052046875	19.93305913	1.4375	38.90625	88.40625	0.363953125	8.666025781	17.49649	DL_LOAD=	17.4315
8	7604	0.29109375	0.140078125	0.005453125	11.11781566	0.28125	23.78125	21.40625	0.226390625	6.213633063	16.81254	DL_LOAD=	16.73526
9	4182	14.41208014	6.83325	0.33149	372.548952	13.56	1269.98	473.78	9.66034004	99.28980448	30.67462	DL_LOAD=	=-0.5213*B9+0.6656*C9+0.8779*D9+0.0137*E9+
10	7614	0.381890625	0.203578125	0.008078125	14.65733259	0.5625	27.90625	24.125	0.31240625	13.28482816	17.00971	DL_LOAD=	0.0065*H9+0.9253*I9+16.0912

- 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?