

Prediction and Fuzzy Logic at ThomasCook to automate price settings of last minute offers



Jan Wijffels:
jwijffels@bnosac.be

BNOSAC - Belgium Network of Open Source Analytical Consultants
www.bnosac.be

July 10, 2009



Introduction to BNOSAC

- ▶ Group of consultants focussed on open source analytical engineering
- ▶ Poor man's BI:
Python/PostgreSQL/Pentaho/OpenOffice/R. . .



- ▶ Expertise in predictive data mining, biostatistics, geostats, python programming, GUI building, artificial intelligence

Business of ThomasCook Belgium

- ▶ Sell holidays (sun and beach in this user case)
- ▶ 70 destinations around Mediterranean and Americas



- ▶ Own planes & bought seats need to be filled with passengers
- ▶ Flight frequency for some destinations up to 4 flights within one day. Some flights can be combined (BRU->ACE->FUE->ACE->BRU)

Introduction to last minute price settings

- ▶ Last minute prices departures Brussels/Liège/Ostend/Lille
- ▶ Up to 2 months before departure
- ▶ People book now to go on holiday e.g. August 10, 2009 to destination X. Can stay 3-28 nights, choose among several hotels, with certain board (All Inclusive, B&B, ...) and certain room type.

e.g. Hurghada (HRG): daily flights from Brussels (BRU)

prices in August: $31 \text{ days} \times 12 \text{ durations} \times 2 \text{ brands} \times 20 \text{ hotels} \times 4 \text{ boards} \times 3 \text{ room types} = \pm 248000 \text{ prices}$

- ▶ Prices can go ↗ or ↘ depending on offer and demand

Business challenge

Business challenge

Fill the planes at the highest prices so that the plane doesn't fill too fast and make sure all seats are filled.

- ▶ Currently **2.9 Mio** promotional prices on the market. Prices change daily.
- ▶ Only cover approaches towards prices of packages (flight + hotel), only price effects of couples (so no children).

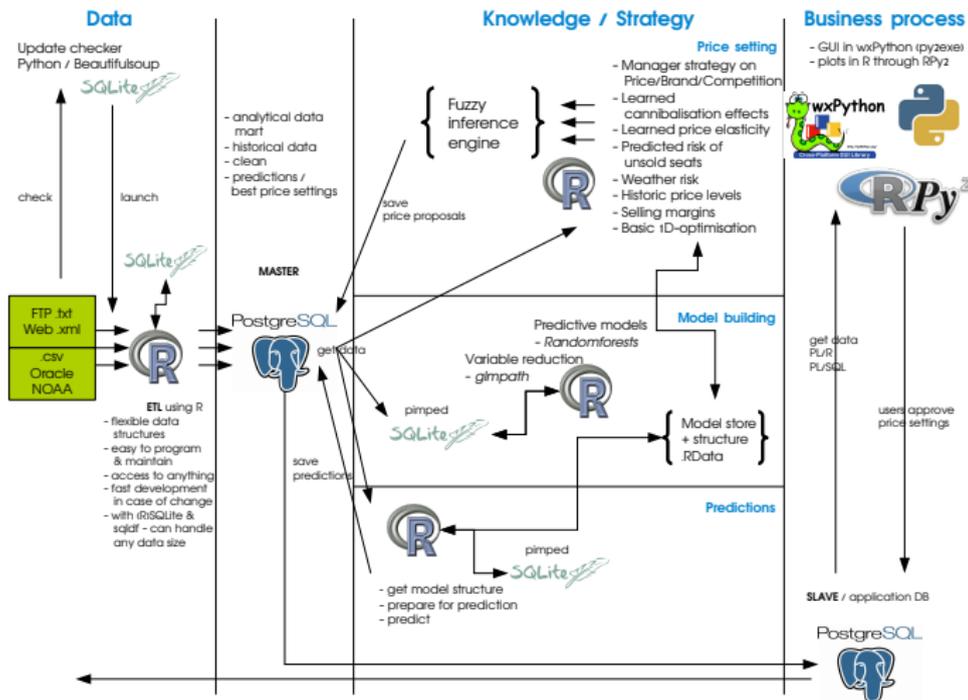
Optimisation problem

- ▶ A lot of factors influencing bookings:
 - ▶ Holiday information / Day of the week
 - ▶ Flight information (hours of departure and of return flights, availability of flights)
 - ▶ Weather
 - ▶ Prices (2 brands, competitor) and price evolution
 - ▶ Cannibalisation (risk of losing passengers to yourself)
 - ▶ prices of similar destinations - last minute customers only want the sun at the cheapest price
 - ▶ prices on similar departure dates (a few days later/earlier)
 - ▶ Days before departure
 - ▶ ... dimensionality is large (> 100000 factors could influence bookings on flight from BRU to HRG on August 10, 2009)
- ▶ Find the best price settings over all these parameters to
 - ▶ optimize margin / minimize risk / optimize market share

Data & speed challenge

- ▶  Data size last year only
 - ▶ own last minute promotional prices: >450 million records.
 - ▶ competitor prices
 - ▶ flight info: ± 60000 flights on the market $\times 365$ days $\pm 21.900.000$ records
 - ▶ weather info at noon:
70 destinations $\times 365$ days \times weather forecasts
- ▶ Speed
 - ▶  "Hello prices" at ± 7 o'clock in the morning (mainframe).
 - ▶  "Hello employees" at ± 8 h30 in the morning
 - ▶ ± 1 h30 to make predictions and give 'best' automatic price proposals

Architectural solution



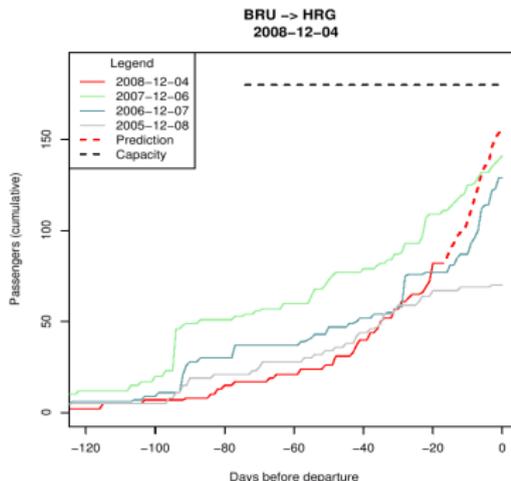
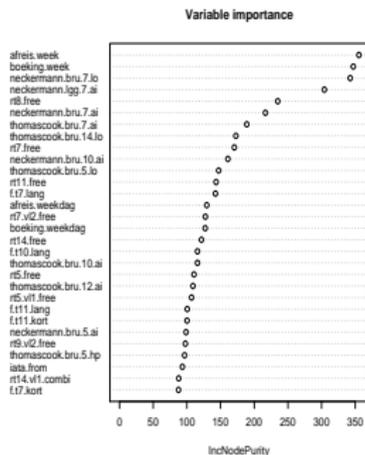
Analytical solution: Predictive modelling

Out of the box solutions exist in R. 'Best practice' approach:

- ▶ Pimp SQLite so that it can handle tables with up to ± 30000 columns. Raw model tables dim 20.000.000 x 30000
- ▶ Data preparation (missing values, split numeric data in categories) - do heavy reshaping/juggling/merging/indexing in (R)SQLite & sqldf, use R for advanced data features
- ▶ Sample depending on CPU/RAM and statistical technique: we have 4 dual cores, 64bit Linux, 32Gb RAM.
- ▶ Reduce: GLM with penalization on the size of the L1 norm of the coefficients $L(\beta, \lambda) = - \sum_{i=0}^n y_i \theta(\beta)_i b(\theta(\beta)_i) + \lambda \|\beta\|_1$ (glmnet package)

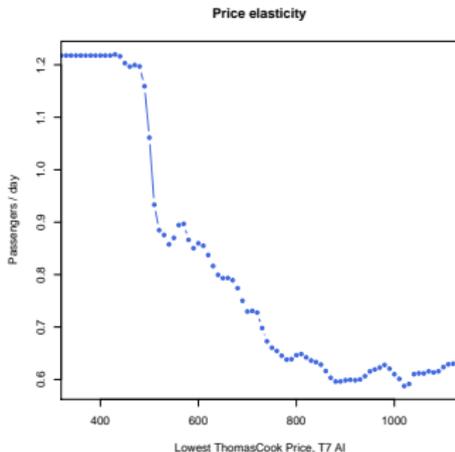
Analytical solution: Predictive modelling cont.

- ▶ Only most important predictors to build randomForest
- ▶ Use randomForest model to predict how fast the flights will fill.

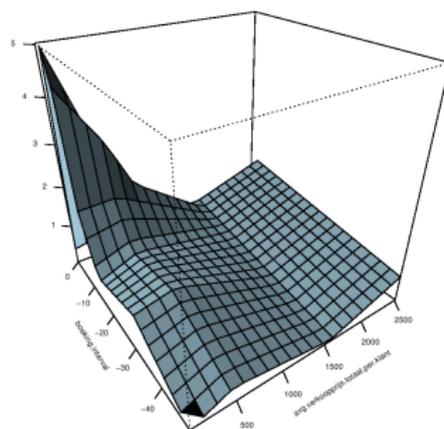


Analytical solution: Predictive modelling cont.

- ▶ Get the price effects from the randomForest model and use it:
- ▶ Do fast 1- or 2-dimensional optimisation to fill seats that will not be filled according to the forecast at the optimal price.



BRU - ACE aantal pax per dag
 Wanneer kopen mensen tegen welke prijs (dagen voor vertrek en verkoopprijs)
 Verkoopprijs cuts 0, 1574, 870, 564, 808, 721, 472 ::: Dag voor afreis cuts 0, -1, -14, -5, -31, -2, -7

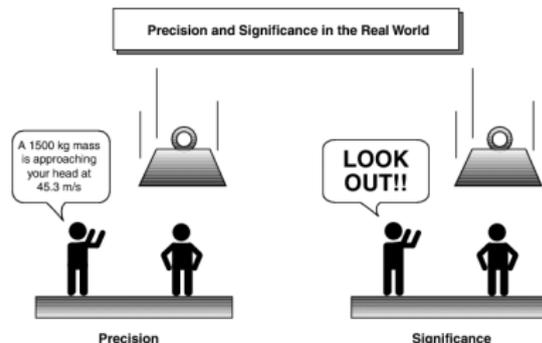


Analytical solution: Fuzzy Logic

Prediction and optimisation is nice but not enough

Managers reason with words/concepts. Mimic them and combine their logic with predictive logic. How?

- ▶ Map business concepts to fuzzy sets.
- ▶ Make fuzzy rule-based engine reflecting how managers/business users decide on price settings
- ▶ Do fuzzy inference to obtain new price settings



Analytical solution: Fuzzy Logic cont.

Map business concepts to fuzzy sets.

- ▶ Listen to the people. Fuzzy concepts have blurred boundaries.
- ▶ Map linguistic variables to a membership degree $\mu(x) \in [0, 1]$
- ▶ sets package (Hornik K., Meyer D., Buchta C.)
- ▶ fuzzy_normal, fuzzy_trapezoid, fuzzy_sigmoid, ...

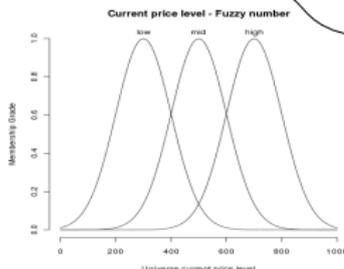
Low-level fuzzy sets/inputs

- Current flight situation

- Randomforest Prediction

- Simulation @ different price levels & timepoints

- Price elasticity in models + importance measures for different competitors
- Overlapping hotels difference
- Overlapping hotels price change
- Other hotels price level
- Other hotels price change
- Price elasticity in models
- Current price levels
- Historic cannibalisation levels



Higher-level fuzzy sets/inputs

Current risk empty seats

Predicted risk empty seats

Elasticity-based optimal Price Move

Competitor risk

Outgoing cannibalisation

Incoming Cannibalisation

Days before departure

Expert opinions & other inputs ...

Fuzzy rule base

Optimal Price Move

Analytical solution: Fuzzy Logic cont.

Make fuzzy rule-based engine, do fuzzy inference & defuzzify.

```
rules <- set(
  fuzzy_rule(predicted_risk %is% low, price_change %is% up),
  fuzzy_rule(predicted_risk %is% high
    & competitor_risk %is% high, price_change %is% down_high)
  ...)
simple.system <- fuzzy_system(variables, rules)
fuzzy.best.price <- fuzzy_inference(simple.system, NEWDATA)
gset_defuzzify(fuzzy.best.price, "centroid")
```

- ▶ Different business strategies can be easily mapped to fuzzy inference engines.

Influence the business process, use visuals, build GUI

Prediction, optimisation and improving on business users
 is nice but not enough, you need to influence the business process

The screenshot displays the 'ThomasCook - Indeed pe wereld - Price Cockpit' application. The main window features a data table with columns for flight details and pricing. Below the table are three charts: 'Dashboard basic prices' (a scatter plot of lowest price vs. departure time), 'Dashboard basic occupation' (a line chart showing occupancy over time), and 'Dashboard basic evolution' (a line chart showing price evolution over time). A bar chart on the right shows the breakdown of prices for different components like Cap, Pax, FO, and Free. A sidebar on the left contains a menu with options like 'Flight situation', 'Prices', 'Weather', and 'Bookings'.

Help	№	Day	Datum	Afreis	Van	Pax	Cap	Bez%	Pax L1	Pax L1w	Pax L2w	SMS AI	Free	T7	T8	T9	T10	T11	T12	T13	T14	Promo	h.c.	Broch	Avail	
(0)	AYT	3	01/07/2009	BRU	480	567	84	4	24	35	706	87	83	5	235	234	161	85	91	220	0	569	0	None	None	Tuvana
(0)	AYT	4	02/07/2009	BRU	557	558	99	0	3	-1	758	1	5	235	234	161	85	91	220	0	667	0	None	None	Tuvana	
(0)	AYT	4	02/07/2009	LGS	134	134	100	0	0	0	None	0	0	0	59	2	0	0	0	0	0	667	0	767	9	HOTEL COLEM G
(0)	AYT	4	02/07/2009	OST	34	34	100	0	0	0	None	0	0	0	56	0	0	0	0	0	0	667	0	767	9	HOTEL COLEM G
(0)	AYT	5	03/07/2009	BRU	511	747	68	3	10	-1	747	236	235	234	161	85	91	220	0	343	0	704	0	839	9	HOTEL PIEDER B
(0)	AYT	6	04/07/2009	BRU	263	516	50	4	8	7	760	233	234	161	85	91	220	0	343	320	0	704	0	839	9	HOTEL PIEDER B
(0)	AYT	7	05/07/2009	BRU	323	525	61	7	16	15	740	202	161	85	91	220	0	0	0	0	0	704	0	839	9	HOTEL PIEDER B
(0)	AYT	7	05/07/2009	LGS	125	127	98	0	4	0	871	2	59	2	0	0	0	0	0	0	0	704	0	839	9	HOTEL PIEDER B
(0)	AYT	7	05/07/2009	OST	55	62	88	0	0	2	881	7	56	0	0	0	0	0	0	0	0	704	0	839	9	HOTEL PIEDER B
(0)	AYT	1	06/07/2009	BRU	386	546	70	0	4	2	723	100	85	91	220	0	0	0	0	0	0	704	0	839	9	HOTEL PIEDER B
(0)	AYT	1	06/07/2009	LGS	121	168	72	0	0	0	753	47	2	0	0	0	0	0	0	0	0	704	0	839	9	HOTEL PIEDER B
(0)	AYT	2	07/07/2009	BRU	159	258	61	8	39	14	763	99	91	220	0	343	320	0	0	0	0	704	0	839	9	HOTEL PIEDER B
(0)	AYT	3	08/07/2009	BRU	218	369	59	7	13	18	693	151	220	0	343	320	0	0	0	0	0	704	0	839	9	HOTEL PIEDER B
(0)	AYT	4	09/07/2009	BRU	158	189	83	0	9	7	None	31	0	0	343	320	0	0	0	0	0	704	0	839	9	HOTEL PIEDER B
(0)	AYT	4	09/07/2009	LGS	44	111	41	0	4	0	842	64	0	0	0	0	0	0	0	0	0	704	0	839	9	HOTEL PIEDER B

PL/R, RPy2, GUI's in R, people

- ▶ PL/R.
 - ▶ Had a lot of shared memory problems while other processes were running. But probably overkilled it (run PL/R script which calls some R code from within R process that uses RdbiPgSQL)
 - ▶ Debugging hell.
 - ▶ R & SQLite is our best choice for heavy data juggling.
 - ▶ PL/R is OK for collecting information on diverse data sources in 1 call from a remote machine.
 - ▶ Useful for plotting purposes in SaaS framework.

PL/R, RPy2, GUI's in R, people cont.

- ▶ User interfaces - developer view
 - ▶ Combining wxPython and R through RPy2 is easy and simple.
 - ▶ py2exe gives easy python binary executables, people only need to have R installed to access its power
- ▶ User interfaces - IT view
 - ▶ IT departments don't like R
 - ▶ R should be SaaS, central server where people can connect to
- ▶ User interfaces - business user point of view
 - ▶ They don't care about R
 - ▶ GUI and plotting the results helped convincing them
 - ▶ Fuzzy logic allowed them to interact and stick to the business.
 - ▶ Combining the results with an improved business process was the most convincing factor.

Questions?

`http://www.bnosac.be`