Forecasting and big data analysis

July 2014

Ron Levkovitz, Ogentech Ltd.



Supply chain forecasting

- Accurate forecasting and demand planning is the basis of efficient supply chain management and execution.
- It is almost impossible to improve supply chain without achieving good level of accuracy.
- SCM Forecast for mid/long term planning is, traditionally, based on the past of the forecasted entity and on the accumulated experience of the planner



Supply chain forecasting – Keep It Simple

- Available information is usually incomplete and unreliable.
- Many important parameters with direct influence on the demand are not used - number of observations is usually low and demand time series may not fit parameters resolution (e.g. monthly data, weekly campaign)
- The statistical algorithms treat fluctuations caused by unknown factors as noise
- The, statistical results are, therefore, simplistic,
- Only the simple algorithms succeed producing meaningful results – the complicated algorithms tend to fit noise
- Result depends on planner's abilities, information and bias
- The process is not captured and there is no learning curve



Real example demonstrating planners' work. What will be the growth in the next \$889 promo?



The planner is expected to know the reasons for changes, for growth, for different behavior during the same promotions.

Supply chain forecasting Keep it simple?

- Many demand planning providers stick to the 'keep it simple' approach.
- Provide the barest forecast expecting the planners to complete the 'complex' part
- Does this approach make sense? Only if we assume the system cannot know what the planner knows/cannot properly interpret the past.
- Q: can we fully automate the process of forecasting?
- Q: can we reshape planners work to provide knowledge instead of numbers?





Now casting

- refers to the sets of techniques devised to make short term forecasts and produce predictions without the need for informal judgment
- Common methods:
 - Improve upon the usual prediction by taking an ensemble of forecasts (for example, forecasts from different lags) using the probability of each instead of the most likely scenario
 - Morphing of model output approach of the long term time series prediction with the very short term prediction (e.g. predictions made by short term time series forecasting, by demand sensing, by the Kalman filter, Bayesian model, etc.)



"Big Data"

Big data (from Wikipedia) : a blanket term for any collection of <u>data sets</u> so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications. The challenges include capture, curation, storage, search, sharing, transfer, analysis and visualization. The trend to larger data sets is due to the additional information derivable from analysis of a single large set of related data, as compared to separate smaller sets with the same total amount of data, allowing correlations to be found to "spot business trends, determine quality of research, prevent diseases, <u>link legal citations</u>, combat crime, and determine real-time roadway traffic conditions."^[1]

- Most companies collect 'gazillions' of data items. Many more are available via Google, Facebook, Twitter, Amazon, etc.
- These data are seldom structured
- Many companies use "Big Data" for manual queries (marketing and sales), to answer research questions etc.
- It is still not common to utilize Big Data automatically and systematically within an algorithmic (forecasting) framework
- We argue that such use will both contribute to both the analysis and to the forecasting (when investigating past results it is very easy to fall into the 'hidden variable' trap)



Two types of data items in "Big Data"

- Continuous information collected and available at known interval resolution.
 For example list price at various retailers. The information is collected and stored every (day, week, month) and as long as a product is sold, it is available
 - The continuous information can be included in the forecast by using traditional methods for multi-variables statistics (e.g. MLR)
- Sporadic information is available only at certain times. For example, hurricane alerts, social network buzz, advertising campaigns etc. At other times it does not exist.
 - The sporadic information is treated in a different way and requires a state analysis to assess its influence. In the upcoming slides we provide the infrastructure for this method.



Precondition, hierarchy and extrapolation Utilizing non continuous 'Big data' information within a forecasting framework

Assume a time series vector a_1, a_2, \ldots, a_t

Parameter b exist only for time buckets j, i, k, $(i < j < k \leq t)$

Parameter c exist only for time buckets i, l ($i < l \leq t$)

We would like to forecast $a_{t+1}, a_{t+2}, \dots, a_{t+n}$. We would like to understand whether b, c are significant variables. The sketch below demonstrates the approach:

- **1**. Back cast $\bar{a}_1, \ldots \bar{a}_t$ using best fit (say, exponential smoothing)
- 2. Note: the back cast is not using all the training one step ahead parameters

3. Structural analysis of error component
$$Err_b = \frac{\sum_{m=j,i,k} \frac{|a_m - a_m|}{a_m}}{3}$$
, $Err_c = \frac{\sum_{m=i,l} \frac{|a_m - a_m|}{a_m}}{2}$, Err_o
$$= \frac{\sum_{m \neq i,l,kj} \frac{|a_m - \overline{a}_m|}{a_m}}{t-4}$$



Precondition, hierarchy and extrapolation

1. If $Err_b \gg Err_o$, $Err_c \gg Err_o$, $Err_b \gg Err_c$ then

Compute Err_b filter parameters, step wise.

- 2. Filter to create $\widehat{a_1}$, ... $\widehat{a_t}$
- 3. Compute $\widehat{Err_c}$, $\widehat{Err_o}$,
- 4. If $\widehat{Err_c}$, $\gg \widehat{Err_o}$ then Compute Err_c filter parameters, step wise.
- 5. Filter to create $\dot{a_1}$, ... $\dot{a_t}$ Compute $\dot{Err_o}$



Expansion of the concept big data and now casting to long term planning - examples

- We collect and archive relevant information items those describing the relevant influences on the current configuration.
- For example: the current configuration consists of (date, per sku and trade partner group: date price, date marketing activities, date availability/on-hand, date presentation, date assortment, date retailer activities)
- The system calculates the transition matrix of these data based on their past instances
- It uses the calculated state parameters to update both the past and the future forecasts
- The approach was tested in two companies, the Whirlpool corporation (mostly non continuous data) and Clalit Health Services (mostly continuous data)
- The Clalit results are described in the talk and presentation by Sharon Hovav.





The Whirlpool Case Study – the company



- Whirlpool Corporation is the world's leading global manufacturer and marketer of major home appliances.
- Annual sales of approximately \$19 billion in 2013,
- 69,000 employees
- 59 manufacturing and technology research centers around the world.
- The company markets <u>Whirlpool</u>, <u>Maytaq</u>, <u>KitchenAid</u>, <u>Jenn-</u>
 <u>Air</u>, <u>Amana</u>, <u>Brastemp</u>, <u>Consul</u>, <u>Bauknecht</u> and other major brand names.







Forecasting targets and characteristics

- Main targets:
 - Accurate forecasting of shipments at all levels (lowest granularity is sales_organization/material/location/trade partner).
 - Measured by 60 and 120 days out mothly WMAPE.
 - Reduced bias

Uses of forecast:

- Manufacturing/ Production Planning
- Reduced inventory @ distribution locations
- Trade partner planning





Characteristics of big data information

- Information is collected and reviewed and trade partner level
- Demand patterns for different trade partners are radically different
 - 1. Trade partner long term success and trends
 - 2. Trade partner retail market segment(construction, clearance, higher end standard, private labels)
 - 3. Economical indicators of trade partner segment
 - 4. Trade partner prices and sales
 - 5. Life cycle/transition information
 - 6. Promotional information
 - 7. Position information



Combined data model example (cont.) WHIRLPOOL ITEM1 Weekly behaviour in sears (from w1, 2012)

Whirlpool

14

 Note that for this product, the planned promotions did not affect actual sales. There are no promotional peaks in the sales data



Results in Summary

- Reconditioning the data at material/Trade Partner level since mid 2013
- Around 2 million different material/location/trade partner combination
- Each of which is reconditioned by the available parameters of its individual state matrix
- Parameters that do not provide improvement value are dropped
- Overall forecasting accuracy at material level 60 and 120 days out has improved by more then 10%
- Research continues to add more data into the approach





References

1. The Ensemble Kalman Filter: theoretical formulation and practical implementation <u>Geir Evensen</u>, <u>Ocean Dynamics</u> November 2003, Volume 53, <u>Issue 4</u>, pp 343-367

2. Automatic time series forecasting:

the forecast package for R, Rob J Hyndman and Yeasmin Khandakar, Department of Econometrics and Business Statistics, Monash University

3. Nowcasting Events from the Social Web with Statistical Learning, Vasileios Lampos and Nello Cristianini, ACM TIST 3(4), n.72, 2012.

4. The Signal and the Noise: Why So Many Predictions Fail — but Some Don't ,September 27, 2012 by <u>Nate Silver</u> (Author) ISBN-13: 978-1594204111

