Leveraging the Expanding Data Universe: Opportunities for Supply Chain Management

Contributed By Dr Barry Adrian Shepherd, National University of Singapore

We hear much regarding how big-data and advanced data analytics can transform the way in which we solve problems. But what specifically can they do for Supply Chain management?

A central theme of big data analytics is the use of data to make decisions. Importantly, this data is not just a few summary statistics but ever expanding amounts of diverse data, frequently raw, real-time and very granular, from a wide range of internal, external, third party and public sources. For supply chain, examples include real-time data from Point-of-Sale transactions, from advanced inventory tagging and tracking systems, from GPS tracking of vehicle fleets, from integrated partner IT systems and from browsing and search logs for the company website. Public and third-party examples include online news, industry and financial market data (commodity prices, exchange rates, bank rates, unemployment rates etc.) weather and road traffic monitoring websites, social media and competitor web sites plus many more.

So how can this big data help? One critical area in the supply chain is inventory management and getting accurate demand forecasts. Commonly used forecasting methods, such as trend estimation, often utilise only a few key variables and use only past sales to predict future sales. While this may make a good first guess it can produce forecasts that are not robust to fast changing consumer demand. If we can feed a richer spectrum of relevant data into the forecast, then we potentially gain much greater accuracy and robustness.

To illustrate, consider Tesco, one of the UKs most successful retailers. Tesco has invested heavily in applying advanced analytics to its supply chain. One big win is a statistical model that improves inventory management by utilising weather data to help predict customer buying behavior. By comparing historical weather data with sales records from over 3,000 stores the model learns which items sell more when hot (e.g. barbeque equipment), cold (e.g. cat litter!) and in between. Other contextual data is also taken into account, for example people are more likely to barbeque when a hot day follows a long cold period and opinions about how hot the day must be before braving the outdoors may differ from the (usually) cooler north to the warmer south of the country. By adding these effects to the model Tesco reduced being out of stock on good weather products by a factor of four.

This leads us to a second major theme of big data analytics - the use of advanced statistical and machine learning methods and text mining techniques to help make sense of these very large and diverse data sets. For example, algorithms that can detect cross-sell opportunities by examining millions or billions of individual SKU sales, and predictive modeling tools that can weight up thousands of variables when making a forecast. While this new predictive power doesn't mean that we should just throw all the data we have or can get into the mix, it does mean we can be expansive in our thinking about what data might be relevant or useful in solving our business problems and how we might gather this data.

For example, many e-retailer sites allow their customers to create personal gift and wish lists which are then communicated to their friends, but this data could be repurposed and also used to help predict future demand. Alternatively, we could use reverse-IP technology to track the zip codes of users visiting the various product pages on our company website to detect possible geographic hotspots of interest. Web search data can also be considered. Google's reuse of their search engine data to geographically track flu epidemics has received much attention – the frequency of flu related web searches in any geographic location is used as an indicator of the flu's prevalence there. We could extend this idea to track consumer interest in products (or product categories) – Google Trends is a public tool that offers such data. Another new source of consumer demand data is the social web. Consumer sentiment and trending product popularity, often referred to as "buzz", can be derived from Facebook postings, Twitter feeds and other social media. For example, in a recent study by IBM, Twitter postings were used to gauge consumer demand and improve demand forecasts for a large camera manufacturer. Commercial social media analytics tools are increasingly available to support this type of analysis. Importantly, these data sources provide up-to-the-minute, often real time, demand information which is vital for accurate forecasting in a fast changing world.

Apart from better consumer insight, robust demand forecasting also requires knowledge of the broader economic and market landscape. For example, information about our competitors, such as their pricing trends and their product popularity trends, that can help infer how much of the future demand pie they might grab. Much of this data is directly available online and much might be inferable, for example, by web mining information published online by competitors. To illustrate, consider Farecast, the airfare prediction system launched in 2007 and acquired by Microsoft a year later. One of its major innovations was the continuous collection and mining of airfare data for key routes from major airline websites. By the time of its launch it had collected over 175 billion airfare observations which were fed into models to predict future airfare price movements.

Predicting future demand in today's fast changing environment requires up-to-the-minute data collected from diverse sources, both internal and external reflecting the full market landscape in which our products will sell. This data will likely be huge in volume, potentially very raw and frequently textual requiring tools that employ advanced predictive modeling and text mining techniques. While this may all sound ambitious today, in a few years it may likely become mainstream and those not doing it could easily fall behind.

By Dr Barry Adrian Shepherd
Senior Member, Business Analytics & Knowledge Engineering Programmes, Institute of Systems Science (ISS) at National University of Singapore

Barry teaches Business Analytics, Data Mining and Knowledge Engineering at ISS and has over 30 years experience in these areas. Before joining ISS, he was based in the US and specialised in web analytics and user response modeling for online ad targeting at Microsoft Display Advertising and in process optimisation and data analytics for e-commerce order fulfillment at Amazon.com.

NUS National University of Singapore | ISS Institute of Systems Science