Big Data Breeds Intelligent Transportation

How unprecedented volume of data tackles transport-related issues?

When app technologies and cloud network structures are maturing these days, people increasingly expect an efficient and intelligent life in almost every aspect. In transportation, which affects everyone every minute, big data becomes a kind of solution which makes people's dream come true.

By Jean Hsu

Seoul is seeking to satisfy the growing demand for public transport. Initially, night bus routes were designed with reference to daytime bus timetables, but they did not actually reflect population movements at night. To overcome this problem, Korea Telecom (KT) and the City of Seoul worked together on a project using KT's big data and the city's public data to enhance the quality of public transport services. Their effort was rewarded, as the project earned the recognition of President Park Geun-hye. In the project, KT analysed over 300 million Call Detail Records (CDR) data together with a variety of Seoul's public data in order to study the movement of citizens around the city at night and identified the specific areas most frequented at night. Weighted distances were calculated between the centre points of populated areas, with relative popularity ranking determining primary stops.

These findings were then related to a heat map of the floating population, grouped by zones. Subsequent analysis then established the optimal locations of night bus stops which satisfy the most number of citizens, ensure citizens' safe journeys, provide economical transportation and maximise the usage of public transportation. Based on the results, bus routes were adjusted to include popular new stops (e.g. Konkuk University), avoid stops little-used at night (e.g. Seoul Art Center) or use routes that are congested by day (e.g. Namsan Tunnel) but clear by night.

Big data is suitable for use with public services like this one for it is based on mass analysis of public transport usage while avoiding issues on privacy and the use of personal data. Better decisions on public transport can be made, as justified by improved service efficiency, transparency, choice and accountability. Seven more night bus routes have been added to the original city plan as a result, and citizens only need to pay KRW2 to travel home by bus, rather than KRW20 by taxi. With this project, users reap real savings and their lives are made more efficient.

In mid July, Tainan people might have seen five foreigners taking buses or riding bicycle around the city, the oldest and historical city in southern Taiwan. Those foreigners were not tourists but representatives of IBM 's "Smart City Consultancy Group" who specialised in cloud computing, big data analysis, mobile computing as well as Integration of Departments or Technologies. They were there to help the government transform Tainan from a hundred-year-old city to a modern smart city of "Sustainability, Low-carbon Life and Livability".

After staying there for three weeks trying to understand more about Tainan's history, IBM representatives came up with a blueprint highlighting three main areas, i.e. Technology, Communication and Governance. Establishing a "Comprehensive Intelligent Transportation Management Centre" is the key of the blueprint which will rely on the use of cloud computing and big data analysis.

The term "Big Data" was made official in 2011 and Year 2012 was called "The First Year of Big Data". Two years later, Big Data has become an inevitable revolution of the IT industry. Some people even regard Big Data as "The Fifth Technology Wave" after Social Media.

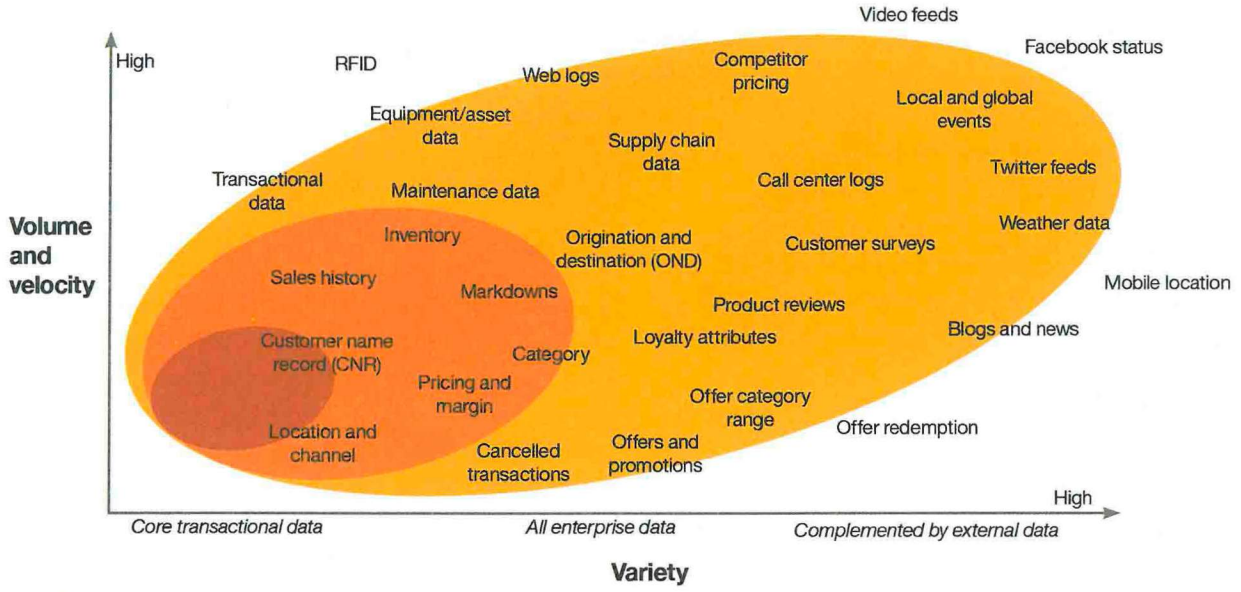

Figure 1: **Travel and transportation companies must consider an increasing amount of information to make strategic decisions.**

(source: IBM)

From Data to Insights

Companies across various segments of the travel and transportation industries – airlines, airports, railways, freight logistics, hospitality etc. – have been handling large amounts of data for years. However, in today's instrumented and interconnected world, unprecedented amounts of data are captured from almost every kind of system or event, and much of it is in a non-standard data format (for example, social, geospatial or sensor-generated data which does not fit neatly into traditional, structured and relational data warehouse). Besides, it is very difficult for enterprises to determine which real-time data, whether captured structured and unstructured, is most valuable for strategic decisions. (Please see Figure 1)

Fortunately, today's advances in analytic technologies and techniques enable enterprises to generate insights from data to reach the levels of sophistication, speed and accuracy previously unachievable. According to IBM's research, data itself has become an important strategic and competitive asset.

Three Major Big Data Business Applications

IBM's recent global big data study revealed that organisations implemented big data projects were 15% more likely to report a significant advantage from information and analytics compared to those relying on traditional analytics alone. Below are three major big data business applications for readers to understand where the market lies.

- *Customer analytics and loyalty marketing:*

Big data and analytics can help companies create a comprehensive, 360-degree customer view, which can dramatically improve customer interaction at every touch point and across the end-to-end customer experience. More complete "personal level" customer profiles can empower organisations to discover new ways to interact with individual customers, as well as enhance service delivery and marketing strategies.

- *Capacity and pricing optimisation:*

Given the capacity constraints travel and transportation companies face, a sharp focus on capacity and pricing optimisation is crucial. The ability to analyse more historical information in higher frequency – or even in near real-time – allows for more dynamic and smarter pricing actions, optimised capacity planning and effective yield management.

- *Predictive maintenance analytics:*

In the asset-intensive travel and transportation industries, success is highly dependent on the safe and reliable performance of those assets. By capturing and analysing more complete operational data, big data and analytics can offer timely reports for travel and transportation companies to manage and maintain their assets properly for greater safety, performance and equipment life.

Understanding Movement of People

Mobility is a major challenge for modern, growing cities,

and the transportation sector has kept innovating to increase efficiency and sustainability. Passengers swiping their RFID-based public transport passes leave useful traces which help dispatchers analyse and direct fleet movements. Companies, road operators and administrations possess enormous databases of vehicle movements based on GPS probe data, sensors and traffic cameras, and they are making full use of these data treasure chests to predict traffic jams in real-time, route emergency vehicles more effectively, or more generally, better understand traffic patterns and solve traffic-related problems.

Drivewise.ly and Zendrive are two California-based startups working on data-driven solutions aimed at making driving better, safer and more eco-friendly. Their assumption is that driving habits and commuting patterns can be recognised or learned by collecting the data captured through the sensors of drivers' smartphones (e.g. GPS, accelerometer) and referencing it to datasets collected elsewhere.

Data, taken in the context derived from a larger community of drivers, is turned into insights which benefit drivers such as "leaving 10 minutes early to reduce commute time by 20 minutes", or "adapting one's driving style to reduce fuel consumption and emissions". Data collected and analysed by transportation apps can attest people for a defensive driving style, which could help in renegotiating insurance premium.

Mobile phone leaves traces of people's mobility too, and these traces are exploited as a resource for transport modeling. This is of particular interest to the industries concerned as available transport-related data is scarce. In the summer of 2012, telecommunication provider Orange, in the 'Data for Development' challenge where "City and Transport Planning" was adopted as one of the themes, gave participants access to anonymised datasets provided by the company's Côte d'Ivoire branch. The dataset contained 2.5 billion records of calls and text messages exchanged between five million users over a period of five months.

Situated on a lagoon with only a few bridges connecting its districts, Abidjan, the capital of Côte d'Ivoire, is experiencing major traffic congestion. As it drafts a new urban transport plan for individual and collective means of transportation, call records serve as an informative set of data on the mobility of the population. By selecting the calls made from a residential area during evening hours (i.e. when people are at home) and monitoring the locations of the calls made on the same phones throughout the following day, the authority concerned can produce data which reveals how many people are commuting, as well as where and at what time; thus producing mobility maps based on which informed decisions on road and transport investment can be made.

On a larger geographical scale, mobile phone data contributes to analysis of migration patterns and is invaluable in crisis management. Global Pulse launched by the Executive Office of the United Nations Secretary-General in the wake of "The Great Recession" is a good example. It is an innovation initiative established in response to the need for more timely information to track and monitor the impact of global and local socio-economic crises. The initiative explores how new, digital data sources and real-time analytic technologies can help policy makers understand human well-being and emerging vulnerabilities in real time, in the interests of better protecting populations from the aftershock of financial and political crises. Global Pulse is a strong advocate of big data for development and humanitarian purposes.

Monetising Network Data Assets

Some telecommunications operators have started exploiting aggregated customer data as a source of income by providing analytics on anonymised datasets to third parties. Long been used exclusively for network management, billing and meeting lawful intercept requirements, communication metadata – data containing information on who sends a message, who receive it, and when and where it is sent – may represent yet another way for telecom players to capitalise on big data during phases like planning, roll-out, operation and upgrade of network infrastructure deployments.

By extracting detailed traffic information in real time, network analytics help telecom providers to optimise their routing network assets and predict faults and bottlenecks before they cause any harm. Based on customer value and behaviour metrics, a customer may dynamically be offered personalised solutions to respond to undesirable situations. Combining real time network insights and complete customer profiles, tailor-made offerings will add value to customer's experience which in turn will increase revenue opportunities and attract and retain customers. Network analytics is also an important means to detect and mitigate denial of service (DoS) attacks.

Conclusion

Big data and analytics are no longer confined to the realm of the technology "sand box." Service providers are making progress in deploying big data and analytics-centric solutions across business functions and capitalising on the resulting business return. Likewise, IT professionals are taking advantage of the

lower data management costs and the alternatives to traditional data warehouse technology offered through big data.

All in all, travel and transportation providers can embrace big data and analytics to more accurately model and optimise demand, capacity, schedules, pricing, customer sentiment, revenue, cost and much more. IDM

Four Dimensions of Big Data

Definitions of big data are somewhat vague, but as different as the application areas described above may be, there exists common characteristics which help in describing big data. Four Vs are often used to characterise different aspects of big data:

1) Volume: Data anytime, anywhere, by anyone and anything

Volume may be the most compelling attraction of big data analytics. Comparing the effectiveness of a treatment on a population-wide base or considering thousands of factors yields far better results than would the same analysis for a dataset of only 100 patients.

How 'big' exactly the data has to be to be qualified as 'big data' is not specified. It is estimated that 90% of the data in the world today was created in the last two years, with machines and humans both contributing to the data growth.

An example of CERN (European Organization for Nuclear Research) has demonstrated that volume can present an immediate challenge to conventional resources, and that volume calls for scalable storage and capacity for distributed processing and querying.

2) Velocity: Every millisecond counts

The speed of decision making – the time taken from data input to decision output – is a critical factor in a big data discussion. Emerging technologies are capable of processing vast volumes of data in real or near real time, increasing the flexibility with which organisations can respond to changes in the market, shifting customer preferences or evidence of fraud. Big data systems also need to be capable of handling and linking data flows entering at different frequencies. Long championed by high-frequency traders in the financial services market, the race for velocity and tight feedback loops is a key part of gaining competitive advantage in a number of industries.

3) Variety: The reality of data is messy

Big data includes any type and structure of data – text, sensor data, call records, maps, audio, image, video, click streams, log files and more. Source data can be diverse, and it may require time and effort to shape it into a form fit for processing and analysis. The capacity of a system to analyse a variety of source data is crucial as it can yield insights not achievable by consulting one type of data in isolation.

4) Veracity: Data in doubt

How accurate or certain is the data upon which we intend to build crucial decisions? Is some data (e.g. sensor data) more trustworthy than other data (e.g. social media data such as twitter)?

Influenced by the three previous Vs, big data tends to hold a lot of uncertainty attributed to data inconsistency, incompleteness, ambiguities and latency. Of course, the level of uncertainty and imprecision may vary, but it must be factored in. Poor data quality constitutes a cost factor. A system therefore needs capabilities to distinguish, evaluate, weigh or rank different datasets in order to maintain veracity.

Table 1: Summary of four big data characteristics

Characteristic	Description	Attributes	Drivers
Volume	The amount of data generated or data intensity that must be ingested, analyzed and managed to make decisions based on complete data analysis.	- Exabyte, zettabyte, yottabyte, etc.	- Increase in data sources - Higher resolution sensors - Scalable infrastructure
Velocity	How fast data is being produced and changed and the speed at which data is transformed into insight.	- Batch - Near real-time - Real-time - Streams - Rapid feedback loop	- Improved throughput connectivity - Competitive advantage - Precomputed information
Variety	The degree of diversity of data from sources both inside and outside an organization.	- Degree of structure - Complexity	- Mobile - Social media - Video - Genomics - M2M / IoT
Veracity	The quality and provenance of data.	- Consistency - Completeness - Integrity - Ambiguity	- Cost - Need of traceability and justification

Adapted from TechAmerica: Demystifying big data

(source: ITU Telecommunication Standardization Bureau)