



**2016 CLM Annual Conference
April 6-8, 2016
Orlando, FL**

Big Data – Big Breakthrough or Big Brother?

I. The Emerging Phenomenon of Big Data

Understanding the Background and Definition of Big Data

With the development of technology and analytics needs, the amount of data in the world has been growing at an exponential rate. Every day, more than 2.5 quintillion bytes of data is generated and stored worldwide, and approximately ninety-five percent of the world's data has been created in the last four years. This big data phenomenon is making a significant impact on all industries, including the insurance industry and the legal sector.

The term “big data” was first used in a paper on computer graphics by scientists at NASA in 1997. Today, we can define it as the creation, aggregation, curation, analytics and consumption of extremely large data sets. It requires a set of non-traditional technologies and techniques to deal with data at unprecedented scale. The majority of the data is digital and stored electronically, and the types of big data formats include but are not limited to video, image, audio, text and numbers. For the insurance and legal industries, the most frequently used data types are text and numbers.

Embracing the Opportunities and Benefits of Big Data

Technology trends support the growth of claims analytics using big data. These include easier and more affordable access to data, faster and more powerful computers and the data mining tools becoming smarter and more advanced. All these improved conditions provide a wealth of new opportunities for the insurance industry to leverage much bigger and richer data sets, to more accurately assess risks, to act on better claims insights and to make business operations more efficient.

Big data can be applied in many different areas for the insurance and legal industries, such as risk avoidance, product personalization, cross-selling and up-selling. Especially for the claim and litigation management arena, big data has been proven to be helpful in

fraud detection, claims management, loss prediction and law firm management. When big data is used appropriately, it can provide valuable strategic benefits, become one of the most critical assets and play a useful economic role for many businesses and professionals in this field, and we will elaborate on this topic in Section II.

Evaluating the Challenges and Risks of Big Data

As big data keeps evolving and brings positive effects to the businesses, it also comes with the challenges and risks that we cannot ignore or underestimate.

Many corporations not only lack the appropriate technology infrastructure to handle the data volume, variety and velocity in the big data era, but also need to learn how to align big data technology with their business initiatives and strategic plans to obtain the maximum value from the technology investment. Also, bigger companies frequently face the legacy system and data silo issues, and smaller companies often have insufficient human capital in the data development and analytics arenas. Because of these challenges, there is still a gap between the potential and the realization of big data value.

In the meanwhile, big data can also expose us to inevitable risks, including cyber security threats and the possibilities of over-reliance on data. We need to carefully take these risks into consideration, while moving forward with big data. In Section IV, we will further discuss these risks and their potential solutions.

II. Big Data Analytics in Claim and Litigation Management

Law Firm Management

When it came to selecting defense counsel for a specific case, litigation and claim management professionals used to rely on word of mouth, league tables and third party rankings to make their decisions. Also, it sounds like a “mission impossible” to compare attorney performance across different states. However, big data has changed this landscape, and now clients can mitigate the inefficiency in the lawyer management process and get more effective results.

A Florida company, Premonition, has assembled the world’s largest database of docket data into a structured dataset and innovatively built an artificial intelligence system that uses predictive analytics and mines big data to find out which lawyers win before which judges for what kind of cases in a given venue.

This company along with its big data brings transparency to the legal professionals and allows insurance companies to easily compare the performances of the lawyers, and it can offer valuable advice on the assignment of defense counsel. Using the similar

technology and experience, insurance companies can also evaluate the plaintiff counsel and have a clearer picture about what they are up against. As a result, claims departments can help their insureds achieve better outcomes and control the losses.

Fraud Detection

Fraud detection is always an important task in the claim management area. Claim managers need to identify suspicious cases that can be potential frauds. The process is sometimes judgement-heavy, time-consuming and labor-intensive. Insurance companies can enhance their abilities to handle claims and improve their performance in detecting frauds, by using big data analytical techniques, including pattern analysis, trend monitoring, and data visualization. For example, big data can simply identify if a certain doctor's clinic or a car-repair shop appears many more frequently than normal and can subsequently drill down within the data to find the cause. In addition, after a sudden catastrophe such as a hurricane, the number of the claims may surge in a short period, and big data can protect the insurance companies from frauds by leveraging its automatic algorithms and fast computing speed to analyze the mountains of claims.

Big data also provides claim management professionals other innovative ways to improve fraud risk decision-making. One new trend is to use the data on behaviors from social media or other online channels. Another trend is to link the profiles under review to any profile that has been determined as fraud by their relationship or similarities. It integrates various sources of data and gives claims staff insight into the parameters of a potentially fraudulent case. These kinds of data are not structured or amendable to traditional data analysis techniques, but provide insights into the increasingly sophisticated world of insurance frauds. New tools for data mining can cope with unstructured data and untraditional data sources to provide insights that were not available in the past.

Loss Prediction

The insurance industry collects a tremendous volume of transactional data, customer exposure information, historic loss, claim data and additional data from vendors and public sources. Although predictive modeling has been used by actuaries for two decades, it has many limitations and big data now allows carriers to implement more advanced predictive models and to process bigger real-time data and multi-dimensional data from a variety of sources in a faster speed.

One important application in predictive models is in evaluating the severity of losses. The loss prediction can facilitate catastrophe planning, severity escalation forecasting and claim reserving. When a natural catastrophe is predicted before its occurrence by predictive models, it can help the claim staff to assign and prepare adjusters and thus accelerate response. Then big data can forecast the severity escalation and help the

managers to set more accurate loss reserves, streamlining the whole process of the claim management with greater accuracy in less time.

Customer Satisfaction

According to a 2015 survey conducted by Capgemini, only 29% of customers say they have a positive customer experience with their insurers. A claim is one of the most critical opportunities for an insurance company to lose or retain a customer and there is a significant opportunity to leverage data and analytics to enhance the customer satisfaction when a claim is filed. Insurance companies can apply big data to accelerate the process of handling a claim, to analyze the behaviors and preferences of the customers, to find the customers better lawyers, to tailor and personalize the communications with the customers, and thus to meet the expectations of today's customers.

III. Big Data Structures and Technologies

Structured vs. Unstructured Data

When actuaries build their models or underwriters play with data for trends or patterns, traditionally they are using clean structured data. To get structured data, it is estimated that about 60 to 80 percent of the time in data analytics is actually spent on curating, cleaning, and indexing the data before it can be used for analytics purpose. For example, Advisen Ltd. transforms publicly available loss data and events to a structured dataset with well-developed taxonomy and linkages to other datasets, so that actuarial modeling and data analysis can be smoothly performed.

Structured data are compatible with the systems that insurers currently use and integrate seamlessly with most analytical tools. When they are accessed by programming languages such as SQL or by actuarial packages in statistical software, the speed is usually limited by the volume of the data. The big data technologies will fundamentally make it faster and easier to store, compute and analyze any big structured datasets.

However, the growth of unstructured data is much faster than structured data. Unstructured data includes raw news stories, social media information, internal emails or uncleaned numbers. These sources can be noisy, imperfect and difficult to analyze, but they can be valuable if correctly used by big data technologies. As discussed before, an example of the use of the unstructured data is to identify whether the claim is legitimate, by monitoring social media channels.

To make it feasible to use unstructured data in the big data era, insurance companies need to apply machine-learning and text-mining techniques and find diagnostic signals within the unstructured data. The machine-learning approach allows a system to analyze

hundreds of variables simultaneously and identify the most important variables. Text-mining techniques provide analysts the opportunity to dig out critical information and find out patterns or correlations from batches of text files.

Although unstructured data may someday replace structured data as the predominant resource for analysis, the most effective interim technique for insurers today is to supplement structured data with unstructured data. This will allow professionals to gain experience with new tools and data sources so that the companies can have a relatively smooth transition from the existing analytical methods and system to the new big data era.

Big Data Infrastructures for Insurance

Despite the tremendous advancement of big data technology, it still entails a high level of complexity to design a big data infrastructure and deploy the tools which can support insurance industry-specific business applications. It is believed that only a small percentage of top tier insurers have the right big data infrastructure in place.

Many insurance companies struggle with data silos, separate storage systems and legacy applications that have kept them in analytical gridlock. With big data, this issue may become very expensive, given that the scale of the datasets is significantly larger than before. Therefore, big data requires a single integrated infrastructure that is sufficiently flexible for all processing tasks in the current systems. There are a few options regarding the infrastructural approach of big data, including Hadoop, NoSQL, Massively Parallel Processing (MPP) and the Cloud.

Hadoop is an open-source software framework for storing data and running applications on clusters of commodity hardware. It provides massive storage for any kind of data, enormous processing power and the ability to handle virtually limitless concurrent tasks or jobs. It is the most frequently used infrastructural approach of big data in the insurance industry.

A NoSQL (non-SQL or non-relational) database provides a mechanism for storage and retrieval of data that is modeled in means other than the tabular relations used in relational databases. NoSQL databases are able to process semi-structured data with low latency in near real-time. It also has more focus on performance and scalability than on atomicity and consistency.

MPP (massively parallel processing) is a computer system comprised of many independent processing units that run in parallel. Cloud computing, is Internet-based, where shared resources, data and information are provided to computers and other devices on-demand. Cloud computing is growing in importance as name-brand companies such as Amazon and Microsoft offer competitive services. These two have their own particular capabilities and can be combined with the first two approaches.

IV. The Dark Side of Big Data

Over-Reliance on Data

Big data can provide insights to insurers and their claims operations, by assembling an ocean of data and analyzing it much faster than before. However, it is naïve to believe that “with enough data, the numbers speak for themselves”. Effective application of analytics relies on man-made algorithms, assumptions, and hypotheses. They greatly impact the outputs of big data analytics, such as the recommendations, predictions and simulations.

Also, the underlying data may contain mistakes and lead to worthless or wrong results if the data collecting process or the pre-analysis data cleaning process are not carried out well. For example, it seems to be the trend to use public data from the internet, such as social media websites, to understand insurance buyers better. In this case, sophisticated users can manipulate the data and essentially influence the data accuracy.

Finally, big data is only a means for insurers to gain insights into their business. We should not expect that big data can rule and will make all decisions, even as some functions become more data-driven. Although data can provide substantial information to support the decision-making process, we cannot deny that we also make decisions based on our experiences and judgement every day.

Data Security and Privacy Issues

Every day, the insurance and legal industries deal with confidential, financial and personally identifiable information. As more data breaches occur and more data environments are moved to the shared space within the cloud, the security of the data systems and privacy of datasets are facing challenges that have to be addressed in the near future.

Without appropriate protection, the risk of the losing big datasets may surpass the potential benefits of using big data. The risks include loss of data assets, business interruptions as well as third party liabilities. If a cyber loss does incur, the severity of this loss will be much higher for big data than ordinary datasets.

Therefore, as the insurance and legal industries invest in big data systems and human resources, they also need to invest in the cyber-security technologies and infrastructures and demonstrate and articulate the value of the latter investment to the stakeholders, otherwise big data may bring new problems rather than resolve existing ones.

V. The Future of Big Data

Fostering and Living in a New Culture

Insurance carriers frequently have a risk-averse corporate culture. Big data, as a new approach to problem-solving and decision-making, will not be fully implemented unless a right culture is built and fostered among executives and team leaders. This point has been confirmed by the lesson of failures – in a 2015 survey conducted by Capgemini, the third top reason of a big data project failure is culture-related (“the ineffective coordination of teams across the organization”).

Even if executives and team leaders in insurance or law believe that their firms can profit immensely from the benefits of big data, the individual employees will not reach for it until a corporate-wide culture of big data has been cultivated. Companies need to educate and train their employees about big data, help them understand its potential value and how to harness it, and make sure they can successfully live in and adapt to the new culture.

Exploring New Analytics and Use Cases

Big data does not only indicate more data and information, but will also lead to creative ways to analyze the data and get new insights. As mentioned previously, we currently see the use cases of big data in fraud detection, claims management, loss prediction and law firm management for the claim and litigation management professional. However, big data today is far from its real potential.

McKinsey & Company estimates that the United States market alone needs additional 140,000 to 190,000 people with analytical expertise and 1.5 million managers and analysts with the skills to understand and make decisions based on the analysis of big data. For the insurance and legal industries, we can expect that more use cases and more analytics will be explored, as more analytical and managerial talents fill in this area and more insurance companies enter the big data stadium. To explore new analytics, we also expect to see an increasing role for data scientists, who use the new data sources and tools and conduct analysis creatively, while actuaries apply traditional techniques using structured claim and exposure data.

Facing the Big Data Competition

Historically, the legal sector does not share as heavy a reliance on data as the insurance industry. However, the big data era will not leave the law firms alone. Big data can allow both clients and lawyers to better understand the performances and bring more transparency and insights to the lawyer selection process.

For the insurance industry, 71% of insurance firms rated big data as a maximum priority by 2018, according to a 2014 survey conducted by BearingPoint. The insurance companies already compete for accurate data and information, and now many companies have realized big data will become the focus in the future competition.

Today pioneering firms are using big data to create value, enhance their competitiveness and outperform their competitors by increasing margins. It will gradually create peer pressure for other players to leverage big data to catch up in the competition. Also, there may be increasing pressure to on smaller players as competing insurers become larger and can afford greater investments in technology and staff. As more insurance companies join the big data competition, the analysis of big data will transition from something nice-to-have to become a must-have. Big data and analytics are one way that successful companies will compete in the next decade.