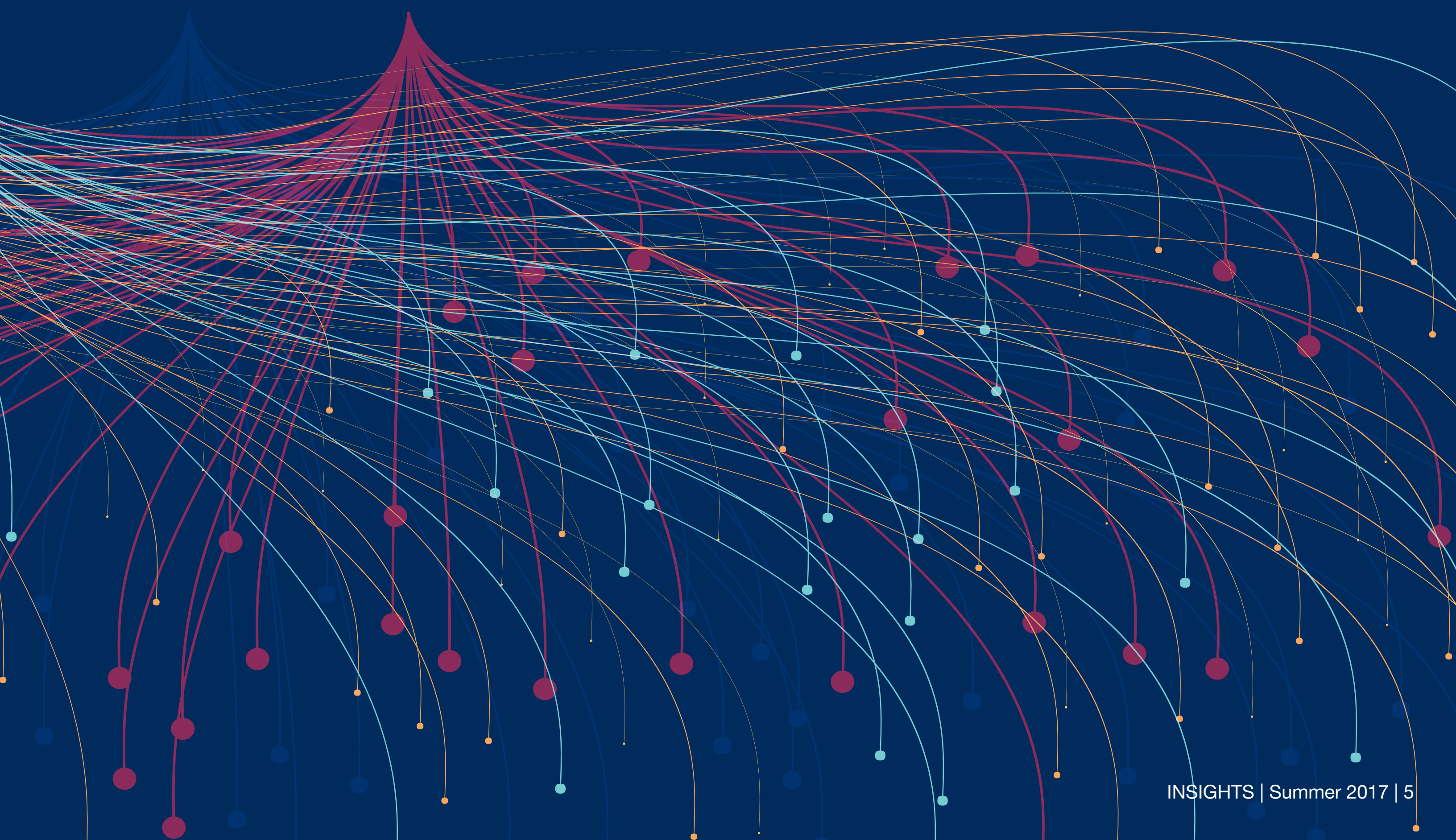
BIG DATA ANALYTICS:
CHANGING THE CALCULUS OF INSURANCE

by Michael W. Elliott

Abstract

Big data, smart technology, advanced analytics, and automation are permanently changing the property-casualty insurance business. Innovative technologies, such as wireless sensor networks and computer vision, are enabling insurers to collect vast amounts of data previously unavailable to them. This, coupled with analytics involving advanced techniques such as machine learning and artificial intelligence, is challenging insurers to transform their organizations into fully digital enterprises in order to increase efficiency, reduce expenses, and remain competitive.



Continued from page 5

Across all industries, reports Accenture, 89 percent of large companies say big data is going to revolutionize business operations, with changes predicted to be on a scale comparable with how the internet changed the way people work in the 1990s.[1] Insurance companies are part of this revolution, which is upending the way they underwrite, handle claims, control losses, develop new products, and service customers.

Big data, as its name implies, involves large sets of data—too large, in fact, to be gathered and analyzed by traditional methods. Although insurers have always gathered and analyzed large amounts of data to make business decisions, the quantity of data available in recent years has increased exponentially because of increased sharing of information and virtually connected objects, through what is called the Internet of Things (IoT).

This increased volume of data challenges insurers to develop new ways to store, access, process, and analyze data. By mining big data for patterns and trends, insurers are able to gain a competitive edge through reduced expenses and improved processes relating to claims, underwriting, and operations.

At the same time, vehicles, buildings, and machines are becoming smarter, through innovative technologies such as wireless sensor networks and computer vision. As a result, many insured objects and workers are safer, causing industry experts to predict that insurance companies will experience significantly lower claims amounts.

Furthermore, technology will foster continuous engagement between insurers and their policyholders, resulting in increased adoption of loss-prevention measures and thereby adding to the downward trend in claims. Given today's technology, it is not hard to imagine your insurance company sending a text to alert you to replace your washing machine's water supply hose or to move your car into the garage because of an imminent hail storm.

> "Many of the underwriting- and claims-related applications being developed by insurers apply machine learning and artificial intelligence to one or more big data analytical techniques"

But it is not only big data and smart technology that are revolutionizing insurance. Recent advancements in analytical techniques, including deep learning, and increasing automation of insurer underwriting and claims processes are also transforming the business. This mix of big data, smart technology, advanced analytics, and automation is permanently changing the calculus of insurance as we know it.

Big Data Analytics

An insurer's big data arises from both internal and external sources and can be structured or unstructured. Insurers' internal underwriting data on losses and exposures is structured because it is organized into databases with rows and columns. An insurer's external data, such as that provided through vehicle telematics, is usually structured in a similar way. By contrast, unstructured data is not organized, with a prime example being text data from claims adjusters' notes.

Analytics is a process that enables insurers to gain deep insight from big data to make effective decisions. Many of the big data analytical techniques employed by insurers are not new, such as exploratory data analysis, which is used to develop a basic understanding of data, and data segmentation, which is used to classify data based on its characteristics. However, automation using these techniques allows insurers to analyze data much more quickly and at more granular levels.

In addition, recent advancements in specific analytical techniques facilitate deep insight into data patterns and trends. A good example is the development of neural networks, which operate in ways similar to the human brain—but more powerfully. Neural networks can simultaneously perform thousands of mathematical calculations on large datasets.[2] Other advanced analytical techniques include text mining, which analyzes words, and social network analysis, which analyzes relationships.

Regardless of the specific technique applied, insight into patterns and trends is greatly enhanced through machine learning, in which computers continuously learn and make decisions based on data. Related to machine learning is artificial intelligence, including cognitive computing, which simulates human thought processes. Many of the underwriting- and claims-related applications being developed by insurers apply machine learning and artificial intelligence to one or more big data analytical techniques.

To achieve organizational goals, some big data analytics applications employ straight-through processing, in which computers make decisions and complete a process without human intervention. A good example is a small auto physical damage claim that, on first notice of loss (FNOL), is automatically scored for potential fraud. If, however, its score indicates that it is a meritorious claim, payment will automatically be made to the claimant. Straight-through processing applies to many other processes, including real-time, online premium quotes for many lines of business.

It should be noted that most big data analytics applications require a human with expertise in the field to analyze model outputs. While these applications serve as a useful tool to help underwriting, claims, or risk management professionals do their jobs, they do not replace the skill and experience of these professionals for most applications.

The Evolution of Big Data Analytics in Insurance

To understand the current influence of big data analytics on insurers, it is important to have a sense of insurers' historical use of data and analytics as well as of recent industry developments.

Data is a fundamental input to insurance, and underwriters have used data and analytics to conduct business ever since the industry started. In the early days, analytics was based on limited recall of events and little claims data. In the twentieth century, industrywide historical claims data was collected and summarized. In the last several decades, actuaries have made great strides using this data for ratemaking with the development of sophisticated analytical models, notably generalized linear models.

The 1990s and early 2000s ushered in what McKinsey & Company refers to as "born through analytics"[3] companies such as Amazon, Facebook, and Google. A distinguishing characteristic of these organizations is that they were built on a digital platform that influenced their structure and the processes they use for key operations, such as pricing, marketing, inventory control, and logistics. As a result, these organizations grew by scaling their data storage and processing systems and by using sophisticated data and analytics for marketing and operations.

Furthermore, customer-service innovations by these data-driven companies have greatly influenced customer expectations in general, including those of insurance consumers. As a result, policyholders often demand quick turnaround on underwriting and claims decisions and twenty-four-hour access to account information through multiple channels, including phones and tablets.

> "...most big data analytics applications require a human with expertise in the field to analyze model outputs"

Throughout the early 2000s, insurers of all sizes experimented with big data analytics projects for specific applications, such as customer segmentation analysis for marketing and link analysis for identifying fraudulent claims. After some early successes, insurers expanded their use of big data analytics and applied it to areas ranging from medical management for workers compensation to reputational risk assessment, using data from social networks. See "Some Common Applications of Big Data Analytics" for more.

To remain competitive and meet increasing customer expectations, insurers are now challenged to go further by transforming their organizations into fully digital enterprises that are largely driven by data for making decisions. Accordingly, many insurers employ large teams of data scientists with high-level math and statistics backgrounds to work alongside marketing, actuarial, claims, and underwriting professionals.

It is imperative that core business users have a basic understanding of the language of data science and predictive modeling so as to ensure good communication among all groups for planning and implementing big data analytics applications. Some large personal lines insurers seem to be the furthest along in these efforts.

Segmenting Big Data to Make Predictions

Data segmentation is central to many big data analytics applications for marketing, underwriting, claims, and risk management. The concept of data segmentation is not new, but its application to big data using machine learning is enabling insurers to discover previously unforeseen patterns and trends, allowing insurers to make better decisions and improve processes.

To make predictions using segmentation, it is important to start with a large sample dataset that has known values for certain

Some Common Applications of Big Data Analytics

Claims	Risk Management
• Complex claims detection • Fraud detection • Return-to-work analysis (workers compensation) • Medical management	• Accident analysis • Reputational risk assessment

Marketing	Underwriting
Customer segmentation	• Risk selection • Pricing

Continued on page 8

Continued from page 7

variables, called explanatory variables, as well as known outcomes representing the event the insurer would like to predict. Examples of explanatory variables from vehicle telematics data are rate of vehicle acceleration and number of left turns, with the outcome being whether the policyholder had an auto accident. Taken by itself, this historical dataset with known accident outcomes is not organized in a way that provides information to predict future accidents.

However, the explanatory variables and known outcomes can be used to train a model that predicts whether a policyholder with certain driving characteristics will have an accident. A computer algorithm applies an iterative process in which various combinations of explanatory variables are sequenced and compared with the outcome data. This process can get quite complex, with the number of explanatory variables sometimes ranging as high as thirty. Then, the results are analyzed to determine how well each variable sequence divides the data into segments that correctly predict outcomes. Data scientists often illustrate the process with a classification tree that shows explanatory variables as nodes, their values as branches linked to other nodes (other explanatory variables), and various outcomes as nodes at the bottom of the tree.

This process of segmenting data using a predictive model is key to many big data analytics applications. When a newly reported claim is identified as complex or when a policyholder is categorized as not renewing a policy, a segmentation process is working behind the scenes to score and categorize the outcome. The segmentation algorithm applied by the model is usually invisible to the business user.

It is important to note that models do not always make correct predictions—in fact, they will often generate a false positive. For example, an injured worker might be categorized as unlikely to return to work based on his or her characteristics and behaviors, when indeed, this is not the case. This underscores the importance of not relying solely on the model and of having someone with expertise in the field review the facts of the case along with the model output. A claims or underwriting professional is a good choice for this work.

> "For many lines of business, insurers need to lessen their reliance on class rating and focus more on pricing risks at individual policy levels"

Using Link Analysis to Discover Hidden Data Relationships

Link analysis (sometimes called association analysis) is another big data analysis technique used for analyzing the associations among various entities to discover patterns that are not obvious by looking at raw data alone. This type of analysis has been successful in analyzing data involving complex relationships among claimants, claims adjusters, and providers—such as auto body shops, home contractors, and doctors—to determine the likelihood of fraud based on patterns of referrals and payments. A computer algorithm typically conducts the link analysis and shows the results in graphical form, known as a link chart.[4]

Data generated through link analysis is often combined with an insurer's other claims data to develop a series of if-then rules to determine whether a newly reported claim is suspicious. However, as with any predictive model, even if a claim is identified as suspicious, it is not necessarily fraudulent. An experienced adjuster should review such claims to determine whether they should be further investigated.

Underwriting With a Risk Pool of One?

For pricing many lines of business, actuaries use classification ratemaking, which bases rates on the average frequency and severity of loss for policyholders within large pools of similar risks. This system has served insurers well for many decades. But with this rating method, how much does an insurer know about the details of a policyholder's risk, including the specific behaviors he or she exhibits that affect the risk level? Does the insurer know how well a policyholder maintains his home appliances? No. Does it know how well a policyholder drives her car and that she avoids congested intersections? No.

Certainly new technologies make it possible for an insurer to collect this risk data, which is unique to each policyholder.

Conventional ratemaking tools are much less precise than the tools that can be developed with modern technology and big data analytics. Connected objects, or the IoT, allow an insurer to collect precise data on a policyholder's risk, and advancements in analytics allow the insurer to quantify the policyholder's characteristics and behaviors and develop a unique premium for that policyholder. In effect, this enables an insurer to place each policyholder in a rating class for which the policyholder is the only risk in the pool.

Industry veterans say that this violates the principle that a well-functioning insurance mechanism is based on risk pooling and the law of large numbers. However, a sea change in automobile ratemaking based on the unique driving characteristics and behaviors of individual policyholders is already taking place through the use of vehicle telematics. Soon there will be additional rating applications based on technological advances that affect other lines of business, most notably homeowners. These changes are reducing the size of risk pools—in some cases,

Continued on page 26

Continued from page 8

down to a level of one. For many lines of business, insurers need to lessen their reliance on class rating and focus more on pricing risks at individual policy levels.

There is ample evidence that traditional analytical methods for risk pricing and selection are being displaced by big data analytics. In addition to vehicle telematics-based rating, there are now by-peril ratings based on big data for some property lines. In the future, the historical exposure and claims data currently used for ratemaking and underwriting many lines of business will have less predictive power than the information provided through big data analytics.

A recent A.M. Best Special Report provides evidence of these changes: "In personal lines, use of data to stratify customers into ever more targeted price groups, and to focus marketing on those groups, is now expected by the market. Companies that have not effectively adopted these technologies find themselves actively selected against." Even commercial lines companies need to pay attention to the influence of data analytics on pricing. That same A.M. Best report states: "Among commercial lines companies, use of data and analytics to determine technical price at a risk level has become critical in maintaining underwriting discipline as conditions have deteriorated."[5]

Preventive Analytics for Loss Control

A relatively new application of big data analytics and technology is preventive analytics, which is an extension of traditional accident analysis. Technological innovations, such as wireless sensor networks and computer vision, allow for continuous monitoring of workers. As a result, when a root cause potentially leading to an accident is triggered, it is immediately dealt with to help prevent the accident.

"In the future, the historical exposure and claims data currently used for ratemaking and underwriting many lines of business will have less predictive power than the information provided through big data analytics"

Sensor-enabled wearable devices, such as smart safety helmets and smart vests, can measure physical and environmental conditions that may lead to accidents. Such conditions include hazardous chemicals, body temperature, hydration, stress, heartbeat, energy output, and surrounding air quality. Motion sensors can detect repetitive motion and alert workers when to change tasks. Sensors in shoes can detect whether a worker is exceeding weight restrictions for lifting objects.[6] The data generated by these smart devices is helping improve root cause analysis and leading to better predictions of worker accident outcomes.

Preventive analytics concepts also apply to physical objects, such as buildings, machines, and vehicles. For example, sensors and analytics are used for preventive maintenance on industrial and commercial buildings, production machines, and commercial vehicle fleets. Sensors monitor prepared food at restaurants to help prevent the spread of food-borne illnesses.

Automating Processes to Generate Cost Savings

It is a well-known fact that the majority of data generated in real time quickly becomes stale, underscoring the need to use it quickly for making decisions. For example, suppose a manager reviews claims once every other day to determine which ones are likely to increase in size and, therefore, should be assigned to an experienced adjuster. An automated process would improve this process by using machine learning and advanced analytical techniques to score claims using a rules-based engine that immediately assigns the claims most likely to increase in size to experienced adjusters. This would occur right after the insurer receives FNOL, providing it with the opportunity to generate substantial savings by quickly deploying appropriate resources for the claim.

To fully capture the potential of big data analytics, a predictive model needs to be continuously applied to a process. A good example is medical management for workers compensation claims whereby data such as utilization review, case management, and pharmacy benefits is collected and assessed as a claim progresses.[7] Machine learning uses this data to continuously score each claim throughout its life and deploy appropriate resources, such as a nurse case manager, at various stages of the claim. Furthermore, the system continuously learns by analyzing past recommendations and new data to optimize the organization's overall claims process from a cost-benefit standpoint.

Automated processes have vast potential to reduce insurer expenses across most lines of business. For example, insurers report substantial savings by reducing the number of motor vehicle reports ordered for personal auto insurance based on analyses of applicant profiles. As another example, automated analysis of vehicle telematics data allows an insurer to determine whether an auto physical damage claim is a total loss immediately upon receiving notice of an auto accident, giving it the opportunity to send the damaged auto directly from the accident scene to a salvage yard rather than to an auto body shop, saving substantial storage and towing charges.

Building a Data-Savvy Team

Big data analytics has the potential to redefine how insurance companies are structured, breaking down traditional barriers between departments like IT, actuarial, data science, claims, and underwriting. A critical component to creating successful and lasting change is for frontline employees in departments like claims and underwriting to develop a data-driven mindset and for data scientists and data managers to better understand the workings of the insurance business.

This data-driven approach to operations must flow seamlessly from data scientists and data managers to every level of the organization, from the C-suite down. That starts with building the right team to oversee the development and adoption of data analytics and predictive modeling processes throughout the company. Finding talent to run big data analytics projects is a huge challenge to implementing big data techniques at many insurers, and integrating big data into existing systems is a further challenge.

Different organizations build these teams differently. Early data science efforts were siloed at most companies, with a few core data analysts sifting through information and getting the preliminary analytical tools ready for more widespread use. Other organizations embed data scientists within operating departments. Both of these structures are efficient in certain aspects, especially early in an insurer's transformation to a digital enterprise.

"Big data analytics has the potential to redefine how insurance companies are structured, breaking down traditional barriers between departments like IT, actuarial, data science, claims, and underwriting"

But as analytical efforts mature at insurer organizations, it becomes increasingly important to build cross-functional teams where data-driven decision making is the primary focus. These steering committees can help guide an organization's overall big data analytics efforts to prioritize the most significant projects and maximize their value. ■

Endnotes

1. Accenture, "Big Success With Big Data," 2014, www.accenture.com/us-en/_acnmedia/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Industries_14/Accenture-Big-Data-POV.pdf (accessed April 13, 2017).

2. The Institutes, *Big Data Analytics for Risk and Insurance*, (Malvern, Penn.: The Institutes, 2016), p. 3.37.

3. David Court, "Getting Impact From Big Data," McKinsey & Company, January 2015, www.mckinsey.com/business-functions/digital-mckinsey/our-insights/getting-big-impact-from-big-data (accessed April 13, 2017).

4. Roosevelt C. Mosley Jr. and Nick Kucera, "The Use of Analytics for Claim Fraud Detection," SAS Paper 1837-2014, support.sas.com/resources/papers/proceedings14/1837-2014.pdf (accessed April 13, 2017).

5. A.M. Best Company, Inc., *Best's Special Report*, "Profitability Slides, Surplus Growth Slows and Competition Intensifies for Property/Casualty Insurers" (Oldwick, N.J.: A.M. Best, January 26, 2017), www.ambest.com/claimsresource/bestconnect/SpecRptJan26_Mar2017.pdf (accessed April 13, 2017).

6. Michelle Kerr, "Putting Wearables to Work," *Risk & Insurance*, June 1, 2015, http://riskandinsurance.com/putting-wearables-to-work/ (accessed April 13, 2017).

7. Jim Harris, "Improving Workers Comp Through Data," *Risk Management*, March 1, 2017, p. 12, www.rmmagazine.com/2017/03/01/improving-workers-comp-through-data/ (accessed April 13, 2017).

Michael W. Elliott, CPCU, AIAF, is a senior director of Knowledge Resources for The Institutes. He is responsible for providing subject matter expertise on educational content for The Institutes' various products and services as well as increasing the industry's awareness of The Institutes' content through presentations and published articles. Before joining The Institutes, Elliott worked for Marsh & McLennan Companies.