SOCIETY OF ACTUARIES

TECHNOLOGY SECTION

CompAct

ISSUE 62 • APRIL 2019

Robotic Process Automation: These ARE the droids you're looking for

By Aaron Hartman

Page 6

CompAct

Issue 62 • April 2019

Published twice a year by the Technology Section of the Society of Actuaries.

475 N. Martingale Road, Suite 600
Schaumburg, Ill 60173-2226
Phone: 847.706.3500 Fax: 847.706.3599
www.soa.org

This newsletter is free to section members. Current issues are available on the SOA website (*www.soa.org*).

To join the section, SOA members and non-members can locate a membership form on the Technology Section webpage at *http://www.soa.org/sections/technology/technology-landing/*.

Publication Schedule

Publication Month: October 2019
Articles Due: July 12, 2019

Letter From the Editors

By Hugh Lakshman and Bill Cember

My fellow members of the Society of Actuaries Technology Section, we would like to welcome you to the first edition of *CompAct* for 2019!

First, we want to thank you all for the feedback on the fall edition of *CompAct*. We will use it to continue adding to and improving the newsletter.

We also want to acknowledge the fine work done by our outgoing editor. Ravi Bhagat has completed his editorial term, rounding up articles and working to ensure that our members receive useful content. Ravi is not going far, though, as we also congratulate him on his new role as vice chairperson of the Technology Section and look forward to his leadership this year.

As a reminder, the goal of our newsletter is **to provide a basis for developing technology proficiency through educating our readers and to promote technology discovery through the exploration of innovative and disruptive topics.** To that end, we have assembled articles that cover a diverse range of topics, with technology as a focus.

The goal of our newsletter is to provide a basis for developing technology proficiency through educating our readers.

Lastly, we encourage you all to continue to collaborate with us by submitting articles or topics that interest you or sharing any other feedback you might have. We really appreciate the contributions and feedback. We can be reached at *hugh.lakshman@ibx.com* and *william.cember@prudential.com*.

In this edition of *CompAct*, we have seven articles that cover a diverse range of topics.

"Robotic Process Automation: These ARE the droids you're looking for." In our first article of this issue, Aaron Hartman of FIS discusses ways to use robotic process automation in an actuarial department. With the ever-evolving nature of our regulatory environment, Aaron sees automation as being necessary for organizations to adapt quickly to these changes and offers an interesting take on improving automation.

"Actuaries, Are You Paying Attention? Global Megatrends in Technology." Blockchain, wearables, the internet of things. In our next article, David Alison and Thomas Bart of KPMG talk about these "global megatrends" that are affecting everyone. They then extrapolate these trends to make five predictions about the future of the life insurance industry.

"Deep Learning and Actuarial Experience Analysis." In this article, Kevin Kuo of RStudio, Bob Crompton of Actuarial Resources Corporation and Frankie Logan of KPMG apply machine learning to a practical actuarial problem and compare their results with the standard actuarial techniques currently being used.

"Transform Your Business With Predictive Analytics." Predictive analytics is having an impact on the way we practice as actuaries. We even have a whole exam dedicated to this topic now! In this article, Martin Snow of Atidot talks about some practical and effective ways he has seen predictive analytics used in our industry.

"Spreadsheet Controls Add Risk Resilience." Every actuary uses spreadsheets, but how do you ensure that the data in your spreadsheet is accurate and complete? In our second article, Diane Robinette, CEO of Incisive Software, offers her insights on controls that can help solve some of the common data integrity issues in spreadsheets. This is the first article of a two-part series on this topic.

"Need for Speed: How to optimize models for maximal run efficiency." Who doesn't wish for faster model run times? In this article, Vincent Xuan, Housseine Essaheb and Benjamin Stirewalt of Prudential discuss ways to minimize the challenges that lead to slower run times. They also tackle the question of multiple models versus a consolidated model approach.

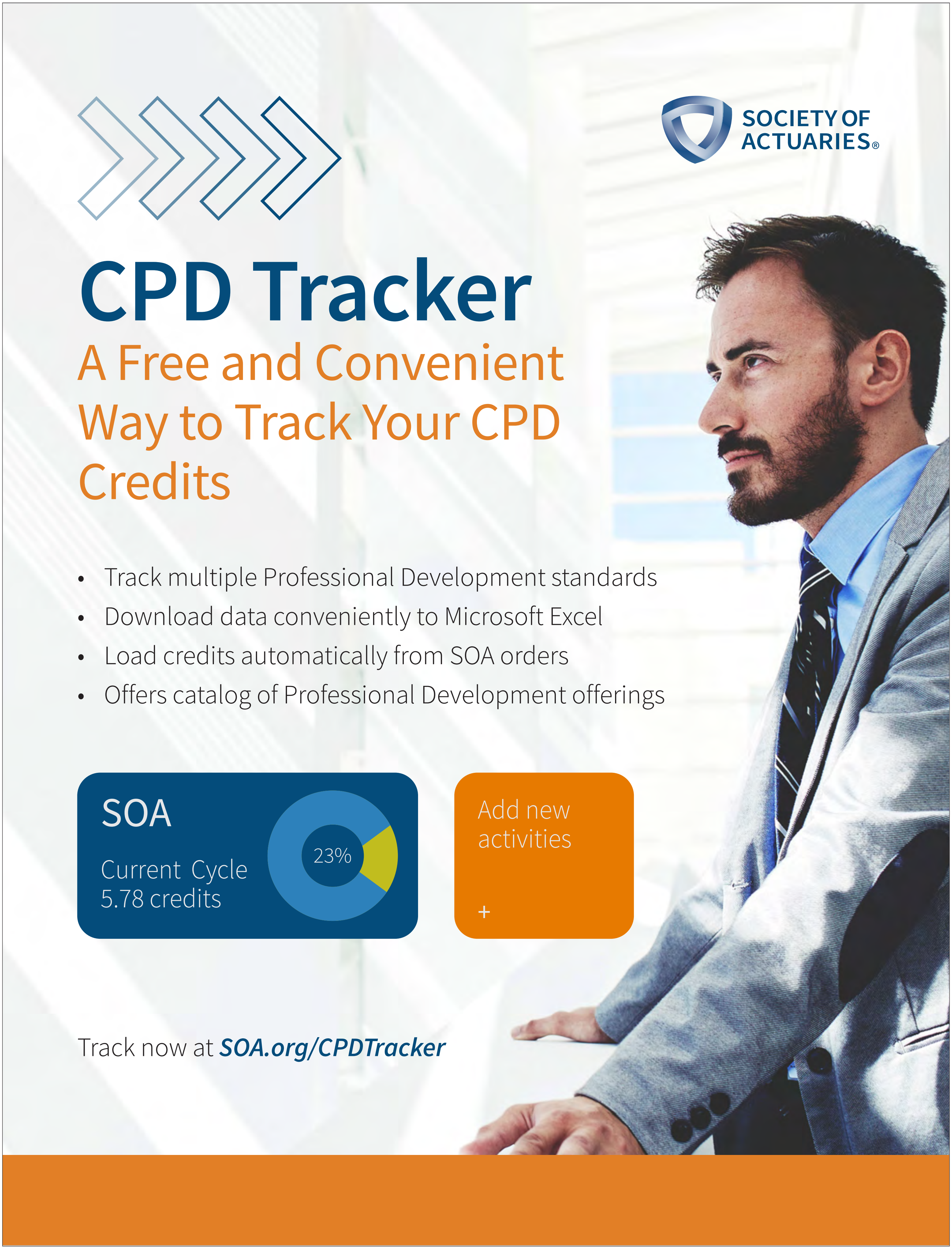
"A Smart Way to Accelerate Model Runs Through In-Force Data Compression." Our final article of this issue begins with a very interesting introduction to cluster analysis. Ramandeep Nagi, Dean Kerr and Xin Yao Li of Oliver Wyman then describe a method to apply cluster analysis to reduce model run time when running computationally intensive models. ■

Bill Cember, FSA, MAAA, is a director and actuary with Prudential Financial. He can be reached at *william.cember@prudential.com.*

Hugh Lakshman, FSA, MAAA, is a director and actuary with Independence Blue Cross. He can be reached at hugh.lakshman@ibx.com.

Special Announcement

"Actuarial Innovation and Technology" is the next Strategic Research Initiative and should launch soon. It will highlight the evolution of technology as it applies to the actuarial profession, industry and population trends. To learn more about Strategic Research Initiatives, click here: *https://www.soa.org/strategic-research/default/?_ga=2.145333556.189334491.1553009745-1918142395.1553009745*


SOCIETY OF ACTUARIES®
CPD Tracker
A Free and Convenient Way to Track Your CPD Credits
• Track multiple Professional Development standards
• Download data conveniently to Microsoft Excel
• Load credits automatically from SOA orders
• Offers catalog of Professional Development offerings
SOA
Current Cycle
5.78 credits
23%
Add new activities
+
Track now at SOA.org/CPDTracker

Robotic Process Automation: These ARE the droids you're looking for

By Aaron Hartman

I'm sure I'm not alone in saying that it seems like for my entire career, actuarial departments have had "improve automation" on their short lists of future goals. Improving automation is, of course, a very reasonable goal for every department. After all, if a computer or machine can do a job just as well or better than a human, why would we want to pay a human to do it?

WHAT IS AUTOMATION?

At its core, automation is the act of programming jobs to be done by computers or machines. These jobs are generally repetitive, manual, rules-based or computational, or they just have low cognitive requirements. These range from assembly lines at automobile plants to complex machines learning algorithms and writing their own code. From an insurance company's viewpoint, the best jobs to automate are repetitive, manual jobs. Computers greatly outmatch humans on these tasks in speed and likelihood of error. Plus, forcing humans to perform these types of tasks lowers focus and inhibits productivity.

Automation has indeed become an integral part of many other departments. Automated voice answering systems have improved call center efficiency, automated claims filing systems have greatly enhanced the claims process, and even underwriting is beginning the first stages of being upgraded to a more automated system. Actuarial departments, however, have been slow in the push toward automation. This is largely due to the complex nature of actuarial work and the relative inability of automation software to perform more complex tasks. As automation software continues to rapidly evolve, the marginal value of implementing that software starts to become more apparent. The opportunity cost to not finding jobs to automate is getting steeper. Having actuaries—whose time is not cheap—focus on automatable jobs takes valuable time away from the work they do that provides more value to the company.

It is at this point in the article where one might begin to ask, "What exactly does automation mean in the context of an actuarial department?" My answer to that question is a type of automation called robotic process automation. A search for "robotic process automation" on Investopedia yields the following definition: "Robotic process automation (RPA) refers to software that can be easily programmed to do basic tasks across applications just as human workers do." This isn't so different from writing a macro in Excel to move data across different tabs or even different workbooks. A major difference between RPA and an Excel macro is that RPA generally exists outside of any one program and is used to coordinate and interact with and across all systems. These automation tools can be integrated easily within an existing framework without impacting the other applications. Because of the ease of integration, companies are observing relatively low upfront investments and low break-even years with RPA.

WHY DOES AUTOMATION MATTER TO ACTUARIES?

OK, so now that RPA has been introduced, it only makes sense to explain the reason to finally incorporate them into actuarial departments. Current and upcoming regulations have made the quarterly close cycle more complicated than ever. The transition from rules-based, formulaic accounting has begun. Whether it be principle-based reserves, long-duration targeted improvements for GAAP, IFRS 17 or another regulatory update, the actuarial close cycle now requires more arduous computing from the actuaries than it previously did. Even with these complex changes, the length of time until close has not increased along with these updates. Actuaries now have less margin for error to close their books every quarter. RPA programs can seamlessly create inputs, kick off runs and format results all at the click of a button. These programs can also alert the user if an input is not available, a process has been stalled for too long or results are outside reasonability parameters. Instead of staying up until 3 a.m. (and risk sleeping under your desk at work that night) just to make sure a model has run and to click some buttons to kick off a new run, actuaries should be looking to employ RPA for these processes.

There have been major regulatory changes before, though, right? Why do these changes make automation more necessary than previous regulatory changes? Good question! These regulatory changes are a catalyst to introduce the next wave of automation to actuarial departments because they are fundamentally changing some major processes. The new GAAP standard, for instance, introduces a new data management challenge, where output from the previous period's run must be used as an input for the current model run. This regulatory change may cause companies to drastically update their valuation systems. Once these systems are in place, actuarial departments can add automation to the process, which will streamline it further. A hot topic in the industry now is end-to-end automation—having a tool automate the entire manual model run process from inputs to outputs—which will allow the actuary to click a button, let the machine run and have results to analyze when the run is finished. Tools exist now that can accomplish these automation goals.

The new GAAP standard is just one example of a regulatory change, but many more major changes are coming. We all know about principle-based reserves and the challenges that brings. Company-specific assumptions can now be selected. The deterministic and stochastic runs must be run now in addition to the formulaic NPR floor calculation. The process is more complicated and simply takes more time to complete. The VM21 withdrawal delay cohort method—a modeling approach that splits an annuity contract into several copies called "cohorts" and models them as separate contracts—will require massive overhauls in inputs to create the cohorts and outputs to manage and integrate results. RPA applications are here to help with these issues. Since RPA has a lower cost of implementation, it makes sense to include it as part of larger valuation-system updates caused by these new regulatory mandates.

WHAT ARE THE CHALLENGES WITH AUTOMATION?

RPA is not perfect and will not be able to be introduced to the industry overnight. Management may have a difficult time becoming comfortable with robots performing all the work for quarterly reporting. To counter this hesitancy, most RPA programs have the functionality to build in manual checkpoints. These checkpoints have many different functions, including the ability to see interim results and approve, the ability to make sure all preceding processes have completed before continuing, and the ability to simply approve that the inputs are correct in the model before running the model. Another issue many in the industry have brought forward is that less-seasoned actuaries may not understand how the models work without running them. There is the potential for a "black box" scenario, where newer actuaries do not understand the inner workings of the models. However, as automation evolves over time, there will be less need for this skill and newer actuaries will likely be learning more critical skills like data analysis and communication.

Actuaries do not have to fear that they will lose their jobs due to these new automation practices, though. The core value of actuaries—interpreting the models and communicating the results of the models—will remain unchanged with this new technology. As technology evolves, so will actuaries. A focus must be placed on skills that computers will never learn (at least computers not named HAL), like critical thinking and communication to nontechnical audiences. This will allow actuaries to work smarter, not harder.

WHERE CAN ACTUARIES GET STARTED?

Two basic methods for implementing RPA within an actuarial department: Build a homegrown RPA tool or purchase RPA software from a vendor. Naturally, there are positives and negatives to both methods. A big advantage for vendor RPA software is that it is oftentimes specifically designed to work with other software an actuarial department might use. For example, the RPA tool we use for Prophet, Prophet Control Center, is specifically designed to work with Prophet and other applications within the Prophet suite. However, a drawback to vendor RPA software—subsequently an advantage to homegrown software—is the lack of control a department has over the system's elements. If a company has a very specific or unique need, homegrown RPA software built to meet this need may be more appropriate.

Other challenges may arise with building and developing homegrown software. Sure, actuaries generally have coding experience, but they are not experts in automation software development. IT departments may have more specific expertise to build and maintain the RPA software, but that has an opportunity cost associated with it as well. Another factor to consider is the time it would take any department to build a fully functioning RPA tool. Insurance companies are already slow in the automation space, and taking the time to build out a homegrown application from scratch may exacerbate that problem. When it comes to total cost of ownership, some companies may prefer to use already-developed vendor software. Each company's actuarial department will have to decide what it values most and act accordingly.

Whatever the specific choice for each department, one idea remains clear: Time is of the essence. The insurance industry already is slow compared with other industries in automation technology. Insurance companies do not have much time before they get lapped by early adopters of automation. Automation technology newer than RPA is fast approaching, and the metaphorical automation "hill to climb" will get steeper if this step is not taken.

WHAT IS THE FUTURE OF AUTOMATION?

This is not the final step for automation in the actuarial workspace, just the next step. After many companies get on board with RPA, the logical next steps are toward intelligent automation. New techniques like machine learning, predictive analytics and natural language processing exist on the horizon for insurance companies. This technology will help bring to light new analysis methods that humans could not formulate on their own.

Regulatory changes can seem burdensome to actuaries. It's natural for everybody to resist change, especially when the change requires so much work. However, these upcoming regulatory changes represent an opportunity—an opportunity to take the next step with automation and maybe make life easier in the long run for the entire department. After RPA becomes standard among all insurance companies, it's all eyes toward intelligent automation. I don't think "improve automation" will ever come off actuarial departments' short lists for future goals, and there's probably a good reason for that. ■

Aaron Hartman, ASA, is a business solutions consultant with FIS. He can be contacted at *aaron.hartman@fisglobal.com.*

Actuaries, Are You Paying Attention? Global megatrends in technology are disrupting the life insurance industry

By David Alison and Thomas Bart

Technological advances in just the last five years have been incredible. Certain sectors such as online retail and social media, even personal banking, are embracing the turmoil and have positioned themselves as leaders in developing and applying new methodologies.

Perhaps the most tremendous aspect of these advances is the sheer scale of the data created by their use. Many industries have already been successful in putting these huge volumes of data to work in real time by using new data architectures, open-source tools and the cloud.

Although data is at the heart of the life actuarial space, ironically, we have yet to see this sector adopt these kinds of technological advances. Other insurance sectors, however, are pushing forward. For example, InsurTech groups are working on solutions to a wide variety of problems, including:

- Automating claims handling and underwriting.
- Testing nonlinear pricing models.
- Testing vast external data sets as a source of signals.
- Scaling up complex calculations using cloud computing.
- Building entirely new technology platforms to handle volumes of data.
- Building chatbots as well as using other complex natural language processing modeling.

At their core, these are massive data-collection exercises. They will need real-time functionality, the cloud, predictive modeling and analytics, and data science. These ongoing innovations show that the life insurance business model is about to change rapidly. To stay in the game, life actuaries had better take note.

GLOBAL MEGATRENDS

As we look at the revolutions in technology occurring across industry borders, we have identified a number of what we call "global megatrends"—technological and societal changes that are expected to affect everyone.

As life insurance actuaries, we can't stay on the sidelines and watch as these changes unfold. Rather, it is our duty to understand these megatrends and prepare ourselves to meet the challenges of harnessing these changes to the benefit of all of our stakeholders.

Wearable Technology

Wearable technology is becoming very popular, with Fitbit-style devices leading the pack. In 2016, 55.2 million fitness trackers were sold worldwide; this figure is expected to increase to 105 million in 2022.[1] Wearables enable users to instantly track and share personal information, such as blood pressure, heart rate, geographic location and sleeping patterns. Active monitoring of these indicators allows individuals to proactively manage their health and time and to take essential and timely measures to live healthier and happier lives. The exponential growth in wearable technology means that **a huge proportion of the insured population not only has access to vital, longitudinal health information, but they are also actively monitoring it themselves.**

The Internet of Things

When your refrigerator reminds you to buy more milk, you'll know you've arrived at the internet of things (IoT). IoT essentially involves connecting everyday items—be it your watch, your toaster or your car—to the internet, allowing everything we interact with to be online and monitoring us. In 2020, it is expected that there will be 24 billion IoT devices (according to *BI Intelligence*), and by 2021, IoT is expected to be an industry worth $1.4 trillion.[2] The primary use here is that connected devices will be able to share and receive data in real time.

Trusted Digital Identities

In the past, you managed your own identity and proved who you were. Now, your online trusted identity vouches for you. Rather than carrying around a driver's license with your photo, your personal information can be stored in an efficient, secure and transparent IT ecosystem. Estonia is leading the way with a transnational digital identity, and e-residents (just like citizens) receive government-issued digital ID and full access to Estonia's public e-services.

You can grant entities access to your digital identity, and various groups—such as medical offices, grocery stores, banks, etc.—would be able to use your digital identity from anywhere in the world with minimal cost and hassle. A secure digital identity allows for the interaction of previously unrelated data sources, thus encouraging deeper holistic understanding of individual behavior and consumer demand.

Blockchain

While this buzzword was popular when Bitcoin was nearing its peak in 2017, the technology itself has massive possibilities for insurance. The practice of using a collaborative approach to validate and store records and transactions has incredible potential for improving connectivity and linking data sources. **By using blockchain technology to integrate health care, financial and other behavioral records, we can construct a "digital health wallet" that can contain an individual's health, financial and demographic information in one secure location.** Blockchain's growth doesn't show any signs of slowing either. In the first half of 2018, blockchain investment in the U.S. exceeded the total investment seen in 2017, according to a recent analysis from KPMG.[3]

Healthy Populations Longevity

It's amazing how much life expectancy has increased in the past 100 years as well as our understanding of the drivers of longevity. A far cry from cataloging headstone records in cemeteries, **actuaries today have access to vast data sets that can contribute to predicting life expectancies in a much more precise manner.** This is especially important due to the aging global population and the impact this cohort has on insurance

> Without aligning to the wider technological trends, historic insurance industry approaches will struggle ...

contracts' profitability. Ultimately, we are getting better at more precisely predicting longevity as we continue to learn more about key drivers of life expectancy, like wealth, physical activity levels and quality of social interactions.

Open-Source Technology

Open-source technology is gaining widespread acceptance in the corporate world, although most actuaries still prefer Microsoft Office to Apache OpenOffice. In fact, some recent acquisitions by some of the software stalwarts (GitHub to Microsoft and Red Hat to IBM) are demonstrating that open source is being adopted by even the most conservative software vendors. Decentralizing software development will enable life insurance **technology needs to be met by combining community knowledge with agile and cost-effective solutions.**

Cloud Computing

Rather than being dependent on your local machine, access to the cloud's vast array of servers that can handle computing power dynamically will **greatly increase the calculations data analysts can perform.** This is the key component to managing, storing and analyzing the exponentially growing volumes of data generated every instant by all the new sensors tracking every imaginable trackable event and phenomenon possible.

Deep Neural Nets

As statistical models traditionally have revolved around regression analysis, neural nets provide a way to find better fits to data with fewer restrictions around initial hypotheses. According to Carlos Meléndez, co-founder and chief operating officer of artificial intelligence and software engineering services company Wovenware, **"The ability of a computer to learn by just analyzing data without having to let the algorithm know what variables are important is unprecedented.** This form of unsupervised learning is drastically changing the role of technology."[4] Taking a life insurance example, actuaries will typically come in with a preconceived idea of what the key factors are when determining mortality (e.g., age, gender, smoking status). This frames our investigations and creates bias in how we consider and group the data. Deep neural nets have the power of being able to analyze all available data and their interdependencies using complex methods/algorithms without this bias, which results in more accuracy.

These megatrends are coming to all industries. Actuaries must take them seriously and begin to think about their impact on the life insurance industry. To help stir some excitement and consideration, we have extrapolated our megatrends through a life insurance lens to generate some predictions on how the industry may develop in the future.

LIFE INSURANCE—SOME PREDICTIONS

Prediction 1: Real-Time Life Insurance

Wearables and improved data processing power will enable more nimble data analysis and monitoring, which in turn will spark demand for a more dynamic life insurance business model. Based on real-time monitoring data, insurance companies can adjust their portfolio-wide reserves dynamically and can, therefore, adjust premiums for customers on a pay-as-you-go wellness score.

Active customer monitoring of customer health and behavior will result in finely tuned longevity predictions and insurance pricing. Could we see best-estimate valuations being derived in real time?

An early iteration of these new models in life insurance can be found in "health nudges," which use hyperconnected wearable technologies to understand policyholder behavior and more closely track longevity risk factors. The primary implication here is that both reserves and premiums can be dynamically adjusted for customers based on wellness scores, which can vary depending on behavior, circumstance and environment.

We note that consumer attitudes toward wearable tech and the interaction with life insurance are generally favorable. According to a GenRe survey, almost 60 percent of people surveyed were willing to allow an insurance company to track wearable data in return for lower premiums.[5] Furthermore, life insurers can offer rewards and improve awareness of threats to longevity goals, aligning our interests with those of the customer.

The major implication of this is that the amount of required data to be collected and processed to support such a business model is **immense**. This data will be necessary to recommend timely health advice, and companies must be able to manage high-volume streams of data to execute their pricing and reserving models. Insurers must ask themselves if their current data architecture is prepared for this surge in data.

Prediction 2: Customer Interaction

Meeting customer expectations will be much more demanding. Many studies have shown that millennials prefer personalized service, which historically has been prohibitively expensive for insurers to provide. Compared with previous generations, millennials are more willing to share their personal data with brands to receive better and more personalized service. They expect you to know all about them whenever they reach out, and the technology will be available to support this expectation based on the trends we are seeing and their expected impact.

The connectivity opportunities of the future will allow longevity professionals to provide health advice to customers, generating a converged customer-centric relationship. Life insurance companies know the drivers behind longevity and can guide customers on ways to achieve their unique life longevity goals. With aligned interests and individualized advice to help them achieve health goals, increased trust between the customer and the life insurer should develop. Life insurance can transition from its traditional role of risk prediction to risk mitigation. By encouraging risk-mitigating behavior for customers and with deep, longitudinal data available to monitor success, life insurance professionals can have much greater influence on the frequency of risks occurring rather than the passive role they currently take in monitoring and prediction.

Prediction 3: Partnership/Adjacencies

Life insurers need to pursue partnerships with data gatherers, distributors and owners. The scope of data availability is changing, and actuaries are no longer necessarily constrained to the policy data they own. More and more avenues are being explored by insurers to try to identify drivers of policyholder behavior outside of traditional underwriting data gathering. The vast amount of data will be too much for life insurers to handle alone.

Many of the new data sources and techniques that will be essential for the future of the life insurance industry will be developed in fields that are totally unrelated to life insurance today. For instance, blockchain was first implemented in cryptocurrency in order to track transactions. Due to the unlimited possibility of new technology sources, in many cases, life insurance companies will be best served by partnering with these new technology firms while they are maturing as opposed to attempting to create tech incubators and develop new technologies themselves.

Life insurers will need to consider who will gather data, who will own the data and who will distribute the data. The increased model complexity to manage multiple personalized risk factors for longevity predictions will be difficult enough for life insurers to take on by themselves. Monitoring various indicators such as physical activity levels, social interactions, opioid use and financial wellness will fall outside a typical life insurer's core strengths. Data sharing with partners and individuals will prove to be an effective way to tackle this problem.

Prediction 4: Changing Actuary Skill Sets

Today, either actuaries need to retrain into predictive modeling and analytics engineering or insurers must hire a team that

includes predictive modeling and analytics engineers. Although actuaries already are typically very comfortable around data, these megatrends will require actuaries to become familiar with ways to manage volumes of data that were previously unimaginable.

Predictive analytics are not new concepts to the insurance industry; the shift in the market has not been the techniques but rather the technology that supports it. Actuarial techniques are the backbone of most predictive modeling approaches; therefore, insurers are more advanced in this space than is often appreciated (for example, generalized linear models are a widely used actuarial concept and an example of such a model-fitting technique). However, due to advances in technology, the way we go about making predictive models is changing; modern predictive modeling is the science of finding patterns in your data in an automated manner using sophisticated coding algorithms.

Data management maturity across the board will also need to be reviewed in consideration of data protection, data and model validation, governance, and controls. Regulation in a high-velocity actuarial environment will to be a challenge that actuaries must spearhead, and they must be proactive in guiding rulemaking to ensure continued innovation while putting data security and risk management first.

Can we expect actuarial engineers and actuarial engineering departments in the future, or will data engineers continue to own predictive modeling and analytics management? These answers are unclear, but we are quite confident that actuaries will need to become more comfortable with massive amounts of data. Failure to do so will result in actuaries finding themselves replaced by data scientists, software engineers and individuals whose positions do not even yet have names.

Prediction 5: Automated Underwriting

In order to meet millennial expectations for speed of decision, proactive life insurance underwriting will become the norm. Customers will be able to grant insurers access to their blockchain digital health wallets, eliminating time wasted filling out forms. With ready access to personalized data, individualized needs and risk assessments can be made with minimal intrusion. While it is likely that manual intervention will still be needed to review outliers and observe the 80/20 rule, automated and proactive underwriting will be able to deliver quotes in real time.

Trusted digital IDs and deep neural nets will be indispensable for continuous fraud prevention, all while enabling insurers to have a more detailed and precise understanding of individuals' longevity expectations than ever before.

Finally, the underwriting interface will become digital and support self-service decision-making. We are already seeing increased use of artificial intelligence-powered chatbots combined with direct-to-consumer distribution networks as possible solutions to the en masse individualized attention demanded by our customer base.

CONCLUSIONS

Without aligning to the wider technological trends, historic insurance industry approaches will struggle to keep up with modern expectations. The exponential growth in computing power and hyperconnectivity means that companies can now process vast volumes of disparate data sources to draw more insightful conclusions.

With an aging population and improved knowledge of longevity risk drivers, a need for competitive differentiation makes such insight ever more valuable, and the easier-to-use software is making it even more accessible for both statisticians and business analysts. As large quantities of available data are no longer owned by insurance companies and statistical techniques become more accessible to nonstatisticians, the pressure to keep up with technological advancement increases.

We cannot be sure how global megatrends in technology will affect our industry, but ultimately, we can certainly be sure that tomorrow will be nothing like today. We are seeing evidence of that already in our Fitbit devices, targeted Google advertising and heavy blockchain investment. We as life actuaries need to carefully think about what changes we can expect from these trends and how we can position ourselves to continue to serve as leaders in the life insurance industry. ■

David Alison, FSA, is a director in KPMG's Actuarial and Insurance Risk practice. He can be contacted at *davidalison@kpmg.com.*

Thomas Bart, ASA, CERA, is a senior associate in KPMG's Actuarial and Insurance Risk practice. He can be contacted at *tbart@kpmg.com.*

ENDNOTES

1 Fitbit—Statistics & Facts. statista.com, *https://www.statista.com/topics/2595/fitbit/* (accessed February 28, 2019).

2 Jones, Kim. Amazing Facts about the Internet of Things. common.org, April 12, 2018, *https://www.common.org/internet-of-things/amazing-facts/* (accessed February 28, 2019).

3 The Pulse of Fintech. home.kpmg, July 31, 2018, *https://home.kpmg/content/dam/kpmg/us/pdf/2018/07/pof-1H-18-report.pdf* (accessed February 28, 2019).

4 Meléndez, Carlos. Top Technology Innovations of Last Three Years: Predictive Analytics is High on the List. wovenware.com, May 31, 2018, *https://www.wovenware.com/blog/2018/05/top-technology-innovations-of-last-three-years-predictive-analytics-is-high-on-the-list/#.XEe05FxKg2w* (accessed February 28, 2019).

5 Purewal, Manraj. Appetite for Life Insurance—A Survey of UK Millennials. genre.com, June 2015, *http://www.genre.com/knowledge/publications/ri15-2-purewal-en.html* (accessed February 28, 2019).

Deep Learning and Actuarial Experience Analysis

By Kevin Kuo, Bob Crompton and Frankie Logan

Deep learning is a type of artificial intelligence that has been successfully applied in areas that involve large amounts of data and have nonlinear relationships between the inputs and outputs. Perhaps the two most widely known areas are image recognition and gameplay. Deep learning has not typically been used in areas with small or medium-sized data sets or in areas where there are strong linear relationships between input and output. Deep learning does not usually provide as much added value to these areas as it does to perceptual tasks with data-intensive nonlinearities.

For this reason, deep learning is not typically a candidate for implementation in standard actuarial work. Much of actuarial work involves linear relationships and small or medium-sized data sets. In addition, much of standard actuarial work is based on robust procedures created from decades of experience. This is certainly true of experience analysis.

However, we wanted to see if it was feasible to implement some desktop version of deep learning for experience analysis. Specifically, we were interested in these parameters:

- Accuracy and consistency of deep learning results compared with standard methods.
- Level of effort in implementation and training.
- Ability to apply deep learning to related and ancillary issues arising from experience analysis.

We have applied deep learning to the 2015 Society of Actuaries (SOA) report on the lapse and mortality experience of post-level premium period term plans (SOA Report). We used the data supplied in the SOA Report. This data is grouped rather than granular at the policy level. We applied our deep learning algorithms against this grouped data yet still obtained results that were surprisingly close to the published data.

A FEW WORDS ABOUT DEEP LEARNING

"Deep learning" is the name of a particular type of artificial intelligence. This name should not be understood to mean that it always generates profound insights. Instead, deep learning refers to neural networks with multiple hidden layers as opposed to a single hidden layer. We define deep learning as follows:

> Deep learning is a statistical technique for classifying patterns, based on sample data, using neural networks with multiple layers.

Case Study

To better understand how deep learning can be applied in the insurance industry, we perform a case study by exploring how it can help us in better understanding and predicting lapse experience to improve risk management and customer retention. In particular, we study shock lapse, which is a phenomenon where insurance companies experience a higher lapse rate post-level premium term. With the increased lapse rate, a book of business can become less profitable as inflow of premium decreases and policyholders that stay with the policies are ones that "need" the coverage. The source code for the experiments is open source and available online.[1]

Data

To create a neural net model, we utilize the publicly available data from the *2014 Post Level Term Lapse & Mortality Report* published by SOA. The data comprise in-force and terminated level term policies from the participating companies. Each row of the data represents a policy block with a unique set of characteristics. For a more detailed description of the underlying data, please refer to the SOA Report. We used policy year 2010 to split out the training (policy year < 2010), validation (policy year = 2010) and testing (policy year > 2010) sets to build our model. The training and validation sets are used to fit and assess candidate models for hyperparameter tuning, while the test set is reserved for final validation at the end of the project.

Model

While deep learning is the focus of this paper, we have attempted to recreate the model used in the follow-up RPG's paper *2015 Post Level Term Lapse & Mortality Report*[2] to the best of our abilities and applied additional machine learning techniques to use as benchmarks.

As with any statistical learning model, we need to encode the categorical factors into numeric values (e.g., how do we represent "risk class is preferred nonsmoking"?). For our model, we apply a mix of one-hot encoding and embedding to transform the categorical variables. The structure of the neural net model includes a dense layer after the inputs, and then it splits off into two branches, each with another dense layer for the two outputs.

There is no "standard" on the network architecture, number of layers or number of neurons per layer. Modelers will often pick an initial model structure and run multiple iterations to arrive at an optimal model structure.

To quickly benchmark against traditional machine learning techniques, we apply automated machine learning (AutoML) to the data. AutoML fits multiple machine learning models (including random forests, gradient boosting machines [GBM], elastic net GLM and feedforward neural nets) with various hyperparameter combinations, within a user-specified time constraint, to determine the model with the best performance. For our case study, a GBM model is selected after five minutes of searching.

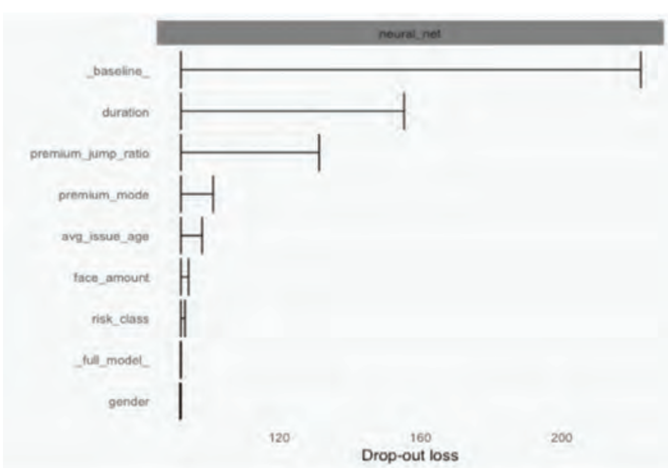
Performance

To measure and compare performance of the models, we use the weighted root mean square error (RMSE) metric applied to actual and predicted lapse rates. The weighted RMSE applies weights to errors of each block on the exposure of that block. The weighted RMSE for each of the models is as follows (lower is better): GLM (0.1722), AutoML (0.1619) and neural net (0.1695). The neural network and AutoML both perform better than GLM. In fact, AutoML performs the best with the least amount of work. We note that since this is an ongoing project, these metrics are calculated using the validation set. As more experiments are performed and we evaluate against the test set, we expect numbers to change. However, at this point, we see that the ML approaches are holding their own against a model built by industry experts.

PEEKING INTO THE BLACK BOX

While the machine learning and deep learning approaches outperform GLM in predictive accuracy, one common objection to implementing ML models in practice is that they are considered

Chart 1
Variable Importances for Neural Network Model



"black box" and impossible to explain. For some use cases, this doesn't matter. For instance, your favorite social media site is concerned more about whether you click on an ad and less about why you do it. The story is, of course, different in regulated industries such as insurance, where transparency is a core requirement.

> Deep learning is a type of artificial intelligence that has been successfully applied in areas that involve large amounts of data.

Even for the same problem, the level of explainability requirements may change with the audience. As an example, for a pricing algorithm, your state regulator may have a higher bar than your underwriting team for transparency. In fact, sometimes they may have completely different definitions for what explainability is.

With the increasing adoption of ML methods in various fields, including "high stakes" applications in medicine and criminal justice, more and more research and software have focused on understanding the behavior of these black-box models.

In our case study, we experiment with a few (out of many possible) model explanation techniques. Some questions we try to answer are "What variables does the model think are important?" "How did the model come up with a particular prediction for the lapse rate?" and "What are the relationships between levels of a categorical predictor?" The plots we show are for the neural network model, although one can also construct them for the GBM and the GLM.

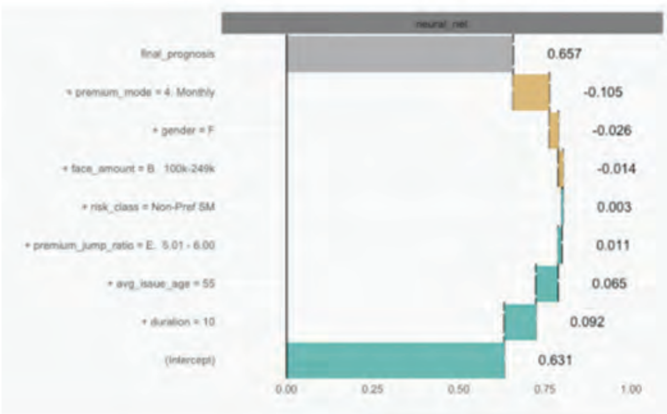
Variable Importance

In a linear model, one straightforward way to obtain predictor importance is to take the estimated coefficients and scale them by standard errors. In complex models such as neural networks, a common way to arrive at such a measure is to permute the values of the predictor of interest (thereby breaking the association with the response variable) and then see how much worse the model performs. In our neural network example, we see that duration and premium jump ratio turn out to be the most important variables, which is as expected. (See Chart 1)

Prediction Breakdown

There are also techniques to "break down" the prediction for a specific data point and approximate the contribution of each

Chart 2
Variable Attribution for a Single Lapse Rate Prediction

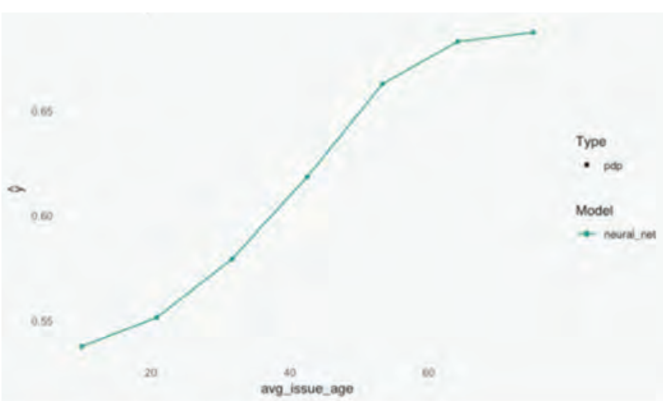


predictor to the predicted response. In this example, we can interpret from the plot that the average prediction for the data set is 63.1 percent lapse rate, but for this particular block, the prediction is higher due to the duration (immediately after the premium shock) and the issue age of 55, and these effects are partially offset by the monthly premium mode, arriving at a prediction of 65.7 percent. (See Chart 2)

Relationship Between a Predictor and the Response

We can construct a partial dependence plot (PDP), which tells us—all else being equal—how a change in one variable affects the response. From the PDP for issue age, we see that the predicted lapse rate tends to increase with issue age, with the effect tapering off at higher issue ages.

Chart 3
Partial Dependence Plot for Model Predicted Lapse Rate and Average Issue Age



All model interpretation techniques are wrong

It's important to keep in mind that model explanations, like the models they attempt to explain, are not exact. Each technique has its pros and cons. As an example, the PDP we show in Chart 3 can fall apart in the presence of highly correlated predictors. Even in the case of linear models like GLM, interpretation can be difficult if the predictors contain nonlinear transforms and interactions, as in the case of the SOA 2015 model.

POTENTIAL FOR DEEP LEARNING IN RELATED AREAS

A couple of insurance areas where deep learning may have some immediate applications are:

- **Data preparation:** This is an area where there has been limited success in automation; the fact that data provided to actuaries is already processed to some extent may make experience data amenable to automated cleansing.

  As just about any practitioner can attest, data preparation is typically the most onerous and time-consuming step in performing experience analysis. Data cleansing is definitely a nonlinear process that requires considerable judgment. Certainly the potential for more efficient and accurate data cleansing makes this a worthwhile area for future investigation.

- **Mortality deterioration:** Deep learning may also provide insight in modeling the extent of mortality deterioration, such as a Dukes-McDonald model.[2] Such a use would be easier to implement than data cleansing but would not provide as much value, since this approach would merely replace any processes that companies currently use for determining mortality deterioration. Nevertheless, deep learning may provide a way to automate these processes.

CONCLUSION

Based on the results we have developed using off-the-shelf deep learning technology, we believe that deep learning is a viable alternative to standard actuarial procedures for experience analysis. In particular, we note that the performance parameters indicate that results using deep learning compare favorably with standard techniques.

> It's important to keep in mind that model explanations, like the models they attempt to explain, are not exact.

In addition, the effort required to implement our model was relatively mild, especially in the feature engineering and selection phase, which was mainly taken care of by the algorithms. In contrast, traditional model building using GLM requires multiple iterations by experts. ■

Kevin Kuo is a software engineer at RStudio and runs Kasa AI, kasa.ai, an initiative for open research in actuarial science. He is based in Seattle and can be reached at *kevinykuo@gmail.com.*

Bob Crompton, FSA, MAAA, is a vice president of Actuarial Resources Corporation of Georgia, located in Alpharetta, Ga.He can be reached at *bob.crompton@arcga.com.*

Frankie Logan is an associate with KPMG's risk consulting practice. He is based in Radnor, Penn. and can be reached at *flogan@kpmg.com.*

ENDNOTES

1 All the analysis is done using R, and the code is open source and available on the GitHub repository, ***https://github.com/kasaai/lapseml***. The neural network is built using the R interface to TensorFlow and Keras, ***https://tensorflow.rstudio.com/***. The AutoML model is implemented using H2O, ***https://cran.r-project.org/web/packages/h2o/index.html***. The model interpretability plots are created using DALEX, ***https://pbiecek.github.io/DALEX/***.

2 Lapse Modeling for the Post-Level Period—A Practical Application of Predictive Modeling. ***soa.org***, ***https://www.soa.org/research-reports/2015/lapse-2015-modeling-post-level/*** (accessed March 1, 2019).

3 1980. *Transactions of Society of Actuaries* 32, 547-584.

Transform Your Business With Predictive Analytics

By Martin Snow

This article first appeared on "Martin's Analytics" *at* https://martinspa.wordpress.com/2019/01/02/transform-your-business-with-predictive-analytics/. *It is reprinted with permission.*

Are you and your peers thinking about predictive analytics and artificial intelligence? You are in good company if you are—as companies, regulators and professional groups—are all paying close attention! For example, we now see "chief analytics officers," an American Academy of Actuaries monograph and the Iowa Insurance Division as a founding partner of the Global Insurance Symposium. Why all the interest? Well, predictive analytics and artificial intelligence are some of the most transformative events in the history of our industry. The revolution has already begun, and predictive analytics and artificial intelligence are reshaping how we do business. Early adopters are poised to achieve major strategic advantages.

Why is predictive analytics important to our industry? The short answer is that predictive analytics can help us (1) improve the dynamics of our business and (2) reduce variability in financial statements. A slightly longer answer is that predictive analytics can enable us to better understand the complex causal relationships that affect the performance of our business. This increased understanding can happen in real time, thereby enabling the exponential strategic advantages that come with real-time influence over the business. In other words, we have the predictive insights in time to act on them. We no longer need to be reactive with our strategy and tactics. Thanks to predictive analytics, we can be proactive!

Before elaborating on this, we insert a word of caution, a quote from the 1970s by the famous statistician George E. P. Box: "All models are wrong, but some are useful." The models used in predictive analytics are no exception. Our aim in this article is to demonstrate the utility of predictive analytics models to our industry and how this utility can be maximized.

We start by showing the value of data with a real-life example. We go back to the 1930s—before the age of computers—so that we can focus on the data itself. One of the major accomplishments of Assistant District Attorney Eunice Hunton Carter was to effectively and efficiently use the data available to her to build a massive prostitution racketeering case against Lucky Luciano, a major organized crime boss. This case was successfully prosecuted, leading to the conviction, imprisonment and deportation of Luciano, the most successful prosecution of organized crime up to that time! Carter was not a data scientist, but she did collect huge amounts of data on prostitution—primarily from people visiting her office—and used the data to clearly demonstrate that organized crime controlled prostitution in New York City and that Luciano was the boss. Data is powerful!

Fast-forward to the 21st century. Most of us have used Amazon to make online purchases. While on the Amazon site, we will be told, "You might also like … ," "Recommended for you … ," or "Customers who bought this item also bought … ." How many times are we amazed that they are recommending exactly the product we need? Clearly, Amazon has effectively mined its data troves and used data science to identify what its customers are likely to need.

Let us look at how our industry compares with those like Amazon that are making effective use of predictive modeling and analytics. For example, I own a life insurance policy from one of the largest life insurers, and automobile and homeowners insurance from one of the largest property and casualty (P&C) writers. My life insurer has never recommended that I buy any product. The closest they have come is to send me a list of all the products they offer and suggest that I spend time discussing my needs with the producer. The P&C insurer (who also sells some life insurance) has done a bit more. Every few years, they recommend that I buy $100,000 of life insurance. But this is hardly personalized! What will it take for our industry to catch up to Amazon?

Now before we go on, we do acknowledge that the health space is using predictive analytics for items such as case finding for medical management programs and the identification of high-cost or high-risk health care patients. Life insurers are also using predictive analytics—at least in certain instances, such as with accelerated (or automated) underwriting and post-level term lapsation and premium setting. However, in other instances, where predictive analytics are used effectively in other industries (e.g., sales & retention), it would appear that many of us are not making as much use of predictive analytics as we could.

Perhaps we draw the conclusion that the experts have looked at it and determined that predictive analytics is unable to help our industry beyond its current uses. Well, when we examine the facts, we see that this is simply not true. For example, suppose

that Company A has poor lapse experience and wants to determine what it can do to improve its persistency. They can call everyone who lapses, find out their issues and try to convince them to reinstate. But, at best, this would be expensive and after the fact. They could call in-force policyholders instead before lapsation happens, but it would be hit or miss on whether they were calling customers at risk; hence, an expensive proposition with dubious results. Worse yet, some policyholders who otherwise would not have lapsed may get the idea from these calls to lapse their policies. Perhaps this partly explains why retention programs are not always given the highest priority by insurers. We are left with the question, "Is there something to do?" Can predictive analytics help Company A improve its retention?

Predictive analytics can supply us with important information that can lead to retention of a policy that otherwise would have lapsed. For example, predictive analytics—without human intervention—has demonstrated that some data, such as the premium payment date (previously thought of by many as important only for administrative purposes), can be significant determinants of lapsation risk. Lower- and middle-income customers who pay their premiums shortly after they receive their paychecks, when they have sufficient funds in their checking accounts, are more likely to keep their policies in force. Those customers whose premium due dates fall a longer time after they receive their paychecks—by which time they may have spent their most recent paycheck—are more likely to lapse. Armed with this knowledge and other discoveries generated by predictive analytics, insurers and producers can know which policyholders to call and when as well as why these customers are at high risk. The machine makes these connections by itself without anyone needing to know them beforehand. Clearly, predictive analytics can be used to improve policyholder retention. As Sir Francis Bacon said in 1597, "Knowledge is power."

Using predictive analytics for retention is particularly powerful, as the model can be extended to related-use cases. For example, once a predictive model is set up to improve retention, it can be further developed to provide more accurate financials with lower variability. Below, we explore how to achieve this by strengthening the assumption-setting process for lapsation.

One big issue in building lapse rate assumptions is the combination of experience from different economic or interest rate environments. For mortality, we routinely combine experience of five years and assume that the year with a particularly harsh winter and a flu epidemic is offset by the year with a particularly mild winter. On the other hand, this is much more difficult for lapses, due to the many different combinations of interest rate environments we can have and the fact that they do not necessarily average out. How do we use predictive analytics to effectively combine lapse experience of periods in close proximity to each other that have different interest rate environments?

One fundamental aspect of predictive analytics is feature engineering. Feature engineering uses domain knowledge of the use case (e.g., the setting of lapse assumptions) to create variables that make the algorithms work. Feature engineering, which can be different for every problem, would tell us to select variables that incorporate the magnitude of interest rates and recent changes in interest rates—and possibly others. Naturally, we would select additional variables that may impact lapsation—based on our knowledge of the business—to include in the model. The model would then identify the appropriate policyholder segments by which to analyze the lapse experience as well as tell us what this experience is for each segment. It goes without saying that we could instruct the model to consider only segments with sufficient credibility.

Based on this, the predictive analytics model would use the selected variables to identify the impact that the interest rate environment (as well as other factors) has on lapses, and the model could effectively identify a base lapse rate vector (or matrix, as the case may be) that is independent of the rate environment. We produce more refined policyholder segments that have been newly identified and are using more data and extended study periods to set credible lapse rate assumptions with lower variability. The lapse assumptions are more accurate than those produced previously, and financial models and results will have lower variability. Clearly, predictive analytics can provide critical support to improve lapse rate assumptions and policyholder retention. Whether this support has strong incremental effects or **exponential** strategic advantages depends on the insurer's implementation. To achieve exponential strategic advantages, the insurer would automate the predictive analytics. The automation would enable expeditious analysis of additional potentially predictive factors that arise from time to time as well as real-time learnings on the impact of behavioral, economic, market and other environmental changes. The insurer can then be proactive—on an ongoing basis—and not reactive in improving policyholder retention and understanding its emerging lapse experience.

Returning to the sales process, let us think about how much valuable information we collect that we do not use. For example, when a policyholder notifies us of a change in address, do we treat it purely as an administrative matter or do we analyze it to see whether the move suggests changed economic or family circumstances and, hence, a need for increased coverage? Do we effectively target our products to customers or prospects who have had life change events? Do we do this in real time? Do we recognize the value in Amazon's "People like you bought …" and its applicability to our industry? Given that many people simply do not buy what a needs analysis says they should buy, perhaps we can start by letting people know how much coverage others in similar circumstances have. This may not solve the entire gap in life insurance coverage, but it is a message that resonates with customers, as Amazon has demonstrated, and it would be a door opener for us to get in and talk to the customers and prospects about their needs. This would be good for business and good for society!

We in the insurance industry have built our businesses by collecting and effectively analyzing huge volumes of data. Let us continue to innovate and use the new tools now available to us. We can effectively revitalize—and, indeed, revolutionize—our businesses using predictive analytics. Exponential strategic advantages are ours for the taking! ■

Martin Snow, FSA, MAAA is vice president & chief delivery officer of Atidot, a predictive analytics firm, and a member of their Advisory Board. He can be reached at *martin@atidot.com*.

Spreadsheet Controls Add Risk Resilience

By Diane Robinette

Data is the foundation of an actuary's success. But what happens when the data, typically located in spreadsheets (from which analysis is based), is inaccurate or incomplete? This article will discuss common causes of data integrity issues relative to spreadsheets. It will also offer insight into controls that can be utilized to help actuaries solve these issues and ensure the accuracy and consistency of the data on which they base their analysis.

DATA IS MIRED IN ARCHAIC, INFLEXIBLE SYSTEMS

Excel spreadsheets have stood the test of time because they continue to meet the analytical needs of actuaries, especially for analyzing and providing evidentiary support for decision-making. For complex calculations where data is continuously changing and for those that require the use of cell functions, Excel is often the go-to tool that actuaries use to get the job done. Yet many of the current techniques and tools used to manage spreadsheet data are either manual processes or homegrown systems. Both approaches are difficult to use and nonresponsive to rapid business change.

For example, rudimentary, last-generation "compliance checkers" often end up as shelfware. Even if compliance checkers are successfully implemented, they cover only a small amount of the total exposure and businesses end up explicitly or implicitly "self-insuring" this ever-growing risk. The exposure, coupled with a restrictive tool set, leaves actuarial teams spending far too much time being risk-reactive, focused on finding and reacting to risk rather than managing and reducing it.

Another issue is outdated data management tools. These archaic tools lack the insight actuaries need to maintain accuracy and compliance and lack the functionality needed to be truly responsive to risk in today's increasingly complex and volatile business environment. Respondents to a recent survey by McKinsey and Company[1] acknowledged that their enterprise had not yet adequately implemented emerging best practices that build their resilience to respond quickly and competently to late-breaking circumstances. Because existing data management systems can't capture the incoming and potentially damaging information, their company has no opportunity to respond when it doesn't know the risks it now faces.

NEW LIFE TO AN AGING TECHNOLOGY

Since their introduction, spreadsheets have provided exceptional insight into data sets, facilitating analysis across each section and category of the document. However, along with an advancing digital age, Excel has been labeled as an aging tool unable to manage increasingly complex data metrics and forms, most of which don't fit neatly into its trusted cells. Not only are traditional spreadsheets not configured to adapt to new data models, but as a stand-alone tool, they are also notoriously fallible.

With the recent addition of automated intelligence (AI) capabilities to Microsoft Office 365, spreadsheets have been given new life. Microsoft has been steadily rolling out new Office 365 AI capabilities—most notably Insights in Excel, which automatically detects and highlights patterns. It analyzes large, complex data within Excel and does so significantly faster than a single human being could. Because it is powered by machine learning, Insights in Excel will provide increasingly advanced analysis as use of the feature grows over time.

While this is exciting news for actuaries, the risk exposure synonymous with Excel remains the same, particularly when building models using AI. Now, however, there is an added risk when relying on Microsoft's models and assumptions—and the assumption that spreadsheets are accurate. Average users will likely click on the Insights button and be thrilled with the new AI-powered charts that Excel now provides, without understanding the possible implications. An equally alarming issue, especially for compliance-driven organizations, is the lack of transparency.

While it's easy to envision the advantages that can be gained using AI, users must proceed with caution. While not all transparency issues can be solved, there are some steps users can take to minimize potential issues—and spreadsheets issues in general. A solid first step is to put controls and technology in place to ensure that the data from which actuaries base their analysis is accurate and complete.

REGAINING CONTROL

Fortunately, advances in technology enable actuaries to overcome many of the aforementioned spreadsheet issues. Automated risk and analysis solutions provide much-needed insight into potential risk and errors that may be hiding in spreadsheets. Yet most organizations don't utilize spreadsheet management solutions simply because they are unaware this technology exists.

Taking a methodical approach to understanding where risks may hide is the first step in managing spreadsheets across an organization. Spreadsheet management solutions offer detailed insight into spreadsheets regardless of where they reside on a network or how many exist. These solutions provide actuaries visibility into who is working on a file, how many people are working on it, when something changes, what changed and who made those changes. Monitoring and tracking this (workflow) information over a period of time provides valuable insight into whether policies are being met. At the same time, it's significantly easier for actuaries to identify potential risks. The ability to document this information enables actuaries to demonstrate that they are following policies and procedures and that they have the right checks and balances in place.

Automation capabilities reduce time-consuming, error-prone manual processes. Spreadsheet management technology features automation capabilities that test for accuracy in both formulas and calculations and seek out documentation that tracks the sheet's functions. They also identify a lack of audit controls, access authority and other critical oversight mechanisms so changes can be made to repair those gaps. With automation, actuaries can count on consistent risk management oversight across all corporate spreadsheets.

ACHIEVING RISK RESILIENCE

Spreadsheet controls add risk resilience, a state in which actuaries are able to quickly iterate processes in a way that boosts flexibility and agility, no matter what is thrown at them. Applying automation gives actuaries speed, scalability and transparency into information. Work is more easily standardized, improving overall performance and alleviating employee fatigue. Out-of-the-box solutions make it easy to adjust practices when requirements and expectations change.

When actuarial teams have the visibility and control they need to manage risks, they are more engaged, productive and valued. By anticipating change-driven needs, actuaries can respond agilely. Using a modernized approach to spreadsheet management and intelligence enables actuarial teams to plant themselves firmly in a risk-resilient posture, which is necessary to succeed and grow.

Want to learn more? In the next issue of *CompAct*, Diane will offer even greater insight into spreadsheets and risk resilience in Part 2 of this two-part series. Stay tuned. ■

Diane Robinette is CEO of Incisive Software. She can be contacted at *diane@incisive.com.*

ENDNOTE

1 How to create an agile organization. mckinsey.com, October 2017, *https://www.mckinsey.com/business-functions/organization/our-insights/how-to-create-an-agile-organization* (accessed March 2, 2019).

Need for Speed: How to optimize models for maximal run efficiency

By Vincent Xuan, Housseine Essaheb and Benjamin Stirewalt

Producing a crystal-clear balance roll-forward on time, obtaining a fresh model output the first time or providing model results for your business partners with a tight turnaround all depend on one thing: how fast the model engine runs.

Why is model run time so important? First, financial reporting is usually not negotiable. Your finance colleagues need to report financial results by a specified date following the quarterly or year-end closing, and they need actuarial analytics to explain the results in a timely manner. Additionally, actuaries need to produce various analytics for senior management's internal management purposes. Capital, surplus and profitability analyses are in regular demand from risk management and treasury partners, and it is critical to be able to provide complex data quickly. Another consideration is that model development requires shorter turnaround time for faster iterations. Finally, long model run time usually means more grid core hours or more cloud computing usage, which results in more modeling and technology staff support. All these factors drive up the computation bill and directly impact the company's bottom line.

Here are some run time challenges and ways to minimize their impact:

- **The size and complexity of the business** is a major contributing factor. The larger the in-force block is, the more policies the model needs to process; therefore, longer run time is required. One potential solution is population compression. For companies with multiple lines of business, modelers need to balance the pros and cons of merging into one model. Within each line of business, companies facing various generations of products can use mapping techniques to avoid modeling each product exactly as described in the product spec, which may significantly drive up the complexity and run time. Modelers can also seek reasonable model simplifications for complex product features.

- **Multiple uses of the same model** can also cause an increase in model run time. In most cases, one model is used for multiple purposes, such as financial reporting, internal financial forecasting, capital management and risk hedging. When modeling multiple reporting bases—including statutory, U.S. GAAP, IFRS and tax—modelers should strive to centralize the common calculation segment as much as appropriate. One possibility is to combine statutory with tax calculations after the recent tax reform. Another run time multiplier is the number of the economic scenarios and assumption sensitivity runs. Model users should be encouraged to trim down the number of scenarios and assumption shocks, especially for ones with muted impact.

- **Model structure inefficiencies** should be regularly examined. The same logic should be programmed and run once instead of multiple times. For example, the liability cash flow generation segment could be calculated once and then shared across different bases for further calculations instead of calculating multiple times for the same outputs. Modelers should consider periodical peer reviews and seek advice from the vendor system on code efficiency. Sometimes an overall run time diagnosis can reveal some unknown run power consumptions.

- **Infrastructure automation and process control** should be considered along with the calculation engine optimization. When redundant manual interventions are involved, it is hard to increase the end-to-end process speed. Try to find ways to eliminate manual feed or handover and instead automate the process. For example, instead of setting up the models manually for different runs, using a batching tool or robotic technology to automate the model runs is highly preferable. Another infrastructure consideration is to optimize parallel run capabilities to improve grid or cloud efficiency. Managing the process control will also help minimize risk instead of creating excessive approval stops.

POPULATION COMPRESSION

There are several approaches for population compression, including randomized selection, clustering and model point creation. These techniques may be used for analytics and model development and testing, even if not for financial reporting.

1. **Random selection:** In this approach, a random subset of the full seriatim in force is selected, and then the calculated results are scaled up. The selection can be randomized in several ways. The simplest way is to sort all the records, for example, by contract number and then select every Kth record. Alternatively, each record can be assigned a random or pseudo-random number between zero and one, and

all records whose numbers are less than K percent can be selected. In either case, results are uniformly scaled up for each record by the compression factor K.

Because the algorithm is rather straightforward, this method is generally easy to implement, in both ad hoc and production settings. The compression ratio is easily controlled by setting the value of the factor K.

This approach is best suited for aggregate results across larger in-force blocks that can take advantage of the law of large numbers. The results may not converge as well for medium-sized blocks or more granular reporting metrics.

2. **Clustering:** In this approach, the contracts are grouped together into clusters based on similar characteristics, such as product type, issue age, gender and moneyness/net amount at risk. A sample is then formed by choosing the best representatives from each cluster. Finally, the results are scaled for each representative based on a measure of the "size" of its parent cluster, which is usually the total account value or benefit face amount.

   Given the complexity of the algorithm, this method can be more complicated to implement in practice. The criteria that define the clusters must be determined, and this generally requires testing several iterations until all criteria are fully specified. As the in-force changes over time, the criteria would also have to be monitored periodically. It can also be harder to achieve a specific compression ratio, as the size of the subset is a function of how strictly one sets the clustering criteria. Several iterations should be tested in order to achieve a desired target ratio. Finally, additional infrastructure components would generally be required to employ this approach in a production setting.

   However, for medium in-force blocks or more granular metrics, the results should converge.

3. **Model points:** This approach begins like the clustering approach, except that once the policy clusters are formed, rather than selecting representatives, all the contracts in each cluster are combined into a single model point and treated as an actual contract. This can be accomplished by adding the seriatim values together within each cluster, and results would not need to be scaled up.

   This method shares the complications of the clustering method previously mentioned. Additionally, it may be harder to trace the integrity of the values comprising the model points.

   As with clustering, this approach should converge even for medium-sized in-force blocks.

ONE VERSUS MULTIPLE MODELS

To consolidate or not to consolidate? This is the question every modeler should ask. Actuarial models are an integral component in the actuarial profession, as they are heavily relied upon for all actuarial work. Actuaries use models for pricing new products, satisfying regulatory requirements for financial reporting and supporting management decisions. The decision whether to use a consolidated model—such as implementing new products, implementing new regulatory requirements or adding new projections capabilities—will depend on its use and actuarial judgment.

> Producing a crystal-clear balance roll-forward on time … or providing model results with a tight turnaround … depends on how fast the model engine runs.

The benefits of having a consolidated model include:

- End-to-end IT infrastructure: A consolidated model will leverage an already fully integrated IT infrastructure, allowing for a shared set of existing input and output facilities. This also allows for centralized aggregation, which uses a common data warehouse to analyze across products and different purposes (e.g., valuation and forecast).

- Implementation efficiency: With a consolidated model, regulatory changes, new reinsurance treaty arrangements or assumption updates need be performed only once rather than duplicating effort into separate models.

- Risk management: Existing modeling controls can be leveraged for additional model implementations rather than creating separate controls specific to different models. For example, a shared business requirement can be used in a consolidated model.

- Cost optimization: Whether the company uses in-house models or vendor software, building upon an existing model reduces costs associated with training, infrastructure support, existing system retrofits and potential fees associated with new vendor models.

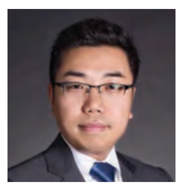
Two examples most actuaries may be familiar with are new product implementation and developing forecasting capabilities.

When a new product line is launched, does a separate stand-alone model need to be created, or should it be consolidated into a main model? How would one go about consolidating? Is it as simple as mapping to a prior product with some tweaks, or should it be built from the bottom up? These are all questions to consider.

For example, if a company decides to roll out a new enhanced death benefit (DB) rider onto a base variable annuity product, there is an opportunity to leverage the original DB rider. Suppose the original DB rider returns the initial deposit to the beneficiary when the insured dies. This type of rider protects in situations where the market depreciates prior to the annuitant's death, resulting in current account value to be lower than the promised DB. The enhanced DB rider resets the death benefit above and beyond the initial deposit, which resets periodically at the highest account value over a certain contractual duration. Given the similarities in the calculation to the original death benefit, the modeler needs to modify only the existing code to accept inputs related to the enhanced death benefit, such as the highest account value, rather than the initial deposit and fees associated with the enhanced DB.

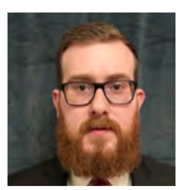
When building a projection model for forecasting capabilities, consider the benefits of consolidating by building off the valuation model versus creating and maintaining a separate model. One key benefit of a consolidated model is the ability to share the same methodology so that the forecast model is always in sync with the valuation financial reporting model. This also enables more sophistication than stand-alone forecast models, which tend to be less complex and use simplification techniques that can potentially cause mismatches between actual and forecast results. The consolidated model will ease the attributions by eliminating model differences.

Renowned computer scientist Dr. Donald Knuth once said, "Premature optimization is the root of all evil." Model consolidation and population compression should be part of all optimization discussions, but before blind pursuit, the pros and cons should be laid out clearly for model users to ensure understanding of limitations and to evaluate the options. ■



Vincent Xuan, FSA, CFA, MAAA, is a vice president and actuary with Prudential Financial. He can be reached at *xu.xuan@prudential.com*.

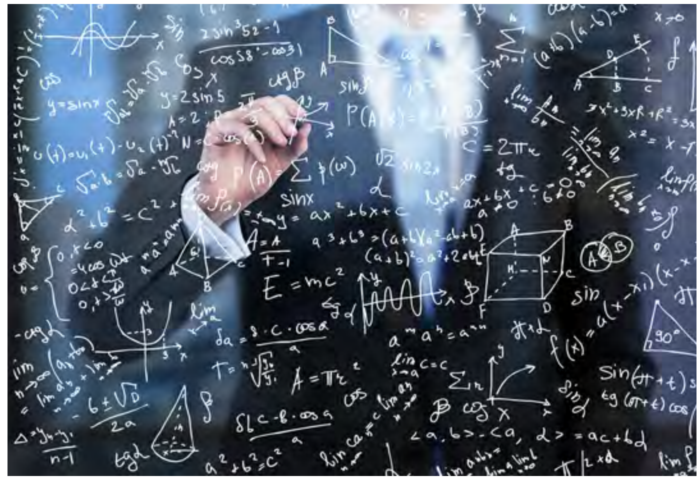
Housseine Essaheb, FSA, CERA, MAAA, is a director and actuary with Prudential Financial. He can be reached at *housseine.essaheb@prudential.com*.



Benjamin Stirewalt is a director and actuary with Prudential Financial. He can be reached at *benjamin.stirewalt@prudential.com*.

A Smart Way to Accelerate Model Runs Through In-force Data Compression

By Ramandeep Nagi, Dean Kerr and Xin Yao Li

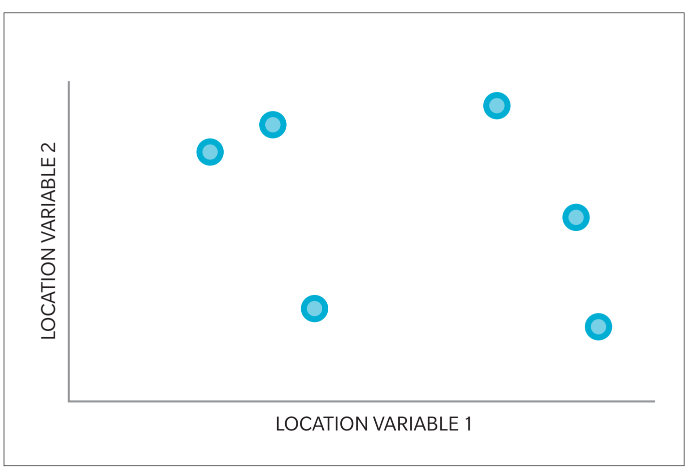


Liability in-force data **compression** is a solution to shorten model runtime by reducing the number of model points. In this article, we will dive into compression approaches, specifically clustering algorithms, and outline how compression can be implemented effectively.

Section 1 provides an overview of **cluster analysis** and describes two common clustering algorithms: K-means and hierarchical agglomerative clustering. Section 2 outlines how to implement a hierarchical agglomerative clustering algorithm. Section 3 illustrates runtime savings achieved by a compression model under different levels of in-force data compression.

Definitions of certain technical terms are provided on page 29; these terms are bolded the first time they are used.

Exhibit 1
Plot of data points based on two location variables



SECTION 1: CLUSTER ANALYSIS

Compression is a type of cluster analysis that groups data points based on a set of characteristics. Clusters can be defined as a group of data points with short **distances** among members or as dense areas in the data space. While clustering algorithms differ in the methodology used to combine data points, all share common properties:

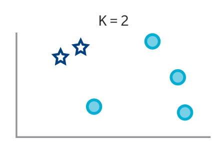
- Clustering is accomplished by setting specific characteristics of data points as **location variables.** (See Exhibit 1)
- The chosen clustering algorithm then iteratively groups data points to optimize a defined objective function.

Clustering Algorithms

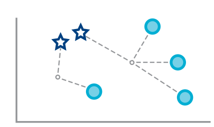
Two common clustering algorithms are K-means and hierarchical agglomerative clustering. (See Exhibit 2)

Exhibit 2: K-means Clustering Algorithm

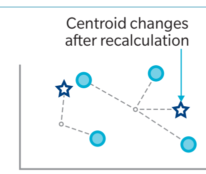
Step 1: Randomly select k data points as **centroids**, where k represents the desired number of clusters.



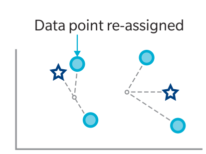
Step 2: Assign every data point to its nearest centroid.



Step 3: Redetermine the centroid of each cluster based on available data points in the cluster.



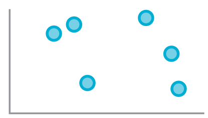
Step 4: Repeat steps 2 and 3 until clusters reach their target state, which is when additional iterations have no impact on the cluster selection.



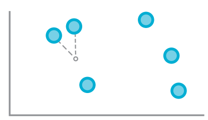
A K-means clustering algorithm is simple to define and illustrate. It partitions the data into a well-distributed set of clusters when k is relatively small. However, this technique can be sensitive to outliers and random initial assignment of the k data points. (See Exhibit 3)

Exhibit 3
Agglomerative Hierarchical Clustering Algorithm

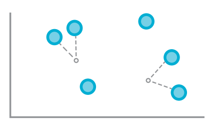
Step 1: Treat every data point as an individual cluster. Calculate the distance between each cluster.



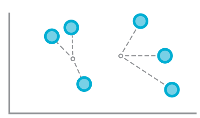
Step 2: Merge the closest pair of clusters.



Step 3: Repeat step 2 until the target clustering level is reached.



Step 4: The result is a set of clusters meeting the target clustering level.



DEFINITIONS

Centroid: The arithmetic mean position of a given set of data points.

Cluster analysis: Data analysis technique that groups data points into clusters.

Compression: Type of cluster analysis technique that compresses large sets of data points into more compact sets.

Compression ratio: Number of data points (e.g., model points) after compression relative to the original number of data points (e.g., seriatim policies).

Distance: Normally the Euclidian distance between two data points in terms of their location variables.

Distortion: Alteration of the original characteristics of the data. As a clustering algorithm executes, distortion is inherently introduced into the data model.

Location variables: Location variables reflect policy characteristics or risk drivers of the underlying policies in the clustering algorithm.

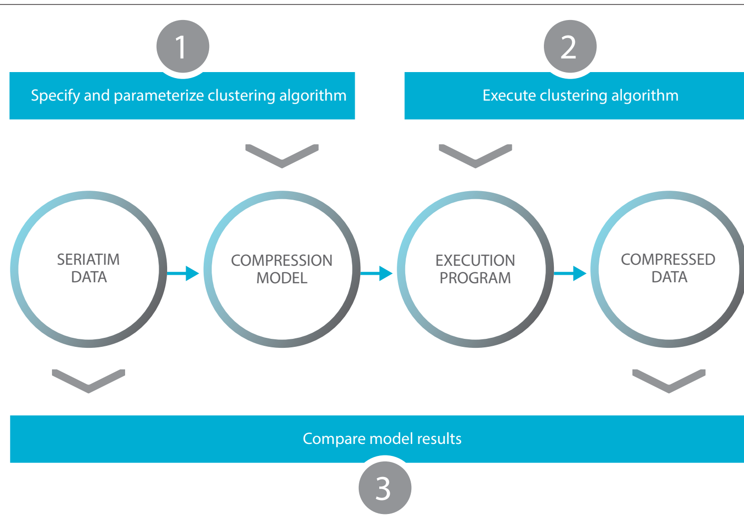
Measure: A metric an actuary attempts to control, or preserve, between the full seriatim and compressed data models (e.g., total reserves).

Weight: Importance assigned to each location variable used to determine the measure metric.

SECTION 2: PERFORMING COMPRESSION

Exhibit 4 outlines key steps involved in compressing in-force data with a hierarchical agglomerative clustering algorithm.

Exhibit 4
Compressing In-force Data

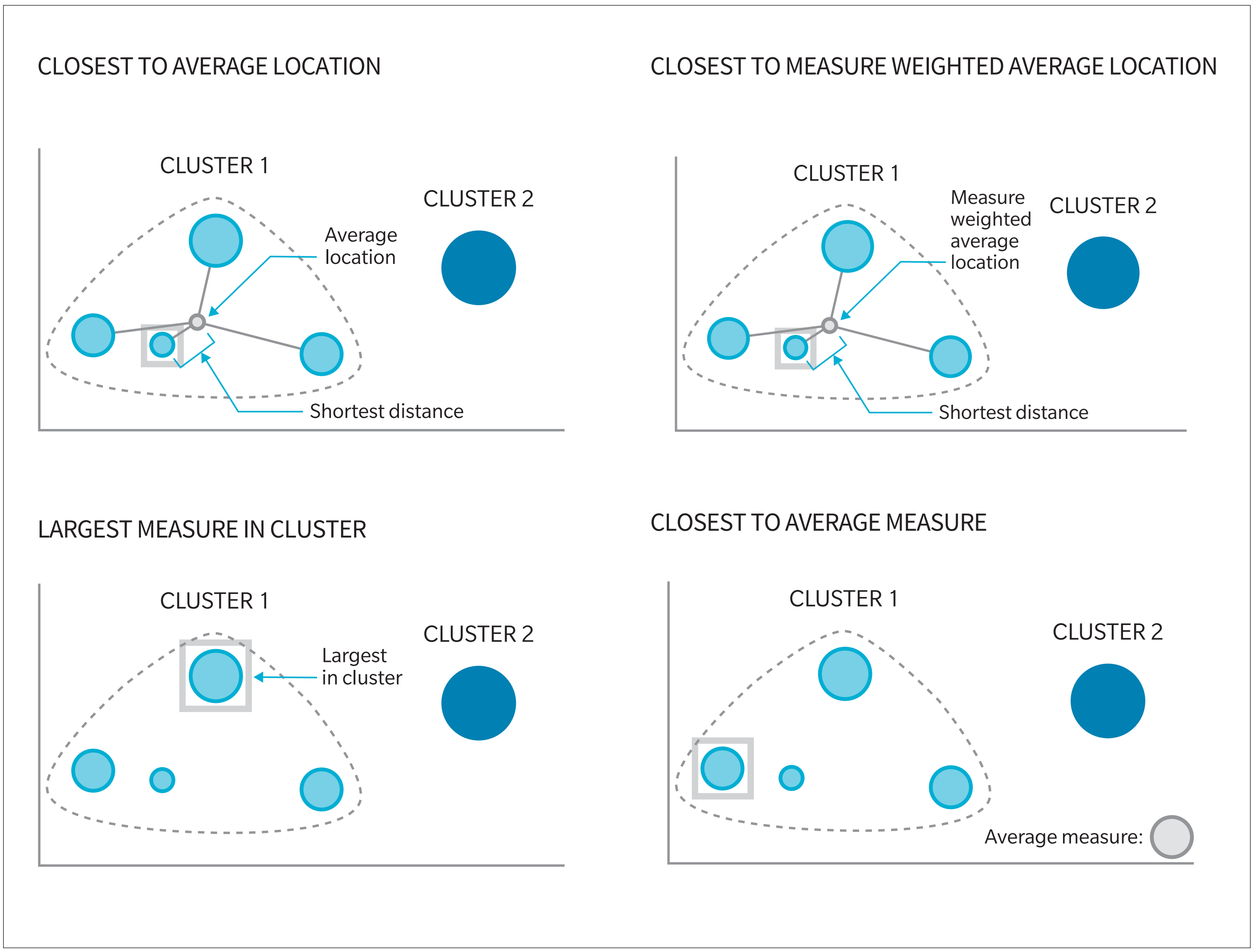


Step 1: Specify and Parameterize Clustering Algorithm
It is important to select a suitable compression algorithm for the problem at hand. K-means has an advantage of being a very fast algorithm but requires predetermination of how many natural clusters exist in the data at the beginning of the process. This information is unknown at the beginning and is generally gained through repetitions of the clustering algorithm. On the other hand, agglomerative hierarchical clustering does not require knowledge of the number of clusters at the beginning of the process but is a much slower algorithm compared to K-means.

The main inputs into the clustering algorithm are full seriatim data, location variables, **weight** of location variables and the **measure**. In addition, data segments can also be defined to achieve better compression results. Segmenting policies (e.g., by major product line, GAAP cohort, gender, etc.) and separately compressing each segment (e.g., different **compression ratios**, location variables, etc.) will generally lead to the best fit of results and decrease the time required to run the clustering algorithm.

Once the clustering algorithm determines which policies are compressed to create a cluster, it becomes important to determine which policy will represent the cluster. This is achieved by creating rules that determine the representative policy for each cluster and its characteristics. A cluster is thus represented by a real policy whose characteristics are already part of the seriatim data. Four possible output linkage rules are shown in Exhibit 5.

Exhibit 5
Examples of Output Linkage Rules



Step 2: Execute Clustering Algorithm

Many ways exist to program and execute a clustering algorithm. In addition to actuarial software, common approaches are to utilize SQL, VBA, R and Python.

Clustering functionality is available in most modern actuarial software platforms. Such functionality can be helpful when compressing model points for inner loop projections. Certain reserving standards (e.g., AG43, VM-20, SOP 03-1) require stochastic calculations. Performing stochastic reserve calculations in an actuarial forecast (often referred to as stochastic on deterministic) significantly increases the computational strain to generate financial results. To overcome this challenge, certain actuarial software platforms offer the functionality to perform reserve revaluations (i.e., inner loop projections) using compressed model points while maintaining the granularity of the main forecast (i.e., outer loop projection) with full seriatim data. This setup improves model runtime proportionally to the compression ratio of the inner loop data model.

A clustering algorithm can also be implemented in SQL, VBA, etc. This may provide additional transparency as a modeler can see the building blocks of the compression algorithm. However, it typically requires programming the clustering algorithms from first principles, which can be time-consuming and may also result in control or efficiency issues.

Finally, due to advancements in data science, clustering algorithms are also available in both R and Python ("scikit-learn" library). The modeler can leverage available libraries for existing code and create modified functions for a range of clustering algorithms.

Step 3: Compare Model Results

The compressed model should be evaluated by comparing model outputs between compressed and seriatim model runs. Experimentation may be necessary to determine optimal parameters: location variables, weights, measure, output linkage rule, segments, and compression ratios.

Careful consideration is required when choosing the location variables. The performance of a compression model depends on how well the location variables represent the underlying policies. For example, for a valuation model, one should choose location variables that drive reserve levels. If policies are not well represented by the location variables and weights, **distortion** will occur even with minimal compression.

Furthermore, once a compression process continues beyond compression ratios supported by the data and attempts to cluster policies that differ more significantly, distortion will increase. This is called over-clustering. As an example, consider the loss of accuracy when attempting to group all policies into a single model point.

Thus, the compression process should involve a tuning phase specific to the intended application. This phase involves selecting location variables and their respective weights based on trial runs and may require several iterations to achieve adequate calibration. However, once a suitable compression model is established, significant efficiency can be achieved without material loss of fidelity in results.

SECTION 3: ILLUSTRATIVE MODEL RESULTS

Compression was performed on an illustrative variable annuity product using a range of compression ratios and compressing on key risk drivers. The following charts show resulting statutory reserves under a range of compression ratios along with the reduced model runtime. (See Exhibit 6)

Exhibit 6
Statutory Reserves Under a Range of Compression Ratios

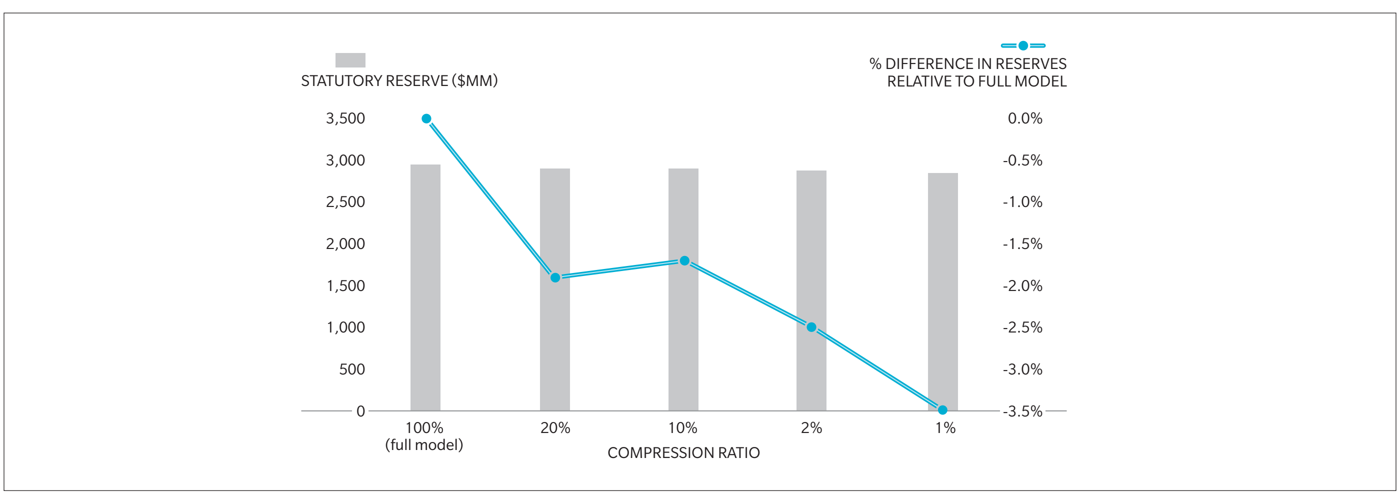


Exhibit 7
Model Runtime Under a Range of Compression Ratios

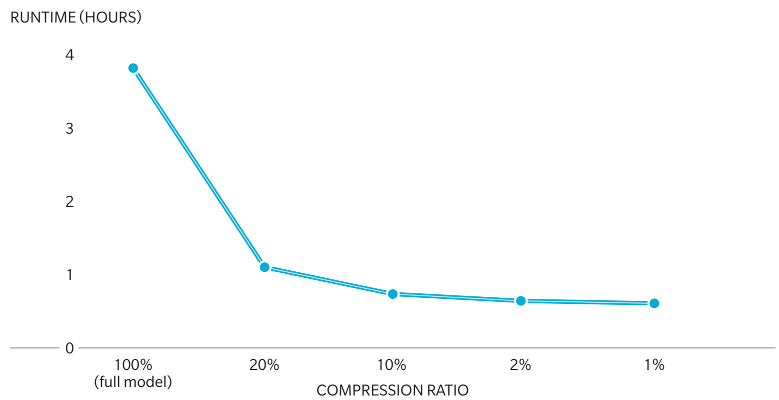


Exhibit 7 illustrates the significant benefit a company may realize by implementing an intelligent clustering algorithm. Valuation (i.e., calculation) runtime is reduced proportionally to the reduction in model points, while calculated reserves deviate by a reasonable margin. Note that overall runtime does not reduce proportionally due to model overhead, such as in-force loading and certain model aggregation and output processes.

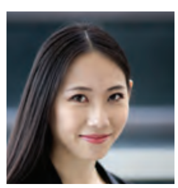
CONCLUSION

In-force data compression provides insurers advanced data clustering techniques and a practical solution to reducing model runtime. For computationally intensive tasks such as stochastic modeling and forecasting, the efficiency achieved by developing a robust compression process could outweigh the loss in model fidelity and upfront development costs. ■

The views or opinions expressed in this article are those of the authors and do not necessarily reflect the official policy or position of Oliver Wyman.

Dean Kerr, FSA, MAAA, ACIA, is a partner at the Actuarial Practice of Oliver Wyman. He can be reached at *dean.kerr@oliverwyman.com.*

Ramandeep Nagi, FSA, MAAA, FCIA, is a senior consultant at the Actuarial Practice of Oliver Wyman. He can be reached at *ramandeep.nagi@oliverwyman.com.*



Xin Yao Li, ASA, is a consultant at the Actuarial Practice of Oliver Wyman. She can be reached at *xinyao.li@oliverwyman.com.*

The Bulletin Board

Updates on important events related to the Technology Section

Announcing the Winners of the 2018 SOA Annual Meeting & Exhibit InsurTech Contest

We are pleased to announce that TCARE and Ostraa are the winners of the 2018 SOA Annual Meeting & Exhibit InsurTech contest. The winners were voted on by all the participants based on originality, impact of InsurTech, InsurTech's reach and maturity stage. A special congratulations to Ali Ahmadi (CEO/co-founder, TCARE) and Amanda Turcotte, FSA (chief product officer, Ostraa), on delivering impressive presentations at the InsurTech Innovation Networking Event and sharing their vision of how technology innovations can be designed to bring efficiency from the current insurance industry model.

We also want to send a big thank-you to everyone who participated in our InsurTech contest and the InsurTech Innovation Networking Event and helped make it a success! ■

SOCIETY OF ACTUARIES®

475 N. Martingale Road, Suite 600
Schaumburg, Illinois 60173
p: 847.706.3500 f: 847.706.3599
w: www.soa.org