SOCIETY OF ACTUARIES
PREDICTIVE ANALYTICS AND FUTURISM SECTION

Predictive Analytics and Futurism

ISSUE 12 • DECEMBER 2015

SOA Launches Predictive Analytics Initiative

By Ian Duncan

Predictive Analytics and Futurism

Issue Number 12 • DECEMBER 2015

Published by The Predictive Analytics and Futurism Section Council of the Society of Actuaries

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

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 Forecasting and Futurism Section Web page at www.soa.org/forecasting-futurism.

From the Editor: Introducing the Predictive Analytics and Futurism (PAF) Newsletter

By Dave Snell

Welcome to your new section! The former Forecasting & Futurism (F&F) section has rebranded and this is our first issue under the new name. We are excited that this name better reflects the work and interests of the various members and is less confusing to SOA members who are trying to choose which sections to join.

How did we arrive at this new name? Why, we did it through a Delphi study, of course!

The section council and friends participated in a three-round Delphi study and narrowed a field of 15 proposed names (including Innovative Tools and Techniques, The Forecasting Section, Alternative Forecasting Methods, etc.) down to a phrase that concisely, yet clearly, describes us: The Predictive Analytics and Futurism Section; or in the SOA's official TLA (three letter acronym) form, PAF.

The three rounds saw much debate, and many passionate arguments supporting the various choices; but a couple dominant themes emerged:

1. Predictive Analytics is a more popular and more comprehensive term for the many quantitative techniques we use such as predictive modeling, generalized linear models, neural networks, genetic algorithms, hidden Markov models, deep learning, fuzzy logic, k-nearest neighbors analysis, singular value decomposition, agent based modeling, and many other algorithms and methodologies that use various analytic approaches to predict outcomes.
2. Futurism is much less numbers oriented; but it recognizes the importance of the qualitative predictive techniques such as Delphi studies, behavioral economics, the actuarial speculative fiction contest, judgmental forecasting, tapping into the wisdom of crowds, and historical perspectives, among other non-numeric methodologies. A recent title of one of our sessions at the Health meeting and the Annual meeting this year was "Behavioral Economics: the Reason Strictly Analytic Models Fail."

In this issue, as in previous ones, we offer you a robust mix of both the technical and the non-technical: the yin and yang of PAF.

Starting us off is our Chairperson's article from Brian Holland, titled "In Good Company." Brian makes the point that the section has become a community beyond the actuarial exams. He also explains that this community building effort is a major strength of the section. It allows us to serve the membership and the profession by facilitating continuing education and forming partnerships with other specialists, such as data scientists and even mathematical oncologists.

Next, Ian Duncan, our outgoing Board Partner, explains in "SOA Launches Predictive Analytics Initiative" how our new name and our continuing focus fits into the SOA initiative to "move Predictive Analytics (PA) front and center for actuaries of the future." This high profile (and funded) campaign is a big plus for our section. The SOA wants to reach out and promote actuaries for predictive analytics opportunities beyond the traditional insurance company roles that actuaries have had in the past. As Ian says about

> "How did we arrive at this new name? Why, we did it through a Delphi study, of course!"

this enhanced marketing by the SOA in 2016, "expect more attention and opportunities for PAF section members."

Some of us longer-term section members remember that we had a name change back in 2009 when we rebranded the Futurism section to emphasize quantitative actuarial forecasting methods. Ben Wolzenski explains one reason why Futurism has remained an important part of our name through the years. A major futurism method we employ is the Delphi method. Ben's article, "Back to the Futurism," summarizes the section's various Delphi studies, from 1989 through the present, and notes that one of them even made a *Wall Street Journal* bulletin item on the front page!

One of the Delphi studies Ben mentions is the one that Steve Easson led 10 years ago: "*A Study of the Use of the Delphi.*" Steve describes the study and its Trend Impact Analysis (TIA) method which was used in conjunction with the Delphi method to derive the quantitative results. You can read about this in his article, "2005 Delphi Study - Reflections 10 Years Later."

Looking forward now, Mary Pat Campbell gives us a head start on the predictive analytics part of our new name. Her article, "Get-

ting Started in Predictive Analytics: Books and Courses," gives us a cornucopia of courses (mostly free) and books (some free) to give a jump-start to your PAF education. You do not need to be an expert in the field to benefit from most of these resources. Many are for the actuary with a casual interest in becoming more PAF literate; and some claim to be for the absolute beginner. Mary Pat provides her perspectives on each of them for us.

Once you get through the basics-to-intermediate coursework described by Mary Pat, you may be ready for the challenge of a certification program in data science, and Shea Parkes shares his experience going through one of the most highly respected online data science programs—the one from Johns Hopkins. His article, "Johns Hopkins Data Science Specialization courses: A review" is from the perspective of an actuary already proficient in data science. He is an official "Kaggle Master" on the site *www.kaggle.com* where data scientists compete on a world-wide basis. Shea relates that he still felt that the nine courses and the capstone project were useful for him, and he writes, "We ultimately deemed it useful enough to make available to all of our staff alongside the actuarial exams and other credentialing opportunities."

Continuing our introduction to predictive modeling, Bryon Robidoux has written an insurance application for us. Read his "Introduction to Predictive Modeling of Fund Manager Behavior for Variable Annuities Riders" to see how an actuary charged with hedging variable annuities can build predictive models that address both short-term and long-term goals of the fund. In essence, he explains the considerations involved to "relieve the tug of war between the fund basis and fund modeling lines … and it improves the accuracy of the Greeks." Bryon's article forms a good primer for any actuary who must work with variable annuities. He explains how to help find the better available information sources, and incorporate them into your models.

On a similar theme of making the "best possible decision with all available information," Kurt Wrobel gives us some practical guidelines in "Predictive Analytics: An Alternative Perspective." Kurt takes us through the conditions we need for a useful analysis, such as accurate historical data, a stable underlying system, a danger of increased sophistication and complexity, and a need to avoid bias in your analysis. He presents a seven point plan to "produce better decisions" and he reminds us that this is not necessarily the same as just adding greater technical sophistication.

Ronald Poon Affat gives us a delightful reminder of the art that accompanies the science in our PAF section. In his article "Actuaries – Personal Time Off," Ronald introduces us to a group he has founded called the Artuaries. Nope, that is not a misspelling of actuaries, the Artuaries are painters, photographers and quilters who use their artistic talents to benefit charities such as the Actuarial Foundation. Ronald represents the futurism side of our section and I felt it was appropriate to include this reprint from the Reinsurance Section newsletter as he inspires us to use our personal talents for worthy causes. His many SOA volunteer efforts have resulted in his 2015 SOA award as an Outstanding Volunteer for the SOA. Congratulations to Ronald!

Continuing the theme of volunteering to help others and the profession, Doug Norris, our outgoing Chair leaves the council with an article "How to get involved: Step one, get involved!" It's great advice. Some actuaries feel unsure of how to start giving back to our profession. Doug offers us a dozen ways to begin; and explains that by helping others you invariably help yourself as well. You can "build your brand," learn and practice management skills, hone your presentation and writing skills, and learn to network with like-minded actuaries across the globe. It's a big world; and volunteering through your sections can help you explore it and enjoy it. Of course, "Big," especially in the phrase "Big Data" is a common term now. The media are overwhelming us with quotes from companies boasting of their big data capabilities and expertise. Yet, big data seems to have a big number of definitions and sometimes it is hard to discern what constitutes "big." In my article, "Big Data or Infinite Data?" I question some of the claims to big data; and I try to put "big" into perspective. Many actuaries are intimidated by the term now, just as previous civilizations found the concept of "many" a challenge. Ironically, actuaries have many of the skill sets to work with big data. Often, we just don't realize it. I try to show that some of the breakthroughs in dealing with many, and with infinite, also apply to big.

Dihui Lai and Richard Xu are pretty comfortable with Big Data. They describe a way to process it in their article "Spark: the Next-generation Processing Engine for Big Data." Spark offers a way to get some dramatic speed and scalability advantages over the Map-reduce methodology whenever you are doing iterative data processing and particularly when you want interactive data analysis capability. Find out why "lazy execution" can be a desirable characteristic when dealing with big data. Hopefully, their examples will whet (spark?) your interest in these new and useful techniques and tools.

The next article takes new and useful and applies it to artificial intelligence (AI). Jeff Heaton, in his research toward a Ph.D. in computer science, and for his book series on AI for Humans, summarizes what is essentially the state of the art in artificial neural networks. His article is "The Third Generation of Neural Networks." Yes, I remember that in the 1980s we thought that AI was going to do wonderful things and we were later disappointed with the limitations of expert systems and neural networks. But in recent years, Deep Learning has changed that and as Jeff writes, "It is a very exciting time for neural network research." If you wish to investigate neural networks, be sure to read Jeff's article and skip the mistakes that early researchers made and instead, use the latest published techniques from recent successes.

We finish this issue with an article that exemplifies how the section is advancing our collective knowledge of predictive analytics. Geof Hileman read the July 2015 newsletter article that Shea Parkes and Brad Armstrong wrote on ridge regression; and Geof utilized it in his article "A Comparison of Risk Scoring Recalibration Methods" where he compared ridge regression to full recalibration and to a residual approach. Geof's analysis supported the assertion made by Shea and Brad for populations of moderate size, but not fully credible. As children we are taught that sharing is a good trait; and in PAF we find that it helps us collectively benefit.

We have come a long way in the past six years. Our newsletter reflects the increased section interest in predictive analytics and in predictive non-analytics. I'm usually not one to give sports analogies (Doug Norris excels at this); but a baseball legend, Yogi Berra, died this year and he was known for many Yogi-isms, such as "It's tough to make predictions, especially about the future." One of my favorites was "The future ain't what it used to be." He was right. The section is not what it used to be either. Welcome to the future. Welcome to PAF! ■

Dave Snell, ASA, MAAA, is technology evangelist at RGA Reinsurance Company in Chesterfield, Mo. He can be reached at *dave@ActuariesAndTechnology.com.*

Chairperson's Corner: In Good Company

By Brian Holland

A few years ago the data science revolution captured the public interest and actuaries' attention. It has even inspired excitement, trepidation and self-questioning. Who are these cool new data scientists? Can actuaries be data scientists too? Wait, wait—are data scientists the same as statisticians? Some data scientists also wonder who we are, and whether we know support vector machines from k-means. I've tried to explain what we do and I describe us as:

Actuary = MBA – most case studies + more statistics + financial risk + sector experience.

This self-examination is tough, but at least we're not alone. We're in very good company with other professions. Attorneys are faced with software to perform automated discovery of documents and emails. Physicians are faced with computational tools supplementing or replacing judgment, such as anomaly detection in scans. For example, I stumbled across *Mathematical Oncology 2013*, a book highlighting developments in a field I'd never imagined. Cooperation is clearly essential between specializations.

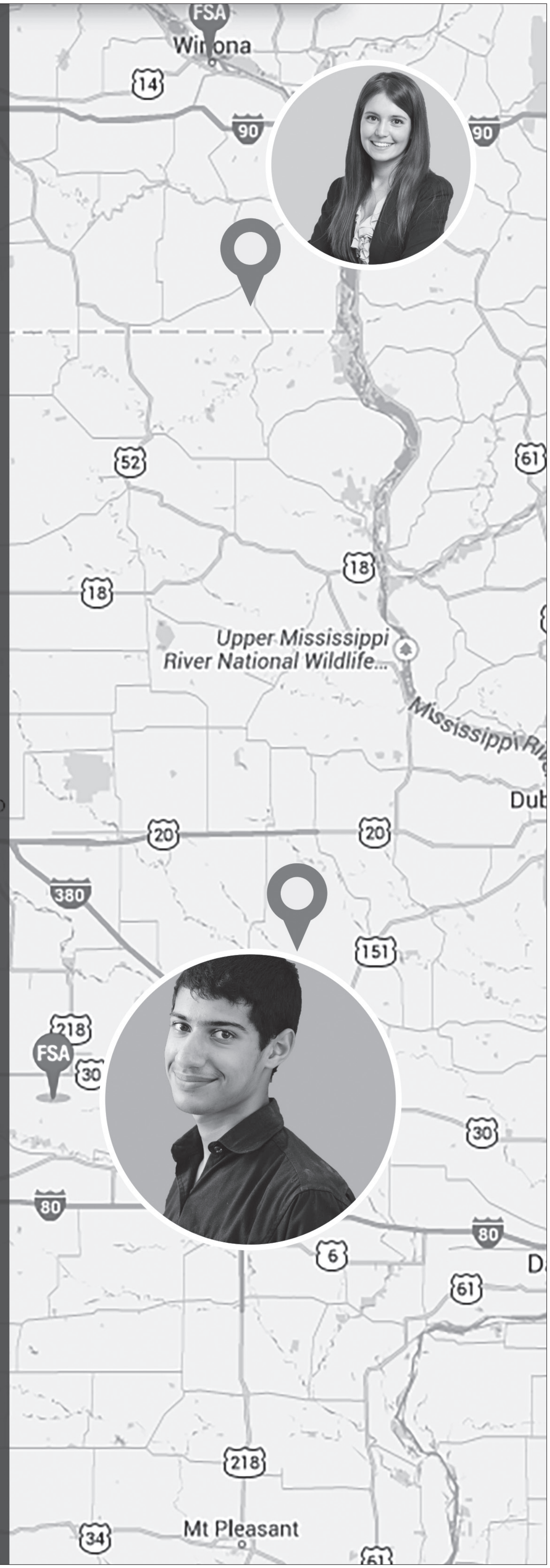
To serve our constituencies, do we focus on continuing education in analytics beyond the fellowship level, or partnering with specialists in data science? The answer must be both. Actuaries have a broad exposure to financial markets, the legal environment, and statistics, but we're not also quants, lawyers or statisticians, apart from a few special individuals. We partner up, as we will with data scientists to get to the end: profitable business in a secure system. As for continuing education, the Predictive Analytics and Futurism Section is here to provide specialized content, application, and community beyond the actuarial exams.

Building a community takes time and effort. Our community has benefitted from the time and efforts of several talented individuals whom I'd like to thank for the excellent position in which they're leaving the council as their terms end: Doug Norris, Dave Snell, and Richard Xu. Doug, our outgoing chairperson, has led us through our own rebranding initiative; provided focus and encouragement for our efforts to create content: the LinkedIn group, SOA meeting sessions, webcasts, our first lunch-and-learn; and generally figured out how time-strapped volunteers can best organize themselves. Dave has presented on webcasts and meeting sessions and edited our newsletter, the main channel for serving members. Thankfully, he will stay on as editor. Richard regularly contributed to the newsletter and SOA sessions and also research. I think right away of the post-level-term study published in 2014. Fortunately for us, they will remain as friends of the council and volunteers, as have past councilors. Ian Duncan, our outgoing SOA board partner, has truly been a partner in keeping a sharp focus and serving our members.

A few words are in order on renaming our section. Our section's efforts have focused for some time on popularizing analytic methods—just review the last several years' newsletters. These methods are applied beyond forecasting. So, with the SOA's support, we've changed our section's name to Predictive Analytics and Futurism to convey our focus when we choose sections. Futurism stays in the name because there are paradigm shifts that wash over predictions and render them obsolete.

I pondered one potential paradigm shift: whether actuaries would even choose assumptions in 10 years. Then I remembered the public emergence of quants in the early '90s. Actuaries had no small amount of interest in these high-flying mathematical risk gurus, with similar questions about sharing the work and expanding our skills. Facing yet another paradigm shift, we will adapt again. We will partner with data scientists and statisticians, some of us will be both, and others will be some of each. The Predictive Analytics and Futurism Section is here when we want to grow and connect. We're in good company with each other in a vibrant, learning and applying community. Please play an active part by attending sessions at SOA meetings, webcasts, joining our LinkedIn forum, and contributing to the newsletter. ■

Brian D. Holland, FSA, MAAA, is director and actuary, of Life and Individual A&H Experience Studies at AIG. He also serves as chair of the Predictive Analytics and Futurism Section Council. He can be reached at *brian.holland@aig.com.*



FSA
Winona
Upper Mississippi
River National Wildlife...
FSA
Mt Pleasant

SOA Launches Predictive Analytics Initiative

By Ian G. Duncan

It is hard to avoid hearing about predictive analytics, big data and data science nowadays. Although this has not been a mainstream focus of actuaries, there are many of us who have been practicing in this area for a number of years, as members of the Predictive Analytics & Futurism Section (PAF) will attest. Recent actions by the SOA leadership and board promise to move Predictive Analytics (PA) front and center for actuaries of the future.

To gauge the prevalence of PA among actuaries, and to learn the type of projects that actuaries are doing, we conducted a survey targeted to members of sections most likely to be involved in PA. The survey showed, among other findings, that:

- Of more than 500 responses, more than half have been involved in predictive analytics over the past 12 months, and of those not involved, 94 percent expressed an interest in learning about predictive analytics;
- Health care is one of largest areas of growing demand from employers followed by life insurers;
- Employer familiarity with actuaries is relatively low compared with other professionals and actuaries appear to be losing relevance as compared to prior research; and
- Actuaries who can combine a study of policyholder behavior, predictive analytics and business intelligence are highly regarded.

A sub-group of the SOA's Cultivate Opportunities Team (COT) has been looking at what is needed to increase actuaries' relevance and recognition in this field. The COT made recommendations to the SOA's board at the June meeting, which were unanimously endorsed, and a sizeable budget has been set aside to promote the initiative among both employers and actuaries. Plans include:

- Identify education/training needs for PA actuaries.

  The survey referenced above evidenced considerable concern that actuaries are inadequately trained in the type of statistics and models that are required for PA. There are many roles in a PA project, however, from data management, through modeling to implementation (where business knowledge and skills are important), so the interested actuary has opportunities that do not necessarily involve advanced statistics and modeling. A workgroup is, however, currently considering (see below) what knowledge and techniques will be required for actuaries in the future.

  The SOA has offered PA continuing education for some time, including the Advanced Business Analytics Seminar. A second seminar specifically aimed at Health Actuaries will be offered beginning in 2016. Continuing education offerings are increasing (the PAF Section being a leader in this regard) and we can expect them to continue to increase in the future.

- Develop a marketing communications campaign to promote actuaries in these roles, target potential employers and inform members of these opportunities.

  A key component of the SOA's PA strategy is marketing the capabilities of actuaries, both to potential clients and employers, as well as to actuaries. Recruiters tell us that there are many opportunities for actuaries in analytical roles—more than there are actuaries qualified to fill them. At the same time, as we move to increase the supply of qualified actuaries we need to ensure that employers are aware of our capabilities when hiring, so that the default action (hire a statistician) becomes a more nuanced decision. The campaign will highlight some of the leading actuaries and their work in the space. Expect to see a number of section members featured as the campaign rolls out!

- Providing educational opportunities for members in predictive modeling:

  - University courses/preliminary exams/fellowship track.

    Many universities teach students the fundamentals of modeling (Time Series, Regression and Generalized Linear Modeling, for example). It is often difficult to fit practical applications of this material into the undergraduate syllabus, given the SOA's and universities' course requirements. Once they graduate and enter actuarial student programs, students frequently do not have opportunities to apply their knowledge of PA, either. Thus the strategy has to be two-pronged: encourage more hands-on modeling at the university level, and encourage more rotations and jobs at actuarial firms that apply these models.

One of the key recommendations of both the COT and the Learning Strategy Task Force (LSTF) was that the Predictive Analytics should be added to the ASA syllabus. A syllabus re-design committee is currently considering this issue, which is not without challenges: PA is not a subject that can be adequately tested in a multiple-choice environment. Simultaneously with the SOA's decision to enhance the syllabus, other actuarial bodies have similar initiatives (e.g., the CAS with its new exam S and the International Actuarial Association's recommendation to include data and predictive modeling on the syllabus for the "qualified actuary" (essentially the SOA's ASA).

The LSTF also made recommendations for the fellowship exams and continuing education. Fellowship exam committees will be encouraged to add practical applications of PA to exam tracks. Actuaries at the fellowship level will not be required to perform the type of modeling that will be expected from associates, but they will be expected to know how the models are applied in practice.

- The SOA has significantly increased its research budget in recent years and this will be directed at projects in PA, particularly in Life.
- Multiple articles: the PAF section is the winner in this regard, by a mile! The recent change in the section's name was due partly to a desire to be aligned with the SOA's strategic direction, and partly to be more transparent to members about the mission of the section. With the enhanced marketing of PA coming in 2016, expect more attention and opportunities for PAF section members.

These are just some of the plans for the Predictive Analytics initiative. If readers have suggestions or questions, please contact me (*Duncan@pstat.ucsb.edu*); Jim Trimble, chair of the Cultivate Opportunities Team (*james.trimble@uconn.edu*); or Courtney Nashan, the SOA staff person who actually does all the work! (*cnashan@soa.org*)

I have been pleased to be involved in this initiative for the past two years, and to have been the board partner for PAF over the past year. It has been a great opportunity to learn about the section and its work, and to move it to the forefront of this new initiative. Section members should be excited by the focus that the SOA is bringing to the initiative—I wish the incoming council every success in the new year, and look forward to staying involved with you as we roll out this important initiative. ■

Ian G. Duncan, FIA, FCIA, FSA, FCA, MAAA, is Adj. Assoc. Professor of Statistics and Applied Probability at University of California - Santa Barbara. He can be reached at *duncan@pstat.ucsb.edu*.

Back to the Futurism

By Ben Wolzenski

The Society of Actuaries once had a Futurism section; that section has seen its name changed twice in the past seven years, but Futurism has remained part of its name, through Forecasting and Futurism to Predictive Analytics and Futurism. A major reason lies in the persistent utility of a technique whose origins date to the middle of the last century, and which rightfully falls under the heading of Futurism: the Delphi method.

The Delphi method (or technique) is an interactive forecasting method that relies on a panel of experts who are provided anonymous feedback to their answers to a series of questions in multiple "rounds." It has been found particularly useful where other, traditional forecasting techniques have struggled. Knowledge about, and assistance in using, the Delphi method has, within the SOA, always resided in the section with Futurism as part of or its entire name. Way back in the 1940s the Army Air Force wanted a forecast of the future effect of technology on warfare. Existing forecasting methods were found to be inadequate. Project RAND was established, which led to the creation of the non-profit RAND Corporation in 1948 and the Delphi method during the 1950s. It has been used for scientific and technology research since the 1960s, and for other fields such as public policy research since the 1970s.

The first recorded use of the Delphi method by the Futurism section was the 1989 Delphi Project, in which members of the Futurism section were asked about "a number of important issues and changes facing our profession and the industry." Participants were asked to opine on what would "most likely be the situation in the year 2000." This led to the development of a "Standard Scenario" for the "Actuarial World of the Year 2000." Among other things, that report predicted that, "Actuaries, still almost exclusively employed by insurance companies and consulting firms, will be faced with more stringent professional standards of conduct … " and that "… all in all, actuarial training will be much the same as in 1989."

The Futurism section sponsored its second Delphi study in 1999 called "Delphi Study 2000." This time there were 25 multiple-choice questions "on varied topics of interest to actuaries" who were "asked to make estimates of future values 10 years from now (2010) and 50 years from now (2050). When the results were published, a media kit was sent to journalists. *The Wall Street Journal* put a business bulletin about it on its front page, and then dozens of articles in the lay and industry press picked up the story. Among the best 2010 predictions were the improvement in mortality and growth of life insurance in force in the U.S.; the worst prediction failed to anticipate the substantial drop in AIDS mortality.

Then in 2005, the Futurism section and the Investment section joined SOA committees to publish "*A Study of the Use of the Delphi Method, A Futures Research Technique for Forecasting Selected U.S. Economic Variables and Determining Rationales for Judgments.*" The selected variables were the annual increase in the Consumer Price Index, 10-Year Treasury spot yields, the S&P 500 total rate of return, and Corporate Baa spot yields. In addition to producing forecasts for the year 2024, the publication contained a list of the events that Delphi panelists thought were mostly likely to influence the outcomes and a blueprint for conducting future Delphi studies.

Another threshold was crossed in 2009 with the publication of "*Blue Ocean Strategies in Technology for Business Acquisition by the Life Insurance Industry*," co-sponsored by the Futurism, Technology and Marketing and Distribution sections. This was the first SOA Delphi study to ask its panel of experts (most of whom were not SOA members) to answer essay questions. After three rounds of questions and responses, the strategies foreseen ranged from the mundane "paperless processing" to the ethereal "virtual world insurance."

Most recently, 2014 saw the publication of "Land this Plane": How the U.S. should deal with the pending LTC crisis, sponsored by the LTC Think Tank and the Forecasting and Futurism section. The objective was no less than producing a consensus about how America should deal with the pending LTC crisis with a comprehensive, integrated solution. Thanks to its members and contacts, the LTC Think Tank recruiting an outstanding panel of experts from the industry, academia, public policy organizations and the SOA. The final report identified six principles for addressing the crisis.

By the time of this newsletter publication, the newly renamed Predictive Analytics and Futurism section will have sponsored a real time Delphi session at the SOA Annual Meeting. Futurism continues to be relevant to actuarial work; our section expects that to continue, into the future. ■

Ben Wolzenski, FSA, MAAA, is managing member at Actuarial Innovation, LLC in St. Louis, Mo. He can be reached at *bwolzenski@rgare.com.*

2005 Delphi Study—Reflections 10 Years Later

By Steven W. Easson

This October marks the 10th anniversary of the release by this Section, called the Futurism Section at the time, of the study titled, "A Study of the Use of the Delphi Method on Economic Variables, A Futures Research Technique For Forecasting Selected U.S. Economic Variables And Determining Rationales for Judgments."[1] As chair of the study's Project Oversight Group (POG), I thought it would be an auspicious time to reflect on this study. I am writing this article to share my experiences and perspectives, mainly with the goal to try to motivate SOA members to further increase their interest and passion in the section's activities.

First a little background. In 1999, I was elected to the Futurism Section chaired by (the late) Mr. Bob Utter. His full-time day job involved futures research and his passion for the field of futurism truly inspired me. I absorbed the 2000 Part 7 Study Note on Applied Futurism by Alan Mills, FSA, and Peter Bishop, Ph.D. with great interest. I also kept up to date on initiatives of the World Futures Society and the "Millennium Project" among others. When I was elected chair of the Futurism Section for 2001-2002, I immediately gathered and read through all of the section's past newsletters since its inception in 1983 (the second section formed, after only the Health Section). By doing so, it became apparent that the section had continuing challenges in finding its mojo. I concluded a solution was application, application, application, à la the real estate mantra of location, location, location. To make a long story short, it took approximately two years to sell the idea, obtain funding,[2] formulate the POG, conduct the RPF process in recruiting our "Principal Investigator," and then approximately 1.5 years to conduct the study.

The Delphi Method, very briefly, is in essence a multi-round controlled debate among experts, preferably as multi-disciplinary as possible. The anonymous feedback of participants' rationales between rounds is the key to success. The goal is to not necessarily derive consensus, rather it is to continue rounds until there is a "stability" of the "fan of plausible scenarios" identified by the participants in the study. This provides management with valuable insights for setting both business and risk management strategies and tactics in the context of multiple plausible scenarios.

From my perspective, the primary goal of the study was to educate SOA members on the Delphi Method through outlining its key characteristics and demonstrating its application to a pertinent topic to actuaries. This in the hope it would motivate actuaries to utilize futures research techniques for many various applications. In my mind the usefulness of the study's results was secondary and within that I viewed the qualitative opinions and rationales for judgments as equally or even more valuable than the quantitative results in many instances.

The study was designed to obtain insights into the rationales and thought processes experts use in making judgments about the long range (20 year) values of four U.S. economic variables: Annual increase in the Consumer Price Index; 10 Year Treasury Spot Yields; S&P 500 Total Rate of Return; Corporate Baa Spot Yields. Round one was sent to the participants in November 2004 and round two in March 2005 (note, at a time of seemingly increasing prosperity).

So, how did we do, both qualitatively and quantitatively? First, two caveats. Upon reflection, I believe the study's results could have been more credible if the forecast period was not as distant as 20 years into the future (i.e., 2024) and if it was not at a point in time.

Rather an average over a period would have been better; perhaps the average over five to 10 years in the future, (i.e., 2009-2014). Secondly, in the interests of brevity for this article, my outline below is necessarily a very small subset of the enormity of the results (as listed in the study's report).

Qualitatively, there were interesting and in some cases very disparate views on many issues, for example among many, the leadership role of the U.S. and its fiscal situation, inflation rates including those for energy, the influence of the Fed to control inflation and avert global recession, currency exchange rates and productivity advances.

The Trend Impact Analysis (TIA) Method was used in conjunction with the Delphi Method to derive the quantitative results. The TIA Method utilized the plausible future developments identified in round one of the study along with round–two–obtained associated probabilities and impacts to produce median estimates and confidence intervals. In essence, the opinions, rationales and judgments from the expert panel were widely separated which led to the wide ranges below. The study's conclusion that the variables are intrinsically (that) uncertain was perhaps not a bad conclusion. Given the recent one–in–many–years or many standard deviation economic and financial results, these wide confidence interval ranges, even at the 80th percentile, do not seem as implausibly wide to me as they did in 2005, and, in fact, in today's environment, the results were perhaps not as extreme as they could have been (e.g., no one foresaw the possibility of negative fixed income yields)!

- CPI: 0.6 percent to 9.9 percent.
- 10–year Treasury Spot Yield: 3.3 percent to 11.4 percent.
- S&P 500 Total Rate of Return: –20.2 percent to 23.1 percent.
- Corporate Baa Spot Yield: 3.8 percent to 14.8 percent.

The study would not have been a success without the efforts of many individuals. At the top of the list was our "Principal Investigator," Mr. Theodore J. Gordon. Mr. Gordon is an acknowledged pioneer in the field with his successes dating back to a Delphi study he co-authored for the RAND Corporation in 1964. The write up on Wiki on the Delphi Method acknowledges him. Throughout the project, Mr. Gordon expressed high enthusiasm and patience with the POG in expanding the scope of his report without a hint of objection to spending more time than targeted without getting paid. Also the POG members (Jack Bragg, Mark Bursinger, Sam Cox, Steve Easson, Doug French, Jack Gibson, John Gould, Phil Heckman, Steve Malerich, Jim Reiskytl, Mark Rowley and Max Rudolph) were highly engaged throughout the project despite its protracted period. Finally, Ronora Stryker and Jan Schuh of the SOA expertly handled the management of the project.

Subsequent to this study's release, there have been a number of successful studies performed by the SOA as follows:

- Blue Ocean: *http://www.soa.org/research/research-projects/life-insurance/research-blue-ocean-strat.aspx*
- Mortality Risk Differentials: *http://www.soa.org/Research/Research-Projects/Life-Insurance/research-ind-mort-risk.aspx*
- Long Term Care: *http://www.soa.org/Research/Research-Projects/Ltc/research-2014-ltp-ltc.aspx*
- Delphi Studies in Pandemic flu research: *http://www.soa.org/research/research-projects/life-insurance/research-impact-pan-influ-life-ins.aspx*

Finally, I would like to set my mind back to 2005 and contemplate the future. My "fan of plausible scenarios" did not foresee the enormous advances the section has made in expanding the scope of futures research techniques and its applications. I have to congratulate all Section Council members over the 10 years for the section's successes. Related, I am surprised the membership of the section has not expanded substantially. Futures research techniques are fascinating and will have increasing relevance, so my current "fan of plausible scenarios" includes this section's membership will be one of the highest among SOA sections over the next few years. ■

ENDNOTES

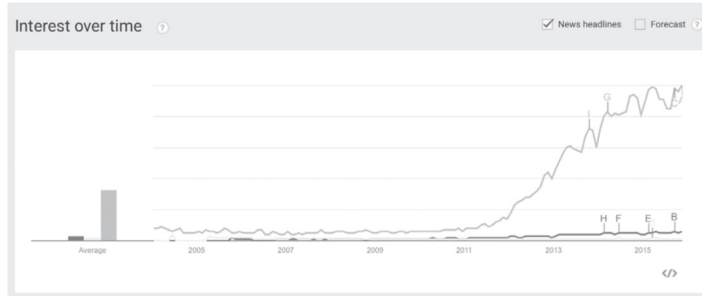
1 The comprehensive (142 page) report on the study can be obtained at:*http://www.soa.org/files/research/projects/delphireport-finalversion.pdf*

2 From the Futurism Section, the Investment Section, and, as they were known at the time, the Committee on Finance Research and the Committee on Knowledge Extension Research.

Steven W. Easson, FSA, FCIA, CFA, is vice president and chief actuary at Canadian Life & Health Insurance Association Inc. in Toronto, Canada. He can be reached at *seasson@clhia.ca.*

Getting Started in Predictive Analytics: Books and Courses

By Mary Pat Campbell



"Data Source: Google Trends (www.google.com/trends)." See https://support.google.com/trends/answer/4365538?hl=en.

Back in September 2009, this section sported a brand new name: the Forecasting & Futurism Section (before it had been the Futurism Section). In the inaugural newsletter that month, introducing the new name, there was also an article introducing Forecasting concepts: "Introduction to Forecasting Methods for Actuaries" by Alan Mills. Alan put together a handy table listing common forecasting approaches in actuarial work, as well as references for those methods.

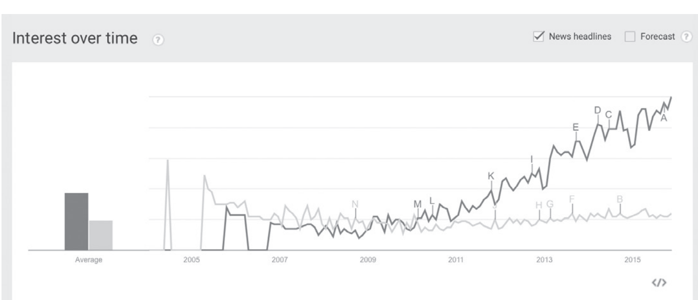
At the time, "Predictive modeling" was relatively new, and he noted it was gaining in popularity.

Here is how Alan described the method:

> "An area of statistical analysis and data mining, that deals with extracting information from data and using it to predict future behavior patterns or other results. A predictive model is made up of a number of predictors, variables that are likely to influence future behavior."[1]

Since that overview article from six years ago, predictive modeling and analytics have taken off—so much so, it's now part of the name for the section!

"Predictive analytics" and "Predictive modeling," have caught on broadly, and in insurance, first being particularly used in property & casualty pricing applications. "Big data" has really risen in popularity as a search term since 2012 … perhaps partly due to the prominence of people like Nate Silver of 538 fame.



"Data Source: Google Trends (*www.google.com/trends*)." See *https://support.google.com/trends/answer/4365538?hl=en.*

Actuaries have the ability to pick up predictive analytics concepts—some of which are not very complicated at all, just being linear regression models from large data sets. But predictive analytics goes beyond Generalized Linear Models, and even with GLMs there are niceties that actuaries should know about.

BUT WHERE TO BEGIN?

Below are some resources for the beginner in predictive analytics … and sometimes a nice way for those already well-versed in many of the techniques to expand to a few more they had not considered.

There are two main threads involved in getting started with predictive analytics:

1. Statistical theory and modeling—understanding the approaches, what each does, and what the strengths and weaknesses are for these; and

2. Computing—specialized software and languages intended for crunching Big Data and performing analytics.

I am going to try to pick resources that combine the two, but sometimes that is not possible. For the most part, I will be highlighting free or inexpensive resources.

BOOKS

STATISTICS (THE EASIER WAY) WITH R BY NICOLE RADZIWILL

Weblinks	Free preview: https://qualityandinnovation.files.wordpress.com/2015/04/radziwill_statisticseasierwithr_preview.pdf
	Amazon link for book: http://amzn.to/1URjyQD
Languages/Topics	R and Introductory statistics—confidence intervals, regression, and statistical inference
Level	Absolute beginner

I partly picked this book because the author is a long-time friend, but also because this is a very easy entry into using R as well as thinking about statistical models. The statistics material in the text is similar to the syllabus of the Statistics VEE, so the topics should be familiar to actuaries.

R is a free statistical software package, and thus is used in many of the predictive modeling texts one finds. However, most statistics texts using R have a large gap in explaining how one uses R … and most R texts have a large gap in explaining the statistics while walking you through how to use R.

Nicole developed this text through her own classes at James Madison University in Virginia (Dr. Radziwill is an assistant professor in Integrated Science and Technology at JMU) geared at undergraduate science majors. As Nicole writes, one of her target audiences was:

> "Smart, business-savvy people who want to do more data analysis and business analytics, but don't know where to start and don't want to invest hundreds or thousands of dollars on statistical software!"

I have gone back to Nicole's text as a reference for doing certain things in R, because she walks through every step. This book is long as a result of the step-by-step R code, but I have found this more helpful than trying to Google "how to do X in R."

DATA SCIENCE FROM SCRATCH: FIRST PRINCIPLES WITH PYTHON BY JOEL GRUS

Weblinks	Joel's site: *http://joelgrus.com/*
	Amazon link for book: *http://amzn.to/1URkqoA*
Languages/Topics	Overview of multiple data analysis techniques, Python, SQL
Level	Beginner

Python is another widely-used language in data analytics. While R was developed originally for statisticians, Python is a more general use programming language. That has led to differing groups of people developing already-created/written code for Python and R. Python is an extremely popular language due to its relative ease in use compared with other languages, and there have been several numerical computing packages developed for Python, such as numpy.[2]

Another disclosure: I am also friends with Joel Grus and previewed this text … I have a lot of friends. Joel is currently a software engineer at Google.

In this text, there is a quick introduction to Python—enough to run and adjust the code in the text. In addition to the linear regression and inference concepts that are also in the Statistics with R text previously, this text covers: clustering algorithms, Bayesian approaches, logistic regression, neural networks, and network analysis. He also covers SQL, because much of the data being used in the data-crunching first originated from SQL databases.

This text just gets you started in these techniques … in some cases, just enough to make you dangerous. While Joel does sometimes cover the pitfalls of certain techniques, his focus is primarily on how one executes certain types of analyses and not how they may go extremely wrong.

AN INTRODUCTION TO STATISTICAL LEARNING WITH APPLICATIONS IN R, BY GARETH JAMES, DANIELA WITTEN, TREVOR HASTIE, AND ROBERT TIBSHIRANI

Weblinks	Book's website: http://www-bcf.usc.edu/~gareth/ISL/book.html
	Amazon link: http://amzn.to/1URmvAL
	Online videos (free!): http://www.dataschool.io/15-hours-of-expert-machine-learning-videos/
Languages/Topics	More rigorous approach to statistical inference/modeling techniques, R
Level	Intermediate

For my last book recommendation, here is a more formal text (though "squishier" than the more advanced The Elements of Statistical Learning by a non-empty intersecting set of authors). It is more expensive than the two prior books, as this is a regular college text, and has the accompanying pricing.

That said: there is a complete set of online videos from a class based on this text. This will provide a link to the online courses I promote below.

I have been very slowly going through this text … the slowness due to me jumping back to other resources on R, so I make sure I understand what I'm doing. That's the weakness with this text—the R is not well-explained for the newbie. I would not start with this text for learning R, but once you've got a founding in R, the exercises in R are not so bad.

What's really nice is that you don't actually have to do any of the sections with R—if all you want are the concepts, you can skip the parts in R and pay attention to their worked-out examples.

Still, I think that doing the hands-on applied exercises in R is important in putting the pieces together.

As this is a "real" college textbook, it has end-of-chapter exercises, divided into "Concept exercises" and "Applied exercises." I really liked the "concept exercises" as they were geared to having the student probe that they really understand what is going on, and these exercises are very much geared towards thinking about which techniques are appropriate for which modeling tasks.

As an example, here is the question and my proposed answer for one of the conceptual exercises:

"4. You will now think of some real-life applications for statistical learning.

(a) Describe three real-life applications in which classification might be useful. Describe the response, as well as the predictors. Is the goal of each application inference or prediction? Explain your answer.

Classification may be useful if you're putting together policyholder data/response:

- underwriting in life insurance—have discrete u/w classes as opposed to more continuous u/w;
- might want to classify policyholders as being reactive/hot money vs. passive—very important in variable annuities; and
- might want to flag claims for possible fraud, but don't want to spend too much resources investigating every claim.

(b) Describe three real-life applications in which regression might be useful. Describe the response, as well as the predictors. Is the goal of each application inference or prediction? Explain your answer.

Regression useful in insurance:

- more continuous u/w as seen in auto coverages;
- if want to consider more continuous life u/w as with John Hancock's Fitbit program; and
- used in putting together projections of exposure in various p/c coverages. Can't observe everything while u/w, but may be able to find key variables.

(c) Describe three real-life applications in which cluster analysis might be useful.

- might be wanting to see if one can come up with new u/w buckets—cluster analysis may help;
- I used cluster analysis to see if there's common asset allocation strategies among life insurers—help tease apart influences; and
- could be used by exam committees to compare current exams against historical, check out various metrics (other than Euclidean) to see if there are clear outliers in exam performance" (COUGH COUGH CAS).

As I said, I've been working through this text, and my notes can be found at my dropbox: *https://www.dropbox.com/s/bf6dxtmtnzat4ny/Exercises%20from%20An%20Introduction%20to%20Statistical%20Learning.docx?dl=0* I have been trying to put in insurance/pension-related applications in answers to conceptual questions, but for some of the topics, it gets to be a bit challenging to think of actuarial applications … but give me time.

My R code for the book's applied exercises can be found here: *https://github.com/meepbobeep/ISLR*

Topics covered in this text: linear regression, classification, resampling/bootstrapping, model selection, dimension reduction in models, nonlinear models, tree-based methods (such as decision trees), support vector machines, unsupervised learning.

Disclosure: I am not friends with any of the authors of the following texts. Yet.

ONLINE COURSES

DATACAMP

Website	*https://www.datacamp.com/*
Languages/Topics	R and data science in general
Level	Absolute beginner to intermediate
Timing	On-demand, very short lessons
Paid features	Access to all courses, statement of completion
Credentials	Statement of completion

Datacamp has online lessons in R, which I originally found out about via a class on edX. Like many of the online courses below, they keep trying to upsell you. The pricing is by time period—either by month or by year (cancel any time!) I have tried only their free content, which tends to be the introductory classes. I suppose they figure if you get a taste, you'll want more.

I thought these lessons were very well-done, taking you step-by-step through R and some of the major tasks one would want to do in R when doing predicative analytics. However, the material I see on the site, even the paid courses, don't get to a very advanced level. However, it does touch on using R in ways the prior texts don't: for prettier graphs and dynamic reporting.

I found these lessons were very rapid to go through, and I'm thinking of paying for the one month of access ... should be no problem to go through 18 available courses over the Christmas break, right?

One of the nice features of the introductory courses is that you do not need to install R yourself—you will be able to run R code in the browser itself.

UDACITY

Website	*https://www.udacity.com/*
Languages/Topics	Data analysis, R, Python, SQL, Hadoop, (and much, much more)
Level	Beginner to advanced
Timing	On-demand, usually takes a few months for a full course (some mini-classes are shorter). Nanodegree and regular degree programs are on a schedule
Paid features	Monthly charge for access to coaches, projects with ongoing feedback, verified certificates and degrees (normal and nano-)
Credentials	Verified certificates, Georgia Tech MS in Computer Science, coming soon: nanodegrees

Udacity has a coding focus, along with applications such as with Front-end development and Data Analysis. For this review, I'm only looking at the courses in the Data Analysis nanodegree.

The classes on Udacity are more like regular classes, with quizzes and assignments. Udacity also has video lectures. Classes are rated for level, the advanced classes tend to have programming experience prerequisites. They have classes with serious Computer Science content, not only about how to program. They have classes built by various well-known tech companies, such as Facebook, Google, Amazon Web Services, Salesforce, and Twitter.

In addition to verified certificates for specific classes, and their partnership with Georgia Tech to provide an online-only M.S. in Computer Science, Udacity has recently created "nanodegrees" in specific areas, one of which is for data analysis. These nanodegrees are intended to be completed in less than a year. It looks like there was great demand, because they increased the fee for the nanodegrees from $150/month to $200/month in the past year, and have restricted enrollment in the nanodegrees to certain times of the year.

To access the classes for free, just click on "Start free course" on the specific class page. You can get to all the material: videos, text files, and even assignments. Within the videos themselves, they often stop for quizzes for immediate checking of understanding. Obviously, there are features you can't access if you aren't paying. The courses that are free are generally available on-demand.

The main place to start for their data analysis courses is Intro to Computer Science, which is mainly about learning to code

in Python. It seems most of their data analysis classes depend on Python.

COURSERA

Website	*https://www.coursera.org/*
Languages/Topics	So very much!
Level	Beginner to advanced
Timing	Most on specific schedules, 4-week to semester-long courses; a very few are on-demand
Paid features	Certifications (see below)
Credentials	Signature Track credential, Specialization certificates from sponsoring universities

I find Coursera the most dangerous of all the websites to go to … because there's so much there, and not all of it is programming. Looking at the list of stuff I've signed up for on this site: The Data Scientist's Toolbox, R Programming, Exploratory Data Analysis, Fundamentals of Music Theory, A Beginner's Guide to Irrational Behavior, Machine Learning, Introduction to Mathematical Thinking, Data Analysis, Comic Books and Graphic Novels, Computing for Data Analysis, An Introduction to Financial Accounting, Exploring Beethoven's Piano Sonatas, The Science of Gastronomy, Coding the Matrix: Linear Algebra through Computer Science Applications, Introduction to Data Science, and Gamification.

That's not necessarily exhaustive.

I obviously don't have enough time to seriously pursue all these courses, especially since, unlike the other sites listed above, most of these classes are built to specific time schedules, with classes starting and ending on particular dates. Usually, I'm only seriously following one class at any given time and downloading all the PDFs, videos, and other supporting documents … completely free. I have used some of the items I've come across to teach my own courses on other topics.

All of the courses on Coursera are backed by accredited institutions, and thus Coursera has a more academic feel than Udacity. Some of the classes come with paid certifications, and some courses have no free version at all. Many of the business-related data analytics courses are like that, I find.

Like Udacity, Coursera has developed something akin to "nanodegrees" called Specializations, which are short tracks of verified courses that take about a year to complete. A few of the Specialization tracks available as I write this article are Machine Learning (University of Washington, six courses), Big Data (UC-San Diego, six courses), Business Analytics (University of Pennsylvania, five courses).

Lots of courses to choose from at Coursera, and my main warning is to check prerequisites. Some of the numerical computing courses assume you know specific languages at particular levels. Some are truly introductory, and will walk you through how to get started in various languages, but many are at intermediate levels or higher for the coding, especially in the data analysis courses, so you want to be careful.

Got any favorite resources for the beginner in predictive modeling and data analytics? Let me know about them—*marypat.campbell@gmail.com*. ■

ENDNOTES

1 Alan Mills, "Introduction to Forecasting Methods for Actuaries." Forecasting & Futurism Newsletter, September 2009. pp 6-9. *https://soa.org/Library/Newsletters/Forecasting-Futurism/2009/September/ffn-2009-iss1.pdfn*

2 *http://www.numpy.org/*

Mary Pat Campbell, FSA, MAAA, PRM, is VP, insurance research at Conning in Hartford, Conn. She can be reached at *marypat.campbell@gmail.com.*

Johns Hopkins Data Science Specialization courses: A review

By Shea Parkes

Data science is a hot buzzword in many industries today, but its definition can be nebulous. Some definitions of a data scientist include:

- A person who is better at statistics than any software engineer and better at software engineering than any statistician;
- A person with an equal blend of computer science, statistics, and domain knowledge; and
- An applied statistician who is rebranding.

Even if nobody agrees on the specifics, the concept of data science can still facilitate a thought exercise in what blend of skills is most useful for data analytics. Actuaries are solidly grounded in statistics and domain knowledge as part of the examination and continuing education process. However, actuaries are traditionally weaker in regards to computer science skills than might be optimal to grow our presence in modern data analytics. This includes some blended skills, such as machine learning and predictive modeling, which require both applied statistics and computer science skills.

Computer science skills can bring a lot of value to a classically trained actuary. These skills can help:

- Make answers more transparent, reproducible, and reusable;
- Answer bigger questions than before;
- Answer smaller questions faster and more efficiently; and
- Present answers more visually and interactively.

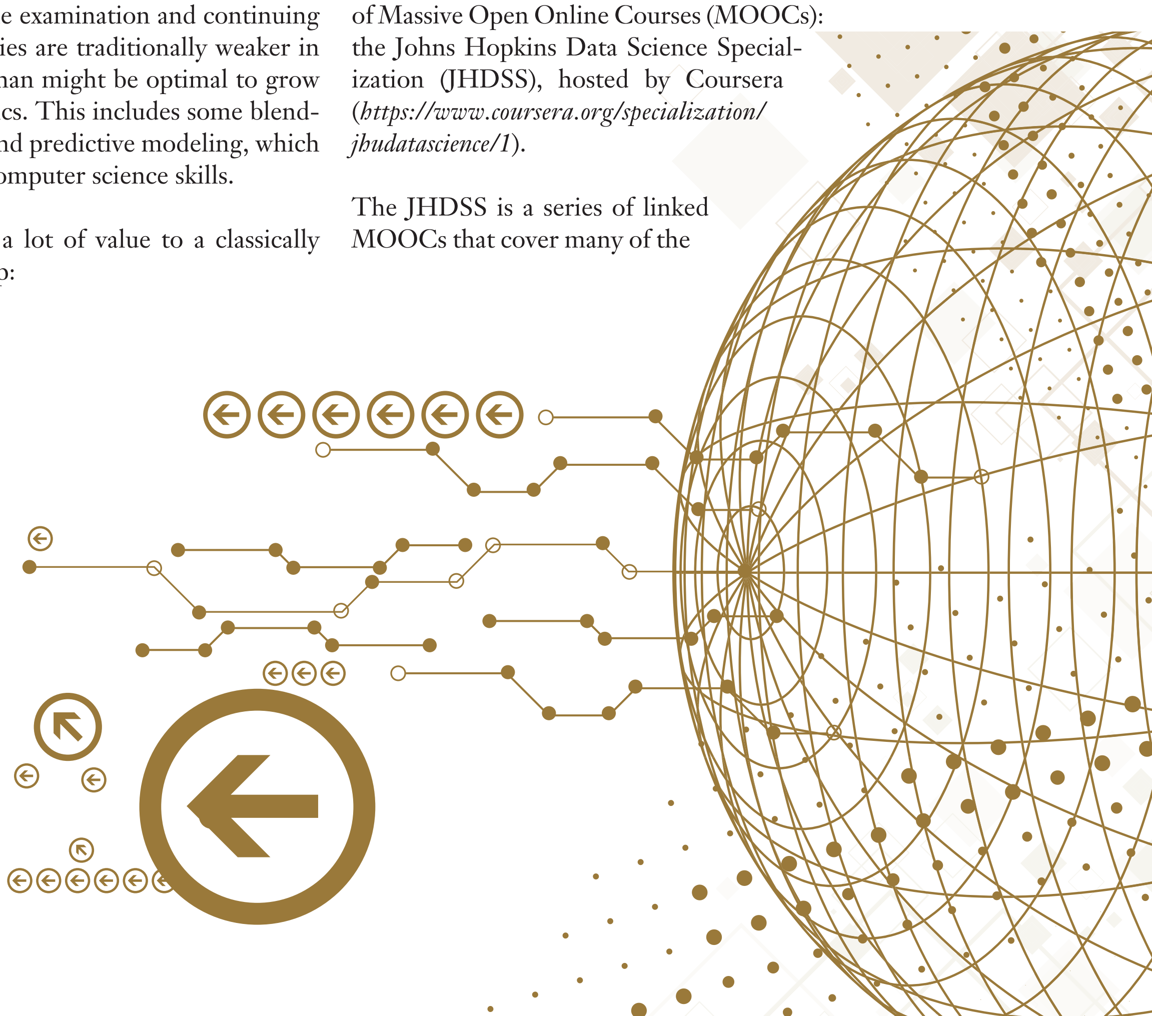
For actuaries interested in rebalancing their skill portfolios toward the data science blend, there are many resources available. I personally was drawn to the data science balance and explored it along a rough path that included:

- Repeatedly attempting (and failing) at cutthroat online predictive modeling competitions (such as those hosted on *http://www.Kaggle.com*) with my coworkers. Every failure was an excellent learning opportunity and after a couple of years we could consistently place in the top 10 as long as we exerted enough effort for a given contest. (Consistently winning was an echelon we never reached.); and

- Forced self-learning while helping carve a new product group out of a large consulting staff. We consumed countless books and other resources on best practices for development of prioritization techniques, software lifecycle management, and gritty details of source control tools and strategies. By the end of the year we reached workable solutions based on ideas such as Scrum, Kanban, Git, and Continuous Integration.

I think this rocky road was actually an excellent way to learn more about machine learning, computer science, and software engineering, but I don't believe it's available or appropriate for everyone. Just about the same time we felt we had found a paved road, a new opportunity was presented in the form of a series of Massive Open Online Courses (MOOCs): the Johns Hopkins Data Science Specialization (JHDSS), hosted by Coursera (*https://www.coursera.org/specialization/jhudatascience/1*).

The JHDSS is a series of linked MOOCs that cover many of the



traditional data science topics in which actuaries might be weakest. The JHDSS courses are not the only MOOCs of their kind, nor are they necessarily the best, but they appeared polished enough to make me interested in trying them. The JHDSS creators are prolific and respected contributors to the data science community in their own blogs and journals.

By the time I had signed up I was already proficient in most of the topics, but I still completed the courses as an external validation of my new skills and also to evaluate them as a continuing education resource for other employees at our company. I completed all of the JHDSS courses with a coworker in a little less than a year. We ultimately deemed it useful enough to make available to all of our staff alongside the actuarial exams and other credentialing opportunities.

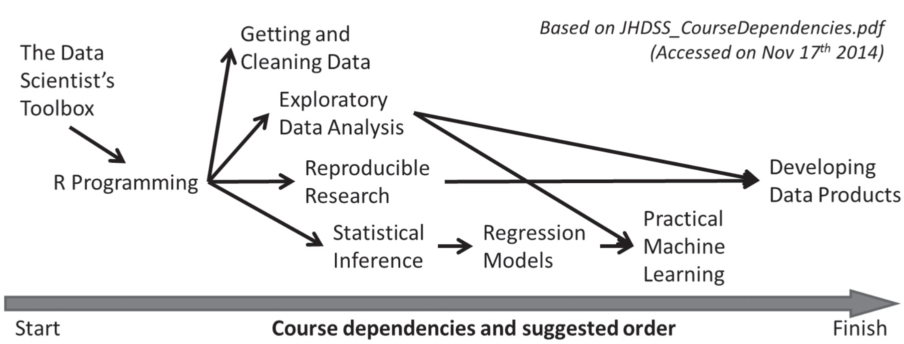
The majority of the JHDSS courses each take a month to complete and require roughly 10 hours of work per week. They include many aspects of standard MOOCs such as:

- Prerecorded lectures with accompanying notes and slides;
- Active discussion forums (an invaluable resource for any student);
- Weekly quizzes; and
- Peer-graded projects on real data.

The peer-graded projects were some of the richest learning opportunities, especially when it was your turn to grade your peers' submissions.

The modest time commitment (10 hours per week) allows for working professionals to complete the specialization somewhat comfortably. The amount of content provided is not enough to make anyone an expert, but it does equip the student with rough tools and ideas that can be sharpened and honed via application and experience.

FIGURE 1: DATA SCIENCE COMPONENT COURSES



Source: *https://www.gitbook.com/book/gdhorne/data-science-boot-camp-survival-manual/details.*

> "The modest time commitment (10 hours per week) allows for working professionals to complete the specialization somewhat comfortably."

The chart in Figure 1 visualizes the names of the component courses and their suggested dependency order.

There is an optional fee to take each of the courses (well under $100 at the time of writing this). The courses can be taken free, but in that case verified "certificates" for resumes are not issued. I personally paid the fee to leverage the sunk cost fallacy and trick myself into committing more to the courses. I also thought the fees were a bargain for a working professional and I wanted to support the content creators who put a lot of effort into a good idea. The opportunity costs of your time will likely be the largest fee.

The remainder of this article presents each course's official tagline and my own brief review of the usefulness and quality of the content.

COURSE 1: THE DATA SCIENTIST'S TOOLBOX

Official tagline: "Get an overview of the data, questions, and tools that data analysts and data scientists work with."

MY REVIEW:

This is a very gentle introduction to some tools that can revolutionize the way you approach solutions. For example, I feel like I'm driving without my seatbelt now if I ever complete work without source control. Source control is a tool that tracks code changes at a very detailed level and greatly facilitates collaboration and quality. The source control tool covered in this course was the very popular GitHub (*http://www.github.com*). Learning a tool like GitHub can be very intimidating, though, and this might serve as a needed boost to get over the initial hump. Some simple text markup language, such as Markdown, introduced in this course, is a necessary companion because classic document solutions like Microsoft Word do not play nicely with most source control tools.

Still, with no prior background or appreciation, this overly simple introduction could lose students' interest because no practical examples are explored. Some of the next few courses do force you to use GitHub and Markdown so you can better internalize what you are exposed to in this course. The later courses just assume you will use source control on your own (and you definitely should).

COURSE 2: R PROGRAMMING

Official tagline: "Learn how to program in R and how to use R for effective data analysis."

MY REVIEW:

The greater difficulty of this course is in sharp contrast to the prior course. Many students might get disheartened if they don't have much prior programming experience. Learning programming is hard, and learning R is harder. However, I agree that R is an excellent domain-specific language (DSL) for data analytics and learning it is worth the effort. I considered myself proficient in R prior to this course, but I learned a few additional aspects of R as a programming language (such as the full nuances of closures). All of the remaining courses depend greatly on this course; you need to be at least somewhat committed to learning R if you are going to complete the JHDSS (and I consider that a good thing).

Because my coworker and I already knew R prior to this course, it is very hard for me to judge how useful this course would be as a beginner introduction to R. It seemed to strike an appropriate balance of explanation, difficulty, and application, but I had a biased view from my place higher up on the R learning curve.

COURSE 3: GETTING AND CLEANING DATA

Official tagline: "Learn how to gather, clean, and manage data from a variety of sources."

MY REVIEW:

This course sustains the high difficulty level of Course 2: R Programming and it continues to teach invaluable data science skills: how to acquire and deal with real data. Coursework intentionally forces you into reading documentation for specific R packages (third-party extensions to R that each add specific functionality) and consulting with Google and Stack Overflow (very good skills to practice).

This course was refreshing compared with the classic style of academic courses that just provide students with already scrubbed data and ask them to perform rote statistical analyses. However, some of the hardest parts of this course were working with data source types that actuaries would be unlikely to dig through. R is great for integrating with traditional data sources such as databases, but this course pushed into some more unusual areas like web services.

COURSE 4: EXPLORATORY DATA ANALYSIS

Official tagline: "Learn the essential exploratory techniques for summarizing data."

MY REVIEW:

I believe exploratory data analysis (EDA) is a chronically underemphasized topic in all forms of education. I have read the classical texts on the subject by John Tukey and William Cleveland and consider them required reading for any aspiring data scientist. We hand out copies of *Show Me the Numbers* by Stephen Few to all new employees at my office and periodically read through it in book clubs. Basically, I loved this course as soon as I read the title. I breezed through the coursework, and I believe it was easier (or at least more innately enjoyable) than the prior courses. They give the subject a respectable treatment and I think any student would benefit from it. I wish they had spent more time with the more advanced tools such as the ggplot2 package for R, but I respect focusing on the theory over the fanciest of tools.

COURSE 5: REPRODUCIBLE RESEARCH

Official tagline: "Learn the concepts and tools behind reporting modern data analyses in a reproducible manner."

MY REVIEW:
I have mixed feelings about this course. I think the concepts of reproducible research are very important and deserve a course of their own. I think the foundational tool-chain they chose (*http://yihui.name/knitr/*) was a solid choice. But I think they went too far when they tried to integrate automatic Web publishing with an unreliable cloud service (*https://rpubs.com/*); stability might improve in the future, but during my course the forums were full of students who had difficulties with the cloud service. I understand why they wanted to go there (theoretical ease of accessibility and "wow" factor), but I believe they should have spent more time covering the advanced capabilities of the fundamental tools instead of trying to layer them into web services.

COURSE 6: STATISTICAL INFERENCE

Official tagline: "Learn how to draw conclusions about populations or scientific truths from data."

MY REVIEW:
I often identify myself as an applied statistician these days (more often than I call myself a data scientist; less often than I call myself an actuary). I find statistics a fascinating topic, but I also find the average teaching of statistics to be rote and formulaic, and this course did not elevate itself above that. I think frequentist statistics has its place, but this course, like many, put it front and center and barely left room to discuss Bayesian viewpoints. I think data scientists should have a firm understanding of statistics, but I believe this course was inadequate to provide that on its own. However, I don't think I could have provided a better grounding in the same amount of time. Statistics is just too big and broad of a subject to dig into as deeply as a data scientist would need to in a single month.

> "The capstone project class will allow students to create a usable/public data product that can be used to show your skills to potential employers."

CHAPTER 7: REGRESSION MODELS

Official tagline: "Learn how to use regression models, the most important statistical analysis tool in the data scientist's toolkit."

MY REVIEW:
Ordinary least squares regression is so far from ordinary. George Box once said "in nature there never was a normal distribution, there never was a straight line, yet with normal and linear assumptions, known to be false, [a scientist] can often derive results which match, to a useful approximation, those found in the real world." Regression theory and models are a great jumping point from applied statistics to predictive modeling and machine learning. I believe this course did a pretty good job of balancing depth of theory while also covering important extensions such as generalized linear modeling. I think aspiring data scientists should spend even more time on this subject to keep a balanced knowledge portfolio, but the next topic (machine learning/predictive modeling) can be quite alluring. I think they could have focused a bit more in this course on relating classical statistical terminology to the corresponding machine learning terminology used later. Making those deep connections really helps understand both topics better.

COURSE 8: PRACTICAL MACHINE LEARNING

Official tagline: "Learn the basic components of building and applying prediction functions with an emphasis on practical applications."

MY REVIEW:
This is a very exciting topic to a large portion of the students that participated, and I think most of them left satisfied. Covering all of the top-tier algorithms is not attempted, nor should be. An appropriately large amount of time is spent focusing on the bias-variance trade-off and model tuning tools such as cross-validation. The exercises force students to build models, but I do think a bit more room could have been allowed for creativity. I introduced some flair into my solutions, but it was not required. I do think the course dependency chart in Figure 1 above is very important, though. This class is a culmination of all that came before and it would be much less without the journey. The courses that come after this are still good ideas, but they take things a subtly different direction (productization).

COURSE 9: DEVELOPING DATA PRODUCTS

Official tagline: "Learn the basics of creating data products using Shiny, R packages, and interactive graphics."

MY REVIEW:
I believe productization is a natural stepping stone in the data science curriculum, but it is a very complicated subject. This class covers a bit of the theory and then jumps into a specific tool (Shiny) used to make responsive web-based applications. Shiny (and its corresponding cloud hosting services) is a promising but young tool that is not without its rough edges. Still, it has the right level of accessibility and "wow" factor; you can learn it and feel proud of your results within the duration of this course. I personally think there should have been more focus on "hardening" advanced modeling code to work stably in a production environment, but that's a much less exciting subject.

CAPSTONE PROJECT: DATA SCIENCE CAPSTONE

Official tagline: "The capstone project class will allow students to create a usable/public data product that can be used to show your skills to potential employers. Projects will be drawn from real-world problems and will be conducted with industry, government, and academic partners."

MY REVIEW:

I believe this was a strong finish to the JHDSS; it was a full two-month project focusing on a single problem. They intentionally introduced an important subject not covered in prior courses (text mining in my sitting) to force you to practice learning something fundamentally new as part of a larger engagement (a common occurrence in the real world). The problem was interesting and the amount of guidance was just right. The ancillary tasks (e.g., quizzes) were surprisingly weak, but that didn't distract from the overall strength of the capstone project. It felt very much like my day job (the fun parts of it), and I think that's the best endorsement I can give it. The difficulty level was quite high, but most participants rose to the challenge.

CONCLUSION

If you, or someone you know, wants to learn more about the data science viewpoint, the JHDSS is a useful means to do so. The largest hurdle might be that participants would need to be committed to learning R, but I consider that a positive aspect of the specialization. Trying to cover these topics without diving deep into an appropriate computer language would have failed to give them the treatment they deserve. The JHDSS is not perfect, but I believe the general content is a really good mix, especially to complement classical actuarial training. ■

Shea Parkes, FSA, MAAA, is an actuary at Milliman in Indianapolis. He can be reached at *shea.parkes@milliman.com*

Introduction to Predictive Modeling of Fund Manager Behavior for Variable Annuities Riders

By Bryon Robidoux

This article is an introduction to how actuaries working with variable annuities (VA) use predictive modeling. The intended audience is for actuaries unfamiliar with daily operation of variable annuity riders, such as Guaranteed Minimum Withdrawal Benefit for Life (GMWBL), and the fund modeling process. This article will explain how predictive modeling is used to model fund manager behavior and its impact on the performance attribution and risk exposure.

The primary goal of an actuary hedging VA is to monitor the behavior of the market, the policyholders, and the fund managers to predict how these will impact the liability and the eventual claims that will be paid. The actual changes to the liability due to market impacts, policyholder behavior, and fund manager behavior are analyzed in the performance attribution. The changes due to market risk exposure are analyzed in the daily risk exposure report. The risk exposure report contains the Greeks which state how the liability will move due to shocks in the underlying risk drivers such as equity, interest rates, and volatility. The underlying account value of each policy is backed by mutual funds selected by the policyholder. Fund modeling is the process of mapping the mutual funds to an array of stock market indices where liquid assets can be purchased to hedge the VA guarantees. The array of stock market indices are used as predictors for the mutual funds' returns. The fund model will produce weights, called betas, which will allocate a certain percentage of the account value to each index. The betas are required to sum to unity. For simplicity, this article will assume that the array of indices used for fund modeling is given. (The indices to use are very company specific. It really depends on the available mutual fund lineup offered by the insurer and the size and type of risk exposures contained in the mutual funds.) Fund modeling is unique in that finding more predictors or different predictors may not help in building a better model if the predictor has no large liquid market to purchase derivatives, not enough exposure in the VA block to warrant hedging, or the block's risk exposure to a predictor is so large that the insurer would dominate the market. Assuming the indices are fixed, the drivers that can change the mapping are the underlying market return, the fund manager's behavior, and the interaction among them.

A great place to get descriptions of mutual funds and their behavior is at Morningstar.com. Morningstar groups the funds by the fund manager's investment strategy, such as U.S. Equity Fund, Allocation Funds, International Equity Funds … Alternative Funds, Commodities, and Sector Equity Funds. These are in order relative to their ease of fund modeling and hedging. (Beyond looking at just the Morningstar group to get a sense of the behavior of a fund, the reader can look at the prospectus to find the amount of cash in the fund and the turnover ratio. The amount of cash in the fund gives an indication of the level of derivatives used. The turnover ratio is the amount of assets that are bought and sold during the year.) The U.S. Equity Funds are the easiest to model because they replicate a major index. The Alternative Funds, Commodities, and Sector Equity Funds consist of a lot of derivatives, have very large turnover ratios, and have high volatility. These characteristics greatly increase the option value of the rider sold and the difficulty in modeling the fund. For these reasons, Alternative Funds, Commodities, and Sector Equity Funds are not rider eligible on VA rider guarantees. The Allocation funds are funds that state in their prospectus they will have a certain proportion of their holdings in equity and the remaining in bonds. Obviously, the aggressiveness of the fund is directly proportional to the amount of equity in the fund. Usually the Allocation funds are fund-of-funds. For proper modeling, the actuary needs to thoroughly investigate the holdings of the fund in its prospectus. For diversification and volatility management reasons, they may contain a certain percentage of their holdings in Alternative Funds, Commodities, and Sector Funds. The higher the percentage to these funds, the more tricky the Allocation fund's behavior can be to model for the reasons already stated.

Even though the Allocation fund has what appears to be an iron clad mandate to its investment strategy, the fund manager does have quite a bit of room to meet the objectives of the fund. The short-term strategy of the fund may be quite a bit different from the long-term strategy, because either the Allocation fund manager or the managers of the underlying funds are trying to take advantage of current trends in the market. If the market is in a straight climb, such as it was for all of 2013 and 2014, then, as time goes on, the fund managers will move more of their holdings to equity so they can beat their benchmark. If there is elevated volatility in the market, such as the fourth quarter of 2011 with the Greek debt crisis or third quarter 2015 with the China equity bubble, then fund managers will allocate more of their holdings to bonds to reduce volatility and minimize losses. These behaviors of the fund manager can greatly affect the decision on what predictive model to use and the behavior of its betas.

Another way to look at fund modeling is to think of it as mapping the individual risk exposures of the mutual funds to equity and

bond indices so that the entire risk of the block can be aggregated and hedged. This implies that the fund modeling has a direct impact on the risk exposure report. How the funds are mapped has a direct result on how closely the Greeks will match the movement of the liability. This in turn affects the amount and the location of the liability's risk exposures. If the risk exposures are poorly mapped, the issue won't be apparent on the risk report, but will instead appear on the attribution report as large unexplained net profit or losses (P&L). The P&L will not be isolated in one location, but bleed throughout the entire report. The fund mapping impacts the performance attribution in the following manner:

1. The fund basis is the difference between actual return and the expected return on the mutual funds. It is essentially the realization of how well the fund model performed over the period. The actual returns come from the underlying fund returns. The expected return is the fund allocation times the index returns. The fund basis line is the difference in the liability valued with expected returns versus the liability valued with the actual returns.
2. The fund modeling update line is the change in the liability due to updating the fund model.
3. Given that the betas are the means by which the account value gets mapped to equity and bond indices, the following are secondary impacts.
   a. The equity and bond exposure is a direct result of the fund mapping, which flows through the equity and interest rate net P&L lines.
   b. The allocation to equity and bond indices determines the amount of volatility exposure in the portfolio. This in turn flows through the volatility P&L line.
   c. The equity, interest rates, and volatility dictate how the velocity of the liability's change and the assets' change due to market forces. This in turn dictates the total market's P&L line.

There is a tug of war between the overall fund basis and the fund modeling update line. In order to reduce fund basis bleed from week to week, a fund model with sensitive weights can be chosen. But during a model update, if the betas significantly shift from equity indices to bond indices or vice versa, this could have a large model update P&L impact because the volatility in the portfolio will change significantly. On the other hand, if the betas are stable during model updates, the model update will have minimal P&L impact, but the fund basis bleed could potentially be large because the model is not responsive enough to the fund manager's short-term behavior.

When managing a VA portfolio, what can be done to deal with the fund manager's dichotomy between short-term incentives and long-term mandates? The easiest thing to do is create two different fund models: one for the short term and one for the long term. In general, the long-term model should have stable betas. This aligns with the principle that the fund manager will meet his fund's stated objectives over the long run. For the short-term model, the prospectus of the fund really needs to be analyzed to determine the proper behavior of the model's weights. In general, the stability of the weights should be inversely related to the turnover ratio, the amount of cash, and the percentage of Alternative funds, Commodities, and Sector Equity Funds contained in the fund.

How does this enhancement affect the attribution and risk report? It helps relieve the tug of war between the fund basis and fund modeling lines of the attribution and it improves the accuracy of the Greeks. The long-term model has the largest impact on the overall liability value because the long term growth rate has the largest impact on the eventual claims that will be paid. The long-term model has very little impact on the Greeks, because the Greeks are an immediate shock to underlying risk drivers. The opposite is true for the short-term model due to similar logic. The fund basis is affected by both the short-term and long-term fund model because it is a direct realization of how well the fund model maps to the mutual fund returns over the life of the liability. The fund basis should be reduced because the two models will do a better job managing the fund manager's dichotomy.

In the liability model during a valuation run, how should the length of time to use the short-term fund model be defined? In our model, it is defined as the stub period, which is just the end of the policy year following the valuation date. This is done for simplicity of the model rather than accuracy. This is counter intuitive because each policy will be using the short-term fund model for different periods of time during the valuation run. But in reality, the length of time to use the short-term fund model should have almost zero impact on the Greeks and long-term liability value because there is a fund model specifically addressing each of these items. It should only have a marginal impact on the fund basis, which I suspect would not be material.

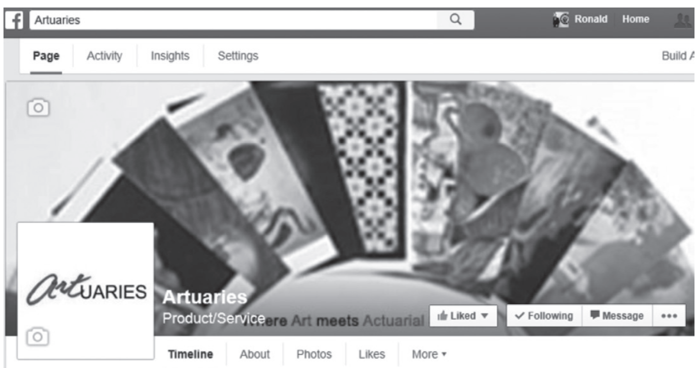
In conclusion, the VA offers a guarantee backed by mutual funds. VA actuaries need to perform fund modeling to map these mutual funds to indices where they can buy cheap liquid derivatives which can be used to hedge the liability. The major objective of fund modeling is to create predictive models, which will allow the actuary to map the funds to common indices to manage the long term risk exposures and growth rates of the account value. With this, the actuary must realize that the fund manager's incentives to outperform the fund's benchmark in the short-run will cause the fund's short-term allocations to equity and bonds to differ significantly from the long-term allocations. The funds that have a higher turnover ratio and allocation to cash are more likely to possess this behavior. To better manage the fund manager's behavior, it makes sense to have a short-term and long-term fund model for each fund. The performance attribution's net P&L should be improved because, when the fund model is updated, it should have less of a P&L shock and the fund basis bleed should be reduced from week to week. The Greeks should be more accurate because they should better reflect the changes due to market risk and fund managers' behaviors. In the liability model, the method used to transition between the short-term and long-term fund model probably should have minimal impact on the P&L, fund basis, and overall liability value. ■

Bryon Robidoux, FSA, CERA, MAAA, is director of Hedging, Global Financial Solution, at RGA Reinsurance Company in Chesterfield, Mo. He can be reached at *brobidoux@rgare.com*

ARTUARIES—WHERE ART MEETS ACTUARIAL SCIENCE

By Ronald Poon-Affat

This article first appeared in the March 2015 edition of *Reinsurance News*. It is reprinted here with permission.



I have brought together a group of super talented actuaries to contribute to a unique actuarial/art/ charity project called the Artuaries.

The present group includes 10 actuaries who are painters, photographers and quilters. The group does include one U.K. actuary, John Gordon, FIA, so we can legitimately claim to be global. The artwork is set out within the Artuaries Facebook page.

A set of greeting cards showcase the group`s artwork. These are available on the Actex website (www. actexmadriver.com) by searching the word "Artuaries. 100 percent of net profits are donated to the Actuarial Foundation; so it's keeping it all in the actuarial family.

HOW DID ALL OF THIS GET STARTED?

Even though I have minimal artistic and marketing talent, I grew up in Trinidad & Tobago during the '60s in a Mad Men environment. My dad was a McCann Erickson Art Director. I eventually brought Dad out of retirement to design the project`s iconic logo.

I came up with the initial idea whilst visiting the home of my long-time friend and actuarial mentor Debi Gero who impressed me with both her deep passion for art history and a prolific contemporary art portfolio. I wanted to share Debi`s art with a wider audience and thought that there must be other artistic gems in the actuarial community. The name Artuaries was the brainchild of Debi.

Very early on, I reached out to Anna Rappaport, past president of the Society of Actuaries (SOA) and a dedicated artist, to discuss how to create a project that would be sustainable and have the greatest impact. Anna was the pragmatic voice who suggested that the first project should be a set of greeting cards that would be timeless (as opposed to a calendar, say), easy to manufacture and distribute and not be too expensive.

WHAT WERE THE CHALLENGES?

The main challenge was to find actuaries who are interested in art and would be interested to show their art. Artistic actuaries are not as boastful as long distance runners, say, so it was quite a challenge to uncover artistic actuaries. Thankfully, we live in an age of social media so Facebook and LinkedIn played an invaluable role in attracting other artists to come out of the woodwork.

To join this merry band, one did not have to be professionally trained; the only qualification was that one had to be an actuary and to have created art that you wanted to showcase.

Just like financial services, distribution is key to a successful operation and Gail Hall of Actex stepped up to the plate volunteering to facilitate the sale of the cards on Actex`s site. We cannot thank Gail enough for her assistance. It was our goal to keep it totally non-profit and actuarial so the Actuarial Foundation was the obvious candidate to be the recipient of our net profits from sales.

NEXT STEPS

A lot has been achieved to date. The artists have been assembled, the cards have been produced, the distribution is in place, the charity has been identified and the artists were profiled in two editions of The Actuary magazine; so what else is there to do?

The present twin goals are to attract more actuarial artists from around the world to the project and to find a tipping point that will substantially increase sales and fund raising. Next steps will be the roll out of a pipeline of projects to proudly display actuarial artwork on calendars, coffee-mugs, t-shirts, caps, etc.

When the project was started the main goals were to raise funds for the Actuarial Foundation, create a network of like-minded actuaries, showcase their art and show the world how cool actuaries really are. On that measure, I think that we are on the road to being a success.

Please like us on Facebook 

Ronald Poon-Affat, FSA, FIA, MAA, CFA, is VP and director with RGA Reinsurance Co. He can be contacted at *rpoonaffat@rgare.com.*

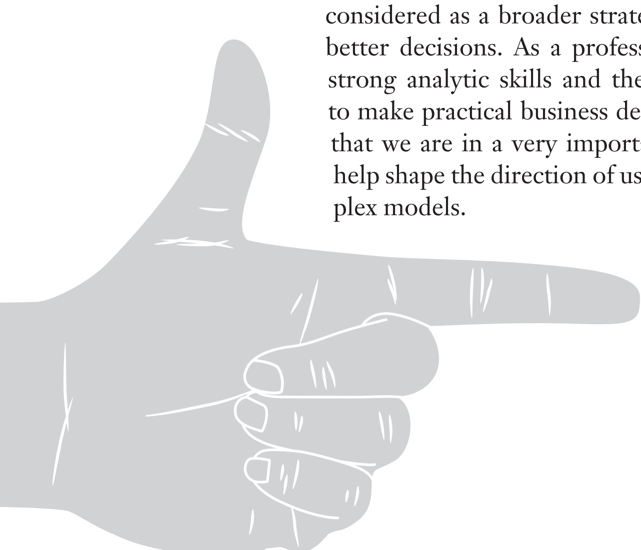
Predictive Analytics: An Alternative Perspective

By Kurt Wrobel

Predictive analytics. The term suggests data, complexity, sophistication, and progress in predicting the future. As suggested by the recent name change in this section, it also represents the general direction of our profession—a move toward more extensive use of data and more complex models. By combining computing power with significantly more data, these analytic processes promise greater accuracy in projecting the future.

There is a great market for this predictive power. Senior managers want to be able to accurately predict the future and set the right expectations for outside stakeholders. Policymakers want to predict the outcomes of policy changes and ensure that these changes are sufficiently funded. IT professionals want to develop a sophisticated infrastructure to help support these data intensive initiatives. Academics want to create even more sophisticated approaches to analyze data. Consultants want to highlight new, but more complex models that have the potential to improve the predictive power over existing models. Considering the many groups advocating for greater complexity, few people stand on the other side of the movement toward more data and greater computing power.

As we move toward more extensive use of predictive analytics and greater complexity, however, I also think that we need to consider the necessary conditions for more sophisticated predictive analytics to be useful and ensure that this tactic is considered as a broader strategy to produce better decisions. As a profession with both strong analytic skills and the responsibility to make practical business decisions, I think that we are in a very important position to help shape the direction of using more complex models.



PREDICTIVE ANALYTICS: THE CONDITIONS NECESSARY FOR A USEFUL ANALYSIS

Before more sophisticated predictive analytics can be proven useful, several conditions should be met before moving to the next step of using a more complex models. These include:

ACCURATE HISTORICAL DATA. Although this obvious step is best characterized by the term "garbage in; garbage out," the potential accuracy of the historical data is often not considered when an analyst makes the next step to introduce a complex model to answer a business question. In many cases, the challenge is that the underlying data is neither completely accurate nor completely wrong, but rather a shade of grey that can be difficult to evaluate. For example, the data could have a selection bias or measurement problem that could impact the accuracy of the model, but the full extent of the impact is difficult to measure. To the full extent possible, an analyst should consider whether these data limitations make a sophisticated data analysis designed to explain subtleties in the data not useful.

A STABLE UNDERLYING SYSTEM WHERE THE HISTORICAL DATA IS A GOOD INDICATOR OF THE EXPECTED EXPERIENCE IN THE PROJECTION PERIOD. If the economic incentives and policies are changing significantly from the historical period to the projection period, the experience and population inherent in the historical data will not necessarily be a good indicator of future performance. The most obvious case of this in our profession has been the pricing development for the Affordable Care Act (ACA). With the change in the underwriting rules and the introduction of income dependent subsidies, the historical experience of a fully underwritten individual population is simply not a good indicator of the future experience for an ACA population. In this case, a sophisticated analysis of the historical data will be much less useful.

THE POTENTIAL ERRORS IN USING A SOPHISTICATED MODELING APPROACH DO NOT OUTWEIGH THE HOPED FOR INCREASE IN ACCURACY. With increased complexity, a model can be become increasingly difficult to understand and more difficult to adequately peer review. The loss of these two important features in a modeling exercise often lead to errors and ultimately decisions that are worse than a model where the results are intuitive and adequately peer reviewed. Although not often explicitly considered, these costs need to be accounted for when moving to a more sophisticated predictive model.

THE PROCESS DOES NOT EASILY LEND ITSELF TO A BIASED ANALYSIS THAT ALLOWS THE RESEARCHER TO PRESENT A PREFERRED OUTCOME. With a more complex analysis, an analyst will have a greater opportunity to "cherry pick" results to present the preferred conclusion in the best possible light. While this problem could be mitigated through adequate review, complex models are much more likely to allow analysts to have this opportunity to skew the final results.

Taken in total, the above conditions are important determinants in whether a complex predictive analytic exercise should be started. Without considering the above factors, we are likely to engage in a costly and time consuming exercise that does little to improve the decision making process and could produce even worse results than a more intuitive approach.

AN ALTERNATIVE APPROACH: GOOD DECISION MAKING

While the term "predictive analytics" has intuitive appeal to many people, its use still needs to produce better decisions that are both accurate and contribute the long term sustainability of the organizations who rely on our estimates. In an effort to highlight a process that produces better decisions rather than a specific tactic—predictive analytics—the following steps outline factors that contribute to better decision-making.

1. CLEARLY DEFINE THE BUSINESS QUESTION AND DEVELOP SEVERAL WORKING HYPOTHESIS THAT COULD CONTRIBUTE TO RESULTS IN THE PROJECTION PERIOD. A clear question with working theories helps focus the analysis and ensure that the research has a well-defined objective.

2. UNDERSTAND ALL ASPECTS OF THE DATA THAT WILL BE USED IN THE ANALYSIS, INCLUDING HOW IT WAS CAPTURED AND ITS POTENTIAL WEAKNESSES. LOOK FOR OTHER DATA SOURCES THAT COULD COMPLIMENT THE DECISION-MAKING PROCESS. Data is the life blood of actuarial analysis and we need to take very seriously its weaknesses as we begin an analysis that presumes that the data are accurate.

3. UNDERSTAND THE SYSTEM BEING PREDICTED AND ENSURE THAT THE HISTORICAL PERIOD DATA CAN ACCURATELY REPRESENT THE EXPECTED RESULTS IN THE PROJECTION PERIOD. While adjustments can be made to the historical experience to better reflect the expected experience in the projection period, more extensive adjustments introduce a greater potential for error in the final estimate. This variability needs to be considered as greater complexity is added to the modeling process.

4. EXHAUST ALL EFFORTS TO ANSWER THE QUESTION WITH SIMPLE DATA ANALYSIS AND QUALITATIVE FACTORS. This high level analysis can help direct the research and ensure that a complex analysis is useful and ultimately passes the high level intuitive test.

5. LOOK TO DISPROVE YOUR THEORY THROUGH ADDITIONAL TESTING OR BY WORKING WITH OTHERS WHO USE ALTERNATIVE APPROACHES. Analysts need to be vigilant about not falling in love with their preferred result and ensure that others adequately test their conclusions.

6. CONSIDER ADDING ADDITIONAL COMPLEXITY THROUGH PREDICTIVE ANALYTICS OR OTHER TECHNIQUES IF MORE SIMPLE TECHNIQUES ARE INADEQUATE AND THE ADDITIONAL COSTS ARE LIMITED. Additional complexity can be costly and the benefits should outweigh the costs.

7. FULLY UNDERSTAND HOW THE RESULTS WILL BE USED AND ENSURE THAT THE RESULTS WILL BE SHOWN IN THE BROADER CONTEXT INCLUDING PRESENTING THE POTENTIAL VARIABILITY ASSOCIATED WITH THE ESTIMATES. We need to be careful to show the likely variability of our estimates and ensure that a point estimate from a highly stable system with less potential volatility is not directly compared with a point estimate from a volatile system.

CONCLUSION

As a profession, our job is to help make the best possible decision with all available information and ensure that our estimates help contribute to the long term sustainability of the institutions that provide health insurance, pensions, and life insurance protection for people at the most vulnerable time of their lives. If a more sophisticated modeling approach or predictive analytics helps contribute to this goal, we should embrace these tactical techniques to help in our mission. That said, predictive analytics is only a potential tactic in a series of steps used to produce the best possible decision. It should not be considered an end in of itself. As our section makes this name change, I hope that we continue to remind ourselves of our broader mission and ensure that our chief goal is to produce better decisions and not necessarily greater technical sophistication. ■

Kurt J. Wrobel, FSA, MAAA, is chief financial officer and chief actuary at Geisinger Health Plan in Danville, Penn. He can be reached at *kjwrobel@thehealthplan.com*

How to get involved: Step one, get involved!

By Doug Norris

As my three-year term on the (now) Predictive Analytics and Futurism section council comes to a close, we stand on the brink of a great opportunity. The marriage of actuaries and predictive analytics has been a long and storied one, and a large part of the actuary's job relates to making best estimates of what the future will bring. However, the business world has the perception that actuaries are behind the wave in this key area.

This perception largely stems from awareness—actuaries have not necessarily promoted themselves as predictive modeling experts, and a lot of the work that we do is proprietary in nature. I have been blessed to work with many other actuaries in this area, and our profession holds a lot of hidden talents. We have a communication conundrum—actuaries have many strengths that align well with predictive analytics, such as the ability to communicate complex topics to a broader business audience. What are the unique qualities of actuaries that we can promote? The beauty of predictive analytics is that the techniques are very accessible, and there are several venues where we can drive perception.

On the other hand, formal actuarial training does not incorporate many of the classic predictive analytic tools and techniques (and it's difficult to present meaningful PA problems in the context of current exam sittings). How do we better prepare the next generation of actuaries to be able to innovate and advance the actuarial PA toolbox?

What this means is that there are many areas where the Predictive Analytics and Futurism section can make a difference, and this is where you come in. As I have said in past articles, my involvement with this section happened largely by accident, and I was fortunate because I couldn't see a cleaner path to participation. One of our section's initiatives over the past few months has been to develop a "microvolunteer" list, where opportunities can be easily identified. Volunteering doesn't always have to be a major undertaking; there are a number of roles that could take just a few hours.

Since you've read this far, you're probably looking for places to get involved. Here are a few:

- Joining the council as a member takes a fair amount of effort, and also requires the winning of a section-wide election. There are only nine spots on our section council, and there are many more roles than nine can provide for. **Becoming a friend of the council** merely requires the expression of interest, and the council can vote to appoint you. You'll be invited to each month's council call, be able to participate in the discussions, and more easily jump in to the activities as they present themselves.
- **Writing a newsletter article** may seem like a heavy lift, but it only requires you to be passionate about a topic that deserves a wider audience. We all have our specialties, we have each seen things that others would find interesting. What's your passion? Writing an article can also lead to …
- **Speaking at an SOA event** is another great way to "build your brand" as an actuary. Our section currently sponsors sessions at the Annual Meeting, at the Health Meeting, at the Valuation Actuary Symposium, and the Life & Annuity Symposium. Session descriptions are usually hammered out well in advance, so if you have an idea for a topic, it's best to talk to a council member right away. SOA audiences have largely moved past the introductory topics, and are looking for information that they can directly apply to their own jobs. Do you have an interesting problem that was solved by a clever predictive modeling application?

Don't have an idea for a fresh topic? We're often looking for speakers on already-scheduled sessions. Predictive analytics lends itself to a lot of material.

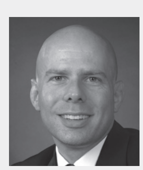
- **Speaking at a section webinar** is similar to the above, but we have more flexibility on topics and scheduling.
- **Coordinating any of the above** is also valuable to the section's goals. Do you know someone in your organization who would be a great speaker or writer? We also need **champions** who can help orchestrate and brainstorm topics.
- Dave Snell has been our newsletter editor for many years, and he does an amazing job. It's a tremendous amount of work, and he would welcome an **assistant newsletter editor** and other support. Keeping track of topics, looking for writers, and keeping Dave generally sane are all things that will help the newsletter become an even brighter light.
- Our section has a burgeoning presence on LinkedIn, and we are always looking for interesting articles, discussion starters, and support for section activities. A **social media coordinator** could do all of these and more, as well as work with other platforms, such as Twitter or the Actuarial Outpost discussion forum.
- The predictive analytics challenge is larger than just our section, and the Society of Actuaries has been working on a

global predictive analytics initiative. This will lead to many opportunities for our session, and **working with the SOA as a liaison** will help us all to be more efficient, and will help to identify new prospects. There will be a lot to do, and this will lead to further volunteer needs.

- The SOA has nearly two dozen sections, and all of them work with predictive modeling on some level (remember, actuaries like to predict things). We've worked with many sections on special projects, including the co-sponsoring of sessions, writing research reports together, and developing special projects. A key part of our future activities will be **collaboration with other sections**, and we need all of the connections that we can find. Are you involved with another section? Would you like to work with us?
- We have also talked about **partnering with outside organizations**, including other actuarial groups.
- We've had a number of **topic champions** in the past, individuals with a special interest (such as Delphi studies, elastic net regression, neural networks, behavioral economics, data improvement techniques, or machine learning). If you have passion and experience with a certain topic, and are interested in promoting it or finding collaboration opportunities, one of these roles could be for you.
- **Member engagement** is going to be a key determinant of how well we succeed in the future. How do we recruit new members? Once they're part of the section, how do we effectively engage them? How do we support the section as members look for ways to contribute? This is an exciting role.

Reading this article, you've probably come up with a few ideas of your own, and you can't believe that I didn't mention those. That's my point—**our section is its members**, nothing more and nothing less. Your ideas are our fuel, and there's nothing that we can't accomplish by working as a team. Volunteer roles don't necessarily have to be large (and tasks don't have to be accomplished solo). If you would like to help, we can find a role that will fit what you're able to give.

Looking back to the start (of the article) and end (of my council term), I am excited at the number of challenges that we have ahead of us. We have an exceptionally strong section council as 2016 dawns, but as you can see, our needs going forward are much larger than nine individuals can reasonably accomplish. I'm somewhat sad that I'll no longer have the role that I've enjoyed, and I'm looking forward to staying involved. Will I have that opportunity? You bet (and so will you). ■



Doug Norris, FSA, MAAA, Ph.D., is a principal and consulting actuary with the health practice of Milliman in Denver, Colo. He can be reached at *doug.norris@milliman.com.*

Big Data or Infinite Data?

By Dave Snell

There was a young fellow from Trinity
Who took $\sqrt{\infty}$
But the number of digits
Gave him the fidgets;
He dropped Math and took up Divinity.
(from One Two Three … Infinity: Facts and Speculations of Science, by Georg Gamow)

One of my favorite books of all time is one I read in high school a half century ago: One Two Three … Infinity by George Gamow. Prior to that time, I had a limited understanding of the concept of infinity. Gamow, an expert in theoretical physics, introduced the idea of infinity by describing a tribe of Hottentots, who had words for one, two, and three; but none for higher numbers. Anything larger than three was considered "many"—our rough equivalent of infinity. Through the tribe analogy, he addressed the issue of how to compare one infinity with another infinity. If you have many beads and many coins, how do you determine which is your larger collection? Gamow related how Georg Cantor, the founder of set theory, compared two "infinite" sets. Cantor proposed pairing the objects of the two collections and see which, if any, ran out first. If each object in the beads collection can be paired with an object in the coins collection, then the two collections are the same size. However, if you arrange them in pairs, and some unpaired objects are left over in one collection, then it is said to be larger, or stronger, than the other collection. Thus, he introduced the "arithmetics of infinity," where the infinite set of all even numbers is the same size, or cardinality, as the infinite set of all odd plus all even numbers. And while you are still wrapping your mind around that non-intuitive result, they both are smaller (less strong) than the cardinality of the set of real numbers, which, in turn, is less strong than the cardinality of the number of geometric curves.

The many years since *One Two Three … Infinity* (I read his 1961 edition; the first edition was published in 1947) have seen a dramatic increase in the number of collections we count and analyze and compare to other collections. Indeed, according to former Google CEO, Eric Schmidt, "Every two days now we create as much information as we did from the dawn of civilization up until 2003." He said that on Aug. 4, 2010 at a Techonomy conference in Lake Tahoe, California and I have to believe the figure today would be even more astounding.

We are clearly into an age of "Big Data"; and it is a term so overused that my Google search for it today yielded 795 million results. Yet, in some respects, we understand this no better than the tribe of Hottentots that George Gamow described in 1947—perhaps no better than how Georg Cantor explained it in 1874. In fact, according to Dan Ariely, the author of *Predictably Irrational*, and other excellent behavioral science books,

> "*Big data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it*"—Dan Ariely

What is Big Data? Some companies brag about being able to handle big data of millions of rows of information. Others claim they process over a billion data items and boast about their big data capability. WalMart was supposedly the first commercial enterprise to store a terabyte of data, in 1992; and then we thought that was truly big data. Now, you can buy a USB thumb drive on Amazon, for your keychain, which stores a terabyte of data. The Human Genome Lab stores petabytes of DNA information. Many database manufacturers claim the ability to store exabytes of data. The NSA stores … oops! That is a classified size, but obviously a lot! Cisco, the multinational technology company that makes and sells networking equipment, predicts that by 2016, about the time you receive this issue in the mail, global internet traffic for the estimated 3.4 billion Internet users will reach a staggering 1.3 zettabytes annually.

What distinguishes Big Data from just large, or very large, or very, very large data?

I'd like to propose a new term: Infinite Data. Infinite data is data that is so large that the mere acquisition of it overwhelms our ability to process it with classical statistical methods.

Take, for example, weather indicators. Our ability to forecast the weather today or tomorrow seems quite good; but our best estimates of weather next month seem unimpressive because the amount of data coming in is so voluminous that the so-called butterfly effect cannot be analyzed in real time.

Another example is the streaming data regarding insurability of a cohort of applicants from the Internet: Tweets, wearables, etc., and other information preclude calculating a classic mean or standard deviation because the data is changing before you even have an opportunity to count it. Like Cantor, we may need to eventually differentiate between Infinite Data of cardinality $\aleph_0$ (read aleph-naught or aleph-zero), the smallest type of Infinite Data, and $\aleph_1$ (aleph-one, a stronger set of infinite data), or $\aleph_2$ (still stronger).

We can also describe big data with more modern terms such as the three Vs: volume, velocity, and variety. Sometimes we add a fourth V, variability, or even a fifth, veracity, to the mix. When these characteristics combine—especially when they are expanding at an increasing rate, we feel that we have Big Data. Yet, actuaries should not feel intimidated by the newer terms. In most cases, we can relate them back to basic techniques we studied years ago under different names.

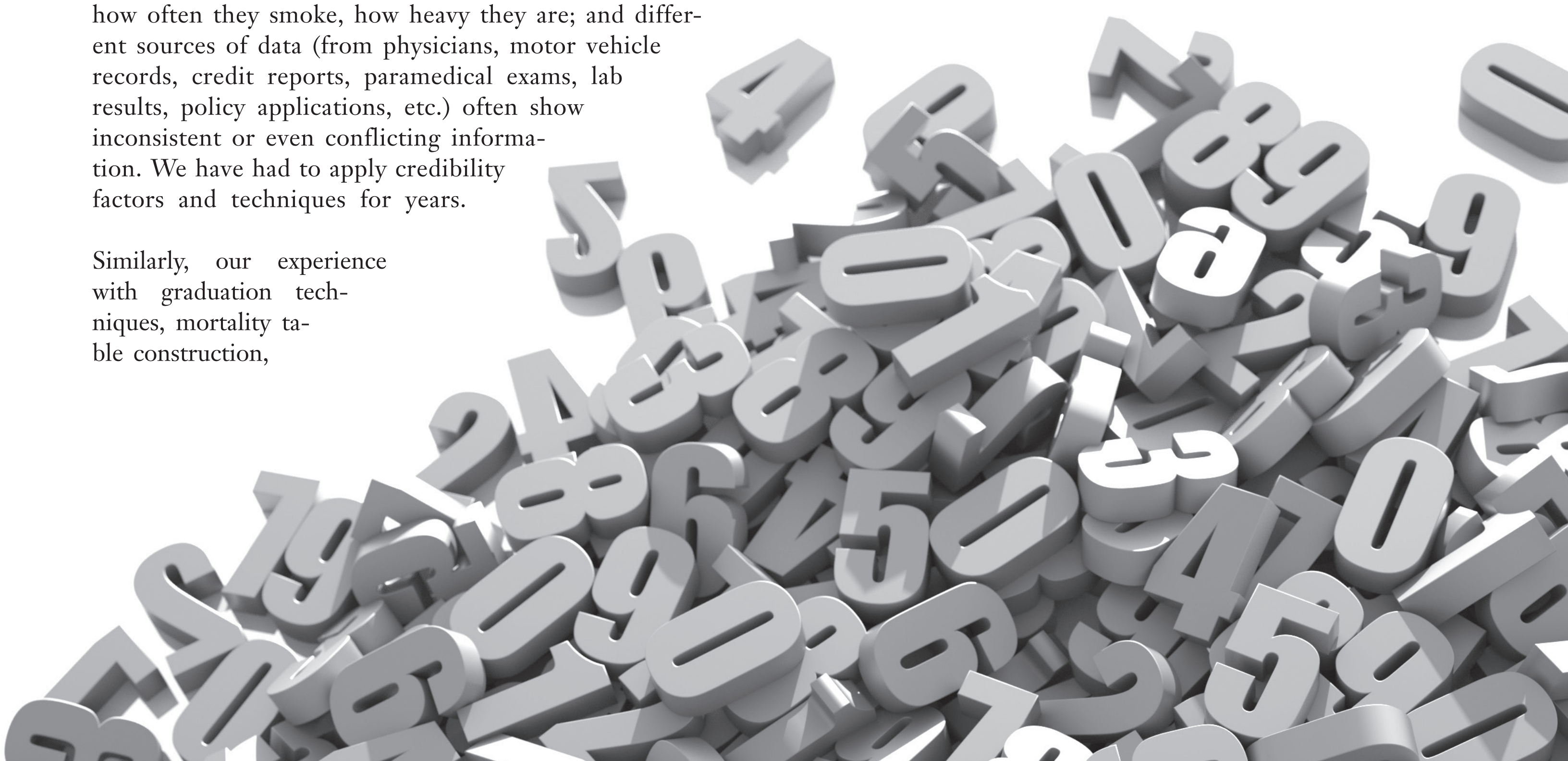
Take the case of velocity. The data may be coming in so fast that by the time we count it, the count has increased. In these situations, we could throw up our hands and say that a mean, a standard deviation, and a random sample are impossible to calculate. Alternatively, we can use stream algorithms, Reservoir sampling, and other algorithms to compute stats on the fly based on the data received to date, and then project the trends. This is conceptually similar to the rolling average that actuaries have used for decades in their experience studies.

Volume has always been a concern for actuaries. Before computers were fast enough to process a block of business on a seriatim basis, we had to employ grouping and sampling techniques. Likewise, the variety of various types of policy benefits (consider disability income policies with their differing benefit periods, definition of disability, elimination periods, waiting periods, occupational classes, etc.) required classification techniques, and some of what data scientists now call feature engineering. Veracity has always been a challenge. Insurance applicants understate the amount of alcohol they consume, how often they smoke, how heavy they are; and different sources of data (from physicians, motor vehicle records, credit reports, paramedical exams, lab results, policy applications, etc.) often show inconsistent or even conflicting information. We have had to apply credibility factors and techniques for years.

Similarly, our experience with graduation techniques, mortality table construction, demography, complex variables, stochastic (and stochastic on stochastic) projections, multivariate contingency analysis, and many other 'standard' components of the actuarial education can be applied to work with big, or infinite, data. Yes, we may have to learn some new names for techniques we already know. Yes, we may have to supplement those techniques with more current research. Yes, we may have to gain a comfort level with some data science tools such as R and Python and others beyond our basic Excel models (although, Excel is a lot more impressive in this arena than most data scientists assume; and actuaries are often experts using it). Yes, yes, yes. We cannot just rest on previous accomplishments and expect to compete on future opportunities. Please read the following article, by Dihui Lai and Richard Xu, about tools such as Spark, to help with the processing speed and volume issues.

The bottom line is that actuaries are entering a new era where they can be pioneers and leaders and highly valued; or they can be followers and Luddites and marginalized. The choice is ours; but only if we are willing to learn to count beyond "many." ■

Dave Snell, ASA, MAAA, is technology evangelist at RGA Reinsurance Company in Chesterfield, Mo. He can be reached at *dave@ActuariesAndTechnology.com.*



Spark: the Next-generation Processing Engine for Big Data

By Dihui Lai and Richard Xu

The concept of "information explosion" was formed more than 70 years ago and the world of big data has evolved ever since. As pointed out by Eric Schmidt (Google CEO), every two days we are creating as much information as we did since the dawn of civilization till 2003. The ever increasing information size has changed the way we store and process data.

Until recently, Hadoop has almost been a paraphrase of big data. The system, famous for its HDFS (Hadoop Distributed File) storage and Map-Reduce processing, has been widely adopted as a tool for big data in IT, health care, financial services, telecommunication and life sciences.

However, the Map-reduce paradigm is not designed for data processing that requires cyclic data sharing, e.g., iterative data processing and interactive data analysis.[1] The invention of Spark, an in-memory data processing engine, seems to bring a solution to the problem.[2] In 2014, Spark was announced as the top-level project of the apache software foundation. Cloudera, a major vendor of Hadoop, considers Spark as a replacement for MapReduce for data execution in their data management system.[3] Spark is also embraced by many big data solution vendors, e.g., Hortonworks, IBM, MapR, etc.

This article gives a general introduction to Spark and shows evidences that Spark could be potentially used as a data processing engine for the insurance industry as well.

DATA STRUCTURES IN SPARK

The core abstraction upon which Spark is built is called Resilient Distributed Dataset (RDD). Basically, RDD is an immutable collection of partitioned records that can be distributed across the nodes of a cluster and operated in parallel. Each partition is a subset of the data it is created from and RDD contains the information on how the partitions are created. If some partitions get lost during a process, they can still be recreated from the original dataset. Therefore, if any nodes in a cluster go down during a large data process, a reconstruction process will be triggered for the lost partition to ensure a successful completion.

Despite its beauty in processing big data, RDD is still a little distal from the data structures that people are familiar with, e.g., SQL schema, data frame. The recent release of Spark introduces DataFrame[4] into its ecosystem. The columnar organized data structure is conceptually similar to a data-frame in R and it also offers relational data operations like SQL. The following are spark codes (in Scala) that create DataFrame from a csv file and perform aggregation on claim counts by states. The 3rd statement appends the claim to the data set by join operations.

```
val df = sqlContext.read.format("com.databricks.spark.
csv").option("header", "true").load("claim_data.csv")

val df_claim_state = df.groupBy("State").agg(count("CLAIM_
CNT"))

val df = df.join (df_claim_state, df("State") === df_
claim_state ("State"), "inner")
```

Equivalently, one can also register the data frame as a SQL table and use SQL-like syntax to do the joining operations, as shown below.

```
df. registerTempTable("claim_data")

sqlContext.sql ("select * from claim_data JOIN
df_claim_state WHERE State claim_data.State = df_

claim_state.State")
```

Both RDD and DataFrame use lazy execution, which means all the operations above will not be executed until some special commands are made, e.g., save, show. The laziness of spark reduces the communication overhead and allows optimization across operations.

SPARK DEPLOYMENT

Spark allows different modes of deployment. It could be deployed with a cluster manager system, e.g., Hadoop YARN, Amazon EC2 or Apache Mesos. Spark also allows a standalone mode by which it can work independent of any cluster management system. It is also possible to run spark on a laptop as a single-node cluster.

The standalone mode is ideal for users to dive into Spark without worrying about the setup of a complicated cluster system. Actually, the standalone mode itself provides a quite powerful tool in dealing with large data of reasonable size. Powerful big data storage systems like HDFS are not necessary for Spark to work. A shared file system, e.g., network file system (NFS) works well for processing data of a large size, e.g., gigabytes as long as the data can fit into a single disk. Spark also integrates well with various databases, including the ones from the NoSQL family, e.g., Cassandra, Hbase and also the ones from the relational database family, e.g., MySQL.

MACHINE LEARNING AND ANALYTICS

As mentioned before, one major limitation of Hadoop's Map-reduce method is that it is not designed for analytics such as machine learning (ML). The in-memory architecture of Spark introduces a nice solution to the problem. By keeping data in memory, Spark allows the users to query data repeatedly and speed up the iterative ML algorithms to a large extent.

Moreover, Spark provides the users with a built-in machine learning library (MLlib). The library covers quite a list of popular ML algorithms, which includes regression (linear/logistic), classification (SVM, naïve Bayes) and clustering (k-means), etc. If one needs an algorithm that is beyond the scope of MLlib, Sparkling-water would be a nice package to add. Sparkling-water is created upon the integration of spark platform and H2O. H2O provides scalable predictive analytics in a wide spectrum, including Generalized Linear Models (GLM), tree algorithms, Gradient Boosting Machine (GBM), Deep Learning, etc.

SPEED AND SCALABILITY

Existing analytical tools such as R or Python do not provide parallel computation for free and are not scalable inherently. Revolution R provides parallelized algorithms but could be very expensive to deploy in a cluster environment. Spark provides an open source platform for analytics and can process large data that could be otherwise hard to handle. Moreover, Spark provides an API for Java, Scala, Python and R. Data scientists who are more familiar with R/Python can dive into the system without much pain.

Generalized linear model (GLM) is widely used in the insurance industry. To understand the potential usage of Spark in insurance, we built GLMs for data of varying sizes and compared the performance difference using terminal server and spark cluster. The terminal server and the spark cluster (7 nodes) had comparable memory sizes in our test. Revolution R showed better performance than regular R regarding the processing speed due to optimized algorithm and parallelism (Table 1). The spark cluster reduced the processing time of the model further due to the involvement of more CPUs. In processing a large data set, the terminal server experienced a memory overflow in processing data 70GB in size while the spark cluster finished the model in about three minutes.

Table 1. Processing data on terminal server and cluster. The processing time of different data on terminal server and spark cluster. Generalized Linear Model is built within two environments on data of different sizes. * we used a GLM routine from the H2O package in the spark cluster.

	Proc Time (Data 1.5 GB)	Proc Time (Data 70 GB)
R (TS)	480.19 s	Memory Overflow
Revolution R(TS)	33 s	Memory Overflow
Spark* (Cluster)	6 s	184 s

To further test the scalability of the spark cluster, we built a GLM model on the same data set while changing the size of the cluster. The processing speed changed depending on the data type and the complexity of the model. Under the given test environment, the spark-cluster showed a close-to-linear scalability where the increase of processing speed was almost proportional to the size of the cluster (Table 2). A 2-nodes cluster failed the task due to memory overflow.

Table 2. Scalability. Comparison between processing time on clusters of different sizes.

Cluster	2-nodes (8 cores)	4-nodes (16 Cores)	7-nodes (28 Cores)
Proc Time	Memory Overflow	300 s	180 s

In summary, spark provides a fast and scalable platform for handling big data. Its in-memory architecture makes it a nice fit for big data analytics. The APIs for multiple languages make it easy to dive-in from various backgrounds. The various deployment modes make it easy to implement into existing big data environments. ■

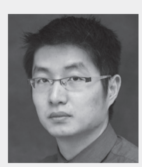
REFERENCES

1 Zaharia M., Chowdhury M., Franklin M.J., Shenker S. and Stoica I., Spark: Cluster Computing with Working Sets. In *HotCloud*, 2010.

2 Zaharia M., Chowdhury M., Das T., Dave A., Ma J., McCauley M., Franklin M.J., Shenker S. and Stoica I.; Resilient Distributed Datasets: A Fault-Tolerant Abstraction for In-Memory Cluster Computing; In NSDI, 2012.

3 *http://www.cloudera.com/content/cloudera/en/products-and-services/cdh/spark.html*

4 Bradley J.K., Meng X., Kaftan T., Franklin M.J., Ghodsi A. and Zaharia M. Spark SQL: Relational data processing in Spark; In Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD '15), 2015.



Dihui Lai, Ph.D., is a data scientist analyst at RGA Reinsurance Company in Chesterfield, Mo. He can be reached at *dlai@rgare.com.*

Richard Xu, FSA, Ph.D., is VP and actuary, head of Data Science at RGA Reinsurance Company in Chesterfield, Mo. He can be reached at *rxu@rgare.com.*

The Third Generation of Neural Networks

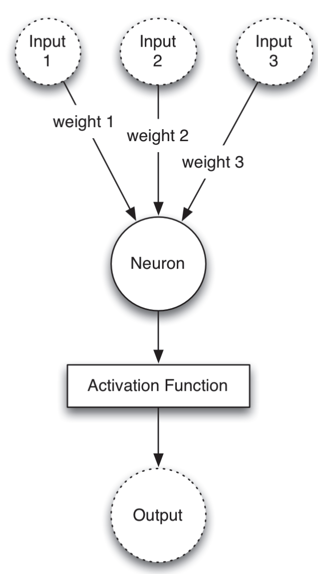
By Jeff Heaton

Neural networks are the phoenix of artificial intelligence. Right now neural networks are rising from the ashes for the third time since their introduction in the 1940s. There are many design decisions that a neural network practitioner must make. Because of their long history, there is a diverse amount of information about the architecture of neural networks. Because neural networks have essentially been invented (and reinvented) three times, much of this information is contradictory. This article presents what most current research dictates about how neural networks should be architected in 2015.

The goal of this article is to provide a tour of the most current technologies in the field of neural networks. A rigorous discussion of why these methods are effective is beyond both the scope, and space requirements, of this article. However, citations are provided to lead you to papers that provide justifications for the architectural decisions advocated by this article.

At the most abstract level, a neural network is still the weighted summation of its inputs, applied to an activation/transfer function, as shown in Figure 1.

FIGURE 1: SINGLE UNIT OF A NEURAL NETWORK



The above unit is still calculated using Equation 1, which has been the same formula since the first generation of neural networks.

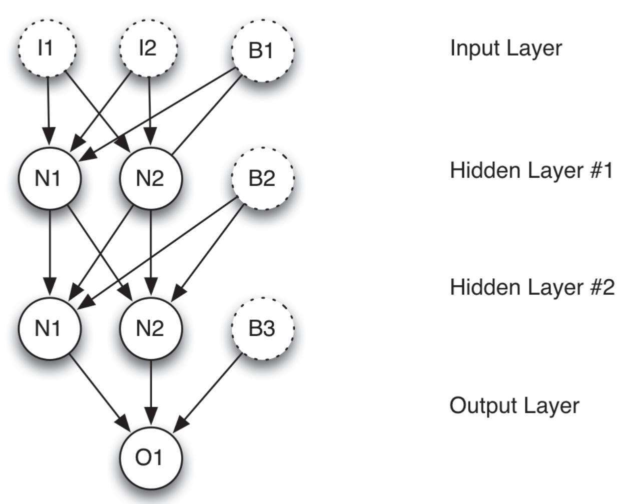
Equation 1: Neural Network Calculation

$$f(x_i, w_i) = \phi(\sum_i (w_i \cdot x_i))$$

The neural network output vector is dependent upon the input vector (x), the weights (w), and choice of activation function (phi, φ). Most implementations also use bias neurons that essentially become the y-intercept. To implement bias, most neural networks add a one to the x-vector and the bias-value to the weight vector. These values are both added at the beginning of these vectors. This is effectively the same as adding the bias/intercept term to the equation with a coefficient of one.

When these units are connected together, third generation neural networks still look the same as before. Figure 2 shows a two-input, single output neural network with two hidden layers.

FIGURE 2: MULTILAYER FEEDFORWARD NETWORK



The above diagram shows how the biases (indicated by B's) are added to each of the layers.

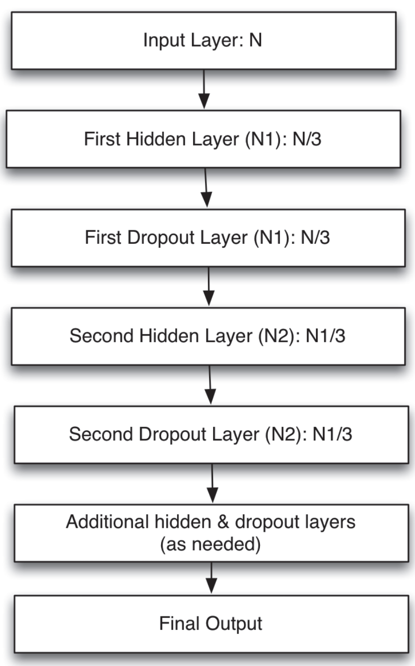
NUMBER OF LAYERS

How many layers and how many hidden neurons has always been the primary question of the neural network practitioner. There is research that indicates that a single hidden layer in a neural network can approximate any function (Hornik, 1991). Because of this it is extremely tempting to use a single hidden layer neural network for all problems. For several years, this was the suggested advice. However, just because a single layer network can, in theory, learn anything, the universal approximation theorem does not say anything about how easy it will be to learn. Additional hidden layers make problems easier to learn because they provide the hierarchical abstraction that is an inherent component in the human neocortex. Additional hidden layers are great, but the problem has been that we had no means of training such deep networks.

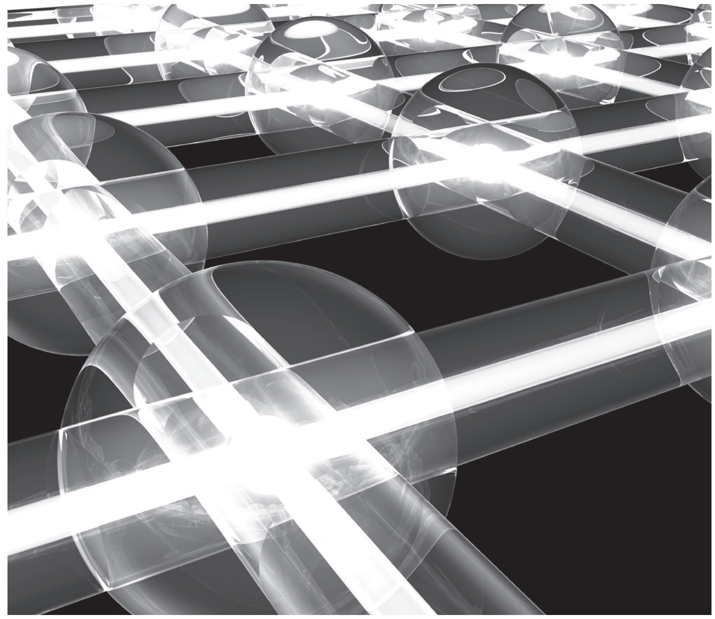
Deep learning is a very general term that describes a basket of technologies that allow neural networks, with more than two hidden layers, to be trained. Initially methods were discovered to train a deep belief neural network (DBNN), using clever techniques based on Gibbs sampling.[1] However, DBNN's can only accept binary inputs for classification. DBNN's showed the potential of deep learning and further research discovered the changes necessary to allow regular deep feedforward neural networks to be trained as well.

A deep modern neural network appears in Figure 3.

FIGURE 3: DEEP NEURAL NETWORK



The above diagram shows how additional pairs of hidden and dropout layers are added. These dropout layers, which help to avoid overfitting, will be discussed later in the article. Hidden layers and dropout layers usually occur in pairs. These hidden layers are often called dense layers, because every neuron is connected to the next layer. Prior to the third generation of neural networks, every layer was dense. Dropout layers are not dense, as will be demonstrated later. You will also notice that the layers of the neural network decrease in their number of neurons. This forces the neural network to learn more and more abstract features of the input as the layers become deeper.



HIDDEN ACTIVATION FUNCTIONS

For years the choice of activation function for the hidden layers of a neural network was a choice between the two most common sigmoidal functions: the logistic and the hyperbolic tangent. Unfortunately, all sigmoidal (s-shaped) activation functions are difficult to train for deep neural networks. Because of this sigmoidal activation functions have largely fallen out of favor for neural networks. The activation function that has replaced them is the rectified linear unit (ReLU). The very simple equation for the ReLU is shown in Equation 2.

Equation 2: Rectified Linear Unit (ReLU)

$$\phi(x) = \max(0, x)$$

There are many papers written that provide more rigorous (Nair & Hinton, 2010) descriptions of the superiority of the ReLU activation function than I will give here. One obvious advantage to the ReLU is that the range of the function is not squashed to values less than one. This frees the practitioner of many of the data normalization requirements typically associated with neural networks. However, the true superiority of the ReLU comes from

CONTINUED ON PAGE 38

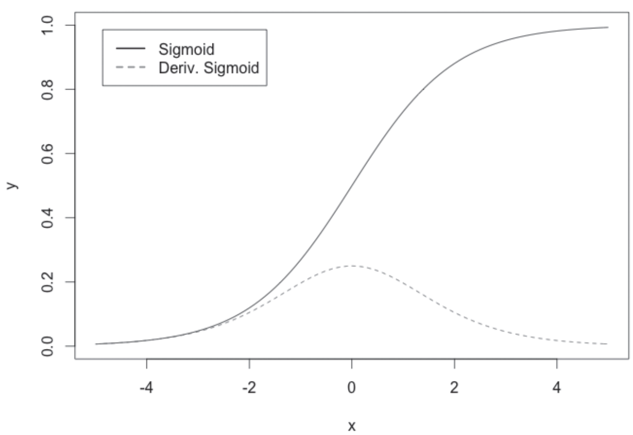
the somewhat contrived derivative of the ReLU, which is shown in Equation 3.

Equation 3: Generally Accepted Partial Derivative of the ReLU

$$\frac{dy}{dx}\phi(x) = \begin{cases} 1 & x > 0 \\ 0 & x \leq 0 \end{cases}$$

Technically the ReLU does not have a derivative at x=0; however, most neural network implementations simply treat this undefined value as 0. Figure 4 shows the derivatives of the less effective logistic and hyperbolic tangent activation functions.

FIGURE 4: SIGMOIDAL (S-SHAPED) ACTIVATION FUNCTION DERIVATIVES



The above graph shows both the logistic (sigmoid) and its derivative. The hyperbolic tangent function would look similar but shifted. The shape of the derivative indicates the problem in both cases. S-shaped activation functions saturate to zero in both directions about the x-axis. This is sometimes referred to as the vanishing gradient problem. This can cause the gradients, as calculated by the derivatives, for these neurons to drop to zero as the absolute values of these x-values become more extreme. Once the gradient for a neuron flattens to zero, the neuron will no longer train. The neural network training algorithms use this gradient to indicate what direction to move the weights.

OUTPUT ACTIVATION FUNCTIONS

Traditionally, the output layer of a neural network would use either the sigmoid, hyperbolic tangent, linear or softmax for the output activation function. Many of these choices have fallen out of favor (A. Maas, A. Hannun, A. Ng, 2014). For a regression model a linear output function should be used, for a classification model, the softmax function should be used. Never use a ReLU as the output layer activation function.

The softmax activation function is very advantageous for a classification problem. Consider a classification with five classes. Such a problem is represented by a five output neuron network. If the neural network were to output the vector [.5, 0.1, 0.75, 0.1, 0.2] you would know that the neural network had selected the third class (indicated by 0.75) as the prediction. However, 0.75 is not the probability, it is simply the largest value. The softmax activation function forces these outputs to sum to one, giving the predicted probability of the data representing each class. The softmax activation function is shown in Equation 4.

Equation 4: The Softmax Activation Function

$$\sigma(\mathbf{z})_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}} \quad \text{for } j = 1, \ldots, K.$$

Essentially you divide the natural exponent of each of the elements by the sum of all natural exponents. The value K above represents the number of output neurons present. For the vector presented above, the logloss would be [0.23 , 0.15 , 0.29 , 0.15 , 0.17]. The following URL provides a utility to calculate softmax.

http://www.heatonresearch.com/aifh/vol3/softmax.html

WEIGHT INITIALIZATION

Neural networks are initialized with random weights and biases. This creates inherently unpredictable results. This can make it very difficult to evaluate different neural network architectures to see which works best for the task at hand. While random number seeds can help produce consistent results, it is still very difficult to evaluate two different networks that have different numbers of weights. One of your candidate architectures might owe its perceived superiority more to its starting weights than the actual structure.

The Xavier weight initialization algorithm (Glorot & Bengio, 2010) has become the standard in weight initialization for neural network. This initialization samples the weights from a normal distribution with a mean of zero and a variance specified by Equation 4.

Equation 4: Xavier Weight Initialization

$$Var(W) = \frac{}{n_{in} + n_{out}}$$

The variance is equal to two divided by the sum of the number of input and output neurons for the layer. The weights resulting from

Xavier create neural networks that converge much faster than other initialization techniques. Additionally, these weight sets produce much more consistent results than many of the other weight initialization techniques.

STOCHASTIC GRADIENT DESCENT TRAINING

Stochastic Gradient Descent (SGD) with Nesterov momentum (Nesterov, 1983) has become the most commonly used training algorithm for neural networks. SGD is very similar to standard batch back propagation. Back propagation works by calculating the partial derivative of the neural network's error function for each weight. The derivatives, called gradients, are scaled by a learning rate and then added to the weights of the neural network. The gradient can be used to maximize the error of the neural network, using gradient ascent. Because we seek to minimize the error of the neural network we use the inverse of the gradient and descend to lower error levels.

Usually these changes to the weights are not applied immediately. Rather, a batch of training set elements is calculated and their gradients are summed. Once the batch is complete the weights are modified. SGD is exactly like regular batch back propagation except that a small batch size of 100-1000 elements is used. This smaller batch size is called a mini-batch. Additionally, the mini-batch is randomly sampled from the training set, with replacement. This random sampling greatly decreases overfitting.

The actual update to the weights is performed using Nesterov momentum. This is a technique that was invented by Nesterov (1983) as a general-purpose gradient descent technique. Geoffrey Hinton later recognized its value to neural network training. Nesterov momentum is a mathematically complex technique that I will not fully describe in this article. Nesterov momentum seeks to limit the damage to the weights that can be done by choosing a particularly bad mini-batch from the training elements.

CROSS ENTROPY

Neural network training algorithms have traditionally calculated error as the difference between the output neuron's actual output and expected output. This is called the quadratic error function. Research from Geoffrey Hinton has caused the quadratic error function to fall from favor. The replacement is the cross entropy function, which is shown in Equation 5.

Equation 5: Cross Entropy Error

$$CE = -\frac{1}{n}\sum_x [y \ln a + (1-y)\ln(1-a)]$$

In the above equation the number of training elements (n), the actual output (a) and the expected output (y) are used. The cross entropy function forces much steeper gradients for larger errors. These larger gradients cause the weights to be adjusted much faster when the error is greater and in turn causes the neural network training to converge to a lower error quicker.

L1 AND L2 REGULARIZATION

Regularization seeks to prevent overfitting by directly adjusting the weights of a neural network. The most common types of regularization are L1, L2 and dropout. The first two, L1 and L2 work by adding the neural network weights, but not the biases, to the error function. This encourages the training to keep the weights lower. This is a form of Occam's razor, in that simple weight structures are likely superior. The only differences between L1 and L2 are how they apply the weight penalty. L1 is shown in Equation 6.

Equation 6: L1 Regularization

$$E_1 = \lambda_1 \sum_w |w|$$

The parameter λ1 represents the relative importance of L1, a value of 1.0 means that the L1 regularization penalty is just as important

CONTINUED ON PAGE 40

as the actual error of the neural network. A value of zero turns off L1 regularization. In practice, L1 values are very low, typically less than a hundredth.

You should use L1 regularization to create sparsity in the neural network. In other words, the L1 algorithm will push many weight connections to near zero. When a weight is near zero, the program drops it from the network. Dropping weighted connections will create a sparse neural network.

Feature selection is a useful byproduct of sparse neural networks. Features are the values that the training set provides to the input neurons. Once all the weights of an input neuron reach zero, the neural network training determines that the feature is unnecessary. If your data set has a large number of input features that may not be needed, L1 regularization can help the neural network detect and ignore unnecessary features.

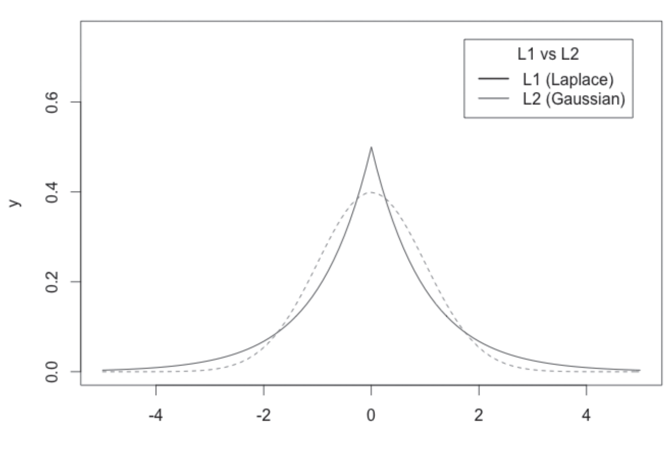
L2 regularization works similar to L1, except there is less of a focus on the removal of connections. L2 is implemented using Equation 7.

Equation 7: L2 Regularization

$$E_2 = \lambda_2 \sum_w w^2$$

The primary difference between L1 and L2 is that L1 uses the absolute value of the weights, whereas L2 uses their square. Both L1 and L2 work differently in the way that they penalize the size of a weight. L1 will force the weights into a pattern similar to a Gaussian distribution; the L2 will force the weights into a pattern similar to a Laplace distribution, as demonstrated by Figure 5.
As you can see, the L1 algorithm is more tolerant of weights further from zero, whereas the L2 algorithm is less tolerant. We will

FIGURE 5: L1 VS L2



highlight other important differences between L1 and L2 in the following sections. You also need to note that both L1 and L2 count their penalties based only on weights; they do not count penalties on bias values.

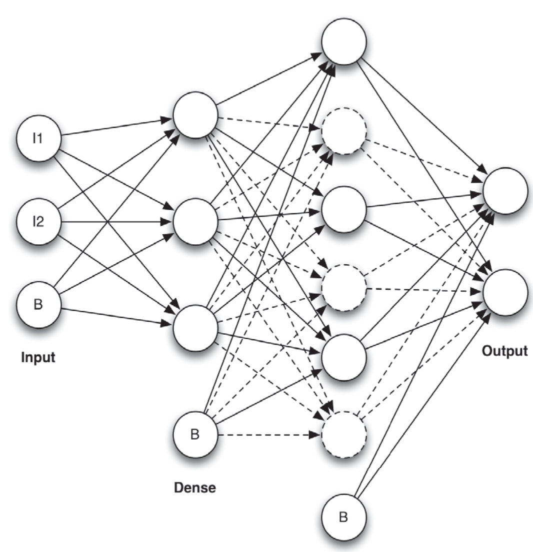
DROPOUT FOR REGULARIZATION

Hinton, Srivastava, Krizhevsky, Sutskever, & Salakhutdinov (2012) introduced the dropout regularization algorithm. Although dropout works in a different way than L1 and L2, it accomplishes the same goal—the prevention of overfitting. However, the algorithm goes about the task by actually removing neurons and connections—at least temporarily. Unlike L1 and L2, no weight penalty is added. Dropout does not directly seek to train small weights.

Most neural network frameworks implement dropout as a separate layer. Dropout layers function as a regular, densely connected neural network layer. The only difference is that the dropout layers will periodically drop some of their neurons during training. You can use dropout layers on regular feedforward neural networks. Figure 6 shows dropout in action.

The above neural network has two input neurons and two output neurons. There is also a dense and dropout layer. For each training

FIGURE 6: DROPOUT



iteration, a different set of hidden neurons is temporally dropped from the dropout layer. The dashed lines indicate the dropped neurons, and their connections. The bias neuron is never dropped. When a neuron drops out, so does its connections. Training is performed as though the dropped out neurons are not present. This forces the neural network to learn to perform even without a full complement of neurons. The neurons become less dependent on each other.

OTHER TYPES OF NEURAL NETWORKS

It is a very exciting time for neural network research. Additional types of neural networks are actively being developed. This article focused primarily upon feedforward neural networks. However, other types of neural networks are very common. Convolutional neural networks (CNN) have become very popular for image recognition. Recurrent neural networks, particularly, gated recurrent units (GRU) have become very popular for deep time-series learning. Additionally, spiking neural networks (SNN) have found great application in the field of robotics ■

Jeff Heaton is the author of the Artificial Intelligence for Humans series of books. He is data scientist, Global R&D at RGA Reinsurance Company in Chesterfield, Mo. He can be reached at *jheaton@rgare.com.*

REFERENCES

X. Glorot & Y. Bengio. (2010). Understanding the difficulty of training deep feedforward neural networks. Journal of Machine learning Research.

K. Hornik (1991) Approximation Capabilities of Multilayer Feedforward Networks, Neural Networks, 4(2), 251–257

A. Maas, A. Hannun, A. Ng (2014). Rectifier Nonlinearities Improve Neural Network Acoustic Models

V. Nair, G. Hinton, G (2010). Rectified linear units improve restricted Boltzmann machines (PDF). ICML.

Y. Nesterov. A method of solving a convex programming problem with convergence rate O(1/sqr(k)). Soviet Mathematics Doklady, 27:372–376, 1983.

N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov; Dropout: A Simple Way to Prevent Neural Networks from Overfitting . Journal of Machine Learning Research (JMLR).15(Jun):1929−1958, 2014.

ENDNOTES

[1] Gibbs sampling is a Markov chain Monte Carlo (MCMC) algorithm for obtaining a sequence of observations which are approximated from a specified multivariate probability distribution. *https://en.wikipedia.org/wiki/Gibbs_sampling*

A Comparison of Risk Scoring Recalibration Methods

By Geof Hileman

In the July 2015 issue of this newsletter, Shea Parkes and Brad Armstrong published an article titled "Calibrating Risk Score Models with Partial Credibility." In this article, they presented an application of the "ridge regression" technique to the calibration of health-based risk scoring models. This calibration process is often undertaken to tailor a risk scoring model to a specific population on which it is being applied.

This article's publication was timely, as we are currently engaged in updating the SOA's periodic study that compares the predictive accuracy of various risk scoring models. This study has been published three times previously (in 1996, 2002 and 2007), with the 2007 study available at *https://www.soa.org/research/research-projects/health/hlth-risk-assement.aspx*.

This new study, while currently still underway, will include a comparison of the accuracy of prospective and concurrent models out-of-the-box, using the weights provided by each vendor with the models. In addition to this out-of-the-box comparison, which is consistent with the way models are frequently implemented, we have also considered various approaches to recalibrating each of the models included in the comparison. This step is an important element of model comparison, as the comparison of recalibrated models gives insight into the potential predictive ability of each model and normalizes for any differences in the populations on which the offered weights are based.

We have considered three approaches to recalibrating each of the models in the study. The first approach under consideration is full recalibration. Full recalibration is the approach used in the 1996 and 2002 studies and is the approach that would be considered to be the most conventional, given an adequately large data source. To perform a full recalibration, the actual scaled cost level is regressed against the complete set of independent variables to determine new model weights. A critical disadvantage of the full re-specification approach is that full transparency into the workings of the model is required. In order to implement full re-specification without losing any of the clinical logic, one would need to know all of the inputs to the model, including any hierarchical logic or combination variables. Not all vendors provide this degree of transparency along with their models.

The second approach we considered is that which was used in the development of the 2007 study. The 2007 approach differs from a full recalibration in that the dependent variable in the regression equation is the residual, rather than the scaled cost variable. Thus, the equation resulting from the regression gives the expected error for each individual and can be added to the originally-predicted risk score. Without any consideration for statistical significance, the estimated coefficients from the residual approach are by definition equal to the difference in the original model weights and the weights that result from the full recalibration. The authors of the 2007 study introduced a credibility weighting where each coefficient is weighted by $(1\text{-}p)^{5.95}$, where p is equal to the p-value associated with that particular coefficient. Accounting for the credibility weighting, the adjustment to each individual's estimated risk score is given by the dot product of three vectors: the estimated coefficients, the credibility weights, and the specific values of each of the independent variables.

The third approach is the ridge regression approach discussed by Parkes and Armstrong. Like the p-value approach used in the 2007 study, this method regresses on the residuals rather than the original dependent variable. However, the blending of the original and the re-estimated coefficients is handled in a less blunt fashion. In an ordinary least squares regression, coefficients are determined to minimize the sum of the squared errors across all observations. In ridge regression, the objective function is modified to incorporate a penalty corresponding to the size of the sum of the estimated coefficients. Thus, the optimal weights strike an appropriate balance between fitting the data and minimizing changes to the original coefficients.

One significant advantage offered by both of the residual approaches is that the details behind the original risk model can remain somewhat obscured. Since both approaches produce an estimate of the expected difference from the original risk score, all of the independent variables that contribute to that risk score do not need to be known. Any variables that are omitted from the re-specification would essentially retain their original weight along with any error that their coefficients contribute toward being absorbed by other variables.

In order to determine the most appropriate approach for our application—comparing commercial risk scoring models—we tested each of the methods in a recalibration of the Clinical Illness & Disability Payment System (CDPS) model. CDPS is an ideal model for testing the recalibration approaches, because it is entirely transparent and, by virtue of the offered weights being based on a Medicaid population, recalibration on a commercial population should lead to different weights. We selected two samples of just under 700,000 adults from Truven Health's MarketScan databases and used one for recalibrating the weights and the second for computing statistics on the recalibrated models.

To evaluate the effects of the three approaches to recalibration, we first compared the coefficients produced by each of the three methods. The coefficients, while all somewhat different from the original CDPS weights, were very consistent across the three methods. As expected, the coefficients resulting from the 2007 approach were identical to the full recalibration approach in cases where the p-value was 0.0 (and the credibility was thus 100 percent). Larger differences were present across the approaches for the higher-severity lower-frequency conditions.

We also compared the degree to which the recalibrated models explained the variation in the cost data. Using the original weights, we calculated an R-squared of 11.24 percent. Both the full recalibration and the 2007 residual approach resulted in an identical R-squared of 13.70 percent, while the ridge regression returned a slightly higher value of 13.72 percent. Additionally, we computed the correlation coefficient among the four sets of predicted values, shown below in Table 1.

TABLE 1: CORRELATION COEFFICIENTS AMONG PAIRS OF PREDICTED VALUES

	Original Weights	Full Recalibration	2007 Residual Approach	Ridge Regression
Original Weights	1.00000	0.90455	0.90455	0.91483
Full Recalibration	-	1.00000	0.99995	0.99652
2007 Residual Approach	-	-	1.00000	0.99648
Ridge Regression	-	-	-	1.00000

Based on this comparison, we concluded that the selection of a recalibration method for large populations does not need to be guided by statistical fit, but rather by the constraints imposed by the particular models that are being worked with. The method described in the July 2015 newsletter was specifically recommended as being worthwhile when "trying to recalibrate a model for a population that is of moderate size, but not fully credible." Our analysis supports this conclusion, in that the approach provides an incrementally better fit, but is not meaningfully different from the more simplistic approaches when applied to a very large population. ■

Geof Hileman, FSA, MAAA, is VP at Kennell and Associates Inc., in Raleigh, N.C. He can be reached at *ghileman@kennellinc.com*

SOCIETY OF ACTUARIES

475 N. Martingale Road, Suite 600
Schaumburg, Illinois 60173