NYU Data Science Community features journalism, research papers, events, tools/software, and jobs for December 18, 2015.

Please let us (Laura Noren, Brad Stenger) know if you have something to add to the newsletter. We are grateful for the generous financial support from the Moore-Sloan Data Science Environment and to NYU's Center for Data Science.

NYU Data Science Community Newsletter Issue 032.

Data Science News

Human-level concept learning through probabilistic program induction

*Science*; Brenden M. Lake, Ruslan Salakhutdinov and Joshua B. Tenenbaum from December 11, 2015

People learning new concepts can often generalize successfully from just a single example, yet machine learning algorithms typically require tens or hundreds of examples to perform with similar accuracy. People can also use learned concepts in richer ways than conventional algorithms—for action, imagination, and explanation. We present a computational model that captures these human learning abilities for a large class of simple visual concepts: handwritten characters from the world’s alphabets. The model represents concepts as simple programs that best explain observed examples under a Bayesian criterion. On a challenging one-shot classification task, the model achieves human-level performance while outperforming recent deep learning approaches. We also present several "visual Turing tests" probing the model's creative generalization abilities, which in many cases are indistinguishable from human behavior.

More from the newspapers:
Humans take note: Artificial intelligence just got a lot smarter, *Los Angeles Times*
A Learning Advance in Artificial Intelligence Rivals Human Abilities, *The New York Times*

Elon Musk Snags Top Google Researcher for New AI Non-Profit

*WIRED, Business, Cade Metz* from December 11, 2015

Tesla founder Elon Musk, big-name venture capitalist Peter Thiel, LinkedIn co-founder Reid Hoffman, and several other notable tech names have launched a new artificial intelligence startup called OpenAI, assembling a particularly impressive array of AI talent that includes a top researcher from Google. But the idea, ostensibly, isn’t to make money.

Overseen by ex-Googler Ilya Sutskever and Greg Brockman, the former CTO of high-profile payments startup Stripe, OpenAI has the talent to compete with the industry’s top artificial intelligence outfits, including Google and Facebook—but the company has been setup as a non-profit. “Our goal is to advance digital intelligence in the way that is most likely to benefit humanity as a whole, unconstrained by a need to generate financial return,” Brockman said in a blog post.

More OpenAI coverage:
AI’s Real Risk, Michael Schrage in *Harvard Business Review*
Why OpenAI Matters, Miles Brundage
Inside Deep Dreams: How Google Made Its Computers Go Crazy
Medium, Backchannel, Steven Levy from December 11, 2015
Why the neural net project creating wild visions has meaning for art, science, philosophy—and our view of reality.

Ten Deep Learning Trends at NIPS 2015
Brad Neuberg, coding in paradise blog from December 13, 2015
I attended the Neural Information Processing Systems (NIPS) 2015 conference this week in Montreal. It was an incredible experience, like drinking from a firehose of information. Special thanks to my employer Dropbox for sending me to the show (we're hiring!)

Here's some of the trends I noticed this week; note that they are biased towards deep and reinforcement learning as those are the tracks I attended at the conference:

1) Neural network architectures are getting more complex and sophisticated

More reports from NIPS 2015:
My takeaways from NIPS 2015, Dan Vanderkam
Interesting things at NIPS 2015, John Langford

Study Finds Economic Stimulus from Research Investments and PhD Recipients’ Earnings
NYU News, Julia Lane from December 10, 2015
A decade ago the late Jack Marburger, a physicist and former college president who served as science advisor to President George W. Bush, challenged academics to come up with scientific evidence on the impact of federal research investment. Even more adamantly, Congress required the National Science Foundation to “better articulate the value of grants to the national interest.”

A study in the latest issue of Science (online Dec. 10, in print Dec. 11) led by Professors Julia Lane of New York University, Bruce Weinberg of Ohio State University and Jason Owen Smith of the University of Michigan demonstrates a significant path by which federally and non-federally funded investment in research make an impact on the economy. Inspired by the Robert Oppenheimer maxim that “The best way to send information is to wrap it up in a person,” the researchers combed an array of new data, combined with Census Bureau information, to trace where doctoral recipients get their jobs and what their earnings are after receiving research training.

Next Steps for Data Visualization Research
Medium, UW Interactive Data Lab from December 17, 2015
Given its youth and interdisciplinary nature, research methods and training in the field of data visualization are still developing. So, we asked ourselves: what steps might help accelerate the development of the field? Based on a group brainstorm and discussion, this article shares some of the proposals we found most promising. We hope that other researchers will join us in ongoing discussion and experiment with new approaches.

More info viz:
Country Collaboration, Nature Index
A Day in the Life of Americans, Nathan Yau at Flowing Data blog
Change how you see the countries of the world: This time with infographics, Dan Goldstein and
The Guinness Brewer Who Revolutionized Statistics

Pricenomics, Dan Kopf from December 11, 2015

One of the greatest minds in 20th Century statistics was not a scholar. He brewed beer.

Guinness brewer William S. Gosset’s work is responsible for inspiring the concept of statistical significance, industrial quality control, efficient design of experiments and, not least of all, consistently great tasting beer.

Music in the brain

MIT News, Josh McDermott from December 16, 2015

Scientists have long wondered if the human brain contains neural mechanisms specific to music perception. Now, for the first time, MIT neuroscientists have identified a neural population in the human auditory cortex that responds selectively to sounds that people typically categorize as music, but not to speech or other environmental sounds.

“It has been the subject of widespread speculation,” says Josh McDermott, the Frederick A. and Carole J. Middleton Assistant Professor of Neuroscience in the Department of Brain and Cognitive Sciences at MIT. “One of the core debates surrounding music is to what extent it has dedicated mechanisms in the brain and to what extent it piggybacks off of mechanisms that primarily serve other functions.”

The finding was enabled by a new method designed to identify neural populations from functional magnetic resonance imaging (fMRI) data. Using this method, the researchers identified six neural populations with different functions, including the music-selective population and another set of neurons that responds selectively to speech.

University of Washington adds data science master’s program to meet demand in job market

GeekWire from December 15, 2015

With the explosion of cloud computing, machine learning and other data-intensive projects throughout the tech field, companies are struggling to find enough data scientists to fill open jobs. A new master’s program from the University of Washington aims to fill that gap.

The new Master of Science in Data Science program was designed by faculty from six UW departments, but the university says it also took into account what companies are actually looking for in their next hire.

Clippy’s Revenge

Medium, Sarah Guo from December 16, 2015

A number of people have written about why messaging matters, from essays on conversational commerce, to lists of invisible apps, to coining the term Assistant-as-a-Service. … It seemed like a good time to jot down some of my thoughts around: why messaging is an increasingly dominant way to engage on smartphones—with people, with service providers, and with software

where I see opportunity for new players vs. platform expansion
whether the economics are good
who's playing for the opportunity now

**UW Announces Two New Positions in Support of Data-Intensive Discovery Initiative**

*UW eScience Institute* from December 14, 2015

Shortly after being named University of Washington Provost, Anne Marie Cauce (now the UW’s president) asked eScience Founding Director Ed Lazowska and Steering Committee member Werner Stuetzle to advise on additional steps to undertake to ensure that the University of Washington was a leader in data-intensive discovery. Under the Provost’s Initiative in Data-Intensive Discovery, the Provost will consider providing 50% support on an ongoing basis for faculty hires who excel both in advancing data science methodology and in data-intensive discovery in some field.

**A business plan, not a curriculum, is key when building a university innovation center**

*Medium, Harvard Innovation Lab* from December 14, 2015

At a recent conference on student entrepreneurship in higher education, I was surprised by how many of my peers were struggling to make university-led innovation programs work within academic settings.

The major problem was that they were approaching their problems like school administrators and not like an entrepreneur.

Higher education initiatives aimed at promoting innovation, entrepreneurship, and small business development need to be run like a venture, not like the English department.

**Data Carpentry to adopt Reproducible Research Curriculum**

*Data Carpentry* from December 17, 2015

Part of the mission of Data Carpentry is to encourage and enable reproducible research. The core Data Carpentry curriculum teaches researchers approaches and skills that are fundamental to reproducible research, such as scripting and data management. We are also adopting a Reproducible Research curriculum that explicitly focuses on reproducible techniques and some of the next steps, including version control and data publishing. This is an update on the efforts on this curriculum so far, and we expect to have it available soon as a Data Carpentry workshop option.

**Events**

**BIDS Docker Workshop**

This hands-on workshop will introduce attendees to the basics of Docker, a containerization technology that can be used for running, wrapping, distributing, and publishing scientific software and workflows. Docker containers provide a lightweight computational framework for all of these things, and we intend that this workshop will provide a friendly and welcoming environment for trying it out.

Thursday-Friday, January 7-8, 2016, at 190 Doe Library

**Steve Blank on Hacking NYC: Beyond the Rise of the NYC Startup Ecosystem**

The current state of NYC’s entrepreneurship ecosystem is widely considered a modern
marvel. ... How much of this success can be traced back to Mayor Bloomberg’s master plan? How has this progress evolved under Mayor de Blasio? How has the NYC story matched or diverged from other startup hotspots like The Valley, Boston and Tel Aviv? What might we see emerge in the years ahead?

Tuesday, January 12, at 6:30 p.m., 620 Eighth Ave. 26th Floor in the Offices of Goodwin Procter.

**JupyterDay Chicago 2016**
We are pleased to announce our second JupyterDay Workshop, in Chicago on February 20th, 2016 from 8:30am-6:30pm.

Saturday, February 20, at Civis Analytics in Chicago

**Deadlines**

**Predicting The Oscars Is Hard. Do You Have The Solution?**
We’re calling for submissions from people who have developed ways to predict the Oscars using data. From the submission pool, we’ll talk to a few modelers who attack Oscar prediction from different angles and then highlight them over the course of our awards season coverage. ... If you’d like to submit an entry, please email the following information to walter.hickey@fivethirtyeight.com by Dec. 31 with the subject line “Oscar Model Pitch”:

- A one-paragraph description of your approach
- A one-paragraph bio of you or your team
- A list of the data that you have assembled for your model so far
- Data you will be able to collect throughout the process
- Data you may need help obtaining before or during the process

Deadline for submission is Thursday, December 31, 2015.

**Announcing the 2016 Moore/Sloan Data Science and WRF Innovation in Data Science Postdoctoral Fellowships**
The University of Washington eScience Institute announces a competition for the 2016 Moore/Sloan Data Science and Washington Research Foundation Innovation in Data Science Postdoctoral Fellowships. We seek outstanding interdisciplinary researchers with expertise in the methods of data science and in a physical, life, or social science.

Deadline for submission materials is Friday, January 15, 2016.

**SIAM Workshop on Network Science (NS16), July 15-16 in Boston**
We invite contributions focused on all aspects of mathematical, algorithmic, and computational techniques in network science. The workshop will feature three formats for contributed presentations: 25-minute talks, 5-minute lightning talks in the “Ignite” format (5 minutes, 20 slides, 15-second auto-advance), and posters.

Submission Deadline: Thursday, February 25, 2016
The Data Science Bowl
Enter the Data Science Bowl, co-sponsored by Kaggle and Booz Allen in partnership with The National Heart, Lung, and Blood Institute (NHLBI), part of the National Institutes of Health. The challenge? Develop an algorithm to empower doctors to more easily diagnose dangerous heart conditions, and help advance the science of heart disease treatment. We’ll award $200,000 to those able to observe the right patterns, ask the right questions, and in turn, help people live longer and spend more time with those that they love.

Competition ends on Monday, March 14, 2016

CDS News
When The United Nations Puts a Call Out to Data Pros...
NYU Tandon School of Engineering, Enrico Bertini from November 23, 2015
Recently, Atefeh Riazi, the Chief Information Technology Officer for the U.N. issued a public appeal. “I invite the global community of data scientists to partner with the United Nations in our mandate to harness the power of data analytics and visualization to uncover new knowledge about U.N.-related topics such as human rights, environmental issues, and political affairs,” she wrote. To that end, the organization mounted a series of challenges, including one aimed at helping observers track the progress that countries have made in meeting Millennium Development Goals and another seeking ways in which to organize and highlight global humanitarian priorities.

Challenges were open to visualization experts from around the world and competition was exceptionally strong. It was thus particularly noteworthy that two of the winning entries came from a single institution—the NYU Tandon School of Engineering—and a single research group—headed by Assistant Professor Enrico Bertini.

Click here to receive the NYU Data Science Community Newsletter OR to have us follow your twitter feed so that our data science twitter bot can easily grab links from your tweets.

To send us an announcement for the newsletter, please email laura.noren@nyu.edu and brad.stenger@nyu.edu by 9 pm Eastern Time on Thursday evenings for inclusion in Friday’s newsletter. We retain curatorial discretion.

NYU Data Science Community Newsletter Issue 032.