NYU Data Science Community features journalism, research papers, events, tools/software, and jobs for February 5, 2016.

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 037.

Data Science News
How convolutional neural networks see the world
François Chollet, The Keras blog from January 30, 2016
In this post, we take a look at what deep convolutional neural networks (convnets) really learn, and how they understand the images we feed them. We will use Keras to visualize inputs that maximize the activation of the filters in different layers of the VGG16 architecture, trained on ImageNet. All of the code used in this post can be found on Github.

[1602.01013] Exploring limits to prediction in complex social systems
arXiv, Computer Science > Social and Information Networks; Travis Martin, Jake M. Hofman, Amit Sharma, Ashton Anderson, Duncan J. Watts from February 02, 2016
How predictable is success in complex social systems? In spite of a recent profusion of prediction studies that exploit online social and information network data, this question remains unanswered, in part because it has not been adequately specified. In this paper we attempt to clarify the question by presenting a simple stylized model of success that attributes prediction error to one of two generic sources: insufficiency of available data and/or models on the one hand; and inherent unpredictability of complex social systems on the other. We then use this model to motivate an illustrative empirical study of information cascade size prediction on Twitter.

The Beckman Report on Database Research
Communications of the ACM from February 01, 2016
A group of database researchers meets periodically to discuss the state of the field and its key directions going forward. Past meetings were held in 1989, 6 1990, 11 1995, 12 1996, 10 1998, 7 2003, 1 and 2008.2 Continuing this tradition, 28 database researchers and two invited speakers met in October 2013 at the Beckman Center on the University of California-Irvine campus for two days of discussions. The meeting attendees represented a broad cross-section of interests, affiliations, seniority, and geography. Attendance was capped at 30 so the meeting would be as interactive as possible. This article summarizes the conclusions from that meeting; an extended report and participant presentations are available at http://beckman.cs.wisc.edu.

Make journals report clinical trials properly
Nature News & Comment, Ben Goldacre from February 02, 2016
... Most researchers maintain a public pose that science is about healthy, reciprocal, critical appraisal. But when you replicate someone’s methods and find discrepant results, there is inevitably a risk of friction.
Our team in the Centre for Evidence-Based Medicine at the University of Oxford, UK, is now facing the same challenge. We are targeting the problem of selective outcome reporting in clinical trials.

Thoughts and Summaries from the Rework Deep Learning Conference

GitHub Pages, Deep Learning Research and Startups by lishali from February 03, 2016

This past week I went to the Rework Deep Learning conference. It was a good two days of talks by both top researchers in DL and companies applying DL. I made summary notes for the talks divided them by ‘Research’ and ‘Companies’, within these two categories they are simply in order of who talked first in the following sections. There was also a Q&A with Andrew Ng which I stuffed under ‘Research’.

The research talks were an excellent line up. All presented on recent work, though if you keep up with the literature, they should be familiar. I won’t mention any in particular here because I think their summaries are all worth reading if you are not yet familiar.

Also:
Team Rework provides its Storify summary of the conference.

A Deep Learning AI Chip for Your Phone

IEEE Spectrum from February 04, 2016

MIT engineers recently presented a chip designed to use run sophisticated image-processing neural network software on a smartphone’s power budget.

The great performance of neural networks doesn’t come free. In image processing, for example, neural networks like AlexNet work so well because they put an image through a huge number of filters, first finding image edges, then identifying objects, then figuring out what’s happening in a scene. All that requires moving data around a computer again and again, which takes a lot of energy, says Vivienne Sze, an electrical engineering professor at MIT. Sze collaborated with MIT computer science professor Joel Emer, who is also a senior research scientist at GPU-maker Nvidia.

Policing the Future

The Marshall Project, The Verge from February 03, 2016

... Over the last five years, Jennings [Missouri] precinct commander Jeff Fuesting has tried to improve relations between officers — nearly all white — and residents — nearly all black — by going door to door for “Walk and Talks.” Fuesting had expressed interest in predictive policing years before, so when the department heads brought in HunchLab, they asked his precinct to roll it out first. They believed that data could help their officers police better and more objectively. By identifying and aggressively patrolling “hot spots,” as determined by the software, the police wanted to deter crime before it ever happened.

HunchLab, produced by Philadelphia-based startup Azavea, represents the newest iteration of predictive policing, a method of analyzing crime data and identifying patterns that may repeat into the future. HunchLab primarily surveys past crimes, but also digs into dozens of other factors like population density; census data; the locations of bars, churches, schools, and transportation hubs; schedules for home games — even moon phases.
New website chronicles tales of collaborative research

Stanford Medicine, Scope blog from February 02, 2016

... Institutes across Stanford support similarly interdisciplinary approaches to solving many of the grand challenges we face today in environmental research, security, economic policy and energy. Technology like virtual reality (above) is being applied to environmental research, questions of empathy, and athletics.

We’ve collected many of these stories and videos of boundary-crossing research on a new website that chronicles the results of venturing outside departmental silos.

Recognizing correct code

MIT News, Martin Rinard from January 29, 2016

MIT researchers have developed a machine-learning system that can comb through repairs to open-source computer programs and learn their general properties, in order to produce new repairs for a different set of programs.

The researchers tested their system on a set of programming errors, culled from real open-source applications, that had been compiled to evaluate automatic bug-repair systems. Where those earlier systems were able to repair one or two of the bugs, the MIT system repaired between 15 and 18, depending on whether it settled on the first solution it found or was allowed to run longer.

Events

NYU Steinhardt Presents Discussion on Video Games as Learning Tools

NYU’s Steinhardt School of Culture, Education, and Human Development will host “Making Waves: Your Brain on Video Games,” a conversation on the promising yet controversial use of video games as tools for learning and to enhance brain functions. The event features Jan Plass, the NYU Steinhardt Paulette Goddard Chair in Digital Media and Learning Sciences and co-director of the Games for Learning Institute. Plass is a leader in improving the ways video games are designed as educational tools.

Tuesday, Feb. 9 at 6 p.m. at NYU’s Global Center for Academic and Spiritual Life (238 Thompson St., 5th Floor).

Moore-Sloan Data Science Lunch Seminar Series

The Lunch Seminar Series takes place on Wednesdays from 12:30 - 1:30 in the Center for Data Science at 726 Broadway, 7th Floor. The first half of this semester’s schedule is:

- February 10: Kyle Cranmer, NYU, Physics
- February 17: Lars Backstrom, Facebook
- February 24: Boris Leistedt, NYU, Cosmology and Particle Physics

NYU CESS 9th Annual Experimental Political Science Conference

We are pleased to announce the Ninth Annual NYU-CESS (New York University Center for Experimental Social Sciences) Conference on Experimental Political Science for Friday, February 19th, 2016 and Saturday, February 20th, 2016.
The Conference is an annual event that we hope will bring together researchers interested in experimental methodology in political science broadly. That is, we welcome the participation of scholars who work in the field and those who work in the lab as well as the participation of political psychologists and political economists.

Friday-Saturday, February 19-20, at NYU.

**The 19th ACM conference on Computer-Supported Cooperative Work and Social Computing**

**CSCW** is the premier venue for presenting research in the design and use of technologies that affect groups, organizations, communities, and networks. Bringing together top researchers and practitioners from academia and industry who are interested in the area of social computing, CSCW encompasses both the technical and social challenges encountered when supporting collaboration. The development and application of new technologies continues to enable new ways of working together and coordinating activities. Although work is an important area of focus for the conference, CSCW also embraces research and technologies supporting a wide variety of recreational and social activities using a diverse range of devices.

Saturday-Wednesday, February 27-March 2, in San Francisco CA.

**Deadlines**

**IK Prize | Tate**

**The IK Prize** celebrates digital creativity in all its forms. Awarded annually for an idea that proposes an innovative application of digital technology, the winning project will enable the public to experience art on display at **Tate Britain** and on our website in exciting new ways.

**Tate** invites entries from creative practitioners around the world in response to the subject of artificial intelligence.

Deadline for submissions is Sunday, February 7, 2016.

**New prize competition seeks innovative ideas to advance open science**

**The National Institutes of Health** has partnered with London-based Wellcome Trust to launch a global science competition for new products or services to advance “open science,” a movement to make scientific research data broadly accessible to the public. Up to six teams of technology experts and researchers stand to win $80,000 each to develop their ideas into a prototype or to advance an existing early stage prototype. The prototype judged to have the greatest potential to further open science will receive $230,000.

First Phase Application Deadline: Monday, February 29, 2016

**Summer Institute in Survey Research Techniques**

The Summer Institute in Survey Research Techniques is a teaching program of the Survey Research Center at the Institute for Social Research. It is located on the central campus of the University of Michigan at 426 Thompson Street in Ann Arbor. The summer courses are select offerings from the Michigan Program in Survey Methodology, and can be used to pursue a
Deadline for completed application is Sunday, May 15. Applicants for Summer Guest admission will be reviewed on the basis of their qualifications and experience. Summer Guests must complete both the enrollment process with the University of Michigan Registrar’s Office and with the Summer Institute.

**Call for submissions: Zika virus related datasets : Scientific Data**

Scientific Data is inviting submissions releasing and describing datasets related to Zika virus and the associated outbreak of microcephalic cases in South America. Submissions may be considered for inclusion in a special article collection on this topic.

To encourage early release of data related to this serious public health issue, all submissions received by 31st May 2016 and ultimately accepted for publication will have their article processing charge waived in full. Submissions after this date are still welcome, and will be considered according to the standard policies of the journal.

**CDS News**

**CDS Fellow Interview: Daniela Huppenkothen**

NYU Center for Data Science from February 01, 2016

The Moore-Sloan Data Science Environment Summit is an annual meeting between the data science centers at New York University, the University of California at Berkley and the University of Washington. We spoke with Daniela Huppenkothen, a Data Science Fellow at NYU’s Center for Data Science about her experience at the 2015 summit.

**Can you tell us about your work and experience as a Data Science Fellow at NYU?**

I spend my time working at the Center for Data Science and the Center for Cosmology and Particle Physics, where I’ve been studying statistical and machine learning methods in Astronomy. At NYU, I found a combination of experts in data science methodologies, and domain experts in Astronomy and Physics.

**Tools & Resources**

**Deep Residual Learning for Image Recognition**

GitHub - KaimingHe/deep-residual-networks from February 03, 2016

This repository contains the original models (ResNet-50, ResNet-101, and ResNet-152) described in the paper "Deep Residual Learning for Image Recognition" (http://arxiv.org/abs/1512.03385). These models are those used in ILSVRC and COCO 2015 competitions, which won the 1st places in: ImageNet classification, ImageNet detection, ImageNet localization, COCO detection, and COCO segmentation.

**Your First Machine Learning Project in R Step-By-Step (tutorial and template for future projects)**

Jason Brownlee, Machine Learning Mastery from February 03, 2016
Do you want to do machine learning using R, but you’re having trouble getting started?

In this post you will complete your first machine learning project using R.

**Wordbank: An open database of children's vocabulary development**

*Stanford University, Michael C. Frank* from January 31, 2016

Wordbank is an open database of information about children's vocabulary growth.

Wordbank archives data from the MacArthur-Bates Communicative Development Inventory (MB-CDI), a family of parent-report questionnaires. Wordbank enables researchers to analyze MB-CDI data in terms of aggregate vocabulary, individual items, demographic variables, and more. It provides interactive visualizations, exploratory reports, and data export tools.

**Berkeley INFO 88 Data and Ethics**

*Anna Lauren Hoffmann* from February 02, 2016

This course provides an introduction to critical and ethical issues surrounding data and society. It blends social and historical perspectives on data with ethics, policy, and case examples—from Facebook’s “Emotional Contagion” experiment to search engine algorithms to self-driving cars—to help students develop a workable understanding of current ethical issues in data science. Ethical and policy-related concepts addressed include: research ethics; privacy and surveillance; data and discrimination; and the “black box” of algorithms. Importantly, these issues will be addressed throughout the lifecycle of data—from collection to storage to analysis and application.

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.