NYU Data Science Community Newsletter features journalism, research papers, events, tools/software, and jobs for June 24, 2016

Data Science News

Galaxy-seeking robots

Gordon and Betty Moore Foundation from June 15, 2016
Moore Foundation grantees at **Lawrence Berkeley National Laboratory** have developed a prototype system, designed as a test for a planned array of 5,000 galaxy-seeking robots.

Dubbed ProtoDESI, the scaled-down, 10-robot system will help scientists achieve the pinpoint accuracy needed to home in on millions of galaxies, quasars and stars with the Dark Energy Spectroscopic Instrument (DESI) planned for the Mayall Telescope at Kitt Peak National Observatory near Tucson, Ariz.

More data-related news from the Foundations who support this newsletter:
- The Conference Scene for Data-Driven Discovery (June 21, Medium, Moore Data, Carly Strasser)
- Four foundations announce support for ASAPbio (June 20, ASAPbio)

Knights Landing Proves Solid Ground for Intel's Stake in Deep Learning

The Next Platform from June 21, 2016
Intel has finally opened the first public discussions of its investment in the future of machine learning and deep learning and while some might argue it is a bit late in the game with its rivals dominating the training market for such workloads, the company had to wait for the official rollout of Knights Landing and extensions to the scalable system framework to make it official—and meaty enough to capture real share from the few players doing deep learning at scale.

Also in hardware:
- "Artificial Synapses" Could Let Supercomputers Mimic the Human Brain (June 20, Scientific American, LiveScience)
- World's Fastest Supercomputer Now Has Chinese Chip Technology (June 20, Bloomberg)
- Barefoot Networks' New Chips Will Transform the Tech Industry (June 14, WIRED, Business)
- HPC Spending Outpaces The IT Market, And Will Continue To (June 22, The Next Platform)

Why Twitter Just Bought an Artificial-Intelligence Start-Up Called Magic Pony

Vanity Fair, The Hive blog from June 20, 2016
With **Twitter's** stock in the gutter and investors clambering for a turnaround, C.E.O. **Jack Dorsey** has embarked on an acquisition spree that could give 2013-era **Marissa**

Mayer a run for her money. Fresh off a $70 million investment in audio-streaming service **SoundCloud**, Twitter announced Monday that it has purchased **Magic Pony**, a British artificial-intelligence start-up, for a reported $150 million. Magic Pony is the third machine-learning start-up Twitter has purchased in as many years.

Also in Internet and big media:
- The Inventors of the Internet Are Trying to Build a Truly Permanent Web (June 20, WIRED, Business)
- Nielsen hopes to bring science to TV casting (June 22, Associated Press)
- NYC Media Lab '16 summit (takes place on Thursday, September 22, at Columbia University)

A path to unsupervised learning through adversarial networks

Facebook Code, Engineering Blog; Soumith Chintala and Yann LeCun from June 20, 2016
At **Facebook AI Research**, we've published a set of papers on stabilizing adversarial networks in collaboration with our partners, starting with image generators using Laplacian Adversarial Networks (LAPGAN) and Deep Convolutional Generative Adversarial Networks (DCGAN), and continuing into the more complex endeavor of video generation using Adversarial Gradient Difference Loss Predictors (AGDL). Regardless of what kinds of images or videos we gave to these systems, they would start learning and predict plausible scenarios of the world.

Using Artificial Intelligence to Humanize Management and Set Information Free

MIT Sloan Management Review, Reid Hoffman from June 14, 2016
Artificial Intelligence is about to transform management from an art into a combination of art and science. Not because we'll be taking commands from science fiction's robot overlords, but because specialized AI will allow us to apply data science to our human interactions at work in a way that earlier theorists like **Peter Drucker** could only imagine.

Also in technology management:
- 50 Smartest Companies 2016 (June 21, MIT Technology Review)
- Baidu Is Using Its Own Data to Measure China's Economy (June 21, Bloomberg)
- What makes a great software engineer? (June 21, Andrew J. Ko, Bits and Behavior blog)

Artificial Intelligence Achieves Near-Human Performance in Diagnosing Breast Cancer

Beth Israel Deaconess Medical Center from June 19, 2016
A research team from **Beth Israel Deaconess Medical Center** and **Harvard Medical School** recently developed artificial intelligence methods aimed at training computers to interpret pathology images, with the long-term goal of building AI-powered systems to make pathologic diagnoses more accurate

Also in clinical research:
- Why Most Clinical Research Is Not Useful (June 21, PLOS Medicine, Essay; John

P. A. Ioannidis)
- Blood-Based Screening for Colon Cancer - A Disruptive Innovation or Simply a Disruption? (June 15, JAMA, Viewpoint; Ravi B. Parikh, MD, MPP and Vinay Prasad, MD, MPH)
- I'm Feeling Yucky :(Searching for symptoms on Google (June 20, Official Google Blog)
- How Artificial Intelligence Is Bringing Us Smarter Medicine (June 20, Fast Company)

Bringing Precision to the AI Safety Discussion

Google Research Blog, Chris Olah from June 21, 2016
We believe that AI technologies are likely to be overwhelmingly useful and beneficial for humanity. But part of being a responsible steward of any new technology is thinking through potential challenges and how best to address any associated risks. So today we're publishing a technical paper, *Concrete Problems in AI Safety*, a collaboration among scientists at **Google**, **OpenAI**, **Stanford** and **Berkeley**. ... We believe it's essential to ground concerns in real machine learning research, and to start developing practical approaches for engineering AI systems that operate safely and reliably.

Meantime, in search and Google:
- The big search upgrade -- and how Amazon could beat Google at its own game (June 21, VentureBeat, Ivan Bercovich)

Tweet of the Week

Twitter, Pinboard from June 20, 2016

Pinboard
@Pinboard

Follow

Machine learning is like money laundering for bias

RETWEETS 778
LIKES 1,111

11:21 AM - 19 Jun 2016

778   1.1K

Félix Fischer @FelixFischer91 · 18h
@Pinboard @vihartvihart Why? In what context? I do not understand

1

View other replies

Pinboard @Pinboard · 11h
@FelixFischer91 @vihartvihart in the context of the Crimson King

Chris @Walshman23 · 1h
@Pinboard Confusion will be my epitaph

1

Events

The 2016 ACM SIGMOD/PODS Conference: San Francisco, USA

The annual ACM SIGMOD/PODS Conference is a leading international forum for database researchers, practitioners, developers, and users to explore cutting-edge ideas and results, and to exchange techniques, tools, and experiences.

San Francisco, CA Sunday-Friday, June 26-July 1. [$$$]

AI Now

The **White House** and **New York University's Information Law Institute**, with support from **Google Open Research**, **Microsoft Research** and the **MacArthur Foundation** will host a major public symposium to address the near-term impacts of AI technologies across social and economic systems.

The focus will be the challenges of the next 5-10 years, specifically addressing four themes: social inequality, labor, healthcare, and ethics.

New York, NY Thursday, July 7, at NYU Skirball Center.

Multisensory Music Hackathon

Explore the relationship between sound and the senses at the upcoming Multisensory Music Hackathon. During this day-long creative event, we welcome you to use

sensors to make music out of motion, temperature, light; or explore interactive ways to connect sound to other sensations and media. This hackathon isn't just for your ears but for your hands, eyes, nose, and taste!

New York, NY Saturday, July 9, starting at 12 noon, 45 W 18th St, 7th Floor

Talk & Poster List | SciPy 2016
The following Talks and Posters have been accepted for SciPy 2016. Additional Talks and Posters will be announced soon.

Austin, TX Monday-Sunday, July 11-17. [$]

DataViz Seattle hackathon
Join us as we host a one-day informal hackathon in Seattle, Washington. The focus will be around using the Graph extension to create interactive graphs, charts, and maps for Wikipedia.

Seattle, WA Friday, July 22. Meet in the Data Science Studio, on the 6th floor of the Physics and Astronomy Tower.

Deadlines

Machine Learning Yearning by Andrew Ng
This is a book I am writing over summer 2016. If you want to get a free draft copy of each chapter as it is finished, please sign up by Friday Jun 24th [today!] for my mailing list. ... Historically, the only way to learn how to make these "strategy" decisions has been a multi-year apprenticeship in a graduate program or company. I am writing a book to help you quickly gain this skill, so that you can become better at building AI systems.

RIT Hockey Analytics Conference
Rochester, NY A one-day workshop on September 10, 2016 showcasing state-of-the-art research in analytics for college and professional hockey

Deadine for submissions is Friday, July 15.

2016 Financial Stability Conference - Innovation, Market Structure, and Financial Stability
Washington, DC The **Federal Reserve Bank of Cleveland** and the **Office of Financial Research** invite the submission of research and policy-oriented papers for the 2016 Financial Stability Conference to be held December 1-2, 2016, in Washington, D.C. The objectives of this conference are to highlight research and advance the dialogue on financial market dynamics that affect financial stability, and to disseminate recent advances in systemic risk measurement and forecasting tools that assist in macroprudential policy development and implementation.

The deadline for submissions is July 31.

International Prize in Statistics: Submitting Nominations

The biannual International Prize in Statistics is stewarded and managed by a foundation comprising representatives of the five major statistical organizations working cooperatively to develop this prestigious award: the American Statistical Association, Institute of Mathematical Statistics, International Biometric Society, International Statistical Institute, and Royal Statistical Society.

Mirroring the successful approach employed by other prestigious scientific prizes, the International Prize in Statistics recognizes an individual statistician or team of statisticians (groups of individuals working on similar ideas as teams of individuals or organizations) for "a single work or body of work."

Deadline for nominations in Monday, August 15.

IBM Watson AI XPRIZE

The IBM Watson AI XPRIZE is a $5 million competition, challenging teams globally, to develop and demonstrate how humans can collaborate with powerful AI technologies to tackle the world's grand challenges. The prize aims to accelerate adoption of AI technologies, and spark creative, innovative and audacious demonstrations of the technology that are truly scalable and solve societal grand challenges. To encourage innovation in any form, the competition is an open challenge in AI. Rather than set a single, universal goal for all teams, this competition will invite teams to each declare their own goal and solution to a grand challenge. The IBM Watson AI XPRIZE is a four-year competition with annual milestone competitions in 2017, 2018, and 2019.

The deadline to apply is December of this year, with detailed project proposals due in March.

Open Context & Carleton Prize for Archaeological Visualization

Increasingly, archaeology data are being made available openly on the web. But what do these data show? How can we interrogate them? How can we visualize them? How can we re-use data visualizations?

We'd like to know. This is why we have created the Open Context and Carleton University Prize for Archaeological Visualization and we invite you to build, make, hack, the Open Context data and API for fun and prizes.

Deadline for submissions is Friday, December 16.

Tools & Resources

Interactive Knowledge Extraction (IKE)

Allen Institute for Artificial Intelligence from June 17, 2016
Recent work on information extraction has suggested that fast, interactive tools can be highly effective; however, creating a usable system is challenging, and few publicly

available tools exist. In this paper we present IKE, a new extraction tool that performs fast, interactive bootstrapping to develop high quality extraction patterns for targeted relations. Central to IKE is the notion that an extraction pattern can be treated as a search query over a corpus.

persistent-rnn: Fast Recurrent Networks Library

GitHub - baidu-research from June 20, 2016
For a GPU, the largest source of on-chip memory is distributed among the individual register files of thousands of threads. For example, the NVIDIA TitanX GPU has 6.3 MB of register file memory, which is enough to store a recurrent layer with approximately 1200 activations. Persistent kernels exploit this register file memory to cache recurrent weights and reuse them over multiple timesteps.

Avoiding reloading layer weights multiple times makes persistent kernels very efficient at low batch sizes.

How we made a VR data visualization

Simon Rogers from June 20, 2016
When we made an interactive guide to the UK's EU referendum, which takes place this Thursday, it seemed like an important opportunity to try out producing our own 360-degree data visualization.

The Stanford Question Answering Dataset

The Stanford Natural Language Processing Group, Pranav Rajpurkar from June 21, 2016
Stanford Question Answering Dataset (SQuAD) is a reading comprehension dataset, consisting of questions posed by crowdworkers on a set of Wikipedia articles, where the answer to every question is a segment of text, or span, from the corresponding reading passage. With 107,785 question-answer pairs on 536 articles, SQuAD is significantly larger than previous reading comprehension datasets.

Text Analytics API Now Available in Multiple Languages

Microsoft Technet, Cortana Intelligence and Machine Learning Blog from June 21, 2016
Although text often contains highly valuable data for companies, extracting meaningful data from it can be a challenge. The field of text analytics utilizes natural language processing to extract meaningful structured data from text, and often includes areas such as sentiment analysis, entity recognition and linking, and text clustering.

The Microsoft Text Analytics API is one of the Cognitive Services that can help you turn unstructured text into meaningful insights.

Careers

Accessibility, diversity, and software teams

Andew J. Ko, Bits and Behavior blog from June 17, 2016

For all of yesterday and half of today, about 25 industry leaders from **Google**, **Microsoft**, **Salesforce**, **Yahoo**, and a few dozen smaller companies and non-profits gathered with the **University of Washington's AccessComputing team**

Scientific Programmer/Science Data Analyst/DevOps Engineer

Association of Universities for Research in Astronomy from June 20, 2016
Tucson, AZ The Science Quality and Reliability Engineering (SQuaRE) team is a Data Management team that focuses in the following areas of project work: building systems for rapid ad-hoc analysis of data to determine image quality and facilitate trouble-shooting.

Join Our Team || Databrary: An Open Data Library for Developmental Science

New York University from June 21, 2016
New York, NY Databrary is a web-based video data library where scientists who collect and analyze research video can store and share their data openly with other researchers. ... We are growing a team to design and build new open source solutions for science using the latest technologies.

Cvpr2016 Jobs Board

Cvpr2016 from June 21, 2016
A long list of open positions.

Projects Data Coordinator: NSF Arctic Data Center at the National Center for Ecological Analysis and Synthesis (NCEAS)

University of California-Santa Barbara from June 22, 2016
Santa Barbara, CA The Projects Data Coordinator manages ecologically and environmentally relevant data from the Arctic Data Center.

Job Listings | SciPy 2016

SciPy from June 22, 2016
One of the goals of the SciPy Conference is to connect the individuals in the SciPy community with organizations that are looking for the best and brightest scientific Python talent. To do so, we offer our Sponsors for the 2016 conference the opportunity to post their job listings on our SciPy Conference website.

Spheryx Solutions - Scientific Software Developer

Spheryx Solutions from June 21, 2016
New York, NY Spheryx Inc. seeks a software developer to participate in ongoing development of analytical software for a new line of cutting-edge scientific instruments.

Training Platform Engineer, NVIDIA Deep Learning Institute

NVIDIA Workday from June 23, 2016
Santa Clara, CA NVIDIA is searching for a world-class platform engineer to lead the development of a next-generation training system and work with expert content

creators to deliver high-quality hands-on GPU programming content around the world, focused in the area of Deep Learning.

Three Research Appointments in Computational Social Science

PoliticalBots from June 21, 2016

Oxford, England We are looking for three talented researchers to join our team of computational social scientists at Oxford. Ideally, we'd like to have one person interested in the sociology of algorithmic production, another interested in the policy implications of bots, and another in the computational analysis of political bots and botnets.