1   Overall

`https://computefest.seas.harvard.edu/symposium`

1.1   Committee

Catherine Chute,Assistant Dean for Professional Programs and Executive Director for the Institute for Applied Computational Science, Harvard John A. Paulson School of Engineering and Applied Sciences

Francesca Dominici, Professor of Biostatistics, Harvard T.H. Chan School of Public Health & Co-Director of the Data Science Initiative, Harvard University

Brian Hayes, Senior Writer for *American Scientist*; Associate in Computer Science, Harvard University; & Y Combinator Research Fellow

Isaac Samuel Kohane, Chair, Department of Biomedical Informatics & Professor of Biomedical Informatics and Pediatrics, Harvard Medical School

Pavlos Protopapas, Scientific Program Director & Lecturer in Computation, Institute for Applied Computational Science, Harvard John A. Paulson School of Engineering and Applied Sciences

Margo Seltzer, Herchel Smith Professor of Computer Science & Director for the Center for Research on Computation and Society, Harvard John A. Paulson School of Engineering and Applied Sciences

1.2   Harvard Data Science Initiative

- Francesca Dominici co-director

- 8 Post-doctoral fellows

- Grants for data science

- 100+ faculty members

- Seminars

- E-mail list

2   9:15 AM Title: "Reinforcement Learning in Healthcare: Challenges and Promise"

Presenter: Finale Doshi-Velez, Assistant Professor of Computer Science, Harvard University

- Right treatment for right person: hard

- optimizing HIV treatment with reinforcement learning

- co-resistance may occur depending on drug combination, sequence of drugs important

- Considering all the past history of treatment

- Partially observable reinforcement learning

- maximizing long-term *rewards* (sum)

- Introduce a *representation* that summarizes the history (parametric assumption?)

- Enst 2005 Parboo 2014, Marivate 2015: simulation model based approaches have not been very successful

- drugs; hidden state exhibits as CD4, viral load, viral mutations (hidden Markov model?)

- Neighborhood-based predictions (Bogojeska 2012): finding similar patients. Works if there are similar patients (no extrapolation required)

- For singleton patients, model-based prediction

- Action space (what to do with drugs need trimming down)

- Policies: random policy, neighbor policy, mode-based policy, mixture did better

- Usable in other settings? Sepsis came up as another example, no single indicator of success like CD4. optimizing BP care is not enough

- Weighted doubly robust? high variance of machine-selected policy. on average better than physician-policy

- Summarize histories and consider long-term effects. AIs are team-members with flaws and strengths

- Machine Learning for Healthcare 2018 August 17-18 Stanford University, Stanford, CA

- Q&A
  - Frequency of observation is not random (MRSA less observations, VIP more observations, etc)
  - Neighbor: what metric of closeness.

- My Thoughts
  - TTT may be modeled this way similarly to HIV care?

- Combining Kernel and Model Based Learning for HIV Therapy Selection

3   9:55 AM Title: "The Algorithm for Precision Medicine"
Presenter: Matt Might, Director of the Hugh Kaul Personalized Medicine Institute, University of Alabama at Birmingham

- Quit White House as of last week

- Precision Medicine: often genome-driven, right drug to right patient

- Data is the best drug, computation is the limiting step.

- mapping a patient to a treatment (not)

- NGLY1 deficiency was found in his son `https://www.ngly1.org`

- Pathogenicity of mutations

- read-through compounds did not work. In biology, theory often does not work in practice

- A blog post describing the symptoms resulted in reporting of 58 patients.

- With CRISPR, knock-out model organisms were created.

- biological model and try out many compounds

- computational assay with deep learning (multi-dimensional data as an assay?)

- PEG Protein-based emzyme med in progress

- Specialized diet (eg. PKU, MPI-CDG).

- additional disfunction on ENGage rescues NGLY deficiency

- deep learning for toxicity

- Scaling up of algorithm

- PEPMA in Precision Medicine initiative

- 5 disease given 12 months (rest of Obama era)

- UAB position

- gain of function find inhibitor; loss of function find activator

- Precision oncology, tumor is sequenced (what in autoimmune diseases?)

- Pharmacogenomics focusing on depression

- Computational medicine

- mediKanren: logical reasoning (miniKanren-based).
  - Graph structured DB of pubmed (via Nlp)
  - Relation describing logic language
  - logical query can be made

4   10:55 AM Title: "Optimizing the patient journey with AI" (big data and the future of medicine)

Presenter: John Brownstein, Professor of Pediatrics, Harvard Medical School & Chief Innovation Officer, Boston Children's Hospital

- MI at Partners Healthcare

- siloed health care data

- no available data, no API, no user app

- BD2K Big Data to Knowledge

- radiology BCH and GE digital tool to distinguish abnormal from normal

- time series in ICU

- Precision medicine service at BCH: pediatrics is where rare mutation

- prediction of patient placement

- digital phenotype in twitter

- trend in disease

- HealthMap. now used by WHO

- digital disease detection NEJM paper

- Wikipedia Usage Estimates Prevalence of Influenza-Like Illness in the United States in Near Real-Time

- What your dinner cancellation on OpenTable says about the flu

- Mapping global environmental suitability for Zika virus

- Online Reports of Foodborne Illness Capture Foods Implicated in Official Foodborne Outbreak Reports

- People are quite open about opioid abuse diversion streetRx - Latest Street Prices For Prescription Drugs

- flu near you

- THERMIA Simplifying Fevers

- Buoy | Understand Your Symptoms

- Flu service via Uber (logistics can be used)

5   11:35 AM Title: "Phenomics is the New Genomics"

Presenter: Marzyeh Ghassemi, Visiting Researcher at Google's Verily & MIT CSAIL Post Doctoral Fellow

- Joining U Toronto and Vector Institute for Deep Learning

- Postdoc with Peter Szolovits

- same patient can receive different recommendation

- phenotyping = targeting + evidence

- 10-20% of treatments are evidence based on RCTs. RCTs are expensive.

- Among asthmatics, only 6% were eligible for RCTs. (only asthma without other morbidities)

- 80% of US hospitals have EHR

- EHR is hard. repurposing (bi-product). heterogenous, sparse (unmeasured, unreported, no follow-up), uncertain (labels, bias, context)

- Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission Having asthma resulted in lower risk of pneumonia in the training dataset (is this index event bias?)

- MIMIC III

- switching state autoregressive model representations (SSAM). latent state gives values of observed variables.

- post-hoc interpretability. can generate state before initiation

- severity of illness very important, but no universal definition

- clinicians rely on notes more than structured data

- topic modeling Latent Dirichlet Allocation Finding scientific topics

- deep learning does not improve beyond LDA

- digital phenotyping (Harnessing Smartphone-Based Digital Phenotyping to Enhance Behavioral and Mental Health)

- Understanding vasopressor intervention and weaning: risk prediction in a public heterogeneous clinical time series database. The latent states from the switching-state autoregressive mode (SSAM) were used as predictors of vasopressor administration and weaning.

6 2:00 PM Title: "Challenges and Opportunities for Machine Learning in Cancer Immunotherapy"
Presenter: Jennifer Chayes, Managing Director, Microsoft New England Research and New York City

- MS partnering with Sloan Kettering Dana Stanford JH

- Adaptive Biotechnologies: T cell profiling

- 70-200 mutations required for cancer

- T-cell clonal expansion after encountering relevant antigens

- checkpoint inhibitors (anti-CTLA4, anti-PDL1; note abatacept is a fusion CTLA4-Fc) highly successful but causes auto-immunity

- mechanistic level: predict immunogenicity of (neo)antigens. high-dimensional feature space

- clinical level predict patient response to checkpoint inhibitor immunotherapy. microbiome trains immunity. high-dimensional feature space but small number of patients

- T-cells collected from 10 subjects, tested against a panel of 1000 antigens

- Sparce matrix completion to predict a new T-cell's response to a panel of antigens

- cancer diagnosis: If T-cells are enriched with clones responsive to GBM, patients may have GBM

- threapy. If a patient has effective T-cell population, upregulation by checkpoint inhibitor likely work

- adverse events: standard self-antigens

- long-term survivors of melanoma

- FDA approved immunotherapy for MSI-high or MMR-deficient tumors (molecular target specific)

- predict T-cell clones in tumor that expand in the blood. surrogate marker

- multifactorial modeling using elastic net. cross-validation

- Ojala & Garriga JMLR 2010permutation test Permutation Tests for Studying Classifier Performance

- 21 patients. 80% explained variance in CV.

- Blum et al COLT 1999. Rules out perverse case where CV known to be unreliable.

- permutation of part of covariates (clinical, etc). (This is breaking correlation among covariates?)

7 2:40 PM Title: "Deciphering the Dynamics of the Anesthetized Brain: A Case Study in Statistical and Mathematical Modeling"

Presenter: Emery Brown, Edward Hood Taplin Professor of Medical Engineering and of Computational Neuroscience, Massachusetts Institute of Technology

- high signal to noise problem

- during anesthesia brain is not turned. activity exists

- drug-induced reversible state comprised of unconsciousness. Mechanism still not known

- unconsciousness barbiturates, propofol, inhalational

- analgesia opioids inhalational

- amnesia benzodiazepines, hypnotics, inhalational drugs

- immobility anticholinergics inhalational drugs

- hemodynamics stability.

- beta oscillation. slow oscillation brain stem inactivated

8 3:45 PM Panel Discussion

Moderater: Margo Seltzer, Herchel Smith Professor of Computer Science & Director for the Center for Research on Computation and Society, Harvard University
Participants: Emery Brown, John Brownstein, Finale Doshi-Velez, Maryzheh Ghassemi, and Matt Might

- NIH NCATS: translational research institute

- Educating physicians: HMS 3rd year electives including research electives. data science elective was very popular.

- Future data apps that could be made

  - Google search history based patient matching.
  - Reinforcement learning used for exercise promotion among diabetics. different prompts give different effects. Learns "you"

- cyberchondria

- World is becoming data driven. High-school-level probability/statistics education is required for literacy.

- More frequent data collection beyond annual checkup is possible using existing data infrastructure.

9 Closing remark
- billing system is artifactual system

- GANS to detect fraud? 1/6 of US economy spent on health care

10 Summary
- Finale Doshi-Velez

  - Combining Kernel and Model Based Learning for HIV Therapy Selection (prediction based on similar patients or model assumptions)

- Matt Might

  - mediKanren Proof-of-concept for reasoning over the SemMedDB knowledge base, using miniKanren + heuristics + indexing. (logic based query of NLP'ed PubMed data)

- John Brownstein

- The Big Data to Knowledge (BD2K) program is a trans-NIH initiative that was launched in 2013 to support the research and development of innovative and transformative approaches and tools to maximize and accelerate the integration of big data and data science into biomedical research.

- Marzyeh Ghassemi

  - Understanding vasopressor intervention and weaning: risk prediction in a public heterogeneous clinical time series database. The latent states from the switching-state autoregressive mode (SSAM) were used as predictors of vasopressor administration and weaning.

- Jennifer Chayes

  - Screening patients' T-cells for reactivity to a panel of antigens. Could be useful in predicting response to checkpoint inhibitor (cancer immunotherapy) response prediction

  - Emery Brown
    * Something about EEG activity during general anesthesia (my brain got anesthetized by this time)