# VA NLP Recommendations

- Drive basic research with end-to-end tasks relevant to your stake holders, develop metrics "outside-in" (overall task metrics drive component-level targets)

- Explore modern statistical techniques, e.g., deep learning, but ensure that ultimate results can be interpreted by the end user

- Invest in Education, Tooling, and Methodologies for your end users

- Improve connection with operations to increase speed, agility, iteration, and actual deployment. Engineering and Research are different skills, but engineers need lots of help to properly leverage and apply NLP tools.

- Publish and advertise research, data, and resources to attract collaborators and top talent. Share as much as possible (open source tools, data, etc)

- Treat physicians as innovation collaborators. Recognize the growing importance of Data Science in all aspects of healthcare

- Look at other commercial research labs for ideas on setting the research agenda and driving technology transfer
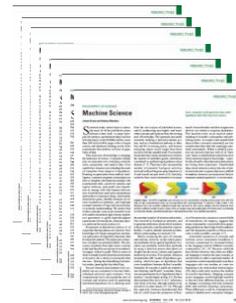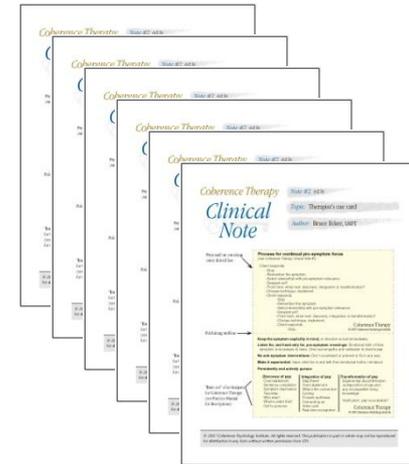
VA NLP

Hoifung Poon

Competitive Advantage: MVP

Panomics
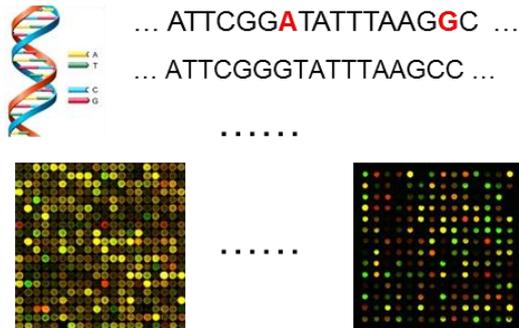
EHR

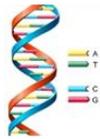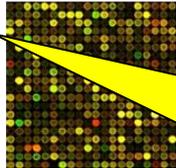**Precision Medicine**
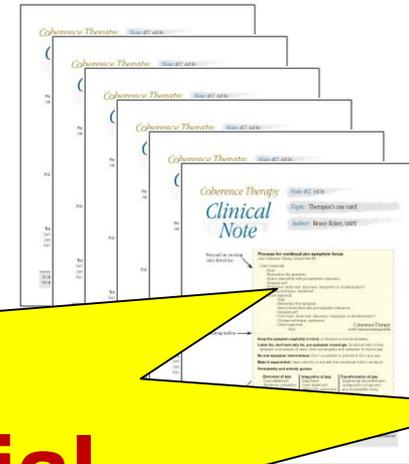
Literature

Shotgun / Rifle

- Scalable phenotyping
- Generalizable features
  - Insulin ~ diabetes
  - <DRUG> ~ <DISEASE>
- Free lunch
  - Prior knowledge: textbook, literature, guideline
  - Disparity in structuredness

Representation

- Grey area: Activity, anger, …
- Probabilistic ontology?
- Molecular phenotyping
  - Cancer: Lung, skin, …
  - Cancer: BRAF, HER2, …

Recommendations for VA NLP Research, a Support Group Approach

John M. Conroy
IDA Center for Computing Sciences
Bowie, MD.

A "Clinical Approach"

- "to abstain from doing harm" --Hippocratic Oath

- Instead of giving recommendations I will share my experience and let you decide for yourself to what extent my experience is relevant to the VA.

- In support groups each member shares from their perspective and it is forbidden to tell another member what they should do.

Institute for Defense Analyses

- Center for Communications and Computing, one Federally Funded Research and Development Center, but three "Centers"

  - Center for Communications Research (CCR),Princeton (formerly Communications Research Division), 1959. Hidden Markov Models developed here in 1960s by Lloyd Welch, Leonard Baum, and others.

  - Center for Computing Sciences, 1985 (formerly Supercomputing Research Center).

  - Center for Communication Research, La Jolla, 1990

What We Do and How

- What: "… creation and analysis of sophisticated encryption methods, high-speed computing technologies, the development of advanced algorithms and their applications, algorithmic and mathematical foundations of cryptology, computer network technologies supporting communications security, information processing technologies supporting cyber security, and *analytical applications for large data sets*."

- How: "a close working relationship with NSA, and ongoing engagement with the broader research community"

  https://www.ida.org/IDAFFRDCs/CenterforCommunications.aspx

More on the How

- "Summer workshops, which draw academics and others to use a concerted "tiger team" approach to tackling several truly difficult problems each summer."

- Internal experts mix with senior researchers, from academia, graduate students work together in a collegial environment. Central tenant is to welcome the newcomer.

- Planning for next summer workshop begins shortly after one has ended.

My Introduction to NLP

- Year long workshop Fall 1999 to Summer 2000 with text summarization as a major topic.

- My previous background was in matrix computations and mathematical modeling.

- Workshop visitors:
  - Daniel Marcu, USC/ISI (Hal Daumé's advisor)
  - Jaime Carbonel, CMU

- Attended Document Understanding Conference planning met with Ed Hovy USC, Kathy McKeown (Ani Nenkova's advisor). In subsequent years Lucy Vanderwend, Karen Spark Jones.

Participate, Collaborate, …

- Community Evaluation: often hosted by National Institute Standards and Technology (NIST)
  - Document Understanding Conferences 2001-2007.
  - Text Retrieval Conference 2002-2003.
  - Multi-lingual Summarization Evaluation 2005-2006
  - Text Analysis Conference 2008-2011, 2014.
  - MultiLing 2011, 2013, 2015.

- Biomedical Literature,
  - Linking Figures with Abstract Sentences, Bockhorst, Conroy, O'Leary, Yu: 2012
  - Summarizing Scientific Literature: Conroy & Davis 2015.

Lessons Learned

- "Let the data speak to you."- para-phrase of John Tukey, but taught to me by Alan Richter, CCR-P.

- "All models are wrong some are useful." George Block, 1978.

- Inter-disciplinary work is challenging, fun, productive.

- Ask questions, even "dumb ones."

- "Ask" the data many questions.

NLP Research & Development Opportunities

Michael E. Matheny, MD, MS, MPH
Associate Director, Advanced Fellowship in Medical Informatics
Tennessee Valley Healthcare System VA

Director, Center for Population Health Informatics
Vanderbilt University Departments of Biomedical Informatics, Medicine, and Biostatistics

"Foundational" NLP

- Patient-level clinical context and disease/symptom/treatment course processing
  - Co-Reference Resolution Improvement
  - Temporal reasoning over time
  - Linking conceptual units across documents and types of care

"Applied" NLP

- Integration of information provenance and author characteristics into context
- Effective concurrent use of different NLP data sources – clinical notes (and types), patient blogs, secure messaging, social media
- Effective integration of 'free lunch' knowledge sources (pubmed, genomic, etc) into population health and point of care delivery
- Systems for increasing automation in merging structured and unstructured text

The "D" in R&D: "Implementation" NLP

- Partnerships with electronic health record vendors and or enterprise data warehouses to support widely used NLP frameworks (UIMA, etc) to allow deployment & execution upkeep of validated NLP tools

- Partner with health care quality improvement, business intelligence, and national clinical initiatives to conduct NLP research in the context of clinical utility

- Evaluation of human-computer interaction and human factors with NLP integration into clinical workflow: unintended consequences, barriers to effective use, key conceptual targets for effective utilization

Next Steps in NLP Research in VA

**Jennifer Garvin PhD, MBA, RHIA, CPHQ, CCS, CTR, FAHIMA
Salt Lake City VA Healthcare System IDEAS 2.0 Center
Department of Biomedical Informatics, University of Utah**

Draft Criteria for "Wins"

- Objectives for "Wins":
    - Accurately extract the data needed for a given use case
    - Develop a system that could be implemented in the VA healthcare system
    - Transform unstructured data to structured data for use in near-real or real-time applications
    - Further the goals of the Blue Print for Excellence
        - Improve Performance
        - Promote a Positive Culture of Service
        - Advance Health Care Innovation for Veterans and the Country
        - Increase Operational Effectiveness and Accountability

Research Aspects Related to "Wins"

- Evaluate readiness for implementation of NLP as well as barriers and facilitators

- Determine evidence-based, effective implementation processes and measures of adoption and impact

- Focus on applied uses of the NLP tools ("Wins") and determine if they are quick, medium, or longer term "Wins" for the VA