New and developing healthcare uses of Natural Language Processing (NLP) in the research, care delivery, and insurance

Friday, March 17th, 2017

Moderator: Joseph Bormel, MD, MPH

Featured Speakers: Joseph Polifroni, Ph.D., and Guy Divita, MS.
The moderator will be our own Joseph Bormel, MD, who will very briefly share several applied case studies, elaborating contracted project work done applying Natural Language Processing and controlled vocabularies groupings to value-based purchasing, in both the public and private sectors. Specifically, he will describe application of NLP to improving documentation of care needed for Medicare Advantage programs to address risk adjustment. He will also describe the relationship of NLP to quality measures that are used in the MACRA program, called electronic clinical quality measures, based on his measures-authoring work for Medicare. Lastly, he will share some recent research on how several government agencies are evolving their bio surveillance to include EHR+NLP derived data to transcend the well-known limits of claims data. Dr. Bormel is a board certified internist, and NLM fellowship-trained informatics professional. He has several decades of executive experience with publicly traded EHR vendors (with Cerner, QuadraMed and the EHRA), program officer experience (with HHS, Office of the National Coordinator for HIT), and start-up experience (with San Francisco-based Healthline Networks and Talix).
The first presentation will be “Promise and Reality: Tales from the Intersection of Big Data, Machine Learning, and Healthcare Analytics”. The presenter, Joseph Polifroni, Ph.D. is the Director, Natural Language Processing for Optum Analytics. His presentation will focus on the importance of understanding the potential and the limitations of medical notes data, the importance of phrasing questions in actionable ways, and the importance of early examination of the data themselves. Dr. Polifroni has been involved in the fields of Automatic Speech Recognition and Natural Language Processing for over 25 years. He began his career as a Research Scientist in the Spoken Language Systems Group at MIT’s Laboratory for Computer Science. He has worked on machine learning for the design and evaluation of spoken dialogue systems, sentiment analysis, named entity extraction, and user modeling in both industry and academe. He also worked on the speech/natural language interface to a home healthcare monitoring system for the elderly deployed by a hospital group in Taiwan.
The second presentation is entitled "The Magic Behind Clinical Natural Language Processing for Epidemiological Studies, Surveillance and Quality Metrics".
It will be given by computer scientist, author, standards developer and industry leader Guy Divita, MS. He is currently a Research Associate at University of Utah Medical School in the Department of Bioinformatics. Divita has lead and published multiple large scale projects including recent work for VA Salt Lake City Health Care System (IDEAS Center) which will be shared at our meeting. Divita has worked in academia, industry and government at the National Library of Medicine, in the field of Big Data and NLP for several decades and is broadly recognized as a leading expert and platform creator/developer.
“IBM Watson ... to eradicate cancer”

Too Good to be True?

MD Anderson Benches IBM Watson In Setback For Artificial Intelligence In Medicine

Matthew Herper, FORBES STAFF

It was one of those amazing “we’re living in the future” moments. In an October 2013 press release, IBM declared that MD Anderson, the cancer center that is part of the University of Texas, “is using the IBM Watson cognitive computing system for its mission to eradicate cancer.”

A Setback for Artificial Intelligence in Medicine?

Is it really a setback, in general, or not?

March 1, 2017

Dale Sanders
Executive Vice President, Software

Dale Sanders’ Conclusions: (paraphrasing)

1. **Watson** - overhyped, overbought, oversold; Naively, not maliciously.
2. Will have a **big impact**.
3. **Ecosystem** is **not ready**, esp the text content
4. Of 1,958 National **Quality Measures**, only 7% represent **outcomes** and < 2% are **PROs**.
5. Adjust strategy and expectation to what is available.
6. [JB adds **Reasonability**, **Path** and **Trust** issues]

---

Forbes: Top 10 Artificial Intelligence Technologies, 2017 Q1

Next Phase (1-3 y)
- Text analytics and NLP (Growth)
- Image and video analysis

Thereafter (3-5 y)
- Speech Recognition, Natural language generation
- Biometrics
- Deep Learning

Moderator’s NLP Framing:

- **Medicare Advantage**- Improving documentation of care needed for programs to address risk adjustment

- **MACRA-** Text for/from quality measures *(electronic Clinical Quality Measures)*

- **Biosurveillance-** agencies, payers and providers are evolving their to include EHR + NLP derived data → BFC
  - Behavioral: Substance Abuse and Mental Health
  - Infectious: CDC; influenza (GFT), SARS, HIV, bioterrorism
  - Cancer:
  - Other Chronic Disease: Vascular, Diabetes, …
  - Safety and Efficacy of Drugs and Devices (includes EHRs and Smart Phones)

https://en.wikipedia.org/wiki/Public_health_surveillance
The basic idea

- Claims data - coded procedures and diagnoses
- Chosen for expediency in reimbursement
- Text notes, now often dictated, contain details not in the claims
- **NLP unlocks specific details**
Billing codes are often inadequately specific for purpose

- predicting expenditures
- assessing quality of care (Value-based purchasing, paying for quality and not necessarily just volume)
- surveillance (incidence and prevalence of conditions, drug and device safety and efficacy, research writ large)
- NLP -> greater meaningful specificity

Example: Diabetes specificity

- Diabetes Mellitus
  - DM2 = 8.3% adults worldwide
  - Billed as
    - 250.00 (ICD-9)
    - E10-E14.xx (ICD-10)
  - Annual expected costs are four times higher with renal or vascular. May need NLP to ‘see’ these patients
  - US cost(2012): $245 B
  - Providers who under code look bad (cost, quality, mortality, access, satisfaction)

Example: NLP and Diabetes

NLP looks for text describing:

- **Symptoms**
- **Medications**
- **Laboratory** findings and procedures
- **Comorbidities** and Complications, such as ketoacidosis coma, amputation, blindness, etc.
- **Quality measures** e.g. “foot exam”
Biosurveillance
in 8 years, 220% change in Oxycodone

Figure 1. Reasons for Drug-Related Emergency Department (ED) Visits, by Year: 2004 to 2011

Table 2: Selected Substances Involved in Drug-Related Emergency Department Visits: 2004, 2006, and 2011

ref: 2013 SAMHSA “Highlights of the 2011 Drug Abuse Warning Network (DAWN) Findings on Drug-Related Emergency Department Visits”
“Real World Evidence” by treatment indication

Key Contributors to Evidence Gaps

- Conceptualization
- FDA Approval
- Clinical Trials
- Off-label indications
- Post-Marketing Studies
- Variances in population characteristics from what was studied
  - Differing age groups – elderly, pediatrics
  - Race, ethnicity, gender variances
  - Unstudied co-morbid conditions
  - Varying severity of disease
  - Differing concomitant medications
  - Varying levels of compliance – i.e. < 80%

Utilization
- Evidence
- Little to No Evidence

Reference Note: As shown at 2011 Quintiles meeting posted by author (Cziraky) on YouTube; original presentation by Mark Cziraky in 2011, downloaded 2016-11: https://www.youtube.com/watch?v=3KkGKFev5f8
**“Real World Evidence” understanding of MH/DR visits**

Visits with Mental Health and/or Drug Related attributes

**Cases We Did Not Know about without Text NLP**

**Cases We Knew about without Text NLP that were “simple”**

**Cases We Knew about without Text NLP, but, with NLP, we now have more relevant context, not present in UB alone (nuanced)**

**“Simple” means that there was no additional information in the visit text that made classification more specific than suggested by the UB coding.**

**“Existing Visibility” means currently identified MH/DR visits with prior approaches (no enhancement from NLP)**

Reference Note: As shown at 2011 Quintiles meeting posted by author (Cziraky) on YouTube; original presentation by Mark Cziraky in 2011, downloaded 2016-11: https://www.youtube.com/watch?v=3KkGKFev5f8
Too Good to be True?
Where Does NLP Fit?

Reasonable Questions
- can they be answered today by humans and/or with enough data

A path to the necessary information

Trust
- e.g. is it safe to document and share what you knew?
- will you be adversely impacted or cheated?

https://www.youtube.com/watch?v=qMsPXSMTpFI
Bormel’s Second Law

“There is always more vital information you don’t have.”

Corollary:

You most need the true intention and context of person providing the information
Where IBM Watson portends the future:

Gen 1: (2011 Jeopardy)
“faster than a human expert for many tasks” We can search and rank hypotheses data+NLP+templates, tuning, and scoring can be magical. We need pick our questions very carefully (see “When to consider” below).

Gen 2: (2014 Wellpoint)
Problems that can be broken down into what humans do (e.g. case managers in IBM/Wellpoint in Oncology) can explode at scale (too many dimensions to engineer with brute force; WSJ 2014)

Gen 3: (2017 MD Anderson)
Contemptuous partnerships, secondary to contract without the resilience of Agile management left parties at odds with each other.

When to consider cognitive solutions

- Clearly specified question to be answered
- Test of fit
  - Productive Path
    - Direct value
    - Guides process
    - Useful failure
    - Specimen text
  - Complicated
    - No answer
    - Inadequate input
    - Poor expectancy
  - Futile Path

Joseph Bormel

JBormel@gmail.com