Clinical NLP and the data dilemma

Philip Resnik, Ph.D.
University of Maryland
CodeRyte, Inc.
Jurafsky and Martin (2009), *Speech and Natural Language Processing*

Graph adapted from Church, K. (2003) “Speech and Language Processing: Where have we been and where are we going,” Eurospeech, Geneva, Switzerland, Adding data from figures in Cardie and Mooney (1999).
Knowledge-based methods informed by large scale data analysis
Sure, Hal, Gingrich does have cool ideas about NASA. What do you think about his position on corporate tax rates?

"I am a HAL 9000 computer, production number three. I became operational at the HAL plant in Urbana, Illinois on January 12th 1997."

- Arthur C. Clarke,
2001: A Space Odyssey (1968 novel)
In today’s NLP, the question is not *whether* you integrate knowledge-based methods with statistical methods, but *how*. 
• **Features**
  – Define what the analysis “pays attention to”

• **Model structure**
  – Defines possible relationships among variables

• **Informative priors**
  – “Softly” incorporate assumptions that can be overridden by enough evidence
• Features
  – Define what the analysis “pays attention to”

<table>
<thead>
<tr>
<th>catheter</th>
<th>catheter</th>
<th>N</th>
<th>3sg</th>
<th>CATHETER</th>
</tr>
</thead>
<tbody>
<tr>
<td>catheters</td>
<td>catheter</td>
<td>N</td>
<td>3pl</td>
<td>CATHETER</td>
</tr>
<tr>
<td>catheterisation</td>
<td>catheterization</td>
<td>N</td>
<td>3sg</td>
<td>CATHETER</td>
</tr>
<tr>
<td>catheterisations</td>
<td>catheterization</td>
<td>N</td>
<td>3pl</td>
<td>CATHETER</td>
</tr>
<tr>
<td>catheterization</td>
<td>catheterization</td>
<td>N</td>
<td>3sg</td>
<td>CATHETER</td>
</tr>
<tr>
<td>catheterizations</td>
<td>catheterization</td>
<td>N</td>
<td>3pl</td>
<td>CATHETER</td>
</tr>
<tr>
<td>catheterise</td>
<td>catheterize</td>
<td>V</td>
<td>INF</td>
<td>CATHETER</td>
</tr>
<tr>
<td>catheterize</td>
<td>catheterize</td>
<td>V</td>
<td>INF</td>
<td>CATHETER</td>
</tr>
</tbody>
</table>
HPI: Atrial fibrillation. This patient is a 56-year-old white gentleman who has had a history of atrial fibrillation on and off since he had his bypass surgery. Patient was originally diagnosed with coronary artery disease as well as mitral valve problems approximately 3 years ago. Dr. Tirona used to take care of him at that time. He had a bypass surgery as well as mitral valve repair done at that time. Postop he had an episode of A-fib which then resolved spontaneously. He remembers somebody talking to him about cardioversion, but then the A-fib resolved spontaneously. So he was started on Coumadin. He would get some occasional episodes, but usually they are very brief, so he never bothered about them. Of late, over the last few months, he has been getting more frequent episodes and duration of these episodes is also prolonged for a few hours. So he saw Dr. Hagan who has referred him here for further evaluation and treatment. The patient states when he does get the A-fib, he feels very weak, tired, and short of breath. He denies any chest pain. Otherwise he is usually very active physically, he works fulltime as an electrician, and has not had any problems as far as doing his day-to-day work.


IMPRESSION: Paroxysmal atrial fibrillation in a patient with prior mitral valve disease, currently having more frequent breakthroughs symptoms.
• Informative priors
  – “Softly” incorporate assumptions that can be overridden by enough evidence

COLD

Chronic obstructive lung disease

Common cold
• Informative priors
  – “Softly” incorporate assumptions that can be overridden by enough evidence

COLD
emphysema
obstruction
airway
pulmonary

Chronic obstructive lung disease
Common cold
So, state-of-the-art NLP depends crucially on learning from relevant data.

That’s a problem.
Uncoded data

One solution

Training Data

Supervised methods  Semi-supervised methods  Unsupervised methods
One solution

ACL Anthology papers
Containing "semi-supervised"

Papers containing
"semi-supervised"
Providers

CHIEF COMPLAINT: Shortness of breath.
HPI: This is a 68-year-old female who presents to the emergency department with shortness of breath going for several days ...

Payers

Coding Tools & Services

12002
873.0

$\$
HPI:
The patient presents with headache and pt here with head injury - jumped and hit head on beam. + LOC, no neck pain, no numbness, visual changes, no vomiting, bleeding controlled at this time, no other injuries. The course/duration of symptoms is constant. Location: occipital. Radiating pain: none. The character of symptoms is throbbing. Associated symptoms: none.

MEDICAL HISTORY:
Medical history: Negative. Surgical history: Negative.

SOCIAL HISTORY:
Social history: Alcohol use: Denies, Tobacco use: Denies, Drug use: Denies.

ROS:
Constitutional symptoms: Negative except as documented in HPI. Respiratory symptoms: Negative except as documented in HPI. Neurologic symptoms: Negative except as documented in HPI. Additional review of systems information: All other systems reviewed and otherwise negative.

EXAM:
Vital Signs: Heart Rate 73 bpm Respiratory Rate 14 breaths/min SBP NBP 101 mmHg DBP NBP 61 mmHg SpO2 95 % General: No acute distress. Head: 5 cm laceration over top of head to sq. does not extend to gales. Neck: Supple, trachea midline, no tenderness. Neurological: Alert and oriented to person, place, time, and situation.
CHIEF COMPLAINT:
Right knee pain.

HISTORY OF PRESENT ILLNESS: The patient is a 61-year-old white female who works part-time at Curves who said she developed sharp severe pain in her right knee beginning in November 2005. She did not have any injury that she knows of but now she is having increasing pain with walking, kneeling down and going up and down stairs. She had sharp severe pain in the medial compartment of her knee. She feels better when she is not moving at all.

PAST MEDICAL HISTORY:
Significant for hiatal hernia and a DVT in left upper extremity.

PAST SURGICAL HISTORY:
Significant for bilateral thoracic outlet surgery and left rib resection as well as a left knee arthroscopy.

CURRENT MEDICATIONS:
Include Crestor.

ALLERGIES:
Codeine, Salactic, iodine, sufa drugs.

FAMILY HISTORY:
Significant for heart disease, stroke, and osteoporosis.

Neurointerventional procedure

Cerebral angiogram, 4/24/2006

Indication: Spontaneous subdural hematoma

Physicians: 1. Dr. John Scott


Discussion: Angiographic examination of the above vessels was performed after consent was obtained. A 5 French catheter was placed in the right common femoral artery, using the Seldinger technique, after Betadine prep and local anesthesia with lignocaine. After the procedure, the catheter was removed and hemostasis achieved at the puncture site.

Findings: Left internal carotid artery: There is subdural mass-effect over the left cerebral convexity manifest by displacement of pial vessels and cortical veins away from the inner table of the skull. There is rightward shift of the anterior cerebral artery complex. There are no arterial or venous vascular abnormalities.

Left external carotid artery: There are no dural vascular abnormalities.

Left vertebral artery: There is medial displacement of the left posterior cerebral artery complex. No vascular abnormalities are identified.

Right common carotid artery: There are no cerebral or dural vascular abnormalities.

Impression: Left convexity mass-effect consistent with known subdural hematoma. Negative cerebral angiogram.
A bigger challenge for healthcare:

With the widespread adoption of EHRs, what happens to natural clinical language?
CHIEF COMPLAINT: Shortness of breath.
HPI: This is a 68-year-old female who presents to the emergency department with shortness of breath going for several days...
“This system is designed for physicians to point and click their way through an entire exam quickly and effortlessly.” (EMR product review)
The clinical narrative

“...In years past, a well-written history and physical, or progress note, would unfold like a story, giving a vivid description of the patient’s symptoms and physical exam at the point of the encounter, as well as the synthesis of the data and the plan of care."

April 14, 2007

CHIEF COMPLAINT: Shortness of breath.

HISTORY OF PRESENT ILLNESS: This 68-year-old female presents to the emergency department with shortness of breath that has gone on for 4-5 days, progressively getting worse. It comes on with any kind of activity whatsoever. She has had a nonproductive cough. She has not had any chest pain. She has had chills but no fever.

EMERGENCY DEPARTMENT COURSE: The patient was admitted. She has had intermittent episodes of severe dyspnea. Lungs were clear. These would mildly respond to breathing treatments and morphine. Her D-dimer was positive. We cannot scan her chest; therefore, a nuclear V/Q scan has been ordered. However, after consultation with Dr. C, it is felt that she is potentially too unstable to go for this. Given the positive D-dimer and her severe dyspnea, we have waved the risks and benefits of anticoagulation with her heme-positive stools. She states that she has been constipated lately and doing a lot of straining. Given the possibility of a PE, it was felt like anticoagulation was very important at this time period; therefore, she was anticoagulated. The patient will be admitted to the hospital under Dr. C.
April 14, 2007

CHIEF COMPLAINT: Shortness of breath.

HISTORY OF PRESENT ILLNESS: This 68-year-old female presents to the emergency department with shortness of breath that has gone on for 4-5 days, progressively getting worse. It comes on with any kind of activity whatsoever. She has had a nonproductive cough. She has not had any chest pain. She has had chills but no fever.

EMERGENCY DEPARTMENT COURSE: The patient was admitted. She has had intermittent episodes of severe dyspnea. Lungs were clear. These would mildly respond to breathing treatments and morphine. Her D-dimer was positive. We cannot scan her chest; therefore, a nuclear V/Q scan has been ordered. However, after consultation with Dr. C, it is felt that she is potentially too unstable to go for this. Given the positive D-dimer and her severe dyspnea, we have waved the risks and benefits of anticoagulation with her heme-positive stools. She states that she has been constipated lately and doing a lot of straining. Given the possibility of a PE, it was felt like anticoagulation was very important at this time period; therefore, she was anticoagulated. The patient will be admitted to the hospital under Dr. C.
April 14, 2007

CHIEF COMPLAINT: Shortness of breath.

HISTORY OF PRESENT ILLNESS: This 68-year-old female presents to the emergency department with shortness of breath that has gone on for 4-5 days, progressively getting worse. It comes on with any kind of activity whatsoever. She has had a nonproductive cough. She has not had any chest pain. She has had chills but no fever.

EMERGENCY DEPARTMENT COURSE: The patient was admitted. She has had intermittent episodes of severe dyspnea. Lungs were clear. These would mildly respond to breathing treatments and morphine. Her D-dimer was positive. We cannot scan her chest; therefore, a nuclear V/Q scan has been ordered. However, after consultation with Dr. C, it is felt that she is potentially too unstable to go for this. Given the positive D-dimer and her severe dyspnea, we have waved the risks and benefits of anticoagulation with her heme-positive stools. She states that she has been constipated lately and doing a lot of straining. Given the possibility of a PE, it was felt like anticoagulation was very important at this time period; therefore, she was anticoagulated. The patient will be admitted to the hospital under Dr. C.
EMERGENCY DEPARTMENT COURSE: The patient was admitted and nontoxic in appearance. **Blood pressure was brought down aggressively. With this combined with BiPAP, she has reversed her respiratory distress promptly.** She has improved significantly. She will not require intubation at this time period. Her family has elected to go back to M, Dr. W. I did discuss this case with Dr. G who is on call for L Cardiology. She has accepted him in transfer; however, there are no PCU or ICU beds at this time period. Will admit here for a brief period until a bed is available at M. I discussed this case with Dr. R who will admit.

Clinicians were trying to determine whether the shortness of breath was due exclusively to her failing heart, or whether she has pneumonia.

Prompt response indicates that pneumonia is not the issue.
“...As EMRs proliferate, and increased Medicare scrutiny looms, medical documentation is evolving from its original goal of recording what actually was going on with a patient, and what the provider was actually thinking, to sterile boilerplate documents designed to justify the highest billing codes.


Text boxes in EMRs don’t solve the problem. We’re at risk of losing the rich language of the clinical record. And if you lose the language, you lose the story.
Take-aways for discussion

• Clinical NLP needs more statistical NLP

• We’ve got a big problem: data availability for clinical NLP R&D

• We, and everyone else, have a far bigger problem: the future of clinical language in electronic health records.
Thanks!