Panel Discussion:
State of VA NLP Research
Accomplishments, Opportunities, Challenges
September 10, 2015
VA’s Unique Advantages I

• Size of intramural research community
  • Huge number of use cases

• Culture of collaboration across academic groups
  • Rewards through positive sum games

• Support for research-operational partnerships
  • Programs such as CREATE, QUERI, COIN
VA Unique Advantages II

• Parallels and synergies between NLP and qualitative research
• Every text project has national scope
• Tons of opportunities to drive transformation
  • Make manual chart review more productive

\[(\text{Brain} + \text{NLP}) > \text{Brain}\]

• Make analysis of electronic data more powerful

\[(\text{NLP} + \text{Computer}) > \text{Computer}\]
What has been accomplished
NLP Researchers New to VA Since 2009

An incomplete list
• Qing Zeng
• Hong Yu
• Wendy Chapman
• Dezon Finch
• Ruth Reeves
• Glenn Gobbel
• Scott Duvall
• Stephane Meystre
• Guy Divita
• Liz Workman
• Olga Patterson
• Lina Bouayad
Consortium for Healthcare Informatics Research (CHIR): Group Photo

CHIR social network graph based on co-authorship
Example Opportunity

• Sections and Templates

OBSecAnnot: An Automated Section Annotator for Semi-structured Clinical Documents

Le-Thuy T. Tran, Guy Divita, Andrew Redd, Marjorie Carter, Joshua Judd, Matthew H. Samore, Adi V. Gundlapalli

Novel Template Identification from VA Text Integration Utility Notes

Andrew M. Redd, Guy Divita, Adi V. Gundlapalli, Le-Thuy Tran, Mathew Samore
Example Challenge

• Create synthetic text documents to facilitate collaboration with partners outside of VA
• Generate documents that retain key distributional properties of real notes (e.g., concept co-occurrences)
• Leverage thousands of template types
Interactive & Active Assisted Annotation and Natural Language Processing of Medical Text

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
Objectives

To develop and validate natural language processing tools capable of near real-time processing employing active & interactive learning techniques

Use Cases:

- Heart Failure Quality Indicator Elements
- Acute Kidney Injury Risk Factors
Online Assisted Annotation Workflow
RapTAT: Annotation/NLP Pipeline

Free Text Doc

XML File

RapTAT Sentence Boundary Detector

OpenNLP / RapTAT Tokenizer

OpenNLP Part-of-Speech Tagger

NLM LVG Lemmatizer

Solution File

RapTAT Concept Mapper

RapTAT Phrase Identifier

RapTAT Context Analyzer
RapTAT for Reviewer Training

Reference Annotations

Reviewer Annotations

RapTAT/eHOST Review Interface

Document Pool

F-Measures
Online Assisted Annotation

Heart Failure Use Case
**Study Design: Docs & Schema**

Documents:

Emergency Department and Inpatient provider and nursing notes as well as Primary care notes from 2007-2008 in 6 VA facilities (404 documents / 171 patients)

Annotation Schema:

<table>
<thead>
<tr>
<th>Concept</th>
<th># Documents containing concept</th>
<th>Example Phrases (Annotated Tokens in Bold)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angiotensin Converting Enzyme Inhibitor</td>
<td>272</td>
<td>&quot;ACEI,&quot; &quot;Altace,” &quot;Vaseretic&quot;</td>
</tr>
<tr>
<td>Angiotensin II Receptor Blocker</td>
<td>107</td>
<td>&quot;ARB,&quot; &quot;Sartans&quot;</td>
</tr>
<tr>
<td>Ejection Fraction</td>
<td>201</td>
<td>&quot;LVEF&quot;, &quot;Ejection fraction&quot;</td>
</tr>
<tr>
<td>Ejection Fraction Quantitation</td>
<td>197</td>
<td>&quot;EF=60-70%,&quot; &quot;EF is about 30%&quot;</td>
</tr>
<tr>
<td>Left Ventricular Systolic Function/Dysfunction</td>
<td>79</td>
<td>&quot;Systolic dysfunction,&quot; &quot;LV function&quot;</td>
</tr>
<tr>
<td>Left Ventricular Systolic Function Value</td>
<td>76</td>
<td>&quot;Systolic function is borderline normal&quot;</td>
</tr>
<tr>
<td>Documented Reason Not on ACE Inhibitor/ARB</td>
<td>40</td>
<td>“Patient refuses to take ACEI,” &quot;Renal disease&quot;</td>
</tr>
</tbody>
</table>
Study Design: Workflow

- **Document Training Batch**
- **Reviewer Training**
- **20 Batch Study Corpus**

**Workflow Steps:**
1. Annotation by Manual Reviewers
2. Manually Annotated Study Corpus
3. Manual Adjudication
4. Reference Study Corpus

**Pre-Annotation of Batch by RapTAT**

- **Annotation & Correction by Assisted Reviewers**
- **RapTAT Training**

**Raptat-Assisted Study Corpus**
Assisted Annotation Time

Seconds per KB of Text

Training Batch

○ Reviewer 1
■ Reviewer 2
Assisted Versus Manual Annotation Rate

- Assisted Reviewer 1
- Assisted Reviewer 2
- Manual Reviewer 1
- Manual Reviewer 2

Median Annotations Per Minute vs. Document Batch
### Inter-annotator agreement

<table>
<thead>
<tr>
<th>Concept</th>
<th>Average IAA (95% Confidence Interval)</th>
<th>Manual</th>
<th>Assisted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angiotensin Converting Enzyme Inhibitor</td>
<td>0.89 (0.86-0.93)</td>
<td></td>
<td>0.93* (0.91-0.96)</td>
</tr>
<tr>
<td>Angiotensin II Receptor Blocker</td>
<td>0.81 (0.72-0.89)</td>
<td></td>
<td>0.97* (0.95-1.00)</td>
</tr>
<tr>
<td>Ejection Fraction</td>
<td>0.86 (0.80-0.93)</td>
<td></td>
<td>0.97* (0.95-1.00)</td>
</tr>
<tr>
<td>Ejection Fraction Quantitation</td>
<td>0.90 (0.85-0.94)</td>
<td></td>
<td>0.88 (0.83-0.92)</td>
</tr>
<tr>
<td>Left Ventricular Systolic Function/Dysfunction</td>
<td>0.82 (0.73-0.91)</td>
<td></td>
<td>0.76 (0.62-0.89)</td>
</tr>
<tr>
<td>Left Ventricular Systolic Function Value</td>
<td>0.85 (0.78-0.93)</td>
<td></td>
<td>0.77 (0.64-0.90)</td>
</tr>
<tr>
<td>Reason Not on ACE Inhibitor/ARB</td>
<td>0.58 (0.46-0.70)</td>
<td></td>
<td>0.54 (0.45-0.64)</td>
</tr>
<tr>
<td>Total (Combined Over All Concepts)</td>
<td>0.85 (0.81-0.88)</td>
<td></td>
<td>0.89* (0.87-0.91)</td>
</tr>
</tbody>
</table>
## RapTAT Performance

<table>
<thead>
<tr>
<th>Concept</th>
<th>Performance Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Precision</td>
</tr>
<tr>
<td>Angiotensin Converting Enzyme Inhibitor</td>
<td>0.97</td>
</tr>
<tr>
<td>Angiotensin II Receptor Blocker</td>
<td>0.99</td>
</tr>
<tr>
<td>Ejection Fraction</td>
<td>0.96</td>
</tr>
<tr>
<td>Ejection Fraction Quantitation</td>
<td>0.77</td>
</tr>
<tr>
<td>Left Ventricular Systolic Function/Dysfunction</td>
<td>0.61</td>
</tr>
<tr>
<td>Left Ventricular Systolic Function Value</td>
<td>0.83</td>
</tr>
<tr>
<td>Reason Not on ACE Inhibitor/ARB</td>
<td>0.36</td>
</tr>
<tr>
<td>Total (Combined Over All Concepts)</td>
<td>0.87</td>
</tr>
</tbody>
</table>
Active Learning: Random vs Uncertainty Sampling

The graph compares the F-Measure against the number of training batches for Random Sampling and Uncertainty Sampling. The blue line represents Random Sampling, while the red line represents Uncertainty Sampling. As the number of training batches increases, both methods converge towards an F-Measure of approximately 0.8, with Uncertainty Sampling generally performing slightly better than Random Sampling.
Near Real-Time NLP

Acute Kidney Injury Use Case
Study Design

• Documents: Stratified sample of emergency department, inpatient, and outpatient provider notes

• 14 Annotated Blocks (112 documents/block [1568 total]): 8 Training Blocks / 6 Testing Blocks

• Results for this presentation only on training data (final performance pending testing data evaluation)
<table>
<thead>
<tr>
<th>Clinical Variable</th>
<th>Attributes *</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Renal Transplant Recipient</td>
<td></td>
<td>“Kidney Allograft”, “Renal Rejection”, “Renal Transplant Recipient”</td>
</tr>
<tr>
<td>Nephrology Care Delivery</td>
<td><strong>Type</strong>: General, Transplant, Dialysis</td>
<td>“Hemodialysis”, “Renal Consult”, “Nephrology Clinic”, “Renal Transplant Clinic”</td>
</tr>
<tr>
<td>NSAIDs</td>
<td></td>
<td>“Meloxicam”, “Ketorolac”, “Celecoxib”</td>
</tr>
<tr>
<td>ACE Inhibitors</td>
<td></td>
<td>“Enalapril”, “Mavik”</td>
</tr>
<tr>
<td>ARB</td>
<td></td>
<td>“Diovan”, “Losartan”</td>
</tr>
<tr>
<td>Diuretic</td>
<td></td>
<td>“Furosemide”, “Spirinolactone”</td>
</tr>
<tr>
<td>Diuresis</td>
<td></td>
<td>“diuresing”, “forced diuresis”</td>
</tr>
<tr>
<td>Intake</td>
<td><strong>Change</strong>: Increase, Decrease, Neutral</td>
<td>“Fluid resuscitate”, “no change in appetite”, “fluid restriction”, “intolerate of PO”, “NPO”</td>
</tr>
<tr>
<td></td>
<td><strong>Fluidity</strong>: Solid, Liquid, Both, Unstated</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Agency</strong>: Provider or Patient Initiated</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Delivery</strong>: IV, Oral, Unstated</td>
<td></td>
</tr>
<tr>
<td>Intravascular Volume</td>
<td><strong>Status</strong>: Low, High, Normal</td>
<td>“Hypovolemia”, “volume contraction”, “isovolemic”, “dry oral mucus membrane”</td>
</tr>
<tr>
<td>Weight Change</td>
<td><strong>Status</strong>: Increase, Decrease, Neutral</td>
<td>“fluctuating weight”, “cachectic”, “weight loss”, “weight gain”</td>
</tr>
<tr>
<td>Nausea/Vomiting/Diarrhea</td>
<td></td>
<td>“NVD”, “N/V/D”, “emesis”</td>
</tr>
</tbody>
</table>

* All variables include standard attributes: Assertion (Negation, Uncertainty), Time Frame (Past, Present, Future), Experiencer (Patient, Non-Patient)
## NLP Performance Summary

<table>
<thead>
<tr>
<th>Category</th>
<th>#</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>Precision (PPV)</th>
<th>Recall (Sens.)</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Drug Exposures</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• ACE Inhibitor</td>
<td>575</td>
<td>553</td>
<td>8</td>
<td>22</td>
<td>0.986</td>
<td>0.962</td>
<td>0.974</td>
</tr>
<tr>
<td>• ARB</td>
<td>149</td>
<td>137</td>
<td>0</td>
<td>12</td>
<td>1.000</td>
<td>0.919</td>
<td>0.958</td>
</tr>
<tr>
<td>• Diuretic</td>
<td>733</td>
<td>684</td>
<td>4</td>
<td>49</td>
<td>0.994</td>
<td>0.933</td>
<td>0.963</td>
</tr>
<tr>
<td>• NSAID</td>
<td>233</td>
<td>201</td>
<td>4</td>
<td>32</td>
<td>0.980</td>
<td>0.863</td>
<td>0.918</td>
</tr>
<tr>
<td><strong>Fluid Status</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Diuresis</td>
<td>118</td>
<td>83</td>
<td>6</td>
<td>35</td>
<td>0.933</td>
<td>0.703</td>
<td>0.802</td>
</tr>
<tr>
<td>• Intake</td>
<td>694</td>
<td>412</td>
<td>46</td>
<td>282</td>
<td>0.900</td>
<td>0.594</td>
<td>0.715</td>
</tr>
<tr>
<td>• Intravascular Volume Condition</td>
<td>527</td>
<td>432</td>
<td>12</td>
<td>95</td>
<td>0.973</td>
<td>0.820</td>
<td>0.890</td>
</tr>
<tr>
<td>• Nausea/Vomiting/Diarrhea</td>
<td>719</td>
<td>674</td>
<td>25</td>
<td>45</td>
<td>0.964</td>
<td>0.937</td>
<td>0.951</td>
</tr>
<tr>
<td>• Weight Change</td>
<td>221</td>
<td>130</td>
<td>14</td>
<td>91</td>
<td>0.903</td>
<td>0.588</td>
<td>0.712</td>
</tr>
<tr>
<td><strong>Radiographic Media Exposure</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Contrast</td>
<td>2095</td>
<td>1858</td>
<td>240</td>
<td>237</td>
<td>0.886</td>
<td>0.887</td>
<td>0.886</td>
</tr>
<tr>
<td>• Potential Contrast</td>
<td>439</td>
<td>255</td>
<td>65</td>
<td>184</td>
<td>0.797</td>
<td>0.581</td>
<td>0.672</td>
</tr>
<tr>
<td><strong>Renal Status</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Anatomical Kidney Status</td>
<td>57</td>
<td>9</td>
<td>4</td>
<td>48</td>
<td>0.692</td>
<td>0.158</td>
<td>0.257</td>
</tr>
<tr>
<td>• Nephrology Care Delivery</td>
<td>210</td>
<td>141</td>
<td>36</td>
<td>69</td>
<td>0.797</td>
<td>0.671</td>
<td>0.729</td>
</tr>
<tr>
<td>• Renal Function Impairment</td>
<td>449</td>
<td>368</td>
<td>44</td>
<td>81</td>
<td>0.893</td>
<td>0.820</td>
<td>0.855</td>
</tr>
<tr>
<td><strong>Total Concept Performance</strong></td>
<td>7231</td>
<td>5661</td>
<td>341</td>
<td>1570</td>
<td>0.921</td>
<td>0.821</td>
<td>0.868</td>
</tr>
</tbody>
</table>
## Concept Assertion Performance

<table>
<thead>
<tr>
<th>Reference Standard</th>
<th>Algorithm</th>
<th>Recall (Sensitivity)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>1736</td>
<td>27</td>
</tr>
<tr>
<td>Negative</td>
<td>142</td>
<td>350</td>
</tr>
<tr>
<td>Uncertain</td>
<td>52</td>
<td>26</td>
</tr>
</tbody>
</table>

| Precision (PPV)    | 90%       | 87%                  | 38% | 85% |

<table>
<thead>
<tr>
<th>Positive vs Negative</th>
<th>TP</th>
<th>FP</th>
<th>TN</th>
<th>FN</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1849</td>
<td>167</td>
<td>358</td>
<td>167</td>
<td>92%</td>
<td>68%</td>
</tr>
</tbody>
</table>
Near Real-Time Deployment

- **National Cardiac Catheterization Cohort:**
  - 158,432 Patients (as of 09/2013)
  - 1,256,685 Documents (Filtered)

Processing Speed: ~1 sec/ document on an single machine installation = 86,400 / day
Conclusions

• Assisted annotation tools can reduce cost of NLP training samples – key requisite for all NLP tool performance assessment

• Focus on accuracy but awareness and optimization for performance in NLP tools necessary to allow near-real time implementation

• Focused NLP tasks, while less generalizable than generalized solutions, are likely to be the first implemented in point of care applications due to optimization and higher accuracy
Acknowledgements

Grant Funding

• Denver VA
  – Tom Maddox
  – Meg Plomondon
  – Tom Tsai

• Salt Lake VA
  – Wendy Chapman
  – Jennifer Garvin
  – Stephane Meystre

• Tampa VA
  – Steve Luther
  – Dezon Finch
  – James McCart

• Nashville
  – Steven Brown
  – Fern FitzHenry
  – James Fly
  – T. Alp Ikizler
  – Shrimalini Jayaramaraja
  – Vincent Messina
  – Sharidan Parr
  – Ruth Reeves
  – Sanjib Saha
  – Edward Siew
  – Dax Westerman
  – Theodore Speroff

• Denver VA
  – Tom Maddox
  – Meg Plomondon
  – Tom Tsai

• Salt Lake VA
  – Wendy Chapman
  – Jennifer Garvin
  – Stephane Meystre

• Tampa VA
  – Steve Luther
  – Dezon Finch
  – James McCart

VA HSR&D IIR 11-292-(Matheny)
VA HSR&D IIR 13-052 (Matheny/Ho)
VA HSR&D CDA 08-020 (Matheny)

VA HSR&D IIR 12-362 (Reeves)
PCORI CDRN Phase 1 & 2 (pScanner Ohno-Machado)

For more information contact:

michael.matheny@va.gov
michael.matheny@vanderbilt.edu

RapTAT Lead Developer & Key Collaborator: Glenn Gobbel
Applying NLP to Clinical Problems

“no family history of colon cancer”

(no (family history (of (colon cancer))))

C00289 negated

• Tell me how I can improve the quality of my colonoscopy exams
• Find patients with X
• Help me use the words in my report that justify fair billing codes
• Help me spend less time documenting
• Help my patient understand her report
Clinical NLP

Methodology Development
Many interesting NLP research problems are far upstream from health care applications.
To have impact, we must go beyond improving f-score to create tools that can be applied to real-world problems.
Clinical NLP

Methodology Development

Application of Methods to Domain Problem
Many new, interesting NLP research problems will arise when working on user-driven development of clinical applications.
Clinical NLP

VA NLP

Methodology Development

Application of Methods to Domain Problem
Effective NLP in the VA = Partnership

Informaticist

NLP Expert

Domain Expert
IE-Viz
Information Extraction and Visualization

Knowledge Authoring
NLP Customizing
Classifier Development
Visualization
“Patient denies a family history of colon cancer”

Disease: colon cancer
Experencer: family
Negation: no
Historical: yes
Modifier Ontology

Types of modifiers

Linguistic expressions

Actions

Translations
<table>
<thead>
<tr>
<th>Medications</th>
<th>Diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>– Type</td>
<td>– Negation</td>
</tr>
<tr>
<td>– Dose</td>
<td>– Uncertainty</td>
</tr>
<tr>
<td>– Frequency</td>
<td>– Severity</td>
</tr>
<tr>
<td>– Route</td>
<td>– History</td>
</tr>
<tr>
<td></td>
<td>– Experiencer</td>
</tr>
</tbody>
</table>

Consistent with other models:
Clinical element models, cTAKES type system,
Common model, FHIR
Domain Ontology for NLP

• Instance of schema ontology
• Clinical elements from a particular domain
Synonyms
Misspellings
Regular expressions
Knowledge Author

• Front end interface for users
• Back end
  – Schema ontology
  – Modifier ontology
• Output
  – Domain ontology
  – Schema for NLP system
African American Adult

Create Person Variable

Person Role

- Patient
- Family Member

Extra Fields

- Race: African American
- Death Date
- Birth Date
- Ethnicity
Ibuprofen

Create Disease Variable

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Link to Reference</th>
<th>Source</th>
<th>Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>ibuprofen</td>
<td>2-(acetylxy)benzoic acid 6-</td>
<td>UMLS</td>
<td>Pharmacologic Substance</td>
</tr>
<tr>
<td></td>
<td>(nitroxymethyl)-2-phenylmethyl ester</td>
<td></td>
<td>Organic Chemical</td>
</tr>
<tr>
<td>ibuprofen allergy</td>
<td>2-(acetylxy)benzoic acid 6-</td>
<td>UMLS</td>
<td>Pathologic Function</td>
</tr>
<tr>
<td>IBUPROFEN INTOLERANCE</td>
<td>(nitroxymethyl)-2-phenylmethyl ester</td>
<td>UMLS</td>
<td>Pathologic Function</td>
</tr>
<tr>
<td>ibuprofen-Zinc</td>
<td>2-(acetylxy)benzoic acid 6-</td>
<td>UMLS</td>
<td>Organic Chemical</td>
</tr>
</tbody>
</table>
Ibuprofen p.o. (per oral)
No family history of colon cancer

Linguistic modifiers

<table>
<thead>
<tr>
<th>Link</th>
<th>Name</th>
<th>Source</th>
<th>Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Malignant tumor of colon</td>
<td>UMLS</td>
<td>Neoplastic Process</td>
</tr>
<tr>
<td></td>
<td>Malignant neoplasm of large intestine</td>
<td>UMLS</td>
<td>Neoplastic Process</td>
</tr>
<tr>
<td></td>
<td>Colon Carcinoma</td>
<td>UMLS</td>
<td>Neoplastic Process</td>
</tr>
<tr>
<td></td>
<td>COLON CANCER (allelic variant)</td>
<td>UMLS</td>
<td>Gene or Genome</td>
</tr>
<tr>
<td></td>
<td>Colonic Neoplasms</td>
<td>UMLS</td>
<td>Neoplastic Process</td>
</tr>
</tbody>
</table>
Select Synonyms

- colon cancer
- cancer of colon
- colonic cancer
- malignant neoplasm of colon
- malignant colon neoplasm
- malignant colonic tumor
- cancer of the colon
- colon cancers
- ca colon
- malignant colonic neoplasm
- cancer, colon
- cancer colons
- cancers, colon
- colon ca
- colon tumor, malignant
- malignant colon tumor
- cancer colonic
- colon neoplasm, malignant
- colonic cancers
- cancers, colonic
Research Questions We Have Addressed

Knowledge Authoring

• Which modifiers are important for interpreting clinical text?
• Does the modifier ontology work with the ConText algorithm for other languages?
• How well do humans agree when annotating modifiers?
• Can we mine text to learn terms used as modifiers?
• Can we assist a non-NLP expert user in developing a domain knowledge base?
• Can we learn synonyms from text?
IE-Viz
Information Extraction and Visualization

Knowledge Authoring
NLP Customizing
Classifier Development
Visualization
NLP Customization

- Domain Ontology
- NLP Tools
- Evaluation Workbench
Use Case: Colonoscopy Quality Measures

Samir Gupta
San Diego VA

Andrew Gawron
SLC VA
Overlapping Measures
Shared knowledge representation enables collaboration
Cumulative Projects in VINCI using Text
NLP Challenges

- Can't see the forest for the trees
  - Clinical NLP initially worked on local data in narrow domains but now needs to handle large and diverse data sets

- Every type of fruit requires a different ladder
  - Most use cases require tailoring and refinement of existing tools
  - Novel use cases emerge

- Fruits are not always low hanging
  - Some use cases are complex: stroke symptoms within 2 hours of ER admission or early signs dementia
My VA CLINICAL NLP Work

NLP Methods

Notes, Instructions, Forms

NLP Ecosystem

Novel Applications

Text Mining for Clinical/Health Services Research
Progress So Far

- We have conducted interviews and workshops to assess strengths and weaknesses in current VA NLP development, and needs relating to ecosystem
- We created a prototype ecosystem cnlpecosystem.org
  - Sample data, Sample tools, VA NLP Bibliography
- Initial analysis of VA records indicate the existence of over a dozen sublanguage groups
My VA CLINICAL NLP Work

- NLP Methods
- Notes, Instructions, Forms
- NLP Ecosystem
- Novel Applications
- Text Mining for Clinical/Health Services Research
Novel Regular Expression Algorithm

• Improve text classification and information extraction
  • REDCL for text classification
  • REDEX for information extraction
Sample Size Prediction
Active Learning

Relation between ACC and uncertainty of the datasets

Relation between ACC and diversity of the examples of the datasets
<table>
<thead>
<tr>
<th>Input</th>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>S: He smokes 1 pack of cigarettes a day</td>
<td>He smokes 1 pack of cigarettes a day</td>
<td>N: He is a non-smoker</td>
</tr>
<tr>
<td>S: He smokes five cigars per day</td>
<td>He smokes one pack a day</td>
<td>N: Non-smoker since 2000</td>
</tr>
<tr>
<td>S: Patient smokes one pack a day</td>
<td>He continues to smoke one-half pack per day</td>
<td>P: He smoked about one pack per day for 50 years</td>
</tr>
<tr>
<td>S: He continues to smoke one-half pack per day</td>
<td>He smokes five cigars per day</td>
<td>U: Do not smoke during oxygen administration</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Alignment</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>He smokes 1 pack of cigarettes a day</td>
<td>He continues to smoke one-half pack per day</td>
<td>He smokes one pack a day</td>
</tr>
<tr>
<td>He smokes five cigars per day</td>
<td>He continues to smoke one-half pack per day</td>
<td>He continues to smoke one-half pack per day</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Key Generation</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>[he smokes, day]</td>
<td>[he, per day]</td>
<td>[pack, day]</td>
<td>[per day]</td>
<td>[pack, day]</td>
</tr>
<tr>
<td>He(s^+s^+{4,5})s+day</td>
<td>He(s^+s^+{5,7})s+day</td>
<td>He(s^+s^+{1})s+smokes(s^+s^+{1})s+pack(s^+s^+{0,2})s+a\s+day</td>
<td>He(s^+s^+{5,7})s+day</td>
<td></td>
</tr>
<tr>
<td>He(s^+s^+{3,5})s+per\s+day</td>
<td>He(s^+s^+{3,5})s+per\s+day</td>
<td>He(s^+s^+{3,5})s+per\s+day</td>
<td>He(s^+s^+{5,7})s+day</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pattern Builder</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>He(s^+s^+{4,5})s+day</td>
<td>He(s^+s^+{5,7})s+day</td>
<td>He(s^+s^+{1})s+smokes(s^+s^+{1})s+pack(s^+s^+{0,2})s+a\s+day</td>
<td>He(s^+s^+{5,7})s+day</td>
<td></td>
</tr>
<tr>
<td>He(s^+s^+{3,5})s+per\s+day</td>
<td>He(s^+s^+{3,5})s+per\s+day</td>
<td>He(s^+s^+{3,5})s+per\s+day</td>
<td>He(s^+s^+{5,7})s+day</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Filtering</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>He(s^+s^+{4,5})s+day</td>
<td>He(s^+s^+{5,7})s+day</td>
<td>He(s^+s^+{1})s+smokes(s^+s^+{1})s+pack(s^+s^+{0,2})s+a\s+day</td>
<td>He(s^+s^+{5,7})s+day</td>
<td></td>
</tr>
<tr>
<td>He(s^+s^+{3,5})s+per\s+day</td>
<td>He(s^+s^+{3,5})s+per\s+day</td>
<td>He(s^+s^+{3,5})s+per\s+day</td>
<td>He(s^+s^+{5,7})s+day</td>
<td></td>
</tr>
</tbody>
</table>
Comparison with SVM
Blood pressure 132/70, weight is 128 pounds. HEENT: Head, ears, nose, throat and eyes are normal.
# Sample Results

<table>
<thead>
<tr>
<th></th>
<th>Weight</th>
<th>METs</th>
<th>Date</th>
<th>Time</th>
<th>Duration</th>
<th>Set</th>
<th>KATZ</th>
<th>Pain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>0.99</td>
<td>0.89</td>
<td>0.97</td>
<td>0.98</td>
<td>0.93</td>
<td>1.00</td>
<td>0.99</td>
<td>0.88</td>
</tr>
<tr>
<td>Recall</td>
<td>0.98</td>
<td>0.92</td>
<td>0.97</td>
<td>0.97</td>
<td>0.96</td>
<td>0.83</td>
<td>1.00</td>
<td>0.97</td>
</tr>
<tr>
<td>F1-score</td>
<td>0.99</td>
<td>0.90</td>
<td>0.97</td>
<td>0.98</td>
<td>0.95</td>
<td>0.91</td>
<td>0.99</td>
<td>0.92</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.98</td>
<td>0.91</td>
<td>0.94</td>
<td>0.95</td>
<td>0.90</td>
<td>0.83</td>
<td>0.99</td>
<td>0.92</td>
</tr>
<tr>
<td># Snippets</td>
<td>968</td>
<td>2701</td>
<td>493</td>
<td>289</td>
<td>169</td>
<td>18</td>
<td>1000</td>
<td>3862</td>
</tr>
<tr>
<td># Reg Ex</td>
<td>98</td>
<td>812</td>
<td>128</td>
<td>54</td>
<td>41</td>
<td>7</td>
<td>35</td>
<td>572</td>
</tr>
</tbody>
</table>
My WORK IN CLINICAL NLP

- NLP Methods
- Notes, Instructions, Forms
- NLP Ecosystem
- Novel Applications
- Text Mining for Clinical/Health Services Research
Cohort Sizes and Overlap

- SD: 6,063
- UD: 101,638
- Total: 38,897
Geographic distribution
Gingko-Warfarin Bleeding Risk

HR = 1.4, p < 0.001

Number at risk
NonGinkgo: 715383 625837 565145 517955
Ginkgo: 8759 1763 1079 818

Days Follow-Up
Survival Probability
My WORK IN CLINICAL NLP

- NLP Methods
- Notes, Instructions, Forms
- NLP Ecosystem
- Novel Applications
- Text Mining for Clinical/Health Services Research
Example

Avoid driving if you are light-headed, drowsy or dizzy

Pre-Processing

Avoid drive if light-head, drowsy or dizzy

Annotation

<negation> <atomic> if light-head, <atomic> or <atomic>

Post-Processing

<negation> <atomic> if <atomic> or <atomic>

Image Composition

Sequence_[$3]_[$4]_[image arrow]_[not $2]

Image Rendering
We could find 5308 patients like you out of which 98.1% were male and the rest where female. The mean age of the patients is 56.5 years. Here are the personal stories of 100 patients:

**Gender: Female  Age: 40yr  Personal Care Aid Provider**

Story: The problem came on very suddenly and with no previous history. Teresa settled down to watch television, experienced sudden and severe palpitations and pain in her shoulders. She tried to stay calm but it was very frightening. She hoped dabigatran would be effective. Despite dabigatran, Teresa did not immediately feel better. It was really difficult for Teresa’s friends to accept Teresa’s passing 8 years and 4 months later.

**Gender: Female  Age: 41yr  Secretary**

Story: The feeling of having atrial fibrillation is debilitating for Susan. Warfarin alleviated Susan’s chest pain. Susan’s physician is optimistic about her outcome.
Patient Information Needs & Patient Centered Outcome
Discussion

• We have developed new methods, tools, and applications.
• Our research has enabled clinical/health services research and led to knowledge discovery
• There is a lot of room for improvement
• We have not been incorporated into VA operation, but hope to...
Patient-Centered Research

Systems for Helping Veterans Comprehend Electronic Health Record Notes

Hong Yu¹,², Balaji Polepalli Ramesh¹,², Jiaping Zheng³, Jinying Chen², Louise Maranda², Cynthia Brandt⁴,⁵, Kathleen Mazor², Donna Zulman⁶,⁷, Thomas Houston¹,²

¹VA Bedford VAMC
²University of Massachusetts Medical School
³University of Massachusetts, Amherst
⁴VA Connecticut Health Care System
⁵VA Palo Alto Health Care System
⁶Stanford University
Background

• Patients reading their EHR notes has the potential to
  □ Enhance medical understanding
  □ Improve healthcare management and outcomes

• Blue Button: The Department of Veterans Affairs (VA)
  Blue Button enables Veterans to view, print, and
download their EHRs, including clinical notes (e.g.,
progress notes).
The Challenge

• Physicians’ notes are difficult to comprehend (Keselman et al)

• Many Veterans have limited health literacy (Schapira et al)

“The patient will be scheduled for a repeat EGD in one year for surveillance purposes of Barrett's esophagus. From a GI standpoint, we recommend to proceed with bariatric surgery. However, he will need to continue daily PPI administration to maximize acid reduction. Otherwise, there are no additional recommendations. The patient was treated with myocardial infarction.”
The NoteAid System

• A system for helping patients comprehend electronic health record notes

• Automatically
  □ Identifies clinically relevant concepts
  □ Links concepts to their definitions and lay language
  □ Links notes to other education material

□ Funded by HSR&D (1I01HX001457) since May, 2015
Research

• Lay language resources and education materials
• NLP for translating EHR to lay language and linking EHR to education materials
Lay Language Resources

• Existing resources
  – Consumer Health Vocabulary (Zeng et al)
  – MedlinePlus, etc
  – However, we found 40%~60% EHR jargon do not appear in existing resources
**Mining Lay Language from Wikipedia**

**Diabetes mellitus (DM)**, commonly referred to as **diabetes**, is a group of **metabolic diseases** in which there are high **blood sugar** levels over a prolonged period. Symptoms of high blood sugar include **frequent urination**, **increased thirst**, and **increased hunger**. If left untreated, diabetes can cause many complications. **Acute** complications include **diabetic ketoacidosis** and **nonketotic hyperosmolar coma**. **Serious long-term complications** include **cardiovascular disease**, **stroke**, **chronic kidney failure**, **foot ulcers**, and **damage to the eyes**.

---

**Polydipsia**

From Wikipedia, the free encyclopedia

For the term formerly used in reference to compulsive drinking of **alcohol**, see **Dipsomania**.

See also: **Polydipsia in birds**

**Polydipsia** is excessive **thirst**. The word derives from the **Greek** πολυδύψις, which is derived from πολύς (polys, "much, many") + δύψα (dipsa, "thirst"). An etymologically related term is **dipsomaniac**, meaning an **alcoholic**. Polydipsia is a nonspecific **symptom** in various medical disorders. It also occurs as an **abnormal behaviour in animals**.
Methods, Evaluation, Results

- Entity Link Frequency (baseline)
- Similarity (word embedding)
- Wiki Pseudo Relevance Feedback
- Similarity+PRF

<table>
<thead>
<tr>
<th>Methods</th>
<th>MAP (Relaxed condition)</th>
<th>MAP (strict condition)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ELF</td>
<td>0.6267</td>
<td>0.2401</td>
</tr>
<tr>
<td>ACS</td>
<td>0.6624</td>
<td>0.2383</td>
</tr>
<tr>
<td>PRF</td>
<td>0.6859</td>
<td>0.2519</td>
</tr>
<tr>
<td>AveR</td>
<td>0.6685</td>
<td>0.2433</td>
</tr>
<tr>
<td>RSC</td>
<td>0.6900</td>
<td>0.2745</td>
</tr>
</tbody>
</table>

Table 1: Mean Average Precision values for Relevance Feedback of 5
EHR notes are frequently long and full of jargon, overwhelming patients.

The other possible etiologies for her nephrotic syndrome could be FSGS, myeloma, or an interstitial nephritis. I will send for urine eosinophil. I will also send for myeloma workup with a urine immunofixation electrophoresis.
Methods

Using CHV terms and CHV familiarity scores to train supervised machine learning models (e.g., SVM regression) to predict EHR term importance and term familiarity

<table>
<thead>
<tr>
<th></th>
<th>Ave. Precision (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TF*IDF</td>
<td>8.83</td>
</tr>
<tr>
<td>C-Value</td>
<td>10.23</td>
</tr>
<tr>
<td>Ranking by importance (UMLS)</td>
<td>19.42</td>
</tr>
<tr>
<td>Ranking by importance + unfamiliarity</td>
<td>23.31</td>
</tr>
</tbody>
</table>
Patient remains in ICU with the following problems: Respiratory failure, hemodynamics, renal failure, status post liver transplant, atrial fib, infectious disease, nutrition.

**Education Material:**
- Respiratory Failure
- **Deep Vein Thrombosis**
- Aspiration pneumonia
- Pulmonary Hypertension
- Kidney Failure
- Atrial Fibrillation or Flutter
- Liver Transplantation
- **Dialysis** - Hemodialysis
<table>
<thead>
<tr>
<th>System</th>
<th>P@10</th>
<th>MAP</th>
<th>Increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>0</td>
<td>0.0091</td>
<td>-</td>
</tr>
<tr>
<td>CHV</td>
<td>5</td>
<td>0.0240</td>
<td>2.6</td>
</tr>
<tr>
<td>LDA</td>
<td>10</td>
<td>0.0489</td>
<td>5.4</td>
</tr>
<tr>
<td>Key (Wiki)</td>
<td>16</td>
<td>0.0851</td>
<td>9.4</td>
</tr>
<tr>
<td>Key (EHR)</td>
<td>16.5</td>
<td>0.0879</td>
<td>9.7</td>
</tr>
<tr>
<td>Key (Wiki+EHR)</td>
<td>18</td>
<td>0.1030</td>
<td>11.3</td>
</tr>
<tr>
<td>Instance Pruning</td>
<td>9.5</td>
<td>0.0316</td>
<td>3.5</td>
</tr>
<tr>
<td>Feature Augmentation</td>
<td>14</td>
<td>0.0684</td>
<td>7.5</td>
</tr>
<tr>
<td>Instance Weighting</td>
<td>21.5</td>
<td>0.1111</td>
<td>12.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>System</th>
<th>Wiki</th>
<th>EHR</th>
<th>Wiki+EHR</th>
<th>IW</th>
<th>IP</th>
<th>FA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>16.27</td>
<td>35.92</td>
<td>33.79</td>
<td>47.59</td>
<td>40.00</td>
<td>46.60</td>
</tr>
<tr>
<td>Recall</td>
<td>26.88</td>
<td>34.09</td>
<td>33.18</td>
<td>34.41</td>
<td>6.02</td>
<td>28.86</td>
</tr>
<tr>
<td>F1</td>
<td>18.74</td>
<td>33.70</td>
<td>32.54</td>
<td>38.32</td>
<td>10.23</td>
<td>34.08</td>
</tr>
</tbody>
</table>
Evaluation 1: Self-Reported Comprehension

- 25 Lay people
- De-identified notes (not own notes)
- With or without NoteAid
- Self-reported comprehension
  - 1 to 5
  - 1: impossible to understand
  - 5: understand completely
## Results

- **Average self-reported comprehension scores**

<table>
<thead>
<tr>
<th>System</th>
<th>Note Alone</th>
<th>MedlinePlus</th>
<th>UMLS</th>
<th>Wiki</th>
<th>Hybrid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score</td>
<td>2.95±0.67</td>
<td>4.12±0.33*</td>
<td>3.63±0.57*</td>
<td>3.85±0.47*</td>
<td>3.92±0.40*</td>
</tr>
</tbody>
</table>

*p<0.01, Non-parametric Mann-Whitney Wilcoxon signed-rank test*
Evaluation 2: Paraphrasing

• **Data:** 5 de-identified progress notes
• **Subjects:** 40 subjects
• **Evaluation Process:**
  – Each subject was presented with 5 notes, one at a time, either with NoteAid (hybrid), or without.
  – Each subject paraphrased the main content of the note
**Examples: Without NoteAid**

- **Note**: The patient is doing okay today. He has some complaints of discomfort around his tracheostomy site but otherwise has no complaints of chest pain, shortness of breath, abdominal pain, nausea, or vomiting. **OBJECTIVE**: Vital Signs : Currently stable. Cardiac : Normal S1, S2. No murmurs, rubs, or gallops are present. Lungs : Clear to auscultation bilaterally. Abdomen : Soft, nontender, and nondistended. Normoactive bowel sounds are present. **LABS** : There are no new labs from today. **ASSESSMENT AND PLAN** : 1. Preoperative assessment from my prior note. 2. Hyponatremia likely to be SIADH related. Continue to monitor for now. 3. Leukocytosis. Likely to be reactive. No clear infectious source. 4. Thrombocytopenia. It is probably chronic in nature and appears to be stable. Although, we would continue to monitor this for now as well.

- **Paraphrase note 1**: “The patient seems to be doing alright despite some minor discomfort. His vital signs are stable and he hasn't had any labs today. The plan going forward is to keep monitoring the patient.”

- **Paraphrase note 2**: “The patient appears to be doing well aside from minor discomfort around the site of where the tracheostomy was performed. Other than that, they do not appear to have any other discomfort. There are other apparent issues like **Hyponatremia** and **Leukocystosis** which appear to be a result of the procedure. The patient appears to be doing well, but should continue being monitored.”
**Examples: NoteAid**

- **Note**: The patient is doing okay today. He has some complaints of *discomfort* around his tracheostomy site but otherwise has no complaints of *chest pain*, shortness of breath, *abdominal pain*, *nausea*, or *vomiting*. **OBJECTIVE**: **Vital Signs**: Currently stable. Cardiac: Normal S1, S2. **No murmurs**, rubs, or gallops are present. Lungs: Clear to **auscultation** bilaterally. Abdomen: Soft, nontender, and nondistended. Normoactive **bowel sounds** are present. **LABS**: There are no new labs from today. **ASSESSMENT AND PLAN**: 1. Preoperative assessment from my prior note. 2. **Hyponatremia likely to** be **SIADH** related. Continue to monitor for now. 3. **Leukocytosis**. Likely to be reactive. No clear infectious source. 4. **Thrombocytopenia**. It is probably chronic in nature and appears to be stable. Although, we would continue to monitor this for now as well.

- **Paraphrase note**: "The patient is doing well although he has some discomfort around his tracheostomy area. The patient has stable vitals. Some conditions will continue to be monitored, including **low salt content** in the blood, **high white blood cell count**, and **low blood platelet count**."
## Evaluation of Summary

<table>
<thead>
<tr>
<th>Terms in 195 Summaries</th>
<th>With NoteAid</th>
<th>Without NoteAid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of medical jargon terms in lay language</td>
<td>268</td>
<td>68</td>
</tr>
<tr>
<td>Number of medical jargon terms copied and pasted</td>
<td>70</td>
<td>169</td>
</tr>
</tbody>
</table>
Patient Evaluation

• Recruit 3 patients, all college graduates
  – One Type I diabetes, two Type II diabetes
  – Being treated for 5~15 years

• Each patient presented with 1~2 own progress notes of his/her most recent clinical visit
  – Ask what they don’t understand
  – Present Note+NoteAid
  – Think-aloud
  – A list of questions
Concepts and Abbreviations

• The 3 patients did not understand 5~10 concepts in his/her own progress note
  – E.g., FT4, auscultation, post-ablative hypothyroidism
Questions

• Do you want to read your EHR notes?  Yes (3)
• When do you want to read your notes?
  – Before and after a clinical visit (3)
• Is there a specific part of notes that you would like to read?
  – All notes (2), current plan (1)
• Would you be willing to share your notes with your friends or family members?
  – No (2), spouse if asked (1)
• Is NoteAid helpful?  Yes (3)
• Do you want to use NoteAid?  Yes (3)
How to Improve NoteAid

- More concepts included by NoteAid (3)
- Easy-to-read definitions (3)
- More education material (3)
Accomplishments: Collaborations

• Collaborations with multiple VA and Universities
  – University of Massachusetts Medical School
  – University of Massachusetts, Amherst
  – VA Connecticut Health Care System
  – Yale University
  – VA Palo Alto Health Care System
  – Stanford University
Accomplishments: Publications

Accomplishments: Dissemination

• Best Student Paper Award
• Major Computer Science, Biomedical Informatics, and Behavior Science conferences
  – SIGIR, EMNLP, ACL, AAAI, MedInfo
• The 2015 HSR&D/QUERI National Conference
• HSR&D Cyberseminars
• Keynote at ICDM BioDM 2013
• Invited seminars at VA
  – Bedford VAMC, West Haven, CWM
  – eHMP
• Invited seminars at Universities
  – UMass-Boston, UMass-Lowell, UMass-Dartmouth, WPI
  – Peking University, University of Science and Technology Beijing
Opportunities

• VA is innovative
• VA is passionate
• VA is collaborative
• VA has excellent leadership
• VA is investing research and researchers
• VA is supportive
• VA is full of opportunities
  – My HealtheVet
    • Blue Button
  – eHMP
Challenges

- IRB application
- EHR data access
- Collaboration with operational partners
- Research fund spending
  - Hiring
  - Student support
  - Inflexibility
- A lot of paper work---fortunately VA is supportive
Thanks and Questions

☐ The NoteAid Prototype Systems:
  ☐ http://www.clinicalnotesaid.org

☐ Contact
  ☐ Hong.yu@va.gov or hong.yu@umassmed.edu
  ☐ Hong Yu, MS, MA, and PhD
    Research Health Scientist,
    Bedford VAMC and
    Professor, Department of Quantitative Health Sciences
    Department of Computer Science (Adjunct)
    University of Massachusetts Medical School
    AS6-2071 ASC-QHS, 368 Plantation Street, Worcester, MA 01605
    508-856-3474/ Fax: 508-856-8993