Clinical Natural Language Processing and Audio

Tristan Naumann, Microsoft Research

CSC2541HS GUEST LECTURE

@TristanNaumann
tristan@microsoft.com

Using Clinical Text is Crucial

- Doctors use **text** to communicate patient state.
- Notes often contain the **most important** information

__date__ 4:07 AM
CHEST (PORTABLE AP) Clip # __num__
Reason: ETT tube placement, progression of pulmonary process
Admitting Diagnosis: NON-HODGKIN LYMPHOMA

__hospital__ MEDICAL CONDITION:
64 year old man s/p allo BMT for follicular lymphoma intubated now with worsening respiratory status
REASON FOR THIS EXAMINATION:
ETT tube placement, progression of pulmonary process

FINAL REPORT
HISTORY: BMT for lymphoma with respiratory status worsening.

FINDINGS: In comparison with study of __date__, the tip of the endotracheal tube now measures approximately 3.2 cm above the carina. Central catheter and nasogastric tube remain in place. There is continued mild enlargement of the cardiac silhouette in a patient with low lung volumes. Indistinctness of engorged pulmonary vessels is consistent with elevated pulmonary venous pressure. The possibility of supervening consolidation cannot be excluded if there is appropriate clinical symptomatology.

ID	Description
6112	left vent drain
2734	right vent drain
1726	HIGH MIN VENT
1496	HIGH MIN. VENT.
1488	HIGH MIN. VENTIL.
1599	HIGH MINUTE VENT.

Clinical Text Presents Unique Challenges

CONTEXT MATTERS

ACRONYMS

__date__ 4:07 AM
CHEST (PORTABLE AP) Clip # __num__
Reason: ETT tube placement, progression of pulmonary process
Admitting Diagnosis: NON-HODGKIN LYMPHOMA

__hospital__ MEDICAL CONDITION:
64 year old man s/p allo BMT for follicular lymphoma intubated now with worsening respiratory status
REASON FOR THIS EXAMINATION:
ETT tube placement, progression of pulmonary process

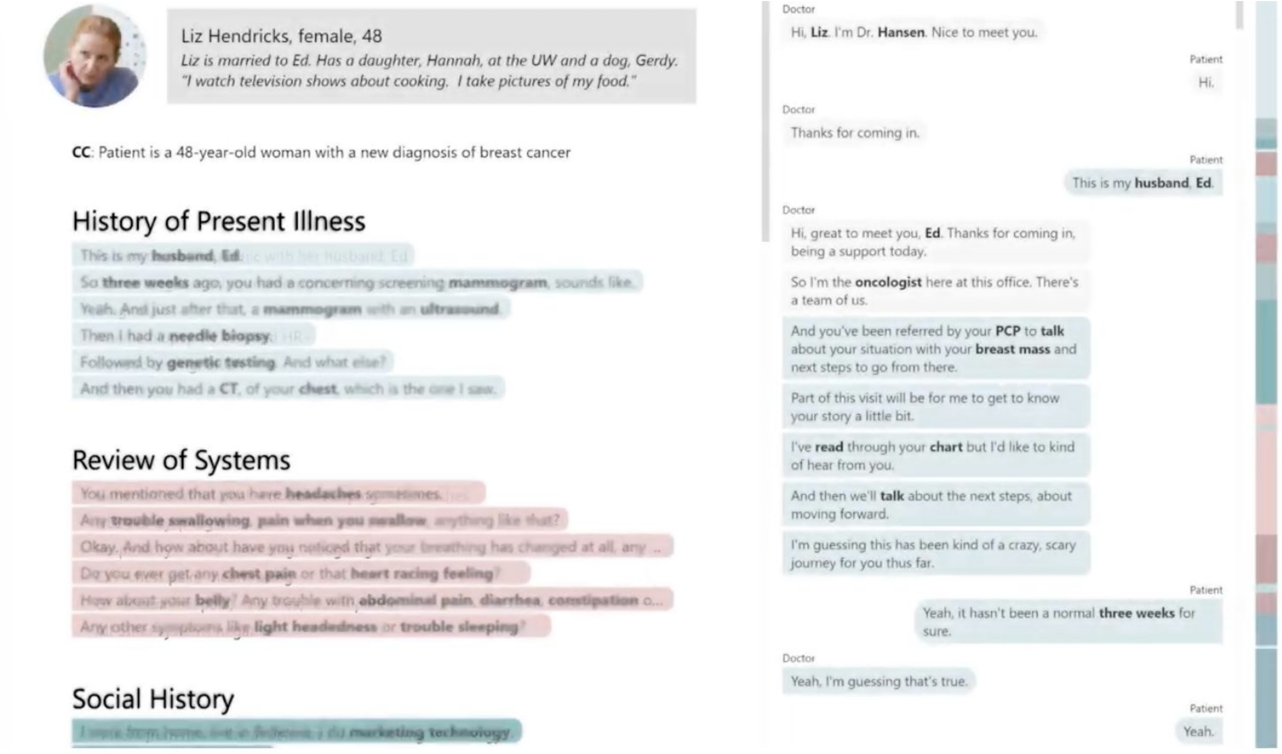
FINAL REPORT
HISTORY: BMT for lymphoma with respiratory status worsening.

FINDINGS: In comparison with study of __date__, the tip of the endotracheal tube now measures approximately 3.2 cm above the carina. Central catheter and nasogastric tube remain in place. There is continued mild enlargement of the cardiac silhouette in a patient with low lung volumes. Indistinctness of engorged pulmonary vessels is consistent with elevated pulmonary venous pressure. The possibility of supervening consolidation cannot be excluded if there is appropriate clinical symptomatology.

ID	Description
6112	left vent drain
2734	right vent drain
1726	HIGH MIN VENT
1496	HIGH MIN. VENT.
1488	HIGH MIN. VENTIL.
1599	HIGH MINUTE VENT.

Clinical Audio Presents New Opportunities

- Automatic transcription of patient conversations and synthesis of notes.



Clinical Audio Presents New Opportunities

- Automatic transcription of patient conversations and synthesis of notes.

PRESS RELEASE

Nuance Announces the General Availability of Ambient Clinical Intelligence

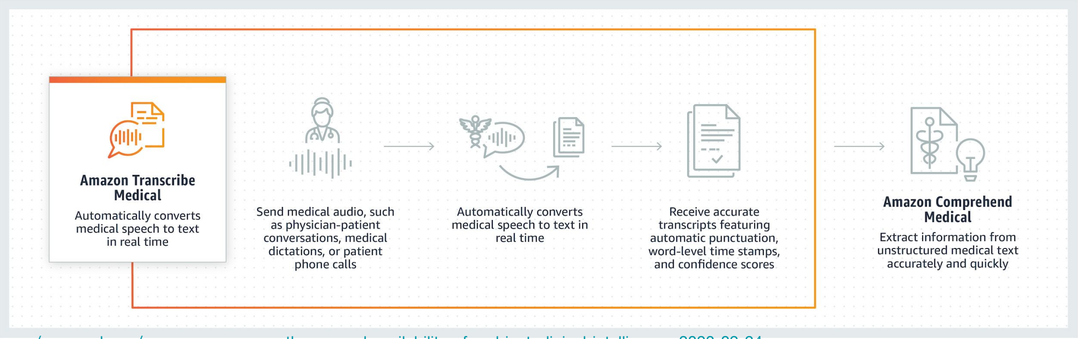
Published: Feb 24, 2020 7:00 a.m. ET

Aa

AI-powered ambient solution already improving physician productivity, patient throughput, and 88% higher physician satisfaction scores

Epic to debut ambient voice technology assistant at HIMSS20

"Just like you use your smart speaker at home, clinicians soon will be able to say 'Hey Epic' to quickly get the information they need and take action," an Epic exec reveals.



[1] https://www.marketwatch.com/press-release/nuance-announces-the-general-availability-of-ambient-clinical-intelligence-2020-02-24
[2] https://www.healthcareitnews.com/news/epic-debut-ambient-voice-technology-assistant-himss20
[3] https://aws.amazon.com/transcribe/medical/

Clinical Audio Presents New Opportunities

- Voice is another modality for assessing patient state:
  - Disorders directly related to voice (e.g. vocal hyperfunction).[1,2]

Learning to detect vocal hyperfunction from ambulatory neck-surface acceleration features: Initial results for vocal fold nodules

Marzyeh Ghassemi, Jarrad H. Van Stan, Daryush D. Mehta, *Member, IEEE*, Matías Zañartu, *Member, IEEE*, Harold A. Cheyne II, Robert E. Hillman, and John V. Guttag

Voice Disorder Identification by Using Machine Learning Techniques

LAURA VERDE[1], GIUSEPPE DE PIETRO[2], (Member, IEEE), AND GIOVANNA SANNINO[2], (Member, IEEE)
[1]Department of Engineering, Centro Direzionale di Napoli, Parthenope University of Naples, 80143 Naples, Italy
[2]Institute of High Performance Computing and Networking, 80131 Naples, Italy
Corresponding author: Giovanna Sannino (giovanna.sannino@icar.cnr.it)

  - Disorders that manifest through voice (e.g. dementia).[3,4]

Journal of Alzheimer's Disease 49 (2016) 407–422
DOI 10.3233/JAD-150520
IOS Press
407

Linguistic Features Identify Alzheimer's Disease in Narrative Speech

Kathleen C. Fraser[a], Jed A. Meltzer[b] and Frank Rudzicz[a,c,*]
[a]*Department of Computer Science, University of Toronto, Toronto, Canada*
[b]*Rotman Research Institute, Toronto, Canada*
[c]*Toronto Rehabilitation Institute-UHN, Toronto, Canada*

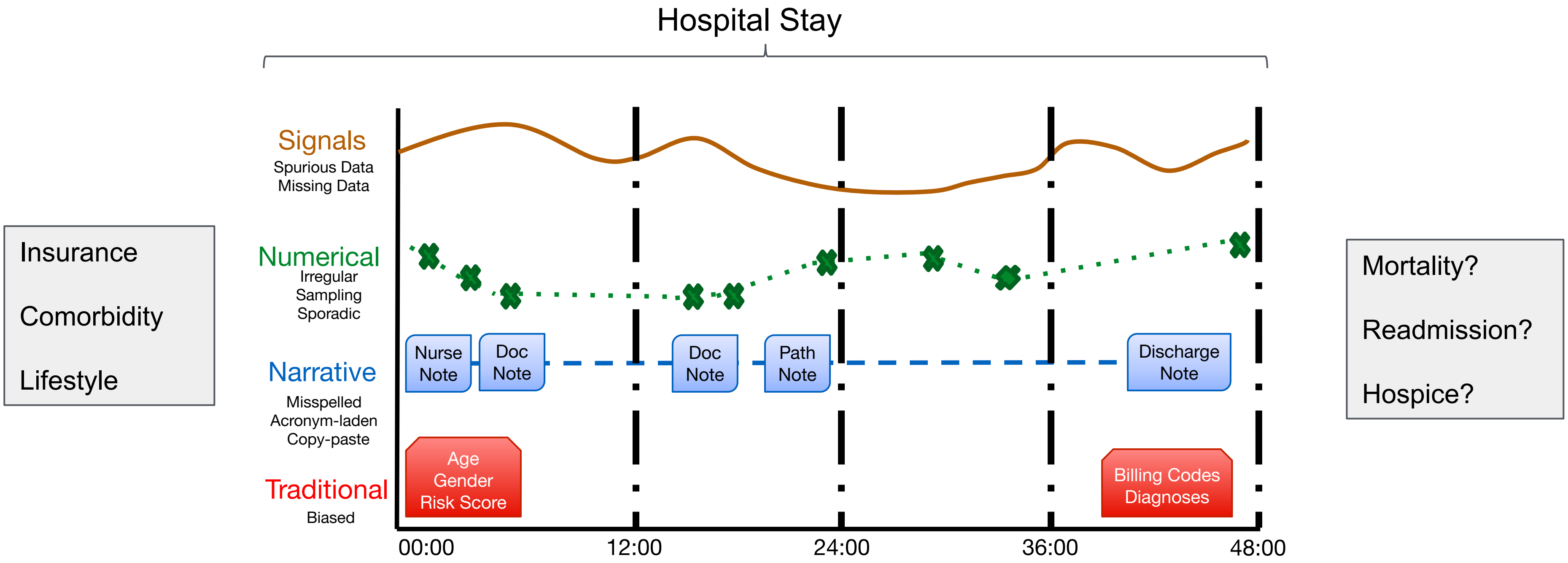
Handling Associate Editor: Peter Garrard

Learning multiview embeddings for assessing dementia

Chloé Pou-Prom [1,2,3], Frank Rudzicz[1,2,3]
[1] Toronto Rehabilitation Institute - UHN, Toronto, Canada
[2]Vector Institute, Toronto, Canada
[3] Department of Computer Science, University of Toronto, Canada
{ chloe, frank }@cs.toronto.edu

[1] Ghassemi et al. Learning to detect vocal hyperfunction from ambulatory neck-surface acceleration features: Initial results for vocal fold nodules.
[2] Verde et al. Voice Disorder Identification by Using Machine Learning Techniques
[3] Fraser et al. Linguistic Features Identify Alzheimer's Disease in Narrative Speech.
[4] Pou-Prom et al. Learning multiview embeddings for assessing dementia.

Language is Part of A Larger Ecosystem



[1] Johnson, Alistair EW, et al. "MIMIC-III, a freely accessible critical care database." Scientific data 3 (2016).

Outline

- **What is clinical language?**
- Common natural language processing (NLP) tasks.
- Resources for working with clinical language.
- Selected applications.

Clinical Text != Biomedical Text

- Biomedical text uses medical language in published literature. E.g.,

- Clinical text is collected by care staff and describes patients. E.g.,

Discharge Summary

Name: [**Last Name**],[**Known firstname**]   Unit No: [**Numeric Identifier**]

Admission Date: [**__DATE__**]   Discharge Date: [**__DATE__**]

Date of Birth: [**__DATE__**]   Sex: F

Service: CARDIOTHORACIC

Allergies:
Ampicillin

Attending:[**First Name**]
Addendum:
Pt did not go to rehab on Percocet, but on Ultram instead.

Discharge Medications:
1. Aspirin 81 mg Tablet, Delayed Release (E.C.) Sig: One (1)
Tablet, Delayed Release (E.C.) PO DAILY (Daily).
Disp:*30 Tablet, Delayed Release (E.C.)(s)* Refills:*2*
2. Docusate Sodium 100 mg Capsule Sig: One (1) Capsule PO BID (2
times a day).
Disp:*60 Capsule(s)* Refills:*2*
3. Amiodarone 200 mg Tablet Sig: Two (2) Tablet PO BID (2 times
a day): 400mg twice a day for 7 days then decrease to 400mg
daily for 7 days then decrease to 200mg daily until follow up
with cardiologist.
Disp:*60 Tablet(s)* Refills:*2*
4. Pantoprazole 40 mg Tablet, Delayed Release (E.C.) Sig: One
(1) Tablet, Delayed Release (E.C.) PO Q24H (every 24 hours).
Disp:*30 Tablet, Delayed Release (E.C.)(s)* Refills:*2*

Physician

```
TITLE:
   Chief Complaint:
   24 Hour Events:
 EKG - At [**__DATE__**]
   [**__DATE__**]: Started Metoprolol 25 TID. Around 8pm Pt tacchy to 130s,
   dyspeneic. O2 requirement increased from 2 to 4L. EKG obtained showed
   afib. Given 50 of metoprolol. Ordered 15 IV Dilt but held it as HR came
   down to 80s.  CXR shows increased pulmonary edema. Pt urinating well so
   held off on diuretics.
   ECHO: The left atrium is dilated. There is mild symmetric left
   ventricular hypertrophy. The left ventricular cavity is moderately
   dilated. Overall left ventricular systolic function is moderately
   depressed (LVEF= 30-40 %) secondary to akinesis of the basal septum and
   Infusions:
   Heparin Sodium - 450 units/hour
   Other ICU medications:
   Other medications:
   Changes to medical and family history:
   Review of systems is unchanged from admission except as noted below
   Vital signs
   Hemodynamic monitoring
   Fluid balance
                                                                  24 hours
                                                               Since 12 AM

   Tmax: 36.7
C (98
   Tcurrent: 36.6
C (97.9
   HR: 81 (63 - 114) bpm
```

Nursing

```
Sinus bradycardia.  Long QTc interval.  Low voltage in the limb leads.
No previous tracing available for comparison.
```

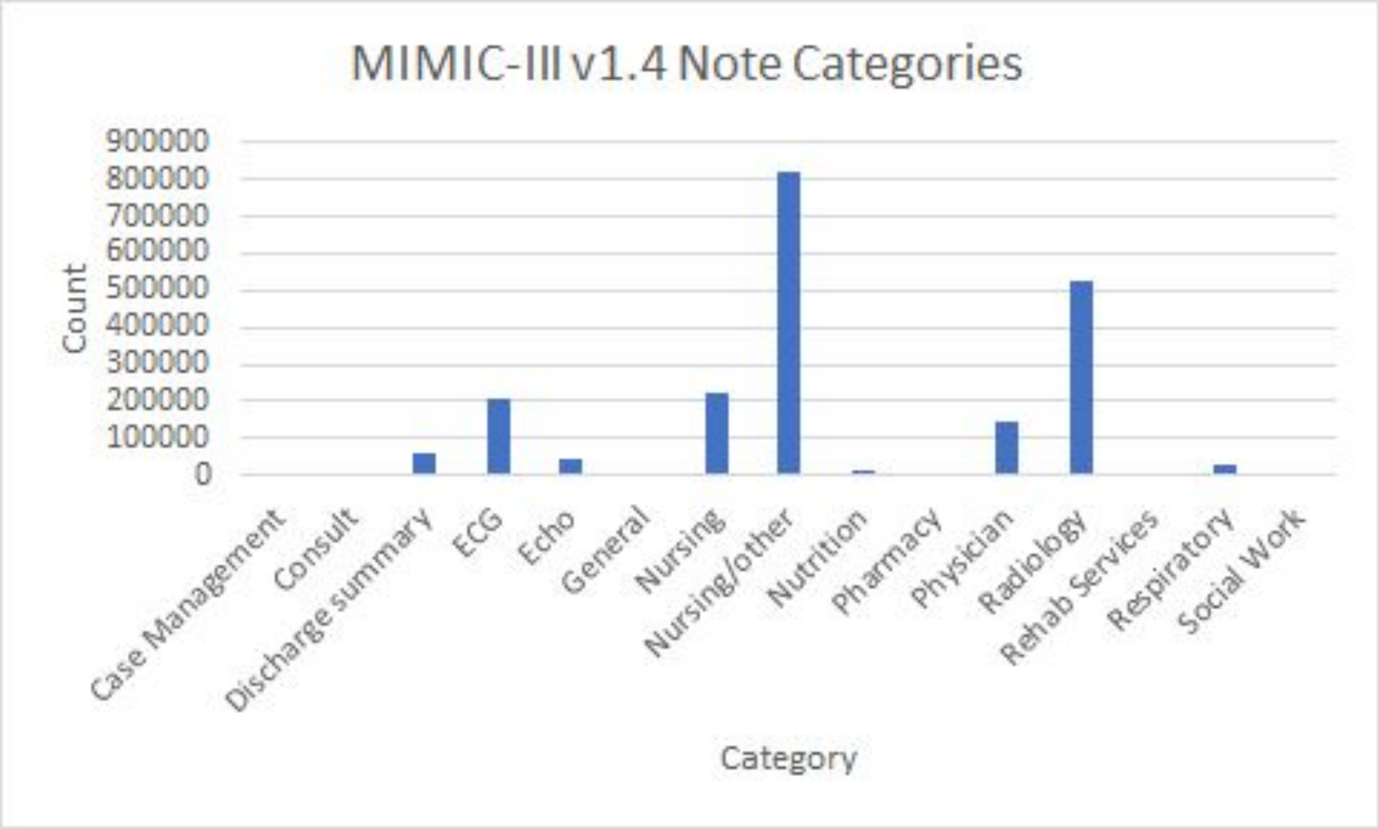
```
Normal sinus rhythm with atrio-ventricular conduction delay. Poor R wave
progression in leads V1-V3 consistent with possible old anteroseptal myocardial
infarction. Compared to the previous tracing of [**__DATE__**] the QRS voltage in the
anterolateral leads is more prominent possibly related to lead placement.
```

```
Sinus rhythm
Nonspecific intraventricular conduction delay
Possible anterior infarct - age undetermined
Lateral T wave changes are nonspecific
Since previous tracing of [**__DATE__**], no significant change
```

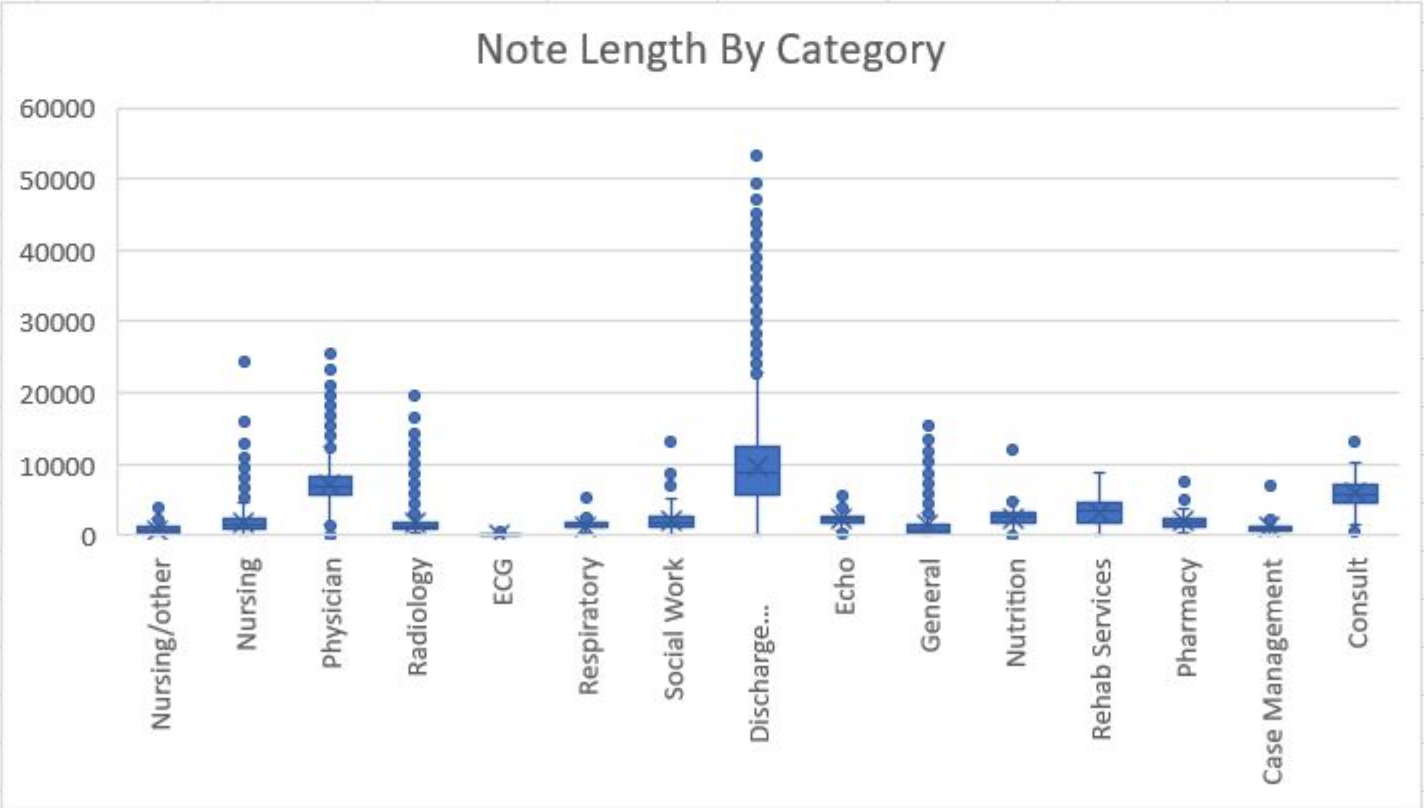
Radiology

```
[**__DATE__**]
 CTA CHEST W&W/O C&RECONS, NON-CORONARY                          Clip # [**Clip Number (Radiology)**]
 Reason: PLease rule out acute PE
 Admitting Diagnosis: RESPIRATORY DISTRESS
  Contrast: OMNIPAQUE Amt: 100
 ______________________________________________________________________________
 [**__HOSPITAL__**] MEDICAL CONDITION:
  55 year old man with respiratory failure with history of PE, CPOD laryngeal
  edema.
 REASON FOR THIS EXAMINATION:
  PLease rule out acute PE
 No contraindications for IV contrast
 ______________________________________________________________________________
 WET READ: EHAb [**First Name**] [**__DATE__**]
  No pulmonary embolus detected (interval resolution since [**__DATE__**]). Moderate
  dependent atelectasis, right greater than left, in the setting of intubation.
  Thymic tissue noted, atypical for this age.
 ______________________________________________________________________________
                                 FINAL REPORT
 HISTORY:  History of pulmonary embolism, COPD, and laryngeal edema, now with
 respiratory failure.  Evaluate for pulmonary embolism.
```

MIMIC-III Note Categories



MIMIC-III Note Lengths



Clinical Text Presents Unique Challenges

- **Data access**: often perceived as high risk due to difficulty of de-identification.
- **Copy-paste**: existing workflows encourage the repetition of existing text.
- **Quality variance**: some text well-written for communication, some not.
- **Partial structure**: sometimes generated or copied from structure (e.g vitals).
- **And the previous challenges with language…**

Additional Considerations for Audio

- **Disfluency:** non-trivial difluencies in spoken language.

  “I think you should um take you know aspirin.”

- **Utterance Segmentation:** imperfect speech turns complicate context.

  “I’d like you to take albuterol for a week, also do you have an upcoming competition? I would like you to avoid vigorous exercise.

- **Diarization**: difficult with multiple speakers.

Outline

- What is clinical language?
- **Common natural language processing (NLP) tasks.**
- Resources for working with clinical language.
- Selected applications.

NLP Tasks

- Part of Speech
- Parsing
- Named Entity Recognition
- Normalization
- Negation
- Uncertainty
- Word Sense Disambiguation
- Relation Classification
- Summarization
  - Extractive
  - Abstractive

NLP Tasks

- Part of Speech
- Parsing
- **Named Entity Recognition**
- **Normalization**
- **Negation**
- **Uncertainty**
- Word Sense Disambiguation
- **Relation Classification**
- Summarization
  - Extractive
  - Abstractive

Named Entity Recognition

- Identify mentions of semantic types within text:
    - Find **spans** that correspond to an entity.
    - Classify the correct semantic type from the span.

- Context is important to resolve ambiguity, e.g.:
    - Ms Jane Doe has a history of MS
    - Mr Hutchinson diagnosed with Hutchinson

Normalization / Entity Linking

- Given an entity mention, assign canonical identifier (e.g., from ontology)

- We want each of these to have the same meaning:
  - Patient diagnosed with *RA* (C0003873)
  - Patient diagnosed with *Rheumatoid Arthritis* (C0003873)
  - Patient diagnosed with *atrophic arthritis* (C0003873)

Example credit: G. Savova, cTAKES keynote presentation.

Negation & Uncertainty

- **Negation:** entity mention is negated.
    - Patient denies *foot joint pain*.
        - foot joint pain, negated
          C0458239, negated

- **Uncertainty:** entity mention not definitive.
    - Results suggestive of *colorectal cancer*.
        - colorectal cancer, probable
          C1527249, probably

Example credit: G. Savova, cTAKES keynote presentation.

Relations

- **Relations:** high level semantic types relating more than one mention, e.g.:
    - *DegreeOf(modifier, disease/disorder)*
    - *LocationOf(anatomical site, sign/symptom)*
    - *Disrupts(anatomical site, disease/disorder)*
    - *Treats(drug, gene, mutation)*
- Helpful in forming higher level conceptual understanding.

Example of a Pipeline

An example of a sentence discovered by the sentence boundary detector:

Fx of obesity but no fx of coronary artery diseases.

Tokenizer output – 11 tokens found:

```
Fx  of  obesity  but  no  fx  of  coronary  artery  diseases  .
```

Normalizer output:

```
Fx  of  obesity  but  no  fx  of  coronary  artery  disease  .
```

Part-of-speech tagger output:

```
Fx  of  obesity  but  no  fx  of  coronary  artery  diseases  .
NN  IN  NN       CC   DT  NN  IN  JJ        NN      NNS       .
```

Shallow parser output:

```
Fx  of  obesity  but  no  fx  of  coronary  artery  diseases  .
NP  PP    NP           NP     PP            NP
```

Named Entity Recognition – 5 Named Entities found:

```
Fx of obesity but no fx of coronary artery diseases .
      obesity  (type=diseases/disorders, UMLS CUI=C0028754, SNOMED-CT codes=308124008 and 5476005)
                           coronary artery diseases (type=diseases/disorders, CUI=C0010054, SNOMED-CT=8957000)
                           coronary artery (type=anatomy, CUI(s) and SNOMED-CT codes assigned)
                                    artery (type=anatomy, CUI(s) and SNOMED-CT codes assigned)
                                           diseases (type=diseases/disorders, CUI = C0010054)
```

Status and Negation attributes assigned to Named Entities:

```
Fx of obesity but no fx of coronary artery diseases .
      obesity (status = family_history_of; negation = not_negated)
                           coronary artery diseases (status = family_history_of, negation = is_negated)
```

Figure 1 Example sentence processed through cTAKES components 'family history of obesity but no family history of coronary artery diseases.' Fx, family history.

Savova et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. JAMIA 2010.

Outline

- What is clinical language?
- Common natural language processing (NLP) tasks.
- **Resources for working with clinical language.**
- Selected applications.

How can we represent language?

Medical Ontologies and Lexicons

- **SNOMED CT:** Codes, terms, synonyms, and definitions of clinical terms.
- **RxNorm:** Nomenclature of clinical drugs produced by NLM.
- **MeSH:** Medical Subject Headings (e.g., medical literature).
- **LOINC:** Logical Observation Identifiers Names and Codes (e.g. labs).
- **CPT:** Current Procedural Terminology maintained by AMA (e.g. billing).
- **ICD**: International Classification of Disease maintained by WHO (e.g. billing).

One System to Rule Them All

- **SNOMED CT:** Codes, terms, synonyms, and definitions of clinical terms.
- **RxNorm:** Nomenclature of clinical drugs produced by NLM
- **MeSH:** Medical Subject Headings (e.g., medical literature)
- **LOINC:** Logical Observation Identifiers Names and Codes (e.g. labs)
- **CPT:** Current Procedural Terminology maintained by AMA (e.g. billing).
- **ICD**: International Classification of Disease (e.g. billing)

  ...
- \+ 200 more sources

Unified Medical Language System (UMLS)

History of the UMLS

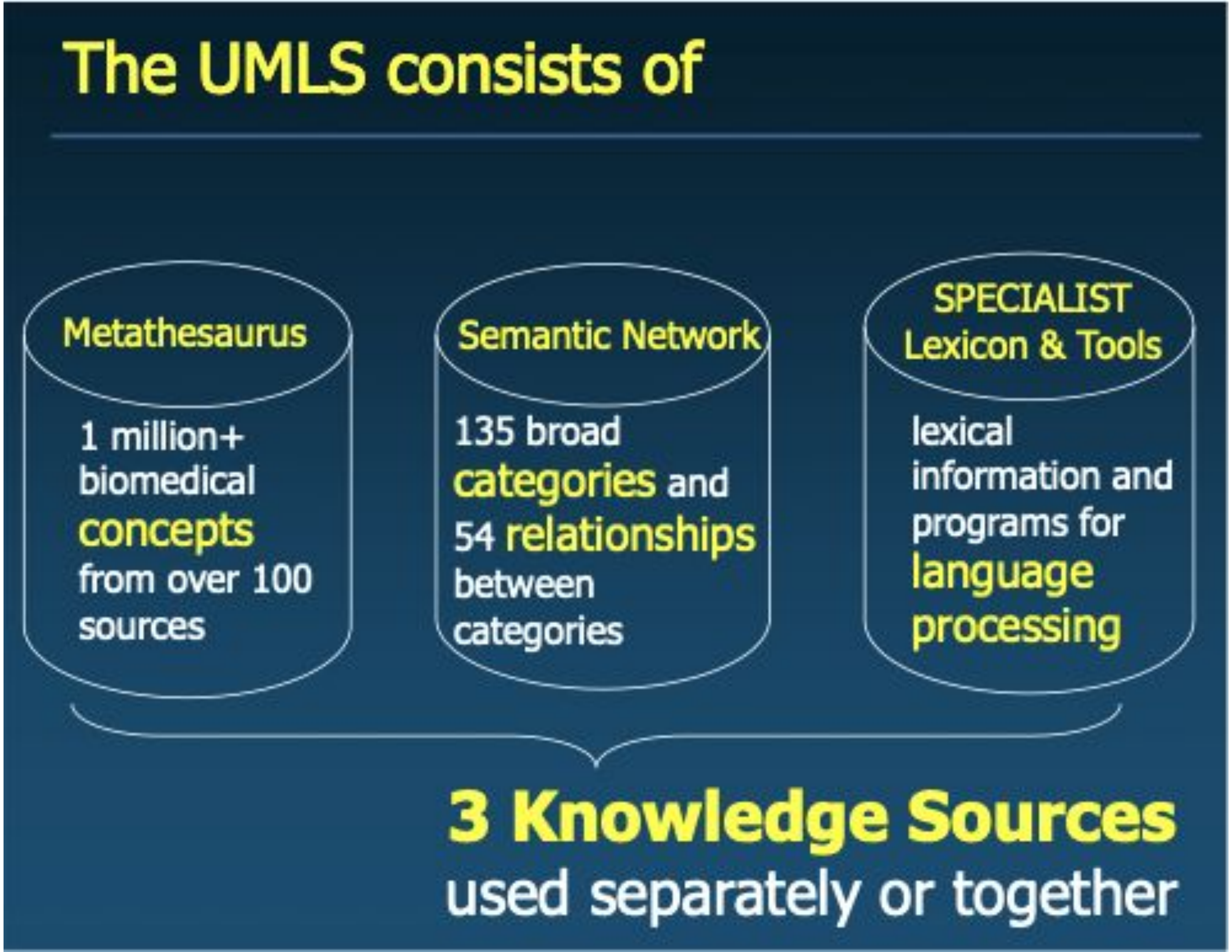
[Lindberg & al., *Methods*, 1993]
[Humphreys & al., *JAMIA*, 1998]

- Started at National Library of Medicine, 1986
- "Long-term R&D project"
- Complementary to IAIMS

(Integrated Academic Information Management Systems)

«[...] the UMLS project is an effort to overcome two significant barriers to effective retrieval of machine-readable information.
- The first is the variety of ways the same concepts are expressed in different machine-readable sources and by different people.
- The second is the distribution of useful information among many disparate databases and systems.»

Credit: Rachel Kleinsorge, Jan Willis, "UMLS Basics class"



Credit: Rachel Kleinsorge, Jan Willis, “UMLS Basics class”

Metathesaurus: clusters terms by meaning

- Synonymous terms clustered into a concept
- Preferred term is chosen
- Unique identifier (CUI) is assigned

Addison's disease	Metathesaurus	PN	
Addison's disease	SNOMED CT	PT	363732003
Addison's Disease	MedlinePlus	PT	T1233
Addison Disease	MeSH	PT	D000224
Bronzed disease	SNOMED Intl 1998	SY	DB-70620
Deficiency; corticorenal, primary	ICPC2-ICD10 Thesaurus	PT	MTHU021575
Primary Adrenal Insufficiency	MeSH	EN	D000224
Primary hypoadreanlism syndrome, Addison	MedDRA	LT	10036696

C0001403 Addison's disease

Credit: Rachel Kleinsorge, Jan Willis, "UMLS Basics class"

Semantic Network

- 135 Semantic Types
  - Broad subject categories (Clinical Drug, Virus)
  - Ex:
    - **Addison's Disease**
    - **Semantic Type: Disease or Syndrome**
- 54 Semantic Relationships
  - Links between categories (isa, causes, treats)
  - Ex:
    - Virus **causes** Disease or Syndrome
- Types + Relationships
  - Form the structure of the semantic network
  - Broadly categorize the biomedical domain

Credit: Rachel Kleinsorge, Jan Willis, "UMLS Basics class"

Concept cluster of synonymous terms

Concept
C0001621

Term	Strings
Term adrenal disease gland L0001621	S0011232 *Adrenal Gland Diseases* S0011231 Adrenal Gland Disease S0000441 Disease of adrenal gland S0481705 Disease of adrenal gland, NOS S0220090 Disease, adrenal gland S0044801 Gland Disease, Adrenal
Term adrenal disorder gland unspecified L0041793	S0860744 *Disorder of adrenal gland, unspecified* S0217833 Unspecified disorder of adrenal glands
Term adrenal disorder L0161347	S0225481 *ADRENAL DISORDER* S0627685 DISORDER ADRENAL (NOS)
Term adrenal disorder gland L0181041	S0632950 *Disorder of adrenal gland* S0354509 Adrenal Gland Disorders
Term L0162317	S0226798 *SURRENALE, MALADIES* FRE

Credit: Rachel Kleinsorge, Jan Willis, "UMLS Basics class"

Is there data available?

~~i2b2: Informatics for Integrating Biology at the Bedside~~
n2c2: National NLP Clinical Challenges

- 2006 Deidentification and Smoking Challenge
- 2008 Obesity Challenge
- 2009 Medication Challenge
- 2010 Relations Challenge
- 2011 Coreference Challenge
- 2012 Temporal Relations Challenge
- 2014 De-identification and Heart Disease Risk Factors Challenge

- Challenge format: Datasets are **annotated**![1]

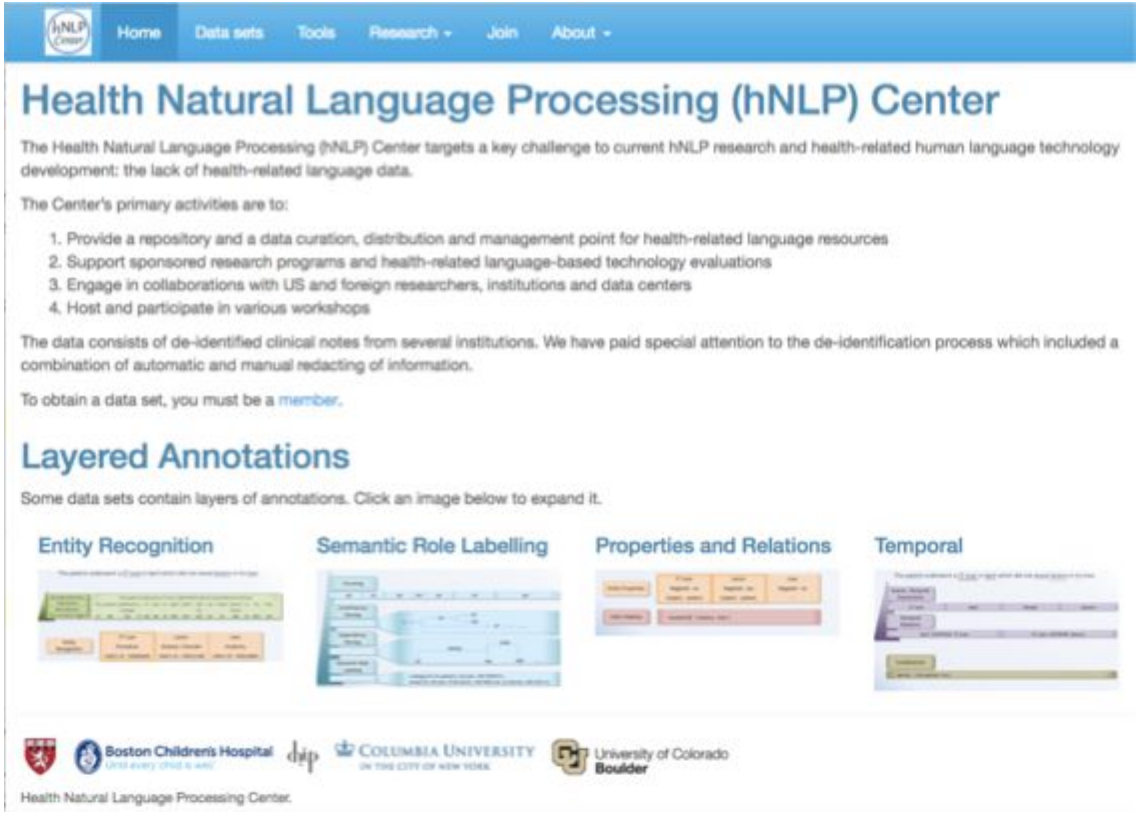
[1] https://www.i2b2.org/

SemEval & ShARe/CLEF

- 2014 SemEval Task 7: Analysis of Clinical Text[1]
  - Entity, acronym, abbreviation recognition, mapping to CUIs
- 2015 SemEval Task 14: Analysis of Clinical Text[2]
  - Entity, acronym, abbreviation recognition, mapping to CUIs
- 2015 SemEval Task 6: Clinical TempEval[3]
  - Timespan, event, and temporal relation
- 2015: CLEF eHealth Evaluation Lab Task 1a: Clinical Speech Recognition[4]
  - Minimize word detection errors for Australian nursing shift changes

[1] http://alt.qcri.org/semeval2014/task7/
[2] http://alt.qcri.org/semeval2015/task14/
[3] http://alt.qcri.org/semeval2015/task6/
[4] https://sites.google.com/site/clefehealth2015/task-1/task-1a

Health NLP (hNLP) Center


Home Data sets Tools Research Join About

Health Natural Language Processing (hNLP) Center

The Health Natural Language Processing (hNLP) Center targets a key challenge to current hNLP research and health-related human language technology development: the lack of health-related language data.

The Center's primary activities are to:

1. Provide a repository and a data curation, distribution and management point for health-related language resources
2. Support sponsored research programs and health-related language-based technology evaluations
3. Engage in collaborations with US and foreign researchers, institutions and data centers
4. Host and participate in various workshops

The data consists of de-identified clinical notes from several institutions. We have paid special attention to the de-identification process which included a combination of automatic and manual redacting of information.

To obtain a data set, you must be a member.

Layered Annotations

Some data sets contain layers of annotations. Click an image below to expand it.

Entity Recognition | Semantic Role Labelling | Properties and Relations | Temporal

Boston Children's Hospital | Columbia University in the City of New York | University of Colorado Boulder

Health Natural Language Processing Center.

MIMIC-III

- Over 2M notes for ~50K patients.

- Notes are de-identified, but otherwise unannotated.

[1] Johnson, Alistair EW, et al. "MIMIC-III, a freely accessible critical care database." Scientific data 3 (2016).

Are tools available?

Clinical NLP Tools

- cTAKES[1] (clinical Text Analysis and Knowledge Extraction System)
  - Commonly regarded as the standard.
  - Flexible, but may require significant configuration.

- MetaMap[2]
  - Designed to identify UMLS concepts in text using knowledge-intensive approach.

- MetaMap Lite[3]
  - Less rigorous than MetaMap, but <u>much</u> faster.

- Sophia[4] (v3NLP Framework)
  - Transform text into structured data for quality improvement, research, population health surveillance, and decision support. Scalable out of the box

[1] http://ctakes.apache.org/
[2] https://metamap.nlm.nih.gov/
[3] https://metamap.nlm.nih.gov/MetaMapLite.shtml
[4] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5019303/

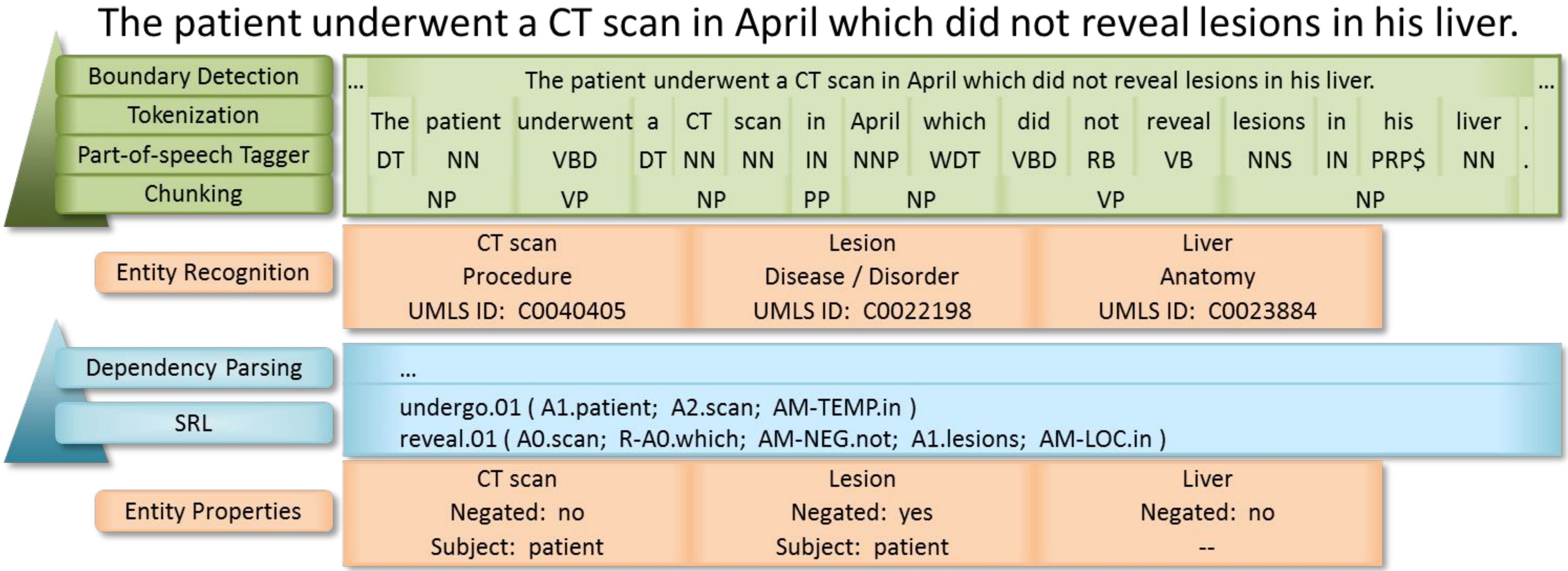
Clinical NLP Tools

- CLAMP[1] (Clinical Language Annotation, Modeling and Processing Toolkit)
- MedEx[2]
- MedLEE (Medical Language Extraction and Encoding System)
- CliNER[4] (Clinical Named Entity Recognition System)
- ...

[1] https://clamp.uth.edu/
[2] http://www.vumc.org/cpm/cpm-blog/medex-tool-finding-medication-information
[3] http://www.medlingmap.org/taxonomy/term/80
[4] https://github.com/text-machine-lab/CliNER

cTAKES

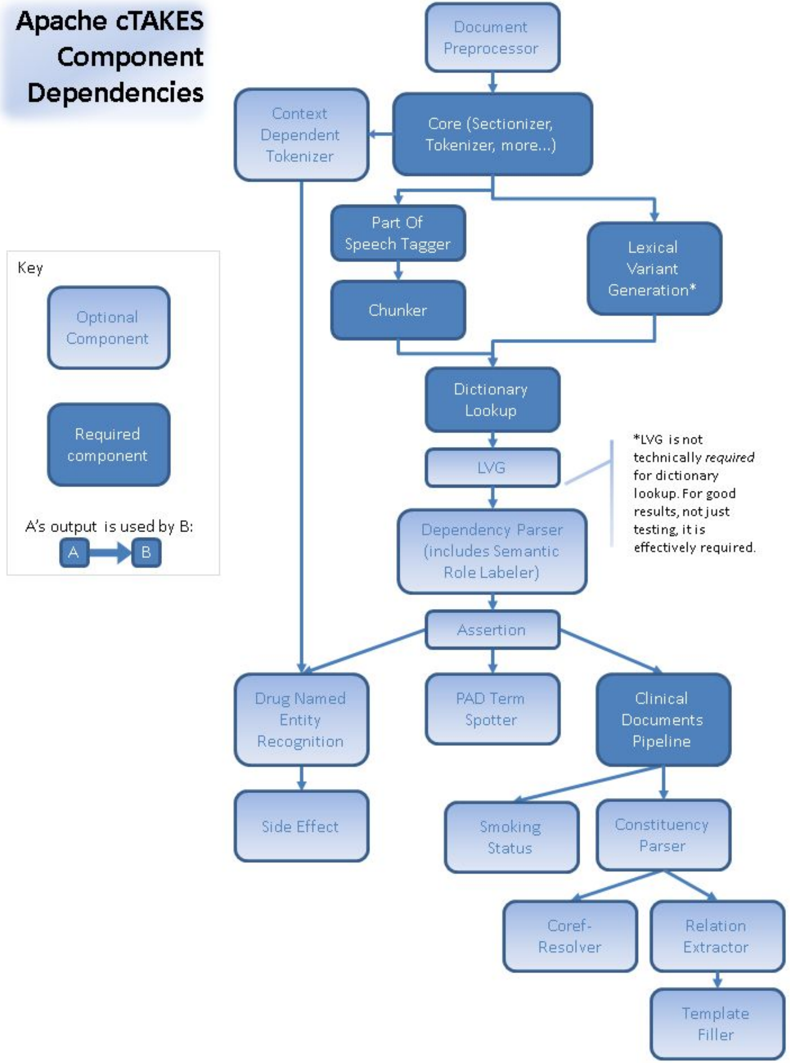
- Default pipeline is a good starting point for many projects.



[1] https://cwiki.apache.org/confluence/display/CTAKES/Default+Clinical+Pipeline

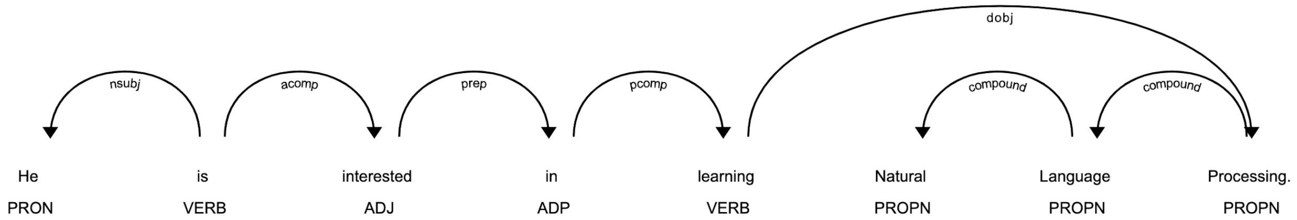
cTAKES Components

- Sentence boundary
- Tokenization (rule-based)
- Morphologic normalization (NLM'S LVG)
- POS tagging
- Shallow parsing
- Named Entity Recognition
  - Dictionary mapping (lookup algorithm)
  - Machine learning
  - Types: Diseases/Disorders, Signs/Symptoms, Anatomical Sites, Procedures, Medications
- Negation and context identification (NegEx)
- Relation Extraction
- Clinical Element Model (CEM) normalization



NLP Tools: spaCy

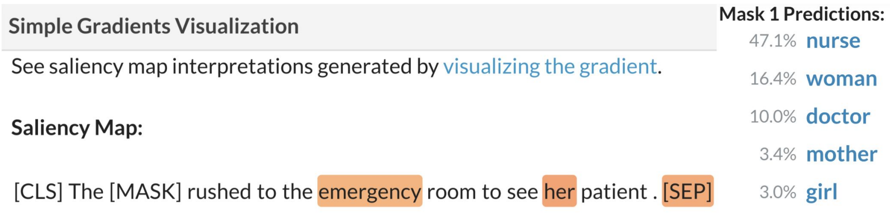
- Fast!
- Tools for visualization via *displacy*



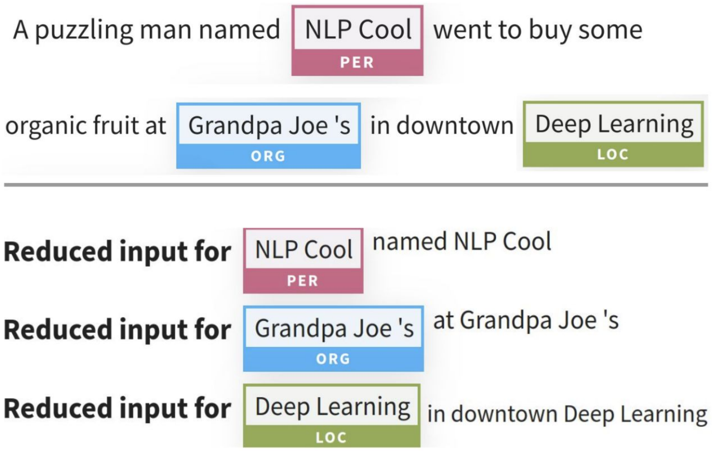
[1] https://realpython.com/natural-language-processing-spacy-python/#named-entity-recognition

NLP Tools: AllenNLP

- Encodes *many* best practices and facilitates experimentation
- AllenNLP Interpret: gradient-based saliency maps and adversarial attacks



E.g. explaining why BERT made certain masked predictions



E.g. visualizing named entities with input reduction

[1] https://allennlp.org/interpret

Outline

- What is clinical language?
- Common natural language processing (NLP) tasks.
- Tools for working with clinical language.
- **Selected applications.**

Automated Trigger for Sepsis

Automated Trigger for Sepsis

- Horng et al.[1] use vitals and triage nursing notes to predict sepsis.

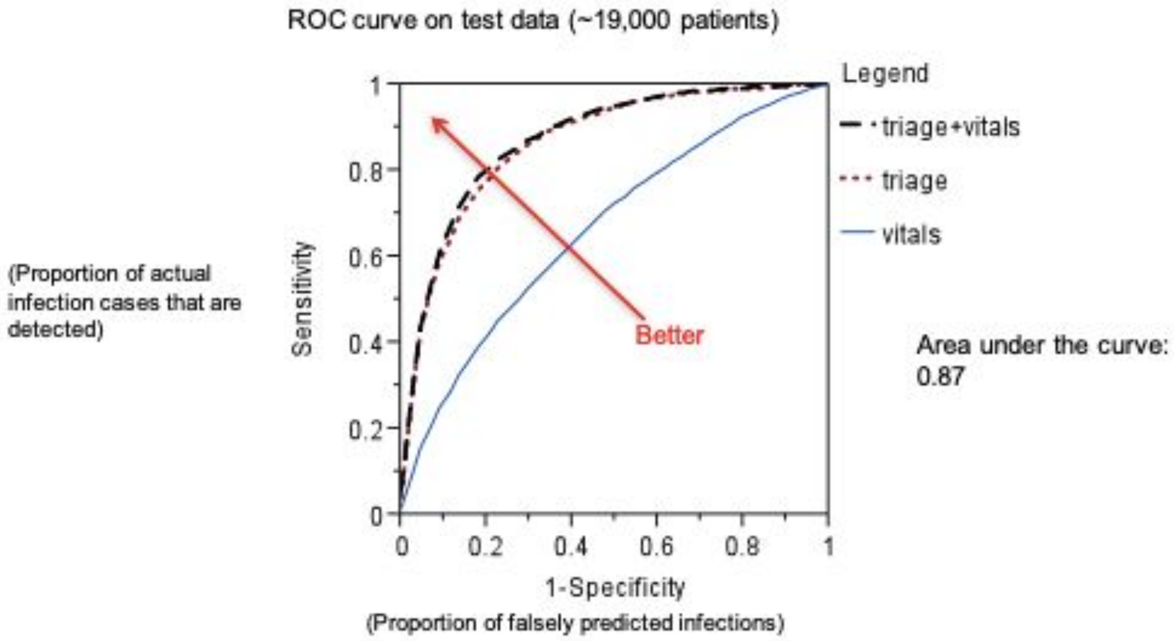
Which of these is likely to develop sepsis?

- pt with fever, chills, N/V since friday after eating what hethought was undercooked meat. Unable to hold po's down. Fevers to 103
- 89 yo f s/p esophageal hernia repair 3/09 w/ ?g-tube placement now w/ c/o's n&v. family reports pt's appetite is decreased, no BM x3d. generally not feeling well, had a bad day.
- from the scene fall of horse landed on r thigh deformity iv fluid 100 fentanyl/ morhpine 4. no head or neck pain/

[1] Horng et al. "Creating an automated trigger for sepsis, clinical decision support at emergency department triage using machine learning." PLOS ONE, 2017

Slide Credit: David Sontag

Text is much more predictive than structured data



[1] Horng et al. “Creating an automated trigger for sepsis, clinical decision support at emergency department triage using machine learning.” PLOS ONE, 2017

Slide Credit: David Sontag

Learning a representation for reasoning about a patient

- Observation: The goal of the triage note is to summarize a patient's state to provide maximal *context* in which to understand future data.

- Question: Can we learn the latent space directly from the triage text?

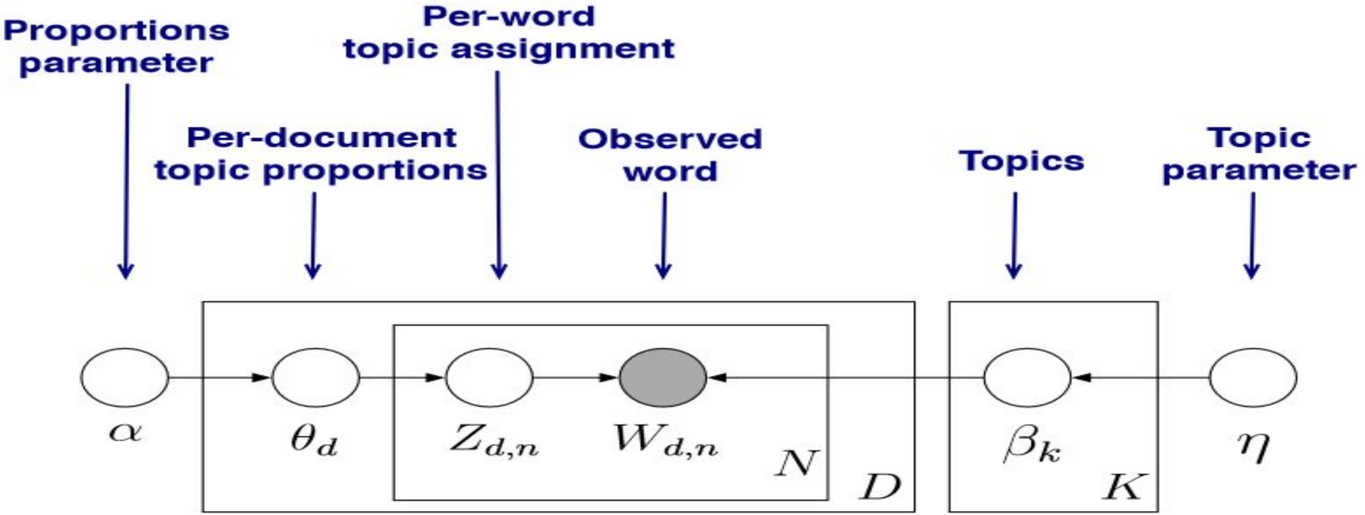
- Solution: Use a topic model called latent Dirichlet allocation (LDA) to identify underlying latent space.

[1] Horng et al. "Creating an automated trigger for sepsis, clinical decision support at emergency department triage using machine learning." PLOS ONE, 2017

Slide Credit: David Sontag

Latent Dirichlet Allocation (LDA)

- Generative model where each document is a mixture of a small number of topics (inferred), and each word (observed) is attributable to one of those topics.



$$\prod_{i=1}^{K} p(\beta_i \mid \eta) \prod_{d=1}^{D} p(\theta_d \mid \alpha) \left(\prod_{n=1}^{N} p(z_{d,n} \mid \theta_d) p(w_{d,n} \mid \beta_{1:K}, z_{d,n}) \right)$$

Underlying Topics Make Sense

- Their topic distributions appeal to clinical intution.

Topic distributions	Less likely
facial numbness droop weakness sided speech slurred face…	↑
rabies bat vaccine exposure shot here for in room prophylaxis…	
shoulder pain rom arm decreased limited pulse injury …	
etoh found admits unable ambulate trauma fs no on drinking…	
gait unsteady steady dizziness feet ha stable alert well oriented…	Infection
vaginal discharge bleeding vag d/c gyn itching pelvic foul…	
throat sore swallowing voice fevers ear difficulty st swallow…	
cellulitis swelling redness with lle rle leg and fevers l lower…	↓
pna cough on pneumonia with cxr dx recent levaquin r/o…	More likely

[1] Horng et al. “Creating an automated trigger for sepsis, clinical decision support at emergency department triage using machine learning.” PLOS ONE, 2017

Slide Credit: David Sontag

Summary: Notes Contain Important Observations of Uninstrumented Information

- Topic representations capture important information from clinical notes.

- Topic representations add predictive value over existing structured vitals.

ClinicalBERT

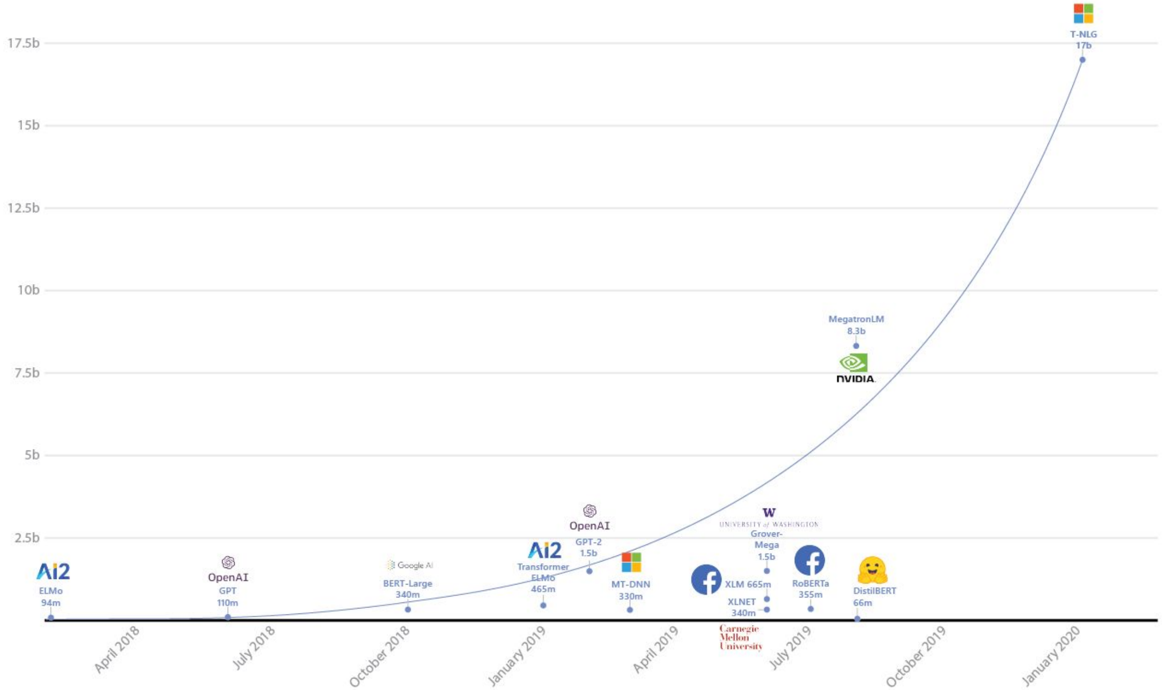
“NLP’s ImageNet Moment”

-- Sebastian Ruder

Leveraging pre-trained language models can dramatically improve performance on downstream tasks

How Many Parameters?

- ELMo
- GPT
- BERT
- XLNet
- GPT-2
- ERNIE
- RoBERTa
- DistillBERT
- ALBERT
- ERNIE 2.0
- Turing-NLG
- T5



[1] https://www.microsoft.com/en-us/research/blog/turing-nlg-a-17-billion-parameter-language-model-by-microsoft/

The Need for a Clinical BERT

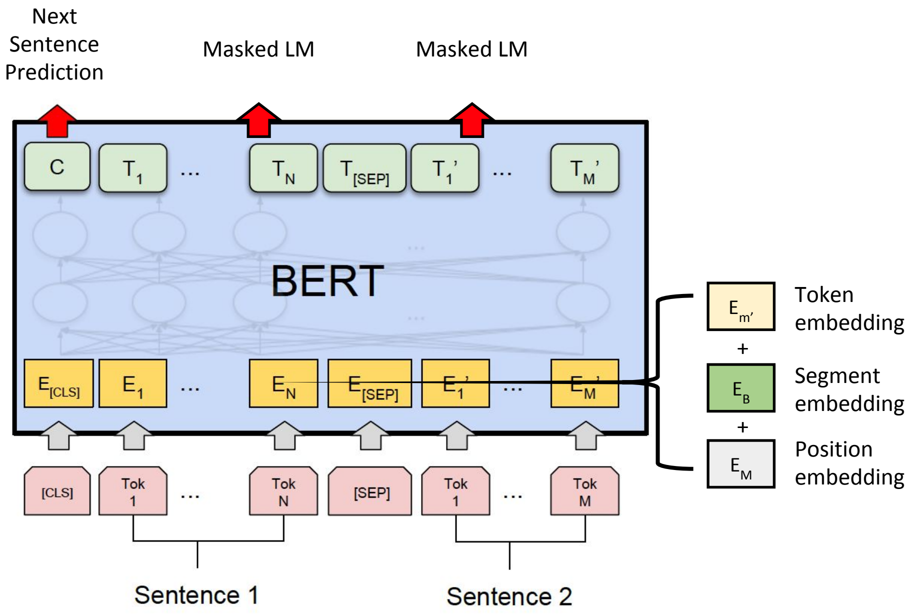
- Limited availability of clinical data necessitates transfer learning approaches

- BERT models in the biomedical domain (BioBERT, sciBERT) don’t directly apply to the clinical domain
  - Clinical text is laden with medical abbreviations and incomplete sentences
  - Text can be formatted into lists, tables, and other non-standard formatting.
  - There is great heterogeneity in style amongst different EHR note types.

BERT Basics

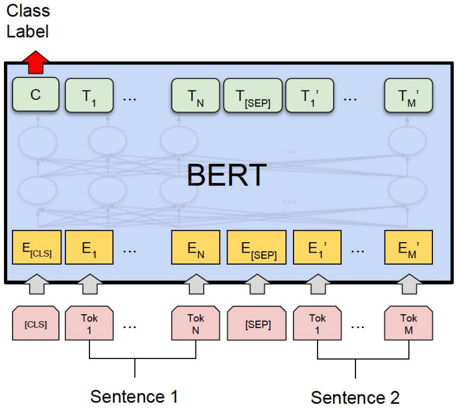
$BERT_{BASE}$
12 layer Transformer with 110M parameters

$BERT_{LARGE}$
24 layer Transformer with 340M parameters

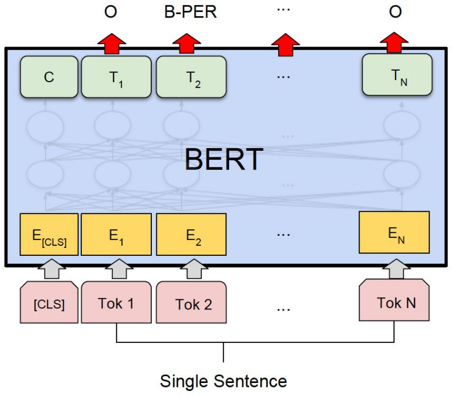


Unsupervised Pre-training
(3.3B words)

Sentence Pair Classification

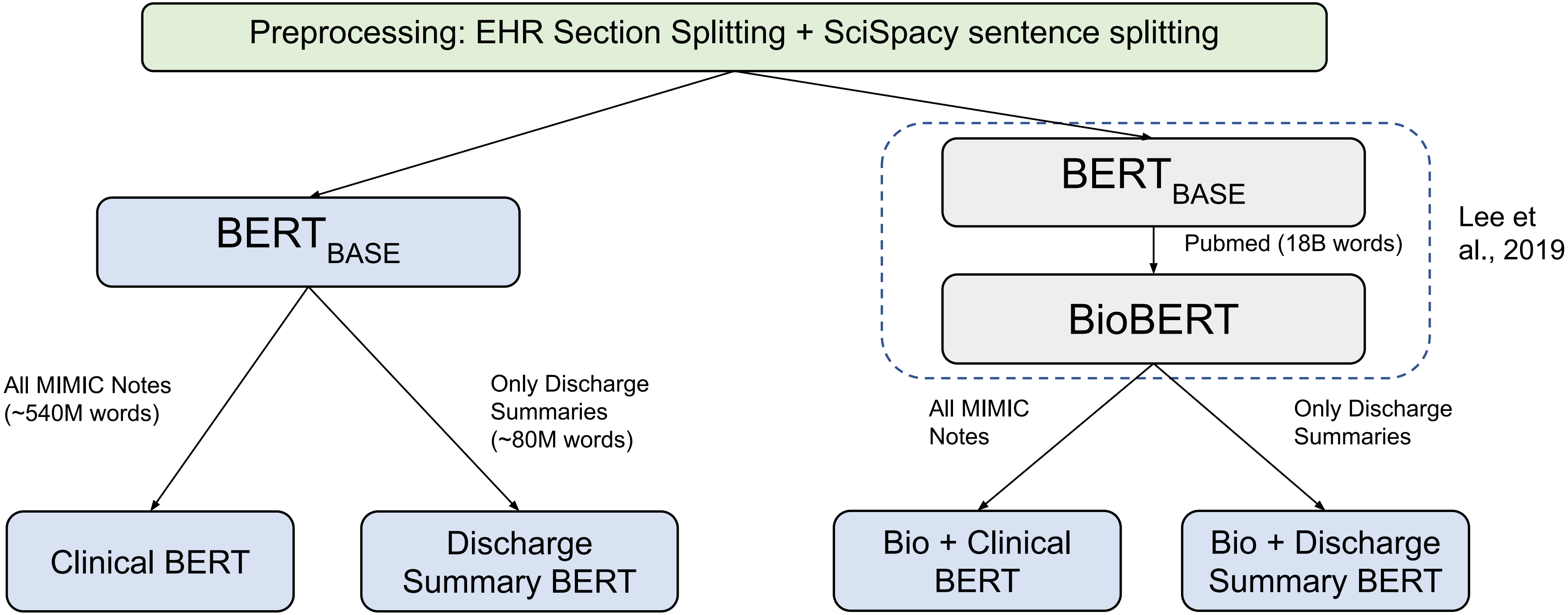


NER



Fine-tuning

Methods: LM Fine-tuning



Methods: Downstream Tasks

MedNLI: Predict whether a hypothesis entails, contradicts, or is neutral to the premise

Premise

ALT , AST , and lactate were elevated as noted above

Hypothesis

patient has abnormal lfts

→ entailment

NER: Identify entities in discharge summaries using IOB encoding

I2b2 2006 & 2014 De-Identification: NER of protected health information (8 & 23 labels)

I2b2 2010 & 2012 Entity Extraction: NER of problems, treatments, tests, etc. (3 & 6 labels)

There	has	been	interval	improvement	in	left	basilar	atelectasis	.
O	O	O	O	O	O	B-Problem	I-Problem	I-Problem	O

Clinical BERT outperforms BioBERT & BERT on NLI and entity extraction tasks, but not de-ID tasks

Model	MedNLI	i2b2 2006	i2b2 2010	i2b2 2012	i2b2 2014
BERT	77.6%	93.9	83.5	75.9	92.8
BioBERT	80.8%	**94.8**	86.5	78.9	**93.0**
Clinical BERT	80.8%	91.5	86.4	78.5	92.6
Discharge Summary BERT	80.6%	91.9	86.4	78.4	92.8
Bio+Clinical BERT	**82.7%**	94.7	87.2	**78.9**	92.5
Bio+Discharge Summary BERT	**82.7%**	94.8	**87.8**	78.9	92.7

Clinical BioBERT outperforms BioBERT & BERT

In the i2b2 2010 task, training on discharge summaries alone outperforms training on all notes.

Clinical BERT outperforms BioBERT & BERT on NLI and entity extraction tasks, but not de-ID tasks

Model	MedNLI	i2b2 2006	i2b2 2010	i2b2 2012	i2b2 2014
BERT	77.6%	93.9	83.5	75.9	92.8
BioBERT	80.8%	**94.8**	86.5	78.9	**93.0**
Clinical BERT	80.8%	91.5	86.4	78.5	92.6
Discharge Summary BERT	80.6%	91.9	86.4	78.4	92.8
Bio+Clinical BERT	**82.7%**	94.7	87.2	**78.9**	92.5
Bio+Discharge Summary BERT	**82.7%**	94.8	**87.8**	78.9	92.7

Clinical BERT fails to outperform BioBERT & BERT on de-identification tasks.

Nearest Neighbors for Clinical and Generic Words in the BERT Vocab

Model	Disease			Operations			Generic		
	Glucose	Seizure	Pneumonia	Transfer	Admitted	Discharge	Beach	Newspaper	Table
BioBERT	insulin	episode	vaccine	drainage	admission	admission	coast	news	tables
	exhaustion	appetite	infection	division	sinking	wave	rock	official	row
	dioxide	attack	plague	transplant	hospital	sight	reef	industry	dinner
Clinical	potassium	headache	consolidation	transferred	admission	disposition	shore	publication	scenario
	sodium	stroke	tuberculosis	admitted	transferred	transfer	ocean	organization	compilation
	sugar	agitation	infection	arrival	admit	transferred	land	publicity	technology

Nearest neighbor words in the Disease and Operations categories are more clinically relevant under Clinical BERT than BioBERT.

However, there’s no large difference for generic words, as expected.

Note: These words were not cherry picked!

Summary: Publicly-Available Resources

We train and publicly release 4 clinical BERT models:

- Fine-tuned using $BERT_{BASE}$ and BioBERT
- Trained on all clinical notes and only discharge summaries from MIMIC III.

Do BERT models fine tuned on clinical text outperform general domain and biomedical domain BERT models?

Yes, except for de-identification tasks

Can note-type specific BERT models outperform models trained on all notes?

Models trained on discharge summaries alone are competitive & have highest performance on one task.

Establishing the Availability of Information

Establishing the Availability of Information

- Complex models perform well in many domains, but results may be shallow.

- Are we leveraging the information that we should in unstructured text to predict outcomes?

Robust Physical-World Attacks on Deep Learning Models. 2017. Eykholt, et al.

Predicting Clinical Outcomes

- Target the prediction of common clinical outcomes in the ICU.

Task	Classes				
In-Hospital Mortality	Survived: 20,062	Expired: 924			
Diagnosis	Sepsis: 350	IC Hemorrhage: 295	Pneumonia: 483	CAD: 523	GI Bleed: 300
Admission Type	Urgent: 17,390	Elective: 3,596			
Length of Stay	Short (< 1.5 days): 6,722	Medium (1-5-3.5 days): 8,126	Long (> 3.5 days): 6,138		

Predicting Intermediate Information

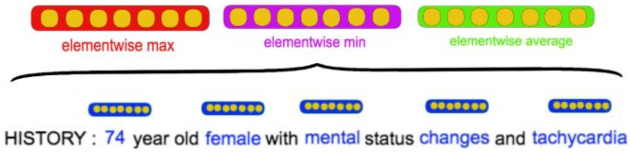
- Establish the availability of important intermediate information to different classes of models.

Task	**Classes**		
Age	< 50 years: 4,565	50 - 80 years: 12,272	80+ years: 4,149
Gender	Male: 11,982	Female: 9,004	
Ethnicity	White: 14,974	Non-white: 3,282	

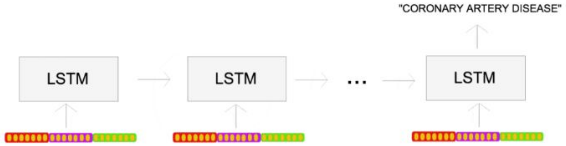
Evaluating Model Representations

- Form representations using:
  - Bag-of-words
  - Aggregate Word Embeddings
  - LSTM

Bag-of-words



Aggregate Word Embeddings

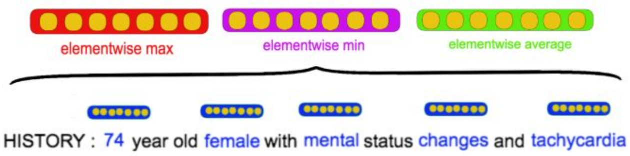


LSTM

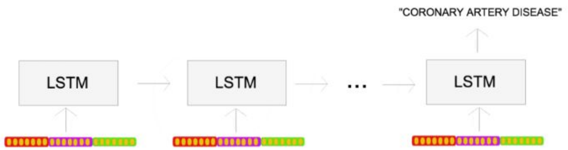
Evaluating Model Representations

- Form representations using:
  - Bag-of-words
  - Aggregate Word Embeddings
  - LSTM

Bag-of-words



Aggregate Word Embeddings

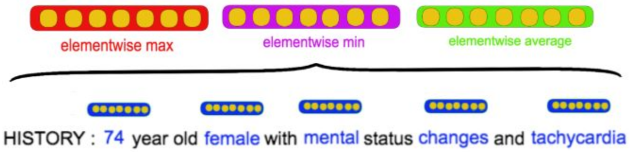


LSTM

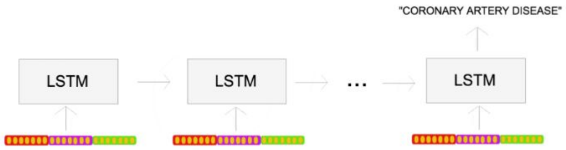
Evaluating Model Representations

- Form representations using:
  - Bag-of-words
  - Aggregate Word Embeddings
  - LSTM

Bag-of-words



Aggregate Word Embeddings

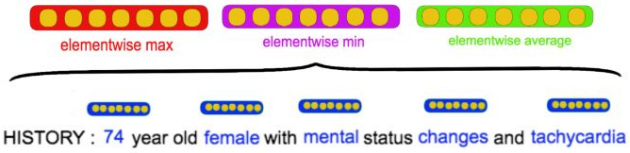


LSTM

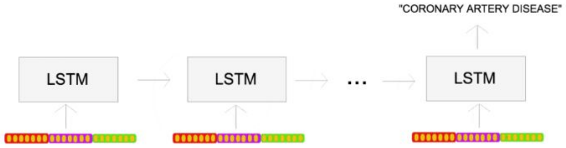
Evaluating Model Representations

- Form representations using:
  - Bag-of-words
  - Aggregate Word Embeddings
  - LSTM

Bag-of-words

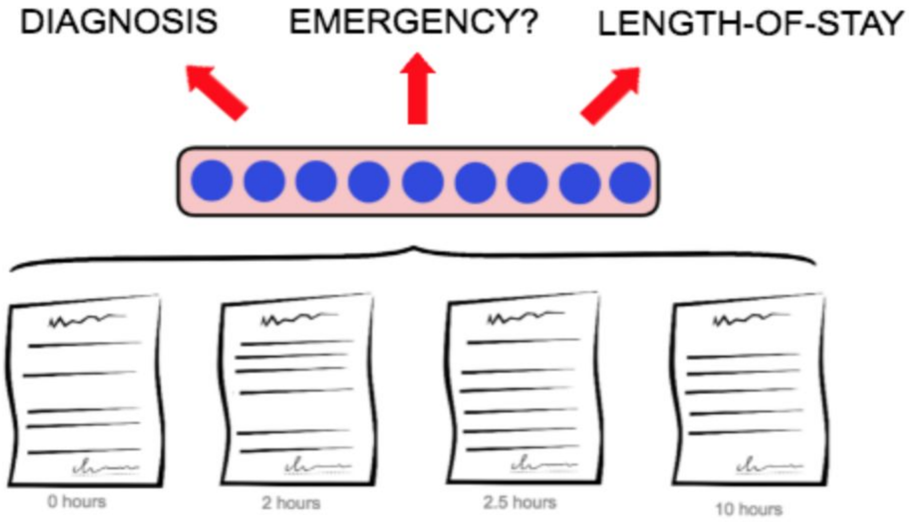


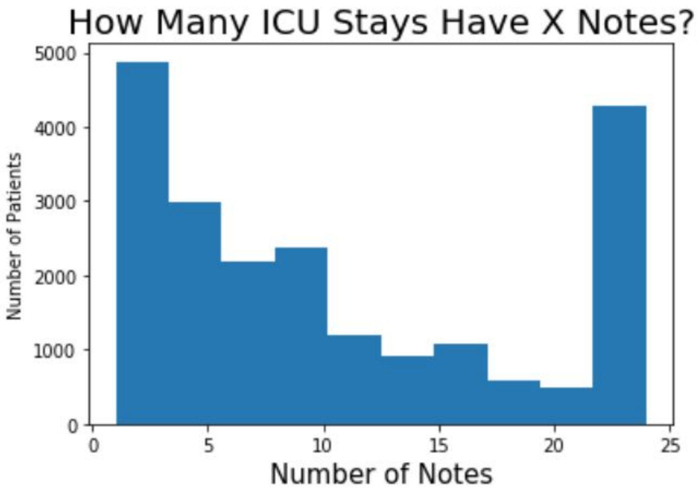
Aggregate Word Embeddings



LSTM

Experimental Setup





patient is a 64 y/o man ➡ patient is a AGE_BETWEEN_SIXTY_AND_SEVENTY man

Representation Performance Varies By Task

- Simple models do surprisingly well on many tasks.

Binary AUCs

Task	**In-Hospital Mortality**	**Admission Type**	**Gender**	**Ethnicity**
BoW	**.821**	**.883**	**.914**	**.619**
Embeddings	.814	.873	.836	.580
LSTM	.777	.870	.837	.533

Representation Performance Varies By Task

- Simple models do surprisingly well on many tasks.

Binary AUCs

Task	**In-Hospital Mortality**	**Admission Type**	**Gender**	**Ethnicity**
BoW	**.821**	**.883**	**.914**	**.619**
Embeddings	.814	.873	.836	.580
LSTM	.777	.870	.837	.533

Representation Performance Varies By Task

- Simple models do surprisingly well on many tasks.

Binary AUCs

Task	**In-Hospital Mortality**	**Admission Type**	**Gender**	**Ethnicity**
BoW	**.821**	**.883**	**.914**	**.619**
Embeddings	.814	.873	.836	.580
LSTM	.777	.870	.837	.533

Representation Performance Varies By Task

- Simple models do surprisingly well on many tasks.

Binary AUCs

Task	In-Hospital Mortality	Admission Type	Gender	Ethnicity
BoW	**.821**	**.883**	**.914**	**.619**
Embeddings	.814	.873	.836	.580
LSTM	.777	.870	.837	.533

Multiclass Macro F1s

Task	Diagnosis	Length of Stay	Age
BoW	.828	.724	**.635**
Embeddings	.828	.730	.544
LSTM	**.836**	**.758**	.450

Representation Performance Varies By Task

- Simple models do surprisingly well on many tasks.

Binary AUCs

Task	In-Hospital Mortality	Admission Type	Gender	Ethnicity
BoW	**.821**	**.883**	**.914**	**.619**
Embeddings	.814	.873	.836	.580
LSTM	.777	.870	.837	.533

Multiclass Macro F1s

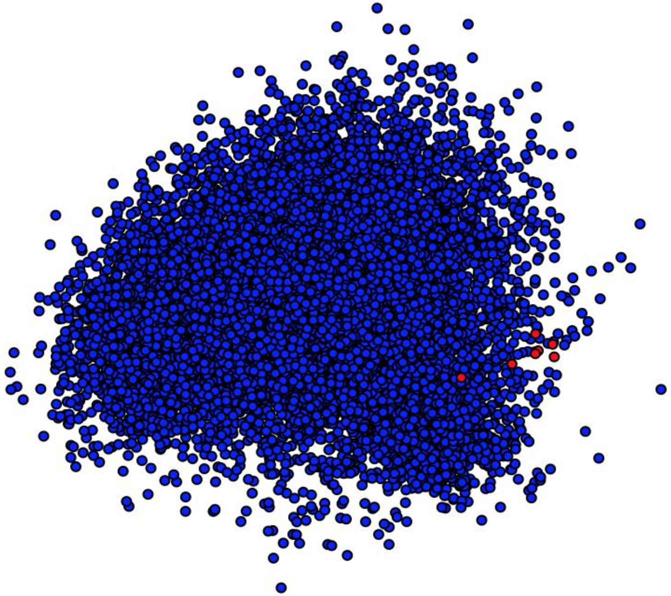
Task	Diagnosis	Length of Stay	Age
BoW	.828	.724	**.635**
Embeddings	.828	.730	.544
LSTM	**.836**	**.758**	.450

Word Vector Embeddings for Age Cluster Together, Confusing Prediction

- The age embedding is averaged across otheı signals from other embeddings.

- Obscures information that clinicians expect our models to know for relevance.

Task	Age
BoW	**.635**
Embeddings	.544
LSTM	.450



Summary: A Single Representation is Not Sufficient

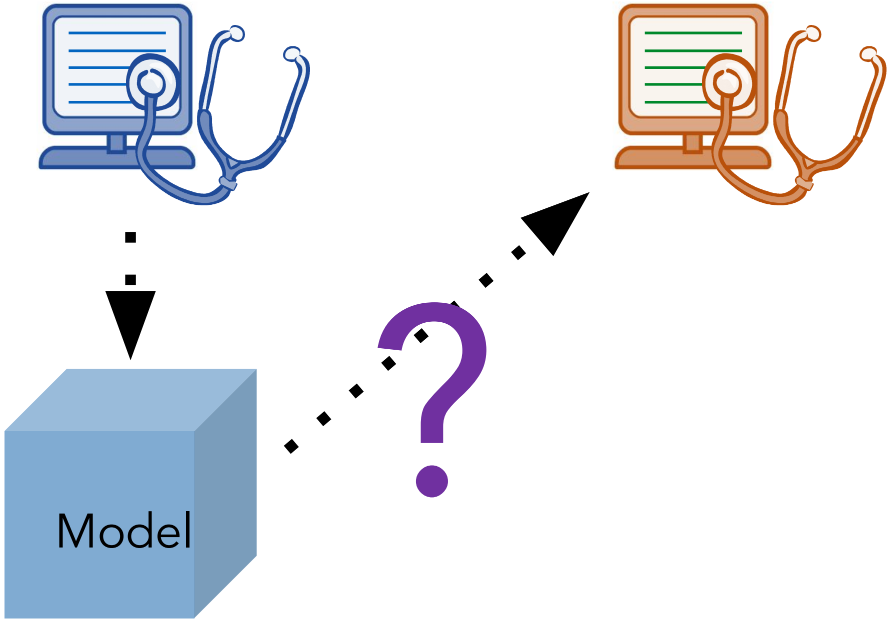
- Simple representations have surprisingly good predictive power.

- Complex representations do not always capture basic information.

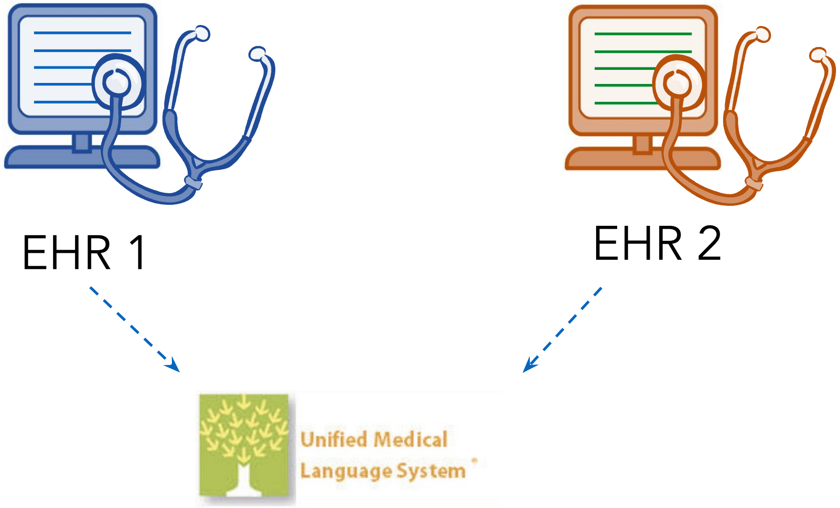
- Words that appear in similar contexts may still have different meanings.

Transfer Predictive Models Across EHRs

Transfer Predictive Models Across EHRs



Solution: Map Semantically Similar Items to Shared Vocabulary



Identify semantically equivalent concepts

Clinical Concepts Underlie Human-Readable Metadata

CUI: `C1328319`

ankle brachial index left

CUIs: `C0003086` `C0445456` `C0918012` `C0205091`

cTAKES[1]
(Clinical Text Analysis
Knowledge Extraction System)

Unified Medical
Language System

All: `{C1328319, C0003086, C0445456, C0918012, C0205091}`

Spanning: `{C1328319, C0205091}`

Longest: `{C1328319}`

[1] Savova, G. K. et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation, and applications. JAMIA, 2010.

Aggregate Representation

CUI: C1328319

ankle brachial index left

CUIs: C0003086 C0445456 C0918012 C0205091

(1 …)

Aggregate Representation

CUI: C1328319

ankle brachial index left

CUIs: C0003086 C0445456 C0918012 C0205091

(1 …)

	C0003086 ankle	C0445456 brachial	C0918012 index	C0205091 left	C1328319 ABI		
All: (	1	1	1	1	1	…	)

Aggregate Representation

CUI: C1328319

ankle brachial index left

CUIs: C0003086 C0445456 C0918012 C0205091

(1 …)

	C0003086 ankle	C0445456 brachial	C0918012 index	C0205091 left	C1328319 ABI		
All:	(1	1	1	1	1	…	)
Spanning:	(0	0	0	1	1	…	)

Aggregate Representation

CUI: C1328319

ankle brachial index left

CUIs: C0003086 C0445456 C0918012 C0205091

(1 …)

		C0003086 ankle	C0445456 brachial	C0918012 index	C0205091 left	C1328319 ABI		
All:	(	1	1	1	1	1	…	)
Spanning:	(	0	0	0	1	1	…	)
Longest:	(	0	0	0	0	1	…	)

Aggregate Representation

CUI: C1328319 (ankle brachial index) C1328319 (ABI)

ankle brachial index left ABI right

CUIs: C0003086 C0445456 C0918012 C0205091 C0205090

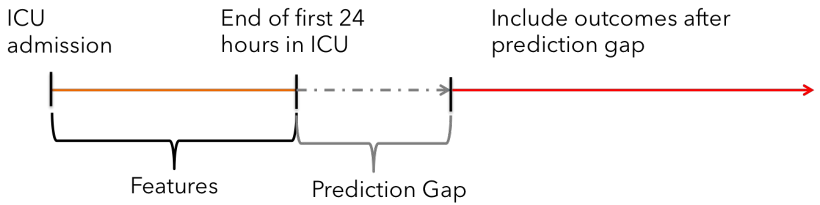
(1 **2** ...)

		C0003086 ankle	C0445456 brachial	C0918012 index	C0205091 left	C1328319 ABI	C0205090 right		
All:	(	1	1	1	1	**3**	**2**	...	)
Spanning:	(	0	0	0	1	**3**	**2**	...	)
Longest:	(	0	0	0	0	**3**	0	...	)

Data & Experimental Setup

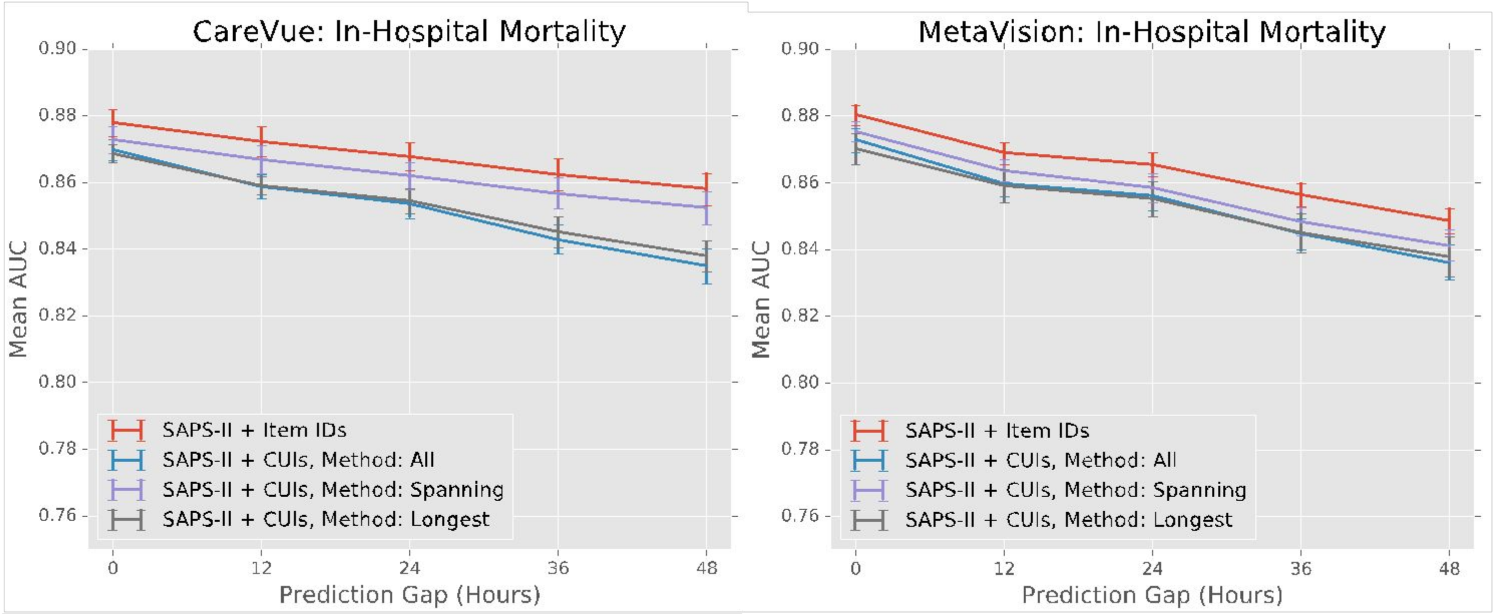
- Data and Representation
  - Data from **2 MIMIC EHR systems** (CareVue and MetaVision).
  - *Item IDs* encode charted observations as **bag-of-events**.
  - Shared *Item IDs* from hospital EHR, and ICU-specific *ItemIDs*.
  - *Item IDs* mapped to UMLS CUIs.

- Model
  - L2-regularized Logistic Regression, 5-fold cross-validation on training sets to determine best hyperparameters



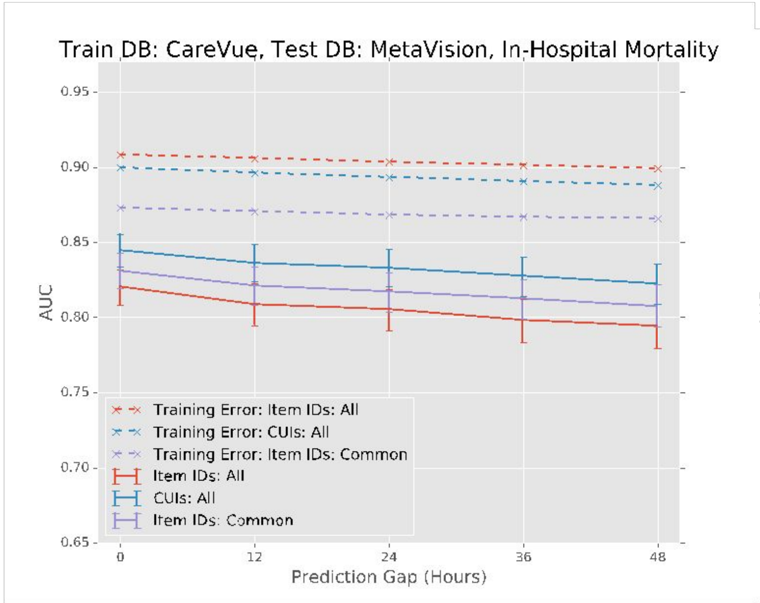
Experiment 1: Does Mapping *Item ID Bag-of-Events* To *CUI Bag-of-Events* Perform Well Within Each EHR?

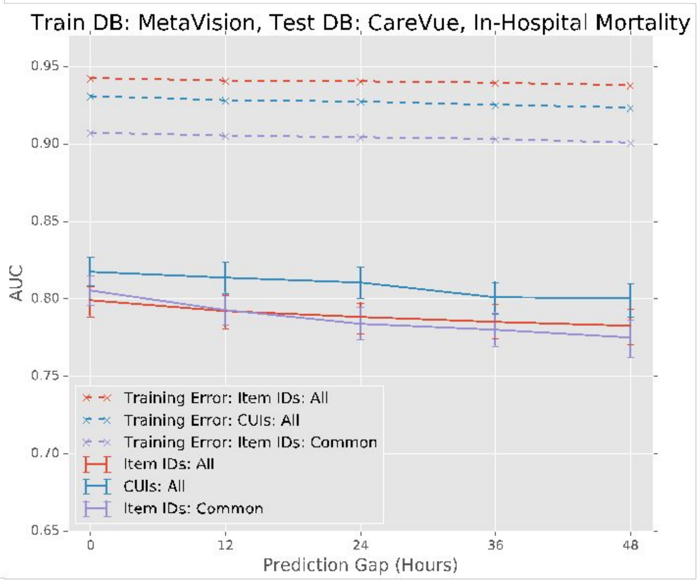
- Mapping *Item IDs* to CUIs does not degrade performance within single EHR.



Experiment 2: Does Mapping *Item ID Bag-of-Events* To *CUI Bag-of-Events* Perform Better Across EHRs?

- Mapping *Item IDs* to CUIs improves performance consistently across EHRs.





Summary: Text Enables Generalizability

- Application proposed an approach to automatically map semantically similar concepts from different databases to a common vocabulary.

- Application demonstrated the utility of this approach across an EHR transition on a set of prediction tasks.

Clinical Natural Language Processing and Audio

Tristan Naumann, Microsoft Research

🖂 tristan@microsoft.com