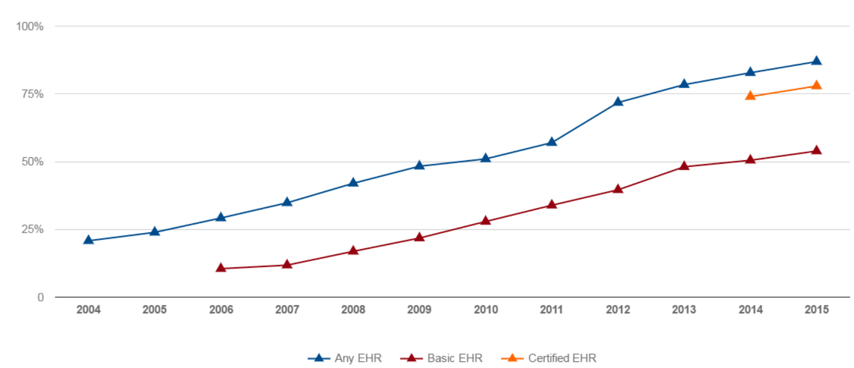
CLINICAL NATURAL LANGUAGE PROCESSING

Mark Dredze

MEDICAL DATA

In the United States there are on the order of 2,000,000,000 doctor-patient encounters per year; that's over 200,000 an hour.

MEDICAL DATA



MEDICAL DATA

"The Electronic Health Record (EHR) is a longitudinal electronic record of patient health information generated by one or more encounters in any care delivery setting. Included in this information are patient demographics, progress notes, problems, medications, vital signs, past medical history, immunizations, laboratory data, and radiology reports. . . ."

National Institutes of Health, National Center for Research Resources, "Electronic Health Records Overview" http://www.ncrr.nih.gov/publications/informatics/ehr.pdf

WHAT'S IN AN EHR?

- Patient demographic data
- Procedure/Treatment/Diagnosis Codes
- Vital signs
- Radiology reports
- Test results
- Clinical text notes

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

Symptoms
Demographics
Tests

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.

WHY CLINICAL NOTES MATTER?

- Improve patient care
  - Find critical information from past visits
  - Suggest diagnosis or medications
  - Ensure followups

SECONDARY USE

If we have access to thousands / millions of medical records representing millions of patients what can we learn?

SECONDARY USE

- Make research advances in medicine
- Discover negative medication interactions
- Identify a cohort of patients for a research study
- Ensure patient safety procedures

ACCESS CLINICAL TEXT

- Most EHR data is structured
  - Relatively easy to export and use
- Clinical text is unstructured?
  - How do we use it?
  - Natural Language Processing

CLINICAL NATURAL LANGUAGE PROCESSING

- The task of applying NLP tools to clinical free text

WHY CLINICAL NLP?

- News
- Social media
- Conversations
- Blogs
- Books



CLINICAL NLP CHALLENGES

- Complex clinical terminology
  - mitral regurgitation, Left ventricular systolic dysfunction, Subarchnoid hemorrhage
- Numerous medical domains
  - Radiology vs. pediatrics vs. oncology vs. …
- Institutional specialization
  - Johns Hopkins Medicine vs. Cheyenne Regional Medical Center
  - Differ in patients, hospital policies and standards
- EHR clinical text storage
  - Different EHR systems/deployments store text data differently

CLINICAL NLP CHALLENGES

- Mixed data types
  - Clinical text notes contain tables, lists, bullets, full paragraphs
- Varying clinical note types
  - Discharge notes: long, boiler plate summaries
  - ER notes: detailed descriptions, wide variety of issues
  - Progress notes: very brief updates (a sentence of less)
  - Radiology reports: descriptive analysis of imagery
- Records in context
  - Clinical record needs to be taken in context of structured data and previous notes

DATA CHALLENGES

- English Wikipedia: 3.7 billion words
- Common Crawl: billions of *pages*
- Social media: 500 million tweets per day
- Europarl: tens of millions of words

DATA CHALLENGES

- Available Clinical data corpora
  - MIMIC 3: 1 ICU million notes
- Data restricted due to HIPAA concerns
  - HIPAA: US law that protects medical records

COMMON CLINICAL NLP TASKS

XXXXX | XXXX | XXXX | | XXXX | 12/01/1998 12:00:00 AM | INCARCERATED UMBILICAL HERNIA | Signed | DIS |
Admission Date: 11/7/1 998 Report Status: Signed

Discharge Date: 1/25/1999

PRINCIPAL DIAGNOSIS: INCARCERATED UMBILICAL HERNIA.

HISTORY: Jane Doe is a 76 year old woman with a complex past medical history including coronary artery disease with a history of MIs times two in the past , a history of DVT back in 1974 , hypertension , rheumatoid arthritis , gout and history of atrial fibrillation and atrial flutter as well as onset adul t diabetes mellitus. She presented to the XXX Community Hospital on the day of admission complaining of an umbilical bulge over the past several weeks. This umbilical bulge had been increasi ng somewhat in size , but had not bothered her and was always reducible. However , over the preceding weekend it became incarcerated and then became somewhat painful. It was not associated with any nausea or vomiting and she reported that she was having normal bowel movements even in the face of this problem. She presented initially to the XXX Coun ty Health Center and was admitted with the diagnosis of incarcerated umbilical hernia.

PAST MEDICAL HISTORY: 1. Coronary artery disease with a history of MI times two in the past with a recent echocardiogram on 1 2/8 showing an EF of 55-60%. 2. History of DVT in 1974. 3. Hypertension. 4. Rheumatoid arthritis. 5. Gout. 6. Atrial fibrillation and atrial flutter on Coumadin. 7. Adult onse t diabetes mellitus.

PAST SURGICAL HISTORY: 1. Status post appendectomy. 2. Status post mitral valve replacement with St. Jude valve. 3. Left hip fracture repair. 4. Status post mitral valve commissurotomy in 1965.

MEDICATIONS ON ADMISSION: Lasix 80 mg a day , sublingual nitroglycerin p.r.n. , Propafenone 225 mg t.i.d. , Lopressor 150 mg b.i.d. , Lisinopril 10 mg a day and Micronase 10 mg b.i.d. , Isordil 40 mg t.i.d. , Coumadin 5 mg a day with 2-1/2 mg every Sunday.

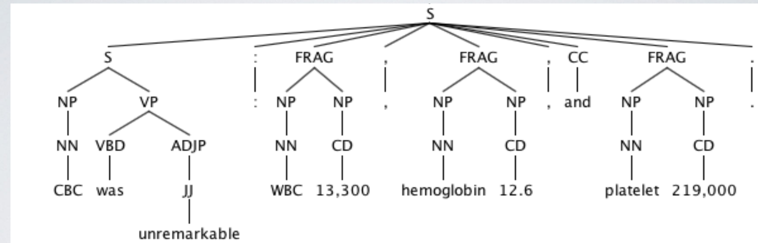
ALLERGIES:: Sheisallergictoaspirinandpenicillin.

PHYSICAL EXAMINATION: She is an extremely pleasant elderly woman in no acute distress. HEENT - showed extraocular movements intact. Pupils equally round and reactive to light. NECK - supple. HEART - regular rhythm. LUNGS - clear. ABDOMEN - soft , nontender , nondistended with approximately 1.5 cm in diameter umbilical hernia to the left of her umbilicus. This hernia was somewhat tender to palpation , but showed no overlying erythema or evidence of necrosis. She had normal bowel sounds.

SEGMENTATION

- Section boundary detection
  - Chief complaint, medical history, assessment and plan, tables, lists
- Sentence boundary detection
  - Identify coherent sentences
    - Many sentence fragments and notes
- Tokenization
  - Complex because of abbreviations, technical terms
  - 75 lbs.   P.O. P O   x-ray   U/ml   left/right
- Abbreviations

SYNTAX



- Part of speech tagging
- Chunking
- Parsing

NEGATION

- When is a concept/statement negated?
  - Additionally, there was no evidence of extension of his infected pseudocyst into the psoas muscle.
  - There is no significant interval change in the 2 large pancreatic pseudocysts.
  - Acute pancreatitis with pseudocyst, with no obvious complications of the pseudocyst at this point in time.

DE-IDENTIFICATION

- Use and sharing of clinical notes requires de-identification (HIPAA)
  - Removal of information that could identify the patient
  - "**Sarah** reported feeling dizzy towards the end of the school day; **left teaching** to come to ER."

NAMED ENTITY RECOGNITION

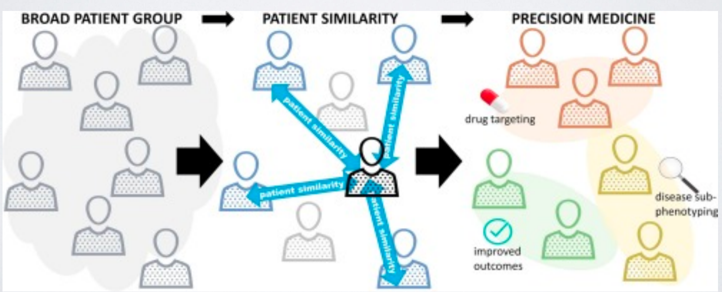
- Identify mentions of named entities in text
  - Medications
  - Symptoms
  - Comorbidities
  - Treatments

SUMMARIZATION

- Produce summaries of clinical notes
  - Single note summarization
  - Cross-note summarization
  - Integrating summaries with structured data

PATIENT SIMILARITY METRICS

- Find other patients like this one
- Applications: precision medicine



AUTOMATED CODING

- Medical data is hand coded
  - ICD10: diseases, symptoms, abnormal findings, complaints, social circumstances, …
  - Creates structured data from free clinical text

Table 2. Revised Wound Care-Related ICD-10 Codes

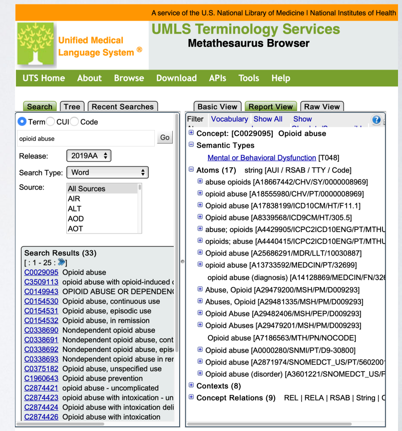
Section 1: Disorders of Metabolism & Lipids	
I63.219	Cerebral infarction due to unspecified occlusion or stenosis of unspecified vertebral artery
I63.239	Cerebral infarction due to unspecified occlusion or stenosis of unspecified carotid artery
I63.333	Cerebral infarction due to thrombosis of bilateral posterior cerebral arteries
I63.343	Cerebral infarction due to thrombosis of bilateral cerebellar arteries
L98.495	Non-pressure chronic ulcer of other sites, with muscle involvement, without evidence of necrosis
L98.496	Non-pressure chronic ulcer of other sites, with bone involvement, without evidence of necrosis
L98.498	Non-pressure chronic ulcer of other sites, with other specified severity
M86.621	Other chronic osteomyelitis, right humerus

SENTENCE SIMILARITY

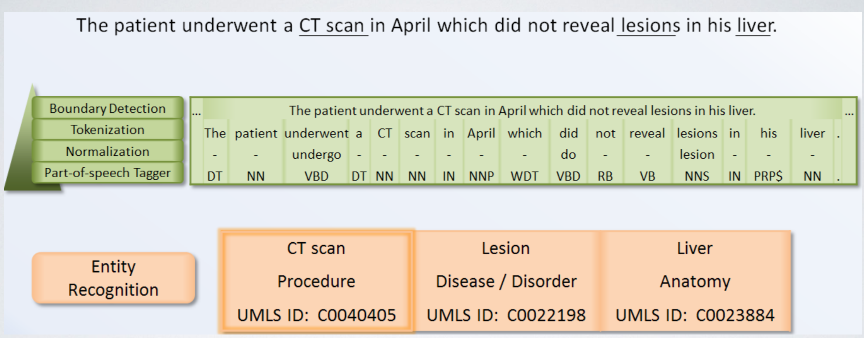
- Are two sentences saying the same thing?
  - No clinically relevant adverse events, such as urinary retention, respiratory disturbances, or wound infections were reported in the M-ADL group.
  - Neither intraoperative nor postoperative clinically relevant adverse events, such as urinary retention, respiratory disturbances, or wound infections, were observed

UNIFIED MEDICAL LANGUAGE SYSTEM - UMLS

- Standardized vocabulary
- Map synonyms to the same concept:
  - C0001969 → 'Alcoholic Intoxication'
  - Synonyms/Atoms: drunkenness, drunk, inebriation, alcohol intoxication
- Normalize to domain ontologies:
  - SNOMED-CT
  - RxNORM
  - LOINC
  - MeSH



CONCEPT LINKING



PHENOTYPING

- Extract patient conditions from free clinical text

387055 |||| 26563 |||| 18146 |||| RADIOLOGY_REPORT |||| 2009-01-26 15:08:00.0 |||| C12 CHEST (PORTABLE AP) |||| |||| Clip # 282-0776 Actual report |||| DATE: [**2009-01-26**] 3:08 PM
CHEST (PORTABLE AP) [**Clip Number (Radiology) 7881**] Reason: r/o ptx, s/p aicd removal

UNDERLYING MEDICAL CONDITION:
57 year old man with pleural effusion.
REASON FOR THIS EXAMINATION:
r/o ptx, s/p aicd removal

FINAL REPORT
INDICATION: 57 y/o male with pleural effusion, rule out pneumothorax, status post AICD removal.

FINDINGS: A single portable chest radiograph is compared with a portable study done earlier today. The right subclavian line is redemonstrated and in appropriate position. There is no evidence for pneumothorax. The AICD has been removed in the interim.

The heart is mildly enlarged when adjusting for technique. There is evidence for mild pulmonary vascular redistribution. There has been interval increase in the size of the right pleural effusion. The patient is status post sternotomy. **The patient has no prior history of smoking.**

IMPRESSION: 1) Status post AICD removal without evidence for pneumothorax.
2) Findings consistent with worsening CHF.
3) Interval increase in the size of the right pleural effusion.

METHODS

ONTOLOGIES / DICTIONARIES

- Medicine has extensive ontologies
  - Structured knowledge repositories
  - List of synonyms, relations, definitions, etc.
- Many clinical NLP incorporate these resources
  - Not typically available in other domains

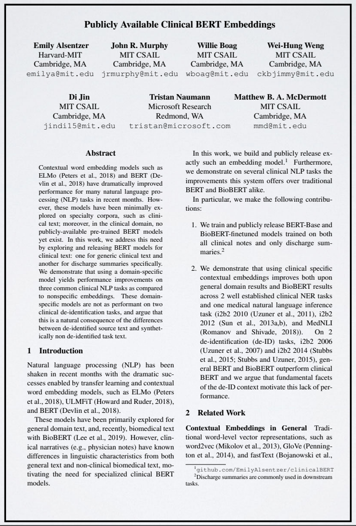
RULES!

- Extensive use of rule based methods in clinical NLP
  - Easier to implement, deploy and understand
  - Medicine requires domain experts, easier for them to understand and create rules
  - Statistical methods require training data: often not present in medicine

STATISTICAL METHODS

- Standard statistical NLP methods
- These work well (often better than rules) but face some challenges:
  - Robustness
  - Training data
  - Interpretability
  - Ease of implementation by domain experts

LANGUAGE MODELS



Publicly Available Clinical BERT Embeddings

Emily Alsentzer, Harvard-MIT, Cambridge, MA, emilya@mit.edu
John R. Murphy, MIT CSAIL, Cambridge, MA, jrmurphy@mit.edu
Willie Boag, MIT CSAIL, Cambridge, MA, wboag@mit.edu
Wei-Hung Weng, MIT CSAIL, Cambridge, MA, ckbjimmy@mit.edu
Di Jin, MIT CSAIL, Cambridge, MA, jindi15@mit.edu
Tristan Naumann, Microsoft Research, Redmond, WA, tristan@microsoft.com
Matthew B. A. McDermott, MIT CSAIL, Cambridge, MA, mmd@mit.edu

Abstract

Contextual word embedding models such as ELMo (Peters et al., 2018) and BERT (Devlin et al., 2018) have dramatically improved performance for many natural language processing (NLP) tasks in recent months. However, these models have been minimally explored on specialty corpora, such as clinical text; moreover, in the clinical domain, no publicly-available pre-trained BERT models yet exist. In this work, we address this need by exploring and releasing BERT models for clinical text: one for generic clinical text and another for discharge summaries specifically. We demonstrate that using a domain-specific model yields performance improvements on three common clinical NLP tasks as compared to nonspecific embeddings. These domain-specific models are not as performant on two clinical de-identification tasks, and argue that this is a natural consequence of the differences between de-identified source text and synthetically non de-identified task text.

1 Introduction

Natural language processing (NLP) has been shaken in recent months with the dramatic successes enabled by transfer learning and contextual word embedding models, such as ELMo (Peters et al., 2018), ULMFiT (Howard and Ruder, 2018), and BERT (Devlin et al., 2018).

These models have been primarily explored for general domain text, and, recently, biomedical text with BioBERT (Lee et al., 2019). However, clinical narratives (e.g., physician notes) have known differences in linguistic characteristics from both general text and non-clinical biomedical text, motivating the need for specialized clinical BERT models.

In this work, we build and publicly release exactly such an embedding model.[1] Furthermore, we demonstrate on several clinical NLP tasks the improvements this system offers over traditional BERT and BioBERT alike.

In particular, we make the following contributions:

1. We train and publicly release BERT-Base and BioBERT-finetuned models trained on both all clinical notes and only discharge summaries.[2]
2. We demonstrate that using clinical specific contextual embeddings improves both upon general domain results and BioBERT results across 2 well established clinical NER tasks and one medical natural language inference task (i2b2 2010 (Uzuner et al., 2011), i2b2 2012 (Sun et al., 2013a,b), and MedNLI (Romanov and Shivade, 2018)). On 2 de-identification (de-ID) tasks, i2b2 2006 (Uzuner et al., 2007) and i2b2 2014 (Stubbs et al., 2015; Stubbs and Uzuner, 2015), general BERT and BioBERT outperform clinical BERT and we argue that fundamental facets of the de-ID context motivate this lack of performance.

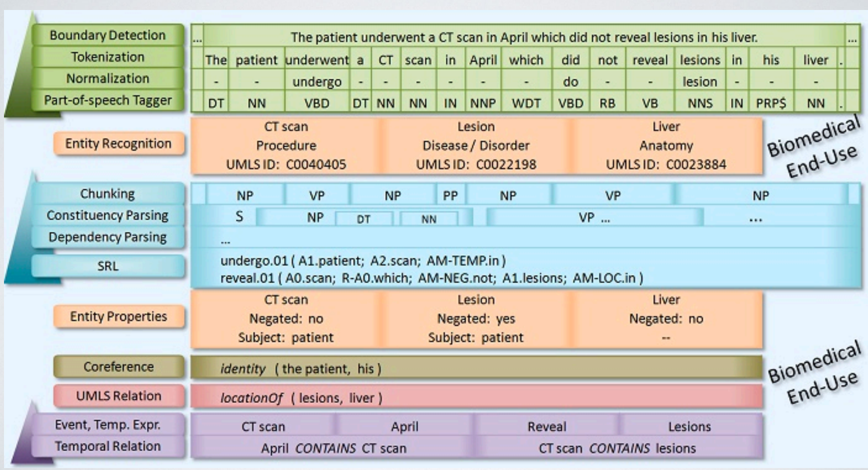
2 Related Work

Contextual Embeddings in General Traditional word-level vector representations, such as word2vec (Mikolov et al., 2013), GloVe (Pennington et al., 2014), and fastText (Bojanowski et al.,

[1] github.com/EmilyAlsentzer/clinicalBERT
[2] Discharge summaries are commonly used in downstream tasks.

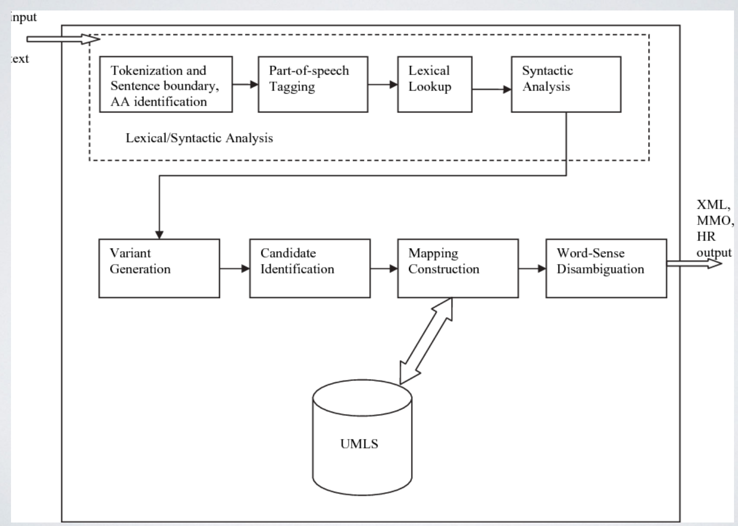
TOOLS

CTAKES



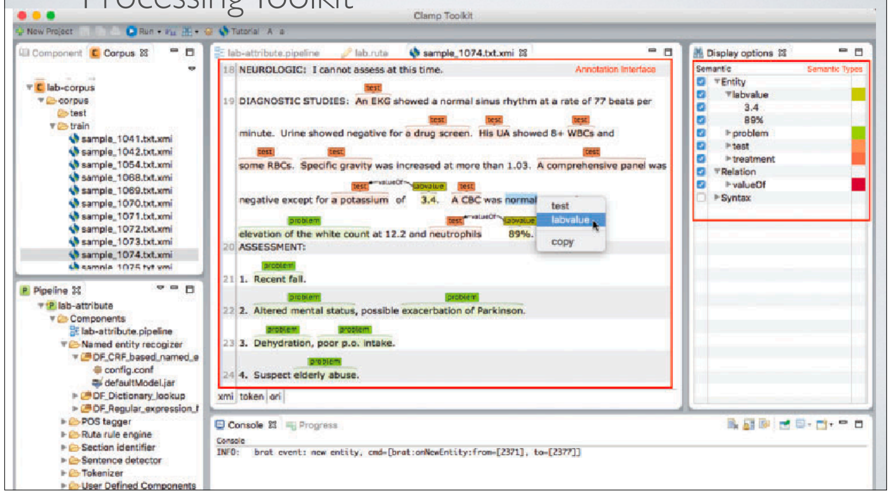
https://ctakes.apache.org/

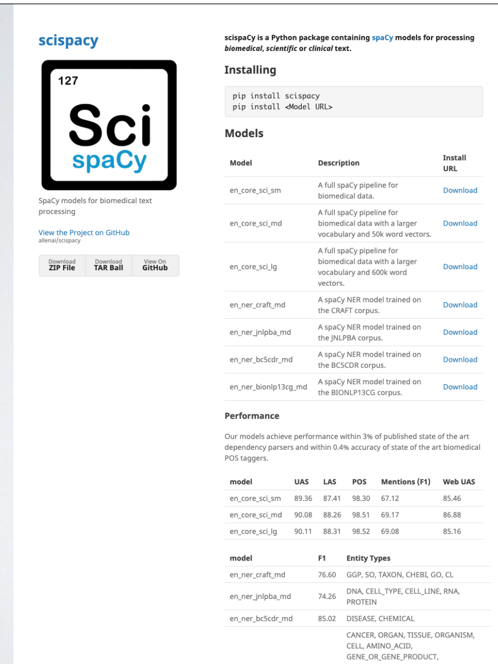
METAMAP



CLAMP

- Clinical Language Annotation, Modeling, and Processing Toolkit





scispacy

127 Sci spaCy

SpaCy models for biomedical text processing

View the Project on GitHub
allenai/scispacy

Download ZIP File | Download TAR Ball | View On GitHub

scispaCy is a Python package containing spaCy models for processing *biomedical, scientific* or *clinical* text.

Installing

```
pip install scispacy
pip install <Model URL>
```

Models

Model	Description	Install URL
en_core_sci_sm	A full spaCy pipeline for biomedical data.	Download
en_core_sci_md	A full spaCy pipeline for biomedical data with a larger vocabulary and 50k word vectors.	Download
en_core_sci_lg	A full spaCy pipeline for biomedical data with a larger vocabulary and 600k word vectors.	Download
en_ner_craft_md	A spaCy NER model trained on the CRAFT corpus.	Download
en_ner_jnlpba_md	A spaCy NER model trained on the JNLPBA corpus.	Download
en_ner_bc5cdr_md	A spaCy NER model trained on the BC5CDR corpus.	Download
en_ner_bionlp13cg_md	A spaCy NER model trained on the BIONLP13CG corpus.	Download

Performance

Our models achieve performance within 3% of published state of the art dependency parsers and within 0.4% accuracy of state of the art biomedical POS taggers.

model	UAS	LAS	POS	Mentions (F1)	Web UAS
en_core_sci_sm	89.36	87.41	98.30	67.12	85.46
en_core_sci_md	90.08	88.26	98.51	69.17	86.88
en_core_sci_lg	90.11	88.31	98.52	69.08	85.16

model	F1	Entity Types
en_ner_craft_md	76.60	GGP, SO, TAXON, CHEBI, GO, CL
en_ner_jnlpba_md	74.26	DNA, CELL_TYPE, CELL_LINE, RNA, PROTEIN
en_ner_bc5cdr_md	85.02	DISEASE, CHEMICAL
		CANCER, ORGAN, TISSUE, ORGANISM, CELL, AMINO_ACID, GENE_OR_GENE_PRODUCT,

CONCEPT LINKING

387055 |||| 26563 |||| 18146 |||| RADIOLOGY_REPORT |||| 2009-01-26 15:08:00.0 |||| C12 CHEST (PORTABLE AP) |||| |||| Clip # 282-0776 Actual report |||| DATE: [**2009-01-26**] 3:08 PM
CHEST (PORTABLE AP) [**Clip Number (Radiology) 7881**] Reason: r/o ptx, s/p aicd removal

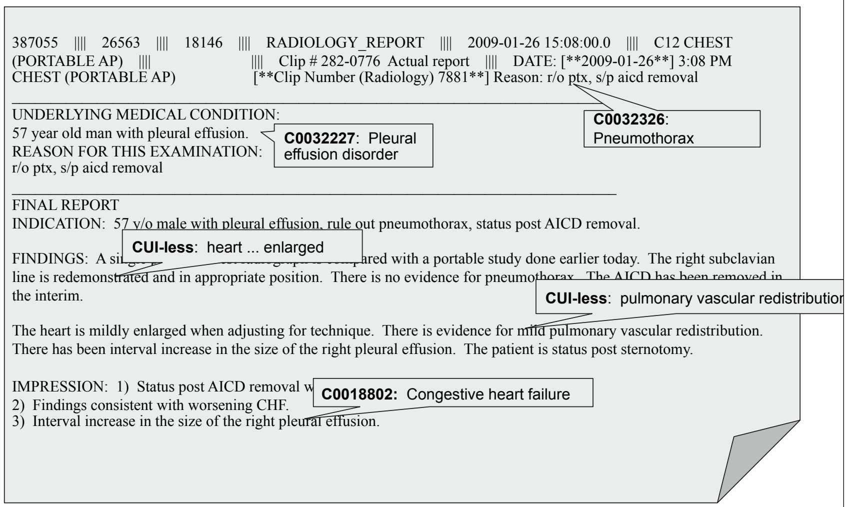
UNDERLYING MEDICAL CONDITION:
57 year old man with pleural effusion.
REASON FOR THIS EXAMINATION:
r/o ptx, s/p aicd removal

FINAL REPORT
INDICATION: 57 y/o male with pleural effusion, rule out pneumothorax, status post AICD removal.

FINDINGS: A single portable chest radiograph is compared with a portable study done earlier today. The right subclavian line is redemonstrated and in appropriate position. There is no evidence for pneumothorax. The AICD has been removed in the interim.

The heart is mildly enlarged when adjusting for technique. There is evidence for mild pulmonary vascular redistribution. There has been interval increase in the size of the right pleural effusion. The patient is status post sternotomy.

IMPRESSION: 1) Status post AICD removal without evidence for pneumothorax.
2) Findings consistent with worsening CHF.
3) Interval increase in the size of the right pleural effusion.



387055 |||| 26563 |||| 18146 |||| RADIOLOGY_REPORT |||| 2009-01-26 15:08:00.0 |||| C12 CHEST (PORTABLE AP) |||| |||| Clip # 282-0776 Actual report |||| DATE: [**2009-01-26**] 3:08 PM
CHEST (PORTABLE AP) [**Clip Number (Radiology) 7881**] Reason: r/o ptx, s/p aicd removal

UNDERLYING MEDICAL CONDITION:
57 year old man with pleural effusion.
REASON FOR THIS EXAMINATION:
r/o ptx, s/p aicd removal

FINAL REPORT
INDICATION: 57 y/o male with pleural effusion, rule out pneumothorax, status post AICD removal.

FINDINGS: A si... ared with a portable study done earlier today. The right subclavian line is redemonstrated and in appropriate position. There is no evidence for pneumothorax. The AICD has been removed in the interim.

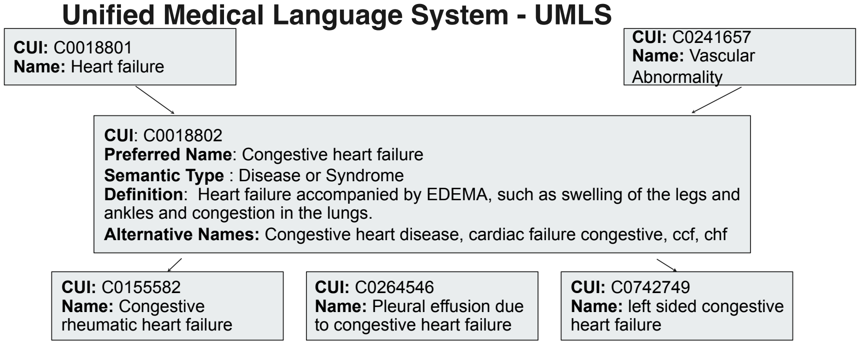
The heart is mildly enlarged when adjusting for technique. There is evidence for mild pulmonary vascular redistribution. There has been interval increase in the size of the right pleural effusion. The patient is status post sternotomy.

IMPRESSION: 1) Status post AICD removal w...
2) Findings consistent with worsening CHF.
3) Interval increase in the size of the right pleural effusion.

Unified Medical Language System - UMLS

CUI: C0018802
Preferred Name: Congestive heart failure
Semantic Type : Disease or Syndrome
Definition: Heart failure accompanied by EDEMA, such as swelling of the legs and ankles and congestion in the lungs.
Alternative Names: Congestive heart disease, cardiac failure congestive, ccf, chf

Unified Medical Language System - UMLS

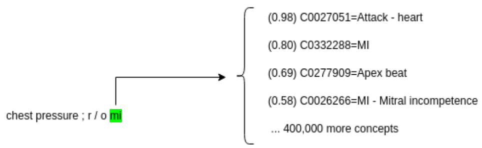


Concept Linking

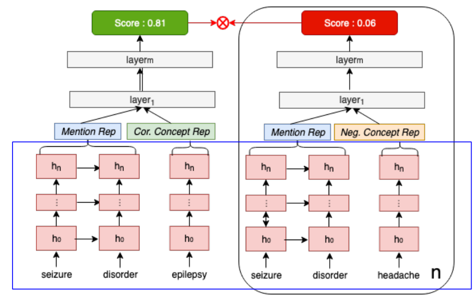
- Task: Given a mention, identify potential links to concepts within the UMLS.
  - We assume gold standard spans
  - Named entity recognition is usually run as a prior step when spans are not provided.
- Current well-known solutions rely on lexical-only methods (e.g. dictionary lookup, abbreviation expansion)
  - MetaMap (Aronson, et al.) - **NIH**
  - cTAKES (Savova, et al.) - **Mayo**

Neural Ranker

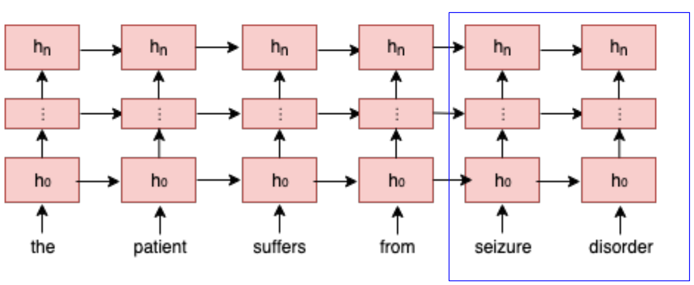
We treat our concept linking task as a learning-to-rank task:



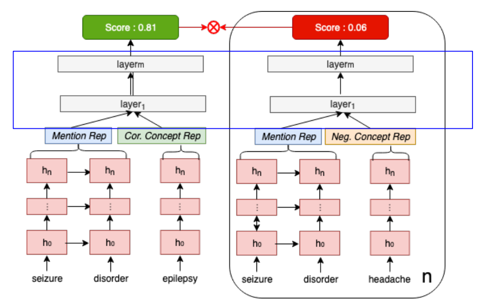
Neural Ranker



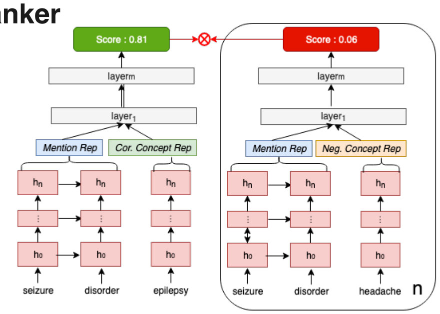
Neural Ranker



Neural Ranker

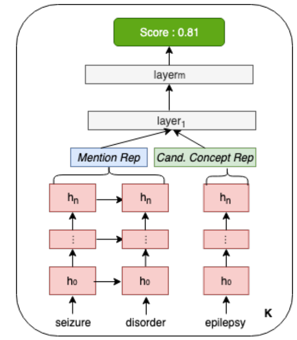


Neural Ranker



$$L(\theta) = \mathbf{max}\{0, \epsilon - (S(\{m, c_+\}; \theta) - \mathbf{max}\{S(\{m, c_{0-}\}; \theta) \ldots S(\{m, c_{n-}\}; \theta)\}\}$$

Neural Ranker



Dataset - N2C2 Task 3

- 50 training clinical notes with ~6,700 training (mention, concept) annotations.
- Concept labels consisted of any CUI among ~430,000 in 2017AB SNOMED, RxNorm subsets.

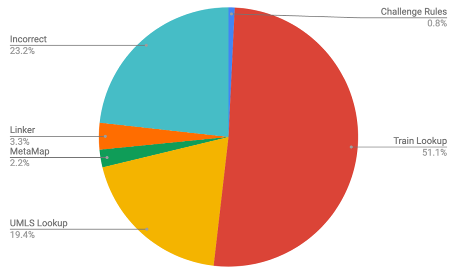
Statistic	Training	Evaluation	Total
Document Count	50	50	100
Mention Instances	6,684	4,235	10,919
Unique Concepts	2,331		3,792

Statistic	Train	Test	Total
Mention Count	5339	1345	6684
Unique Concept Count	4058	1035	2331
Disjoint Mentions	87	18	105
CUI-less count	120	31	151

Lexical Rules

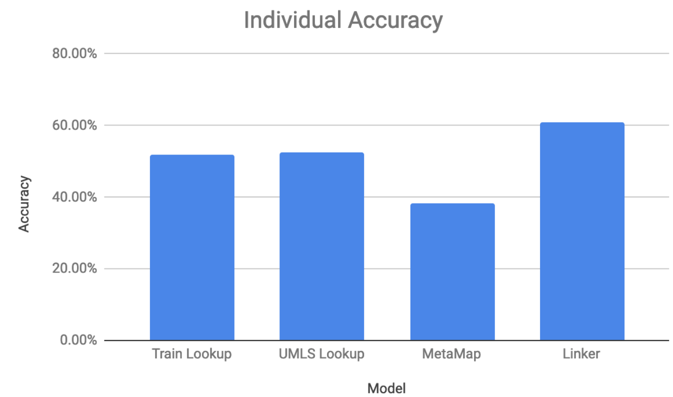
- First, we apply two lookup procedures.
  - If a mention was previously seen in the training set, we assign it the same concept
  - If a mention name matches exactly to one entry in the UMLS, we assign it that concept
- These rules work well for this dataset due to annotation guidelines
  - Annotators were requested to try to break apart mentions to be easily matched
- If none of these rules match to a single concept, we use the neural mention ranker to identify the best concept.

Results



	Accuracy	Labeled
Challenge Rules	0.77%	0.77%
Train Lookup	51.83%	53.69%
UMLS Lookup	71.23%	79.22%
MetaMap	73.43%	83.57%
Linker	76.75%	100.00%

Results



Outline

1. Introduction to Clinical NLP
2. Concept Linking
3. **Clinical Semantic Textual Similarity**
4. Phenotyping

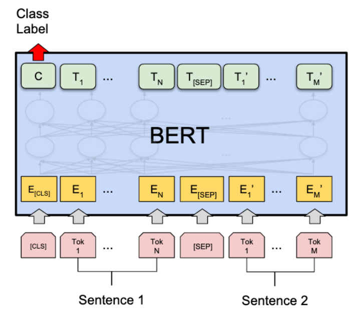
N2C2 2019 STS Data

Albuterol [PROVENTIL/ VENTOLIN] 90 mcg/Act HFA Aerosol 2 puffs by inhalation every 4 hours as needed.	Albuterol [PROVENTIL/VENTOLIN] 90 mcg/Act HFA Aerosol 1-2 puffs by inhalation every 4 hours as needed #1 each.	5
Cardiovascular assessment findings include heart rate normal, Heart rhythm, atrial fibrillation with controlled ventricular response.	Cardiovascular assessment findings include heart rate, bradycardic, Heart rhythm, first degree AV Block.	3
The risks and benefits of the procedure were discussed, and the patient consented to this procedure	The content of this note has been reproduced, signed by an authorized	1

Clinical STS

- Task: Given a pair of sentences from a clinical note, predict the degree of content overlap
- Current well-known solutions
  - Include hand-crafted features
  - InferSent - A sentence encoder from GloVe embeddings

Model



From “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”, Devlin et al 2018

BERT for Clinical Text

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

Table 2: Accuracy (MedNLI) and Exact F1 score (i2b2) across various clinical NLP tasks.

From “Publicly Available Clinical BERT Embeddings”, Alsentzer et al. 2019

Results

Model	**Data**	**Pearson Correlation ρ**
Levenshtein Distance	--	.680
Clinical BERT	STS	.771
Clinical BERT	MedSTS	.849
Clinical BERT	STS + MedSTS	.854

Outline

1. Introduction to Clinical NLP
2. Concept Linking
3. Clinical Semantic Textual Similarity
4. **Phenotyping**

387055 |||| 26563 |||| 18146 |||| RADIOLOGY_REPORT |||| 2009-01-26 15:08:00.0 |||| C12 CHEST (PORTABLE AP) |||| |||| Clip # 282-0776 Actual report |||| DATE: [**2009-01-26**] 3:08 PM
CHEST (PORTABLE AP) [**Clip Number (Radiology) 7881**] Reason: r/o ptx, s/p aicd removal

UNDERLYING MEDICAL CONDITION:
57 year old man with pleural effusion.
REASON FOR THIS EXAMINATION:
r/o ptx, s/p aicd removal

FINAL REPORT
INDICATION: 57 y/o male with pleural effusion, rule out pneumothorax, status post AICD removal.

FINDINGS: A single portable chest radiograph is compared with a portable study done earlier today. The right subclavian line is redemonstrated and in appropriate position. There is no evidence for pneumothorax. The AICD has been removed in the interim.

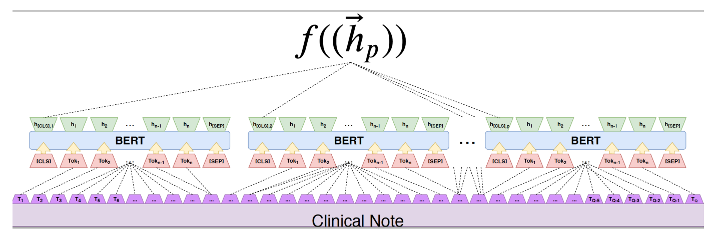
The heart is mildly enlarged when adjusting for technique. There is evidence for mild pulmonary vascular redistribution. There has been interval increase in the size of the right pleural effusion. The patient is status post sternotomy.

IMPRESSION: 1) Status post AICD removal without evidence for pneumothorax.
2) Findings consistent with worsening CHF.
3) Interval increase in the size of the right pleural effusion.

Phenotyping

- Task: Identify patient trait from a clinical note
- We cannot rely on only the first part of the document as in some tasks (e.g. topic detection)
- Current well-known solutions
  - For the i2b2 2006 Smoking dataset, the best performing system uses handcrafted regular expressions, rules, and features to train a SVM binary classifier
  - For the i2b2 2008 Obesity dataset, the best performing system consisted only of hand-engineered rules
  - Other approaches have looked at using neural architectures (e.g. CNNs) with worse results

Phenotyping



Classification models

- We explored several inputs into our FNN for our document given a set of `[CLS]` embeddings.
  - Dimension-wise max over all CLS embeddings
  - Input all embeddings into a FNN with padding
  - Input all embeddings into Transformer
  - The final state of an LSTM over all embeddings

Dataset

- 2006 Smoker Identification
  - Past Smoker
  - Current Smoker
  - Non-Smoker
  - Unclear
- 2008 Obesity Identification
  - Obesity
  - 14 Co-Morbidities (e.g. congestive heart failure)
  - As a note can have more than one label, we train a classifier for each label.

Results

Table 1: Phenotyping results (micro-averaged F_1) with our architectures compared to the top shared task system and recent deep learning based systems.

	I2B2 2006: Smoking	I2B2 2008: Obesity
f_{max}	60.0	74.7
f_I	82.9	81.3
$f_{\text{Transformer}}$	75.9	87.9
f_{LSTM}	**97.5** (97.1 ± .48)	94.5 (93.9 ± .59)
Shared Task 1st Place	90.0	95.0
CNN [16]	77.0	–
CNN + Rules [7]	–	96.2

THANK YOU

- This presentation based on slides from
  - Brant Chee, Masoud Rouhizadeh, Elliot Schumacher, Philip Resnik