Extracting CHF information from clinical text using CLAMP

Hua Xu, PhD

pSCANNER 2016
What is CLAMP - Clinical Language Annotation, Modeling, and Processing?

• A general purpose clinical NLP system – “CLAMP CMD”
  – Built on proven methods
  – Good performance, high speed
• An IDE (integrated development environment) for building customized clinical NLP pipelines via GUIs – “CLAMP GUI”
  – Annotating/analyzing clinical text
  – Training of ML-based modules
  – Specifying rules
• An enterprise solution for NLP needs in healthcare organizations – “CLAMP Enterprise”
  – Task management
  – Visual analytics
CLAMP CMD – built on proven methods

<table>
<thead>
<tr>
<th>NLP Tasks</th>
<th>Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Named entity recognition</td>
<td></td>
</tr>
<tr>
<td>2009 i2b2, medication</td>
<td>#2</td>
</tr>
<tr>
<td>2010 i2b2 problem, treatment, test</td>
<td>#2</td>
</tr>
<tr>
<td>2013 SHARE/CLEF abbreviation</td>
<td>#1</td>
</tr>
<tr>
<td>UMLS encoding</td>
<td></td>
</tr>
<tr>
<td>2014 SemEval, disorder</td>
<td>#1</td>
</tr>
<tr>
<td>Relation extraction</td>
<td></td>
</tr>
<tr>
<td>2012 i2b2 Temporal</td>
<td>#1</td>
</tr>
<tr>
<td>2015 SemEval Disease-modifier</td>
<td>#1</td>
</tr>
<tr>
<td>2015 BioCREATIVE Chemical-induced disease</td>
<td>#1</td>
</tr>
</tbody>
</table>
CLAMP GUI - Efficiently build customized NLP pipelines for individual applications!

[Diagram of CLAMP GUI interface showing resource, corpus, and pipeline sections with various tools and components.]
The patient is an 80 year old female with breast cancer, status post lumpectomy / radiation therapy / Tamoxifen (2000), hypertension, hyperlipidemia, multiple urinary tract infections who presents with a four day prodrome of dry cough, rhinorrhea, coryza, malaise, chills, headache, decreased p.o. intake,
Specifying rules

```
TYPESYSTEM ClampTypeSystem;
   //Auto generated by rule editor

BLOCK(ForEach) Sentence[FEATURE("segmentId", "medications")]{
    BaseToken[REGEXP("Tamsulosin") -> UNMARK(ClampNameEntityUIIMA, true),
    CREATE( ClampNameEntityUIIMA, 1,1,"semanticTag" - "treatment")];
}
```

1. Tamsulosin 0.4 mg Capsule, Sust. Release 24HR Sig: One (1) Capsule, Sust. Release 24HR PO HS (at bedtime)
Extracting CHF information using CLAMP built-in components
CHF information

• CHF Terms/Concepts
  – Any terms mentioned regarding to Congestive Heart Failure
  – E.g. “congestive heart failure”, “systolic heart failure” or “diastolic heart failure” etc.

• Lab tests
  – Laboratory test aiming to evaluate if patient has Congestive Heart Failure
  – E.g. “BUN”, “RBC” etc.

• Medications
  – Any medications to treat Congestive Heart Failure
  – E.g. ACE inhibitors, Angiotensin receptor blockers etc.

• Image tests
  – image tests to evaluate if patient has Congestive Heart Failure
  – E.g. “Ejection Fraction”, “chest x-ray” etc.
CLAMP built-in components

• Basic name entity recognition
  – Regular expression based NER - Numbers
  – Machine learning based NER - problems, treatments and tests
• Medication and signature identification
  – Wrap the rule-based MedEx-UIMA as a CLAMP component
• Lab test and value identification
  – Rule based module
• UMLS Encoding
  – Assign UMLS code to previously identified name entities
• Customize it to CHF using post-processing rules
  – Filtering by CUIs and names
CLAMP screen shots
Evaluation

• MIMIC-III corpus
  – 10,000+ patient with CHF diagnosis code
  – Randomly selected 30 discharge summaries from these patients
  – A nurse annotated CHF related information

• Six categories of name entities: CHF Term, Lab test name, Lab test name + value, Image name, Drug name, Drug name + signatures

• Matching criteria: exact vs. inexact
• Metrics: precision, recall, and F1
## Preliminary Results

<table>
<thead>
<tr>
<th>Category</th>
<th># of Gold</th>
<th># of Sys</th>
<th>Recall(%)</th>
<th>Precision(%)</th>
<th>F-1 measure(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHF Term</td>
<td>80</td>
<td>79</td>
<td>80.0/93.7</td>
<td>81.0/94.9</td>
<td>80.5/94.3</td>
</tr>
<tr>
<td>Lab test name</td>
<td>318</td>
<td>331</td>
<td>89.6/95.6</td>
<td>86.1/91.8</td>
<td>87.8/93.6</td>
</tr>
<tr>
<td>Lab test name + value</td>
<td>318</td>
<td>331</td>
<td>84.5/92.7</td>
<td>81.2/89.1</td>
<td>82.8/90.8</td>
</tr>
<tr>
<td>Image</td>
<td>136</td>
<td>139</td>
<td>85.2/99.2</td>
<td>83.4/97.1</td>
<td>84.2/98.1</td>
</tr>
<tr>
<td>Drug name</td>
<td>388</td>
<td>392</td>
<td>83.5/96.6</td>
<td>82.6/95.6</td>
<td>83.0/96.0</td>
</tr>
<tr>
<td>Drug name + signature</td>
<td>637</td>
<td>642</td>
<td>81.4/95.7</td>
<td>80.8/95.0</td>
<td>81.1/95.3</td>
</tr>
<tr>
<td>All</td>
<td>1693</td>
<td>1691</td>
<td>84.7/94.3</td>
<td>84.8/94.5</td>
<td>84.8/94.4</td>
</tr>
</tbody>
</table>
Adapting VA EF-Extractor to CLAMP
EF-Extractor by VA

• **Left ventricular ejection fraction (EF)** is a key component of heart failure quality measures used within the Department of Veteran Affairs (VA)

• **EF-Extractor** use regular expressions and rules to capture the EF;

• It is UIMA based and can run on UIMA AS;

• We integrate the **EF-Extractor** as an ‘user-defined-component’ into CLAMP;
Integrate EF_Extractor into clamp
Further Improvement with clamp

Add more rules and Ruta script to further improve the performance;

<table>
<thead>
<tr>
<th>Name</th>
<th>Component</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DF_Detect_sentences_by_newline</td>
<td>Sentence detector</td>
<td>Detect sentences by Newline (\n)</td>
</tr>
<tr>
<td>DF_Clamp_tokenizer</td>
<td>Tokenizer</td>
<td>Rule based tokenizer</td>
</tr>
<tr>
<td>DF_OpenNLP_POS_tagger</td>
<td>POS tagger</td>
<td>OpenNLP based pos tagger</td>
</tr>
<tr>
<td>EF_Extractor</td>
<td>User Defined Components</td>
<td>Regular expression based labtest-value extractor</td>
</tr>
<tr>
<td>DF_Ruta_script_file</td>
<td>Ruta rule engine</td>
<td>Ruta script</td>
</tr>
<tr>
<td>DF_Relation_connector_after_ruta</td>
<td>User Defined Components</td>
<td>set relation name after ruta script;</td>
</tr>
</tbody>
</table>

During her admission at CMHH her ejection fractions during this time ranged from 30-40%.

```
ClampNameEntityUIMA{ FEATURE( "semanticTag", "ef::Measurement" ) }
BaseToken{ REGEXP( "during" ) }
BaseToken{ REGEXP( "this" ) }
BaseToken{ REGEXP( "time" ) }
BaseToken{ REGEXP( "ranged" ) }
BaseToken{ REGEXP( "from" ) }
ClampNameEntityUIMA{ FEATURE( "semanticTag", "ef::NumericValue" )
  -> GATHER( ClampRelationUIMA, "entFrom"=7, "entTo"=1 ) };
```
Test on UT dataset

• Extract 200 sentences from UT notes that contain keywords like ‘ejection fraction’ or ‘EF’.
• Evaluation:
  – Concept level: Recognize all EF mentions and related values;

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Dataset</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>VA *</td>
<td>95%</td>
<td>88.9%</td>
<td>0.919</td>
</tr>
<tr>
<td>Original</td>
<td>UTHealth</td>
<td>83.0%</td>
<td>84.7%</td>
<td>0.838</td>
</tr>
<tr>
<td>Customized</td>
<td>UTHealth</td>
<td>98.2%</td>
<td>89.1%</td>
<td>0.934</td>
</tr>
</tbody>
</table>

*Garvin JH et al JAMIA 2012
CLAMP – It’s all about transportability

• Building customized pipelines for different applications using default components
• Integrating existing components/tools into CLAMP
• Exporting CLAMP pipelines to other NLP platforms