Comparison of ACM and CLAMP for Entity Extraction in Clinical Notes

FATEMEH SHAH-MOHAMMADI, WANTING CUI, JOSEPH FINKELSTEIN

CENTER FOR BIOMEDICAL AND POPULATION HEALTH INFORMATICS, ICAHN SCHOOL OF MEDICINE AT MOUNT SINAI, NEW YORK, NY.

Overview

- Why is Entity Extraction needed?
- Clinical Entity Extraction Tools:
  - Clinical Language Annotation, Modeling and Processing (CLAMP)
  - Amazon Comprehend Medical (ACM)
- Dataset
- Evaluation Metrics
- Results & Discussion
- Conclusion

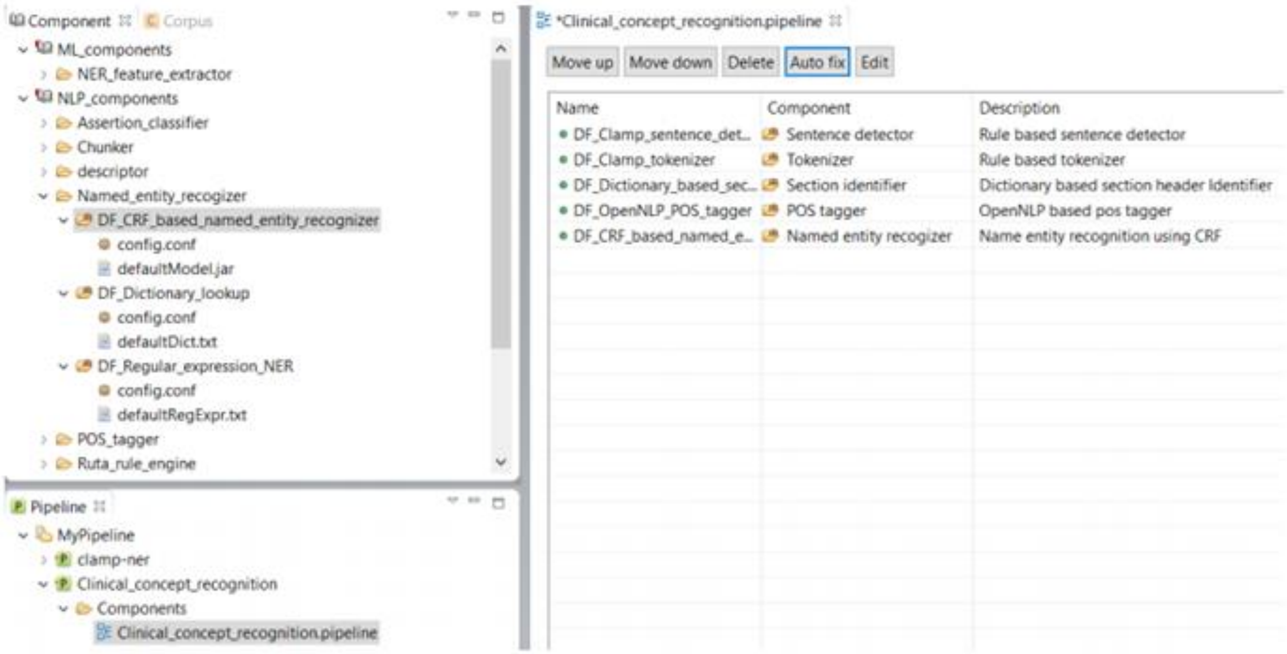
Why is Entity Extraction needed?

- Clinical Notes recorded in unstructured format
- Clinical Notes contain vast amount of information
- Information needs to be extracted for further utilization and analysis in daily healthcare setting
- Extracted information also form basis for other tasks (disease correlation and classification)

Roque FS, et al. Using electronic patient records to discover disease correlations and stratify patient cohorts. PLoS Comput Biol. 2011 Aug 25;7(8):e1002141.
Yildirim P, Çeken Ç, Hassanpour R, Tolun MR. Prediction of similarities among rheumatic diseases. J Med Syst. 2012; 36(3):1485–90

Tools: Clinical language annotation, modeling and processing tool (CLAMP)

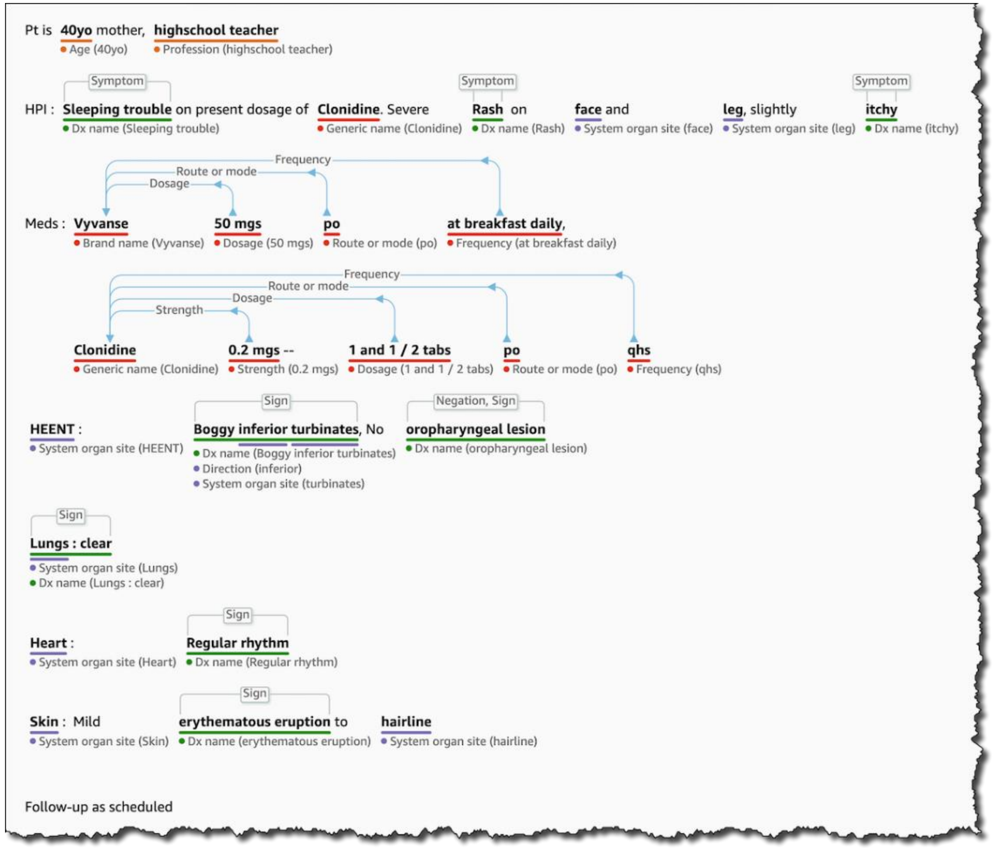
- NLP-based clinical entity extraction tool
- Developed by: University of Texas, Health Science Center at Houston
- Provides interactive development environment (IDE) for building customized clinical NLP solutions
- Presents a pipeline-based architecture that builds NLP systems from multiple components



Soysal, Ergin, et al., "CLAMP–a toolkit for efficiently building customized clinical natural language processing pipelines," Journal of the American Medical Informatics Association 25.3, pp: 331-336, 2018.
[CLAMP | Natural Language Processing (NLP) Software (uth.edu)]

Tools: Amazon Comprehend Medical (ACM)

- A deep neural network-based entity extraction tool
- Developed by Amazon Web Service (AWS)
- Uses deep learning based system (Long Short Term Memory (LSTM) network and Transfer Learning)



P. Bhatia, B. Celikkaya, M. Khalilia and S. Senthivel. Comprehend Medical: A Named Entity Recognition and Relationship Extraction Web Service, 18th IEEE ICMLA. 2019;1844-1851
https://aws.amazon.com/blogs/aws/new-amazon-comprehend-medical-adds-ontology-linking/

Dataset

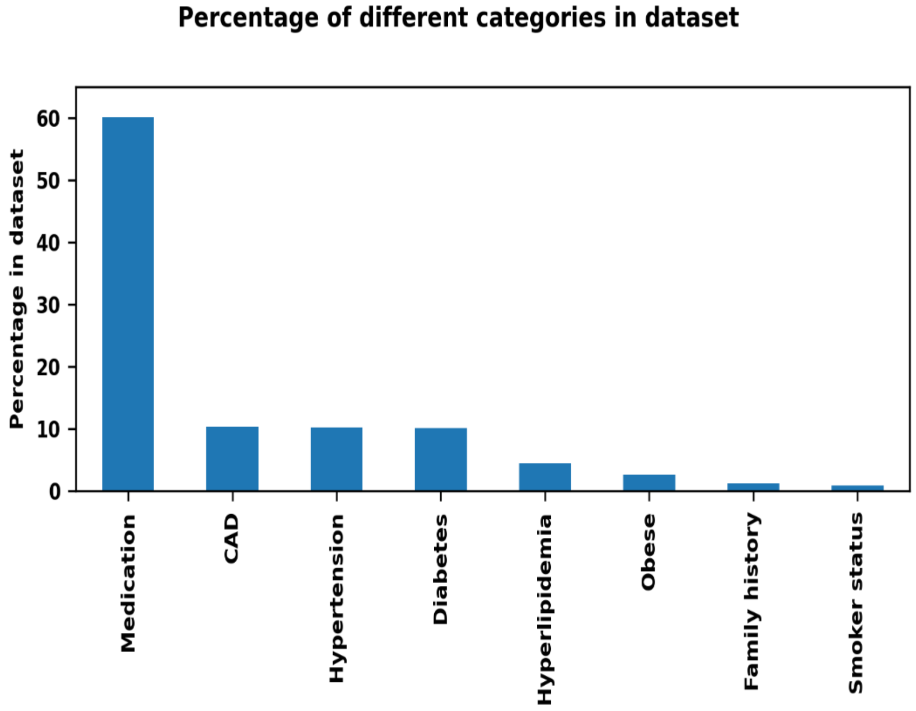
- The 2014 i2b2 heart disease and its associated risk factors identification dataset
- Consists of 521 medical records with distribution of 8 disease risk factor categories and 38 associated indicators

Category
Hypertension
Hyperlipidemia
Diabetes
Obese
Coronary Artery Disease (CAD)
Medication
Family History
Smoking Status

Indicator
Hyperlipidemia
Dyslipidemia
Hypercholesterolemia
High Cholesterol

Dataset

- Both IE systems link their extracted entities to normalized concept identifiers from the RxNorm database.
- RxNorm provides normalized names and unique identifiers only for medicines and drugs.
- Amongst 8 categories, we only considered entities categorized as “medication”.
- Entities tagged as “medication” account for around 60% of the annotations.



Evaluation Metrics

- Expert annotation considered as a gold standard for evaluation
- Data cleaning pipeline:
  - Records in XML format
  - Separated actual narrative text from the annotations
  - Imported annotations into a relational database
- Evaluation metrics: Recall, Precision, and F-score

id	start	end	text	tag
M0	1339	1346	ZESTRIL	MEDICATION
M3	1400	1407	LIPITOR	MEDICATION
M6	1272	1275	ASA	MEDICATION
M9	1174	1180	ATENOLOL	MEDICATION

Results & Discussion

20 entities has been selected for comparison

Entities annotated by experts	Frequency of occurrences	CLAMP			ACM		
	(sample size equals 1251)	Recall	Precision	F_score	Recall	Precision	F_score
Atenolol	211	1	0.91	0.95	1	0.93	0.96
Norvasc	60	0.80	1	0.89	1	0.90	0.95
Lipitor	185	1	0.99	0.99	1	0.84	0.91
Aspirin	195	0.99	1	0.99	1	0.94	0.97
Metoprolol	69	0.72	1	0.84	0.67	1	0.80
Glucophage	60	0.85	1	0.92	1	1	1
Toprol	36	0.50	1	0.67	0.50	1	0.67
Lisinopril	225	1	0.89	0.94	1	0.86	0.92
Pravachol	23	1	0.92	0.96	0.39	1	0.56
Zocor	34	0.82	1	0.9	1	0.83	0.91
Nifedipine	23	0.91	1	0.95	0.83	1	0.91
Zestril	53	0.96	1	0.98	1	0.81	0.89
Lovastatin	4	1	1	1	1	1	1
Pravastatin	34	0.82	1	0.90	1	0.92	0.96
Isosorbide	7	1	0.88	0.94	1	0.88	0.94
Labetolol	8	1	0.80	0.89	1	0.80	0.89
Zebeta	2	1	1	1	1	1	1
Coreg	7	0.86	1	0.92	0.86	1	0.92
Accupril	3	0.33	1	0.50	0.33	1	0.50
Glucotrol	12	1	1	1	0.67	1	0.80
Average		0.88	0.90	0.91	0.86	0.94	0.87

Results & Discussion

Entities annotated by experts	Frequency of occurrences	CLAMP			ACM		
	(sample size equals 1251)	Recall	Precision	F_score	Recall	Precision	F_score
Atenolol	211	1	0.91	0.95	1	0.93	0.96
Norvasc	60	0.80	1	0.89	1	0.90	0.95
Lipitor	185	1	0.99	0.99	1	0.84	0.91
Aspirin	195	0.99	1	0.99	1	0.94	0.97
Metoprolol	69	0.72	1	0.84	0.67	1	0.80
Glucophage	60	0.85	1	0.92	1	1	1
Toprol	36	0.50	1	0.67	0.50	1	0.67
Lisinopril	225	1	0.89	0.94	1	0.86	0.92
Pravachol	23	1	0.92	0.96	0.39	1	0.56
Zocor	34	0.82	1	0.9	1	0.83	0.91
Nifedipine	23	0.91	1	0.95	0.83	1	0.91
Zestril	53	0.96	1	0.98	1	0.81	0.89
Lovastatin	4	1	1	1	1	1	1
Pravastatin	34	0.82	1	0.90	1	0.92	0.96
Isosorbide	7	1	0.88	0.94	1	0.88	0.94
Labetolol	8	1	0.80	0.89	1	0.80	0.89
Zebeta	2	1	1	1	1	1	1
Coreg	7	0.86	1	0.92	0.86	1	0.92
Accupril	3	0.33	1	0.50	0.33	1	0.50
Glucotrol	12	1	1	1	0.67	1	0.80
Average		0.88	0.90	0.91	0.86	0.94	0.87

In comparison with ACM, CLAMP showed better performance by around 2% for the average recall and 4.6% for the average F-score.

Results & Discussion

Entities annotated by experts	Frequency of occurrences	CLAMP			ACM		
	(sample size equals 1251)	Recall	Precision	F_score	Recall	Precision	F_score
Atenolol	211	1	0.91	0.95	1	0.93	0.96
Aspirin	195	0.99	1	0.99	1	0.94	0.97
Lisinopril	225	1	0.89	0.94	1	0.86	0.92
Lovastatin	4	1	1	1	1	1	1
Zebeta	2	1	1	1	1	1	1
Accupril	3	0.33	1	0.50	0.33	1	0.50

- Amongst the three least frequent entities, both tools were able to perfectly identify two of them.
- For three most frequent entities, Average recall for CLAMP: 0.997, for ACM: 1.
- ACM performs better in identifying the most frequent entities.

Conclusion

- Need for automated entity extraction tools
- Two such tools: CLAMP and Amazon Comprehend Medical (one is general purpose)
- CLAMP showed better performance by around 2% for the average recall and 4.6% for the average F-score, in comparison with ACM.
- In the future: evaluate performance in extracting entities belonged to the other remaining categories.

Thank you