Entity Extraction for Clinical Notes, a Comparison Between MetaMap and Amazon Comprehend Medical

FATEMEH SHAH-MOHAMMADI, WANTING CUI, JOSEPH FINKELSTEIN

ICAHN SCHOOL OF MEDICINE AT MOUNT SINAI

NEW YORK, NY

Overview

- Why is Entity Extraction needed?
- Clinical Entity Extraction Tools:
  - MetaMap (MM)
  - Amazon Comprehend Medical (ACM)
- Dataset
- Evaluation Metrics
- Results
- Discussion
- Conclusion

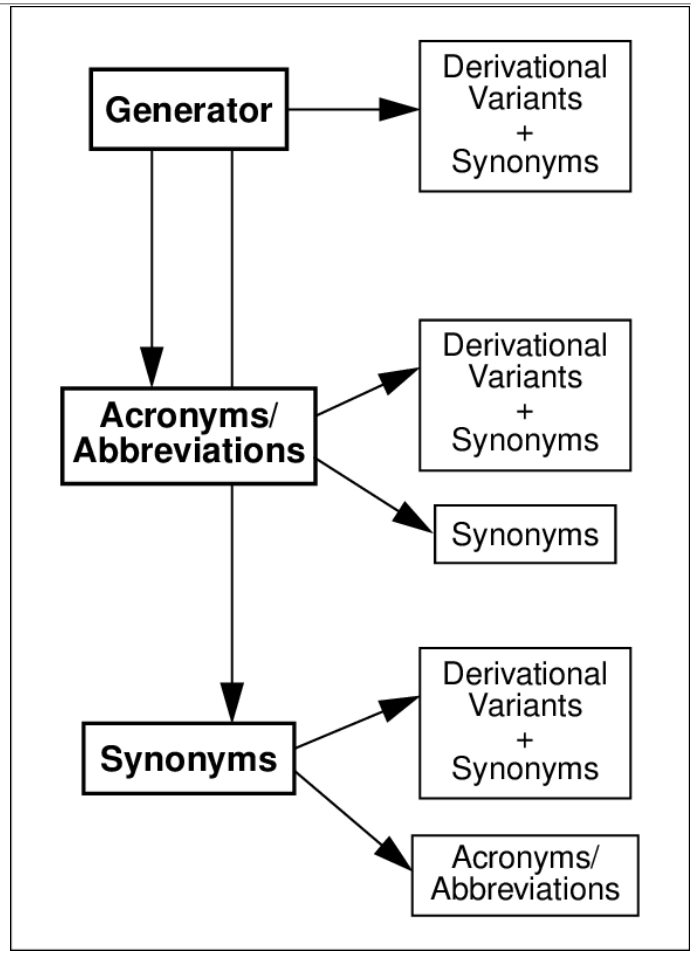
Why is Entity Extraction needed?

- 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: MetaMap (MM)

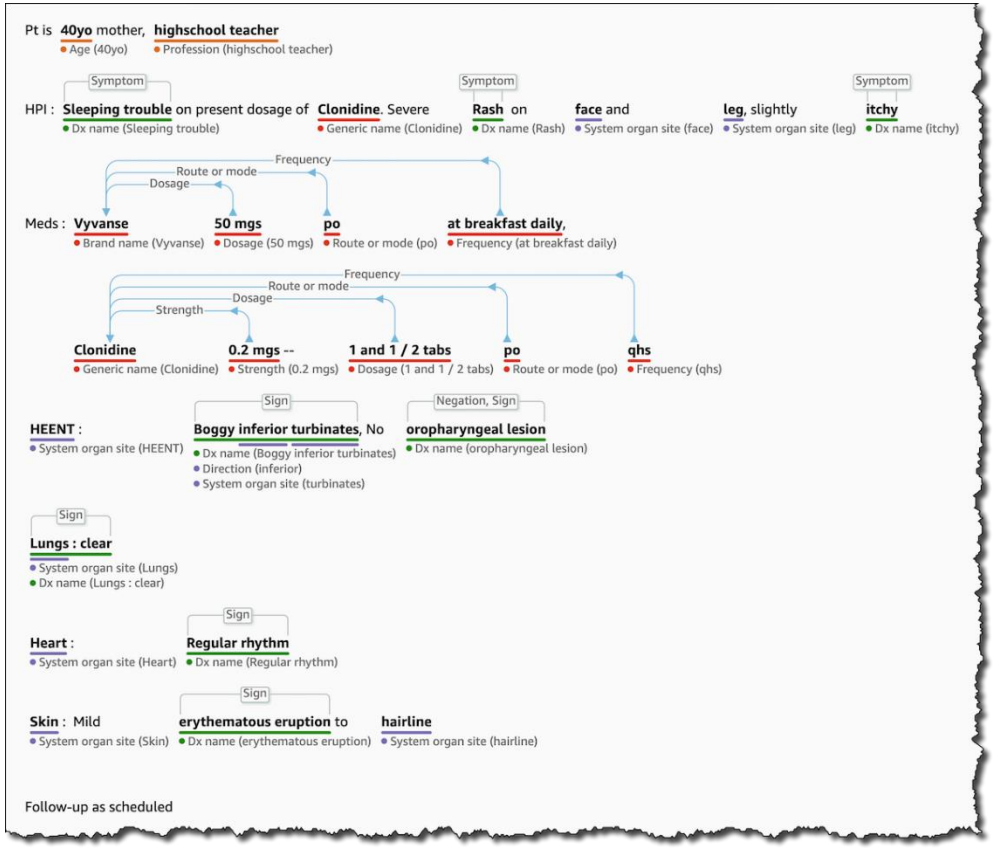
- A rule-based entity extraction tool
- Developed by National Library of Medicine (NLM)
- Maps biomedical texts to UMLS concepts
- Uses hybrid approach: NLP, computational linguistic techniques and knowledge-intensive approach



Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. InProceedings of the AMIA Symposium. 2001;17-21.
Aronson AR, Lang FM. An overview of MetaMap: historical perspective and recent advances. JAMIA. 2010;17(3):229-36.

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

Indicator
Hyperlipidemia
Dyslipidemia
Hypercholesterolemia
High Cholesterol

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	PLAVIX	MEDICATION

Results

30 entities has been selected for comparison

Entities annotated by			Evaluation			
experts and	MM			ACM		
frequency of occurrences	R	P	F	R	P	F
Hypertension (264)	1	0.74	0.85	1	0.93	0.96
Hypertensive (14)	0.29	1	0.44	1	0.68	0.76
htn (352)	1	0.78	0.88	1	0.8	0.89
Hyperlipidemia (166)	1	0.59	0.74	1	0.86	0.92
Dyslipidemia (24)	1	0.69	0.81	1	0.86	0.92
Hypercholesterolemia (3)	1	0.66	0.8	1	0.98	0.99
High Cholesterol (12)	1	0.67	0.8	1	0.92	0.96
Diabetes Mellitus (4)	0.75	1	0.86	1	1	1
Diabetic (17)	0.51	1	0.69	1	0.59	0.74
DM (268)	1	0.94	0.97	1	0.92	0.96
Insulin Dependent Diabetes Mellitus (1)	1	1	1	1	1	1
Non Insulin Dependent Diabetes Mellitus (1)	1	1	1	1	1	1

Results

	MM			ACM		
Obesity (70)	1	0.75	0.85	1	0.96	0.98
Morbid Obesity (13)	1	0.75	0.87	1	0.69	0.81
Coronary Artery Disease (104)	1	0.71	0.83	1	0.89	0.94
Coronary Artery Bypass Surgery (7)	0.72	1	0.83	0.57	1	0.73
Myocardial Infarction (41)	1	0.8	0.89	1	0.76	0.86
MI (68)	0.55	1	0.71	1	0.68	0.81
Chest Pressure (7)	1	1	1	1	0.47	0.63
Zestril (56)	1	0.53	0.76	1	0.81	0.9
Lipitor (201)	1	0.64	0.78	1	0.91	0.95
Verapamil (19)	1	0.79	0.88	1	1	1
Beta-Blocker (26)	0.39	1	0.56	0.77	1	0.87
AVERAGE	0.88	0.83	0.82	0.97	0.86	0.90

ACM resulted in better performance in comparison with MM with 10% higher average recall, 4% higher average precision, and 10% higher average F-score.

Discussion

- Poor recall performance of MM: stems from its inability in identifying multi word phrases as concepts, unless exact matches can be found in the dictionary.
- ACM is a neural network-based tool, its training dataset included a wider range of vocabularies.

Tag name	Entities annotated by experts	MM	ACM		
	and frequency of occurrences		R	P	F
Hyperlipidemia	High Chol (1)	nan	1	1	1
	~~Increased~~ Cholesterol (1)	nan	1	1	1
Diabetes	Insulin Dependent Diabetes (1)	nan	nan	nan	nan
	Insulindependent Diabetes (5)	nan	nan	nan	nan
	Insulin Requiring Diabetes (1)	nan	nan	nan	nan
Obese	Morbidly Obese (7)	nan	1	1	1
	Severely Obese (2)	nan	nan	nan	nan

Conclusion

- Need for automated entity extraction tools
- Two such tools: MetaMap and Amazon Comprehend Medical (with different computational capability)
- ACM resulted in better performance in comparison with MM with 10% higher average recall, 4% higher average precision, and 10% higher average F-score.
- ACM is a neural network-based tool, its training dataset included a wider range of vocabularies.
- Future use: Amazon Comprehend Medical

Thank you