S58: Clinical Word Sense Disambiguation with Interactive Search and Classification

Yue Wang[1], Kai Zheng[2], Hua Xu[3], Qiaozhu Mei[1]

University of Michigan[1]
University of California, Irvine[2]
University of Texas Health Science Center at Houston[3]

Disclosure

- Funding support

  NLM 2R01LM010681-05

  NSF IIS-1054199

- No relevant financial relationships with commercial interest

- A fan of machine learning, NLP, and applications in medical domain!

Word sense disambiguation (WSD)

- Words have multiple meanings (senses)
  - Ice: (1) frozen water; (2) methamphetamine; (3) caspase-1
  - PD: (1) police department; (2) peritoneal dialysis; (3) posterior descending …
- Given ambiguous word in context, assign a sense
  - A classification task: $f_{word}(context) = sense$
- Critical step for many clinical NLP applications
  - Text indexing & categorization
  - Named entity extraction
  - Computer-assisted review

Current state of the art

- Approaches
  - Unsupervised learning (Liu'01,Xu'12)
  - Supervised learning (Liu'04, Xu'06,Wu'15)
  - Semi-supervised learning (Liu'02, Yu'06)
  - Active learning (Chen'13)

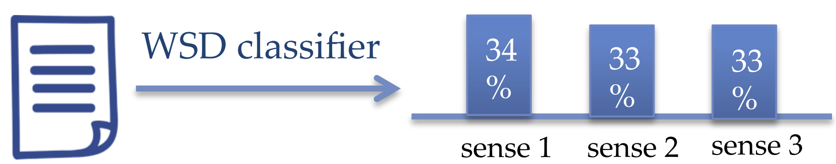
- Key to better performance:
  more high-quality labeled data

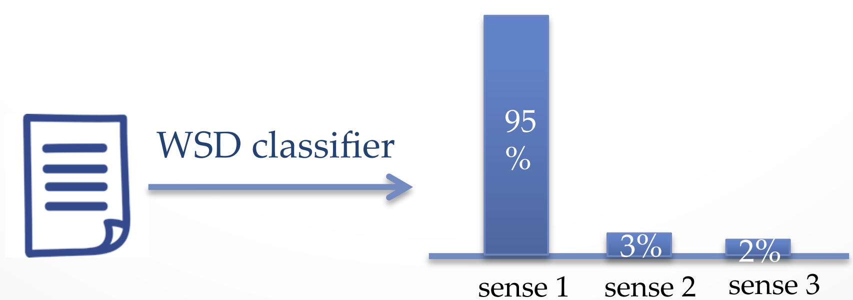
Our goal: Train an accurate WSD model with as few labels as possible

Asking for labels

- Random sampling

- Active learning (Settles'08, Chen'13)
  - Machine asks instance with biggest uncertainty





Expert knowledge > labels

- Without looking at examples, experts already know typical contexts for a sense
  - AB (blood type): “blood”, “positive”, “negative”, …
  - AB (influenza type): “influenza”, “flu”, “seasonal”, …

- Algorithm may not ask the right question
  - Especially at the very beginning: labels are few

Invite experts to search!

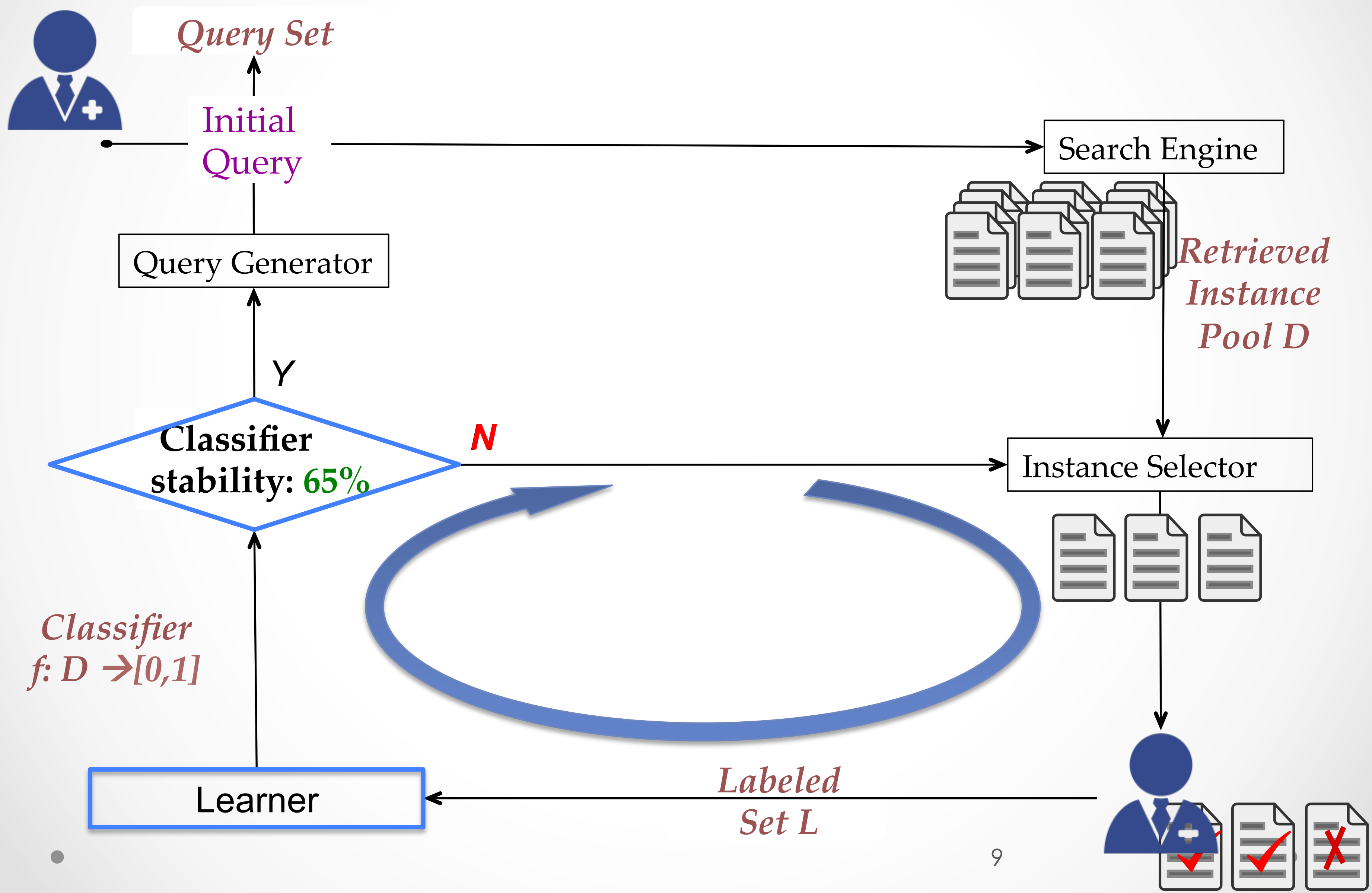
blood type AB

Go

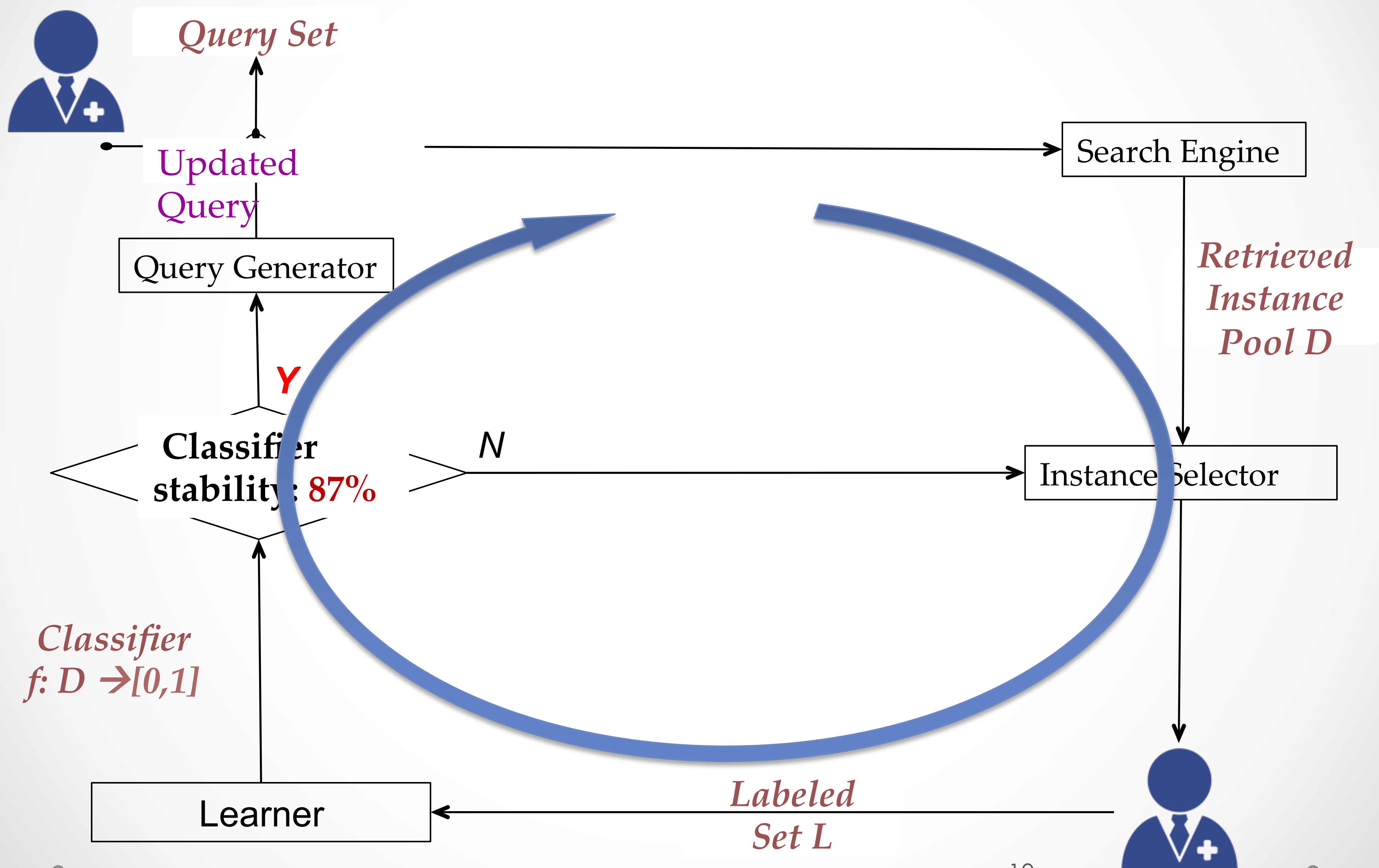
… The patient is **blood type AB** positive …
… maternal **blood type** was **AB** negative …
… She is **blood** group **AB** positive …

- Include context words in queries
- Kick off classifier training with typical examples
- Efficient when working with big data sets
- Increase the chance of seeing rare senses

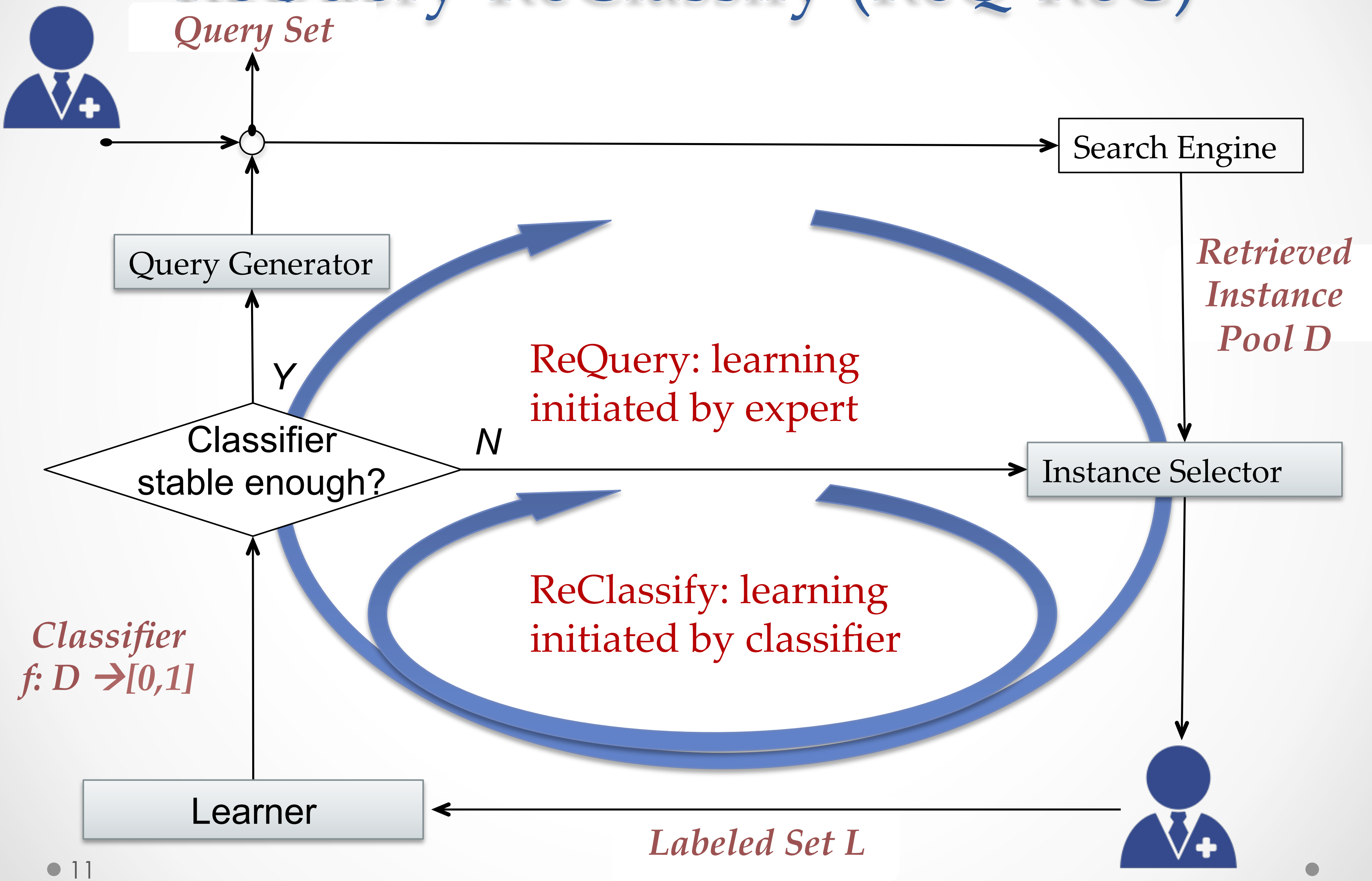
Contents



Query Set
Initial Query
Search Engine
Query Generator
Retrieved Instance Pool D
Y
Classifier stability: 65%
N
Instance Selector
Classifier f: D →[0,1]
Labeled Set L
Learner


Query Set
Updated Query
Search Engine
Query Generator
Retrieved Instance Pool D
Y
Classifier stability: 87%
N
Instance Selector
Classifier
f: D →[0,1]
Learner
Labeled Set L

ReQuery-ReClassify (ReQ-ReC)



Implementing ReQ-ReC

- ReQuery loop
  - Search engine: Apache Lucene fulltext indexer
  - Query generator: variant of Rocchio's method in information retrieval
  - Expert: approve, edit, or compose queries

- ReClassify loop
  - Classifier learner: one-vs.-rest logistic regression model
  - Document selector: margin active learning
  - Expert: provide labels

Contents

- Background & Motivation
- A new interactive learning framework
- Evaluation
  - Data Sets
  - Methods
  - Performance metric
  - Results
- Conclusion

Evaluation datasets

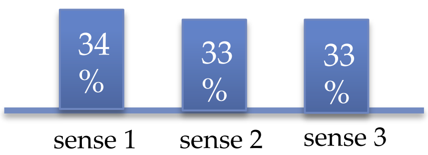
- MSH: MEDLINE abstracts
- VUH: clinical notes, Vanderbilt U Hospital
- UMN: clinical notes, U of Minnesota Fairview Hospital

	#ambiguous words	#sense/ word	#contexts/ word	#tokens/ context	majority guess accuracy
MSH	198	2.1	190	203	54.0%
VUH	24	4.3	194	19	78.3%
UMN	75	5.4	500	61	73.8%

Learning methods

- Random sampling

- Active learning methods
  - **Least confident**
  - **Margin**
  - **Label entropy**

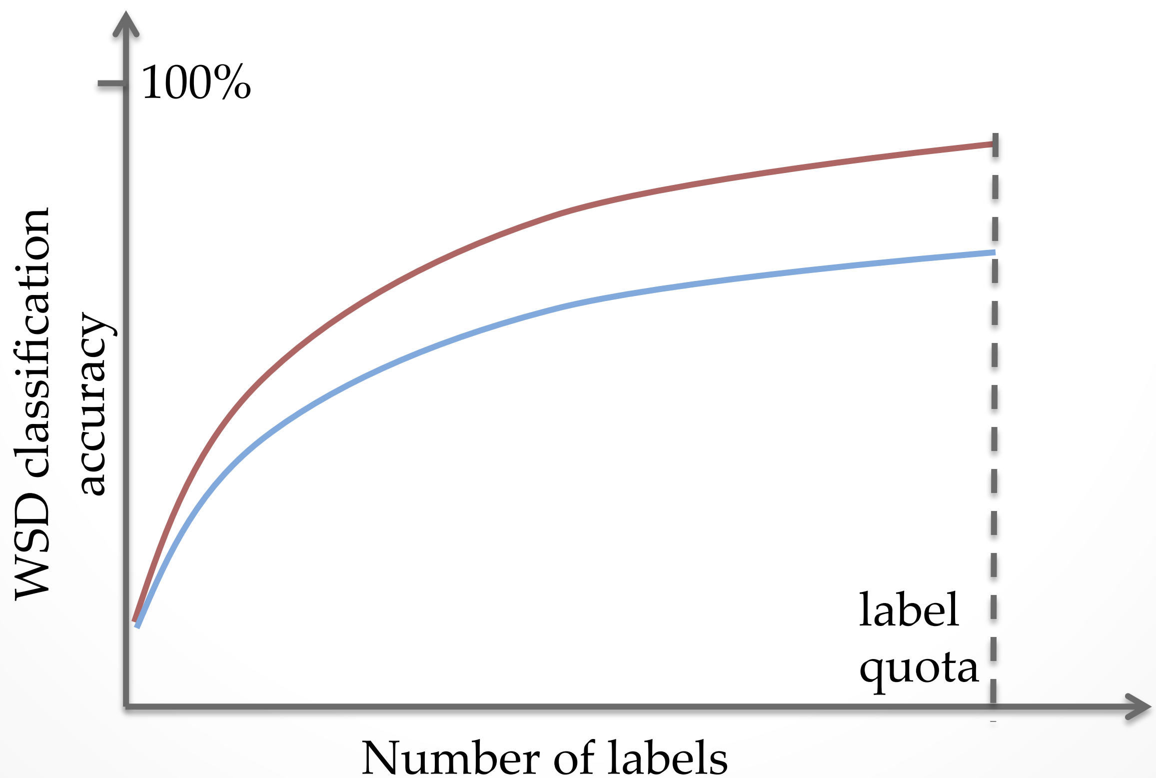


- ReQ-ReC
  - **Worst-case**: always accept machine-generated query

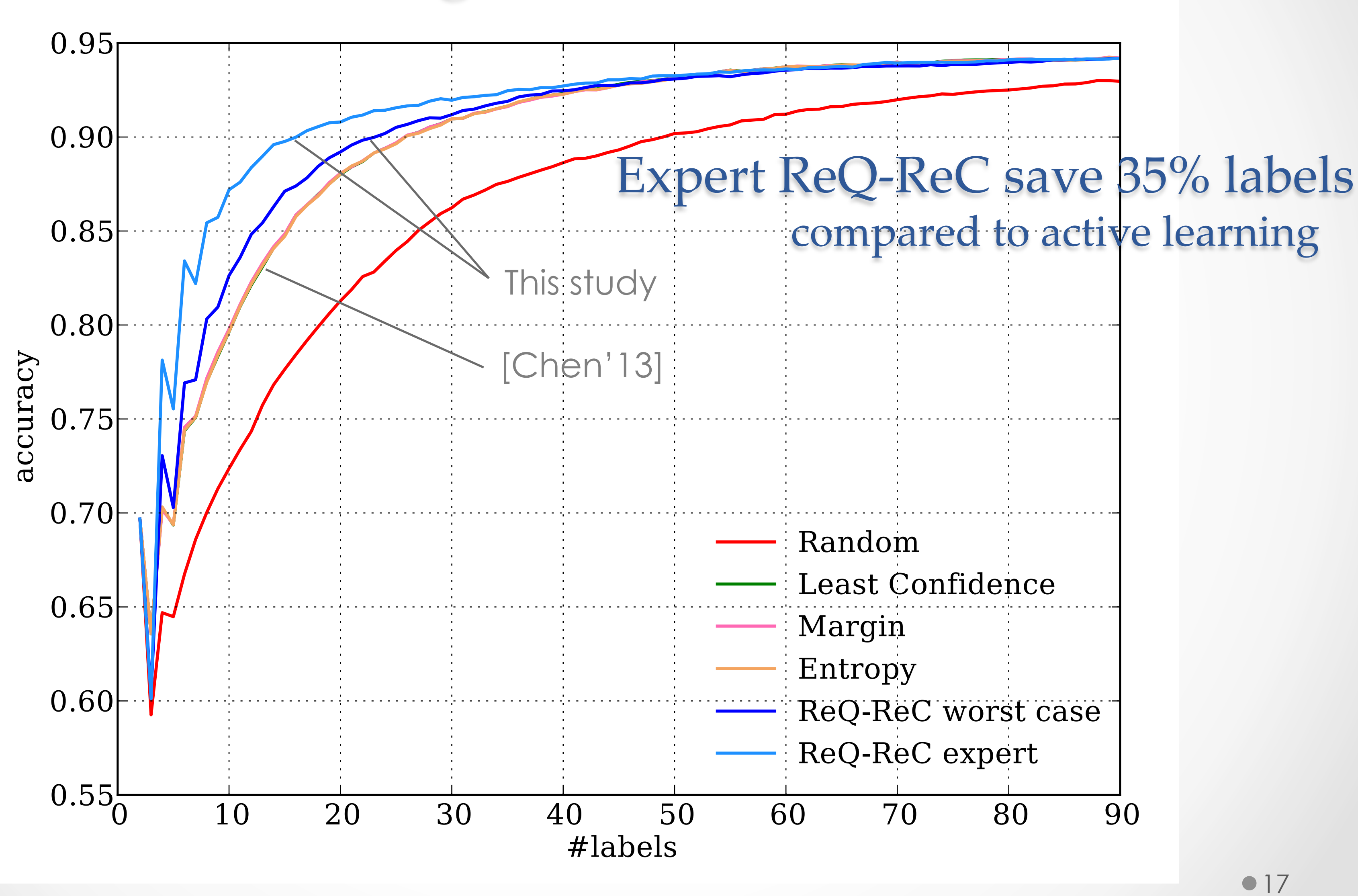
  - **Simulated Expert**: query = informative words (measured by info. gain)

Evaluation metric

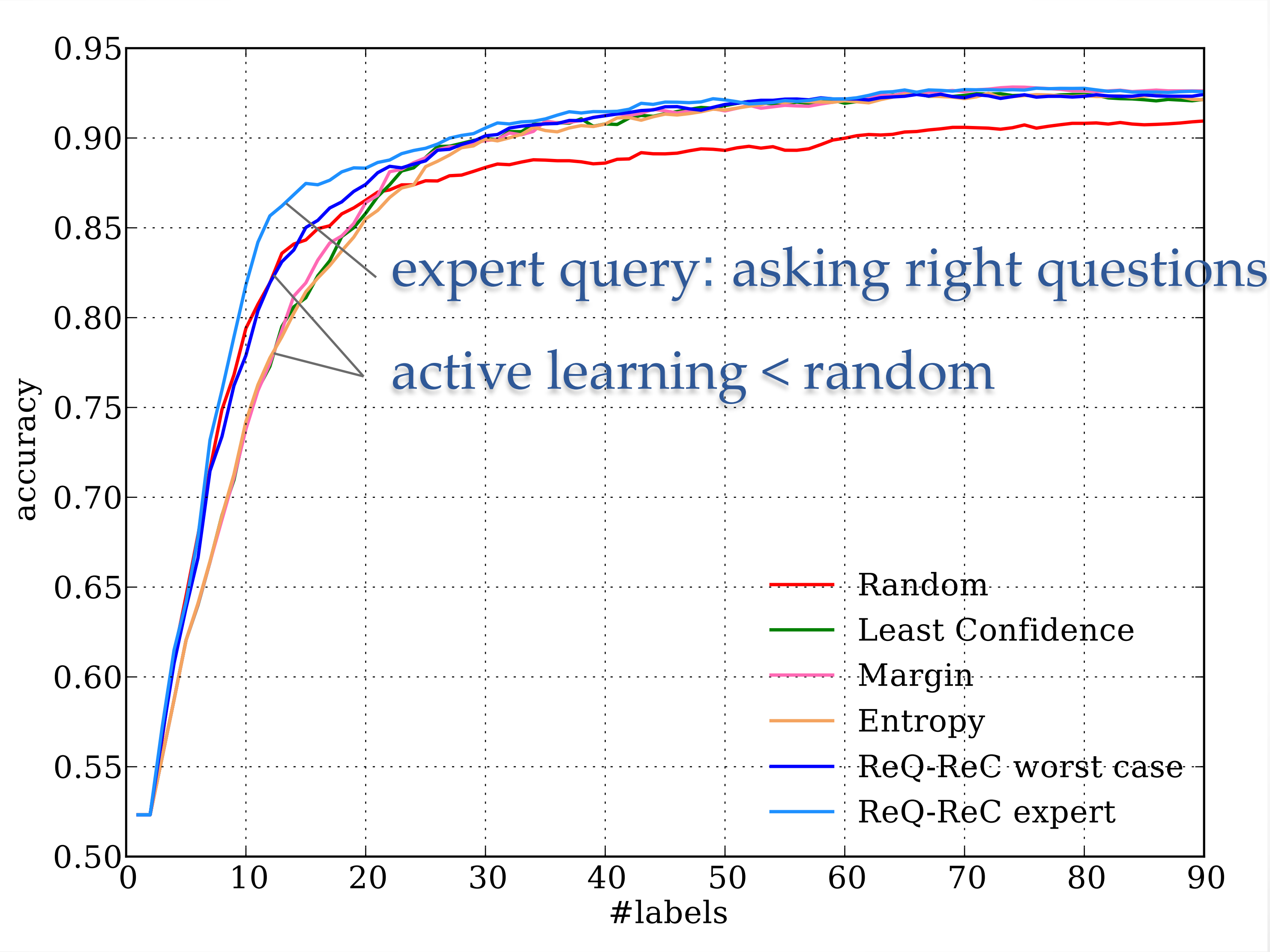
- Learning curve
- Area under learning curve (ALC)



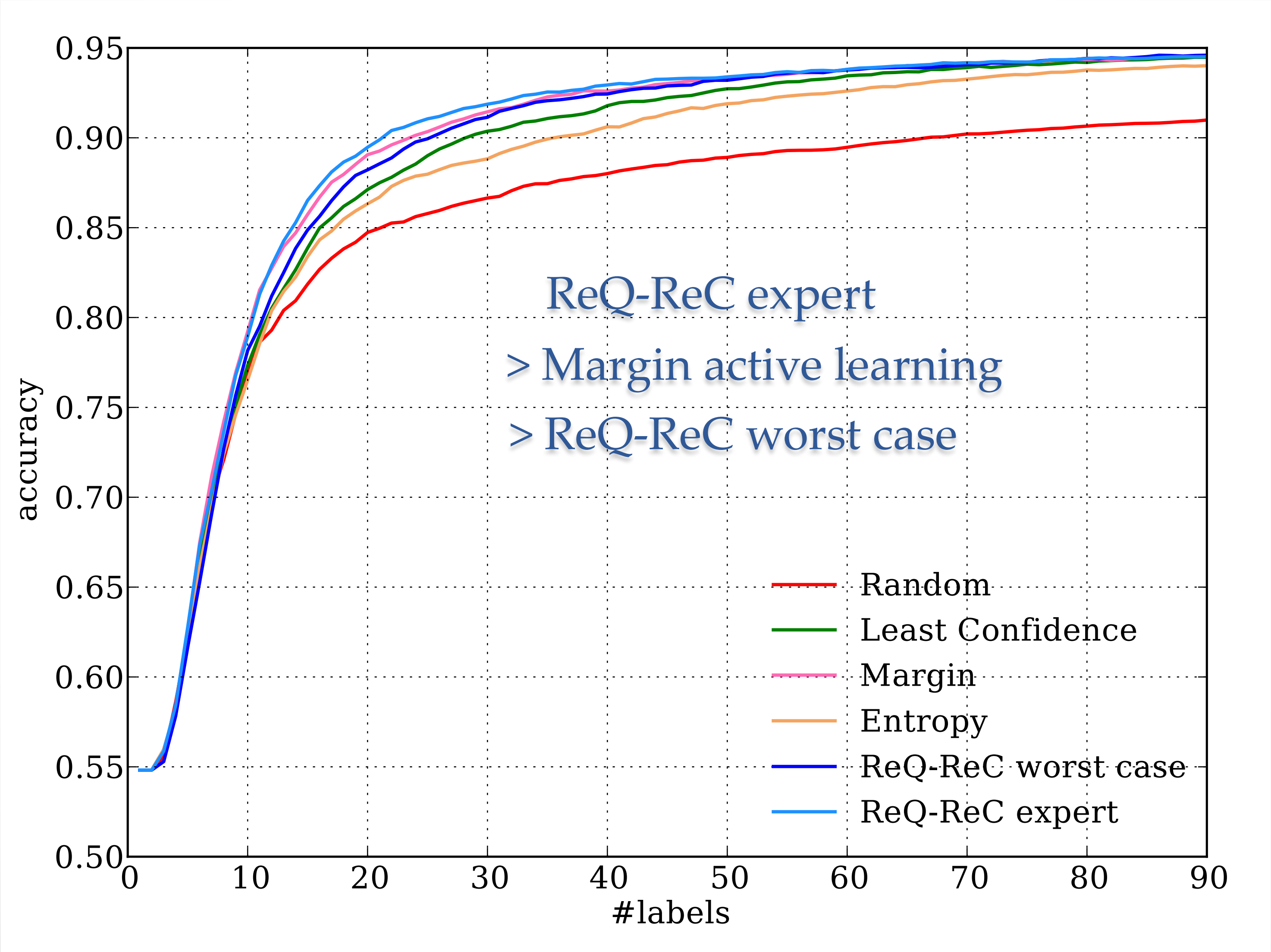
MSH: 198 ambiguous words in MEDLINE



VUH: 24 clinical abbreviations

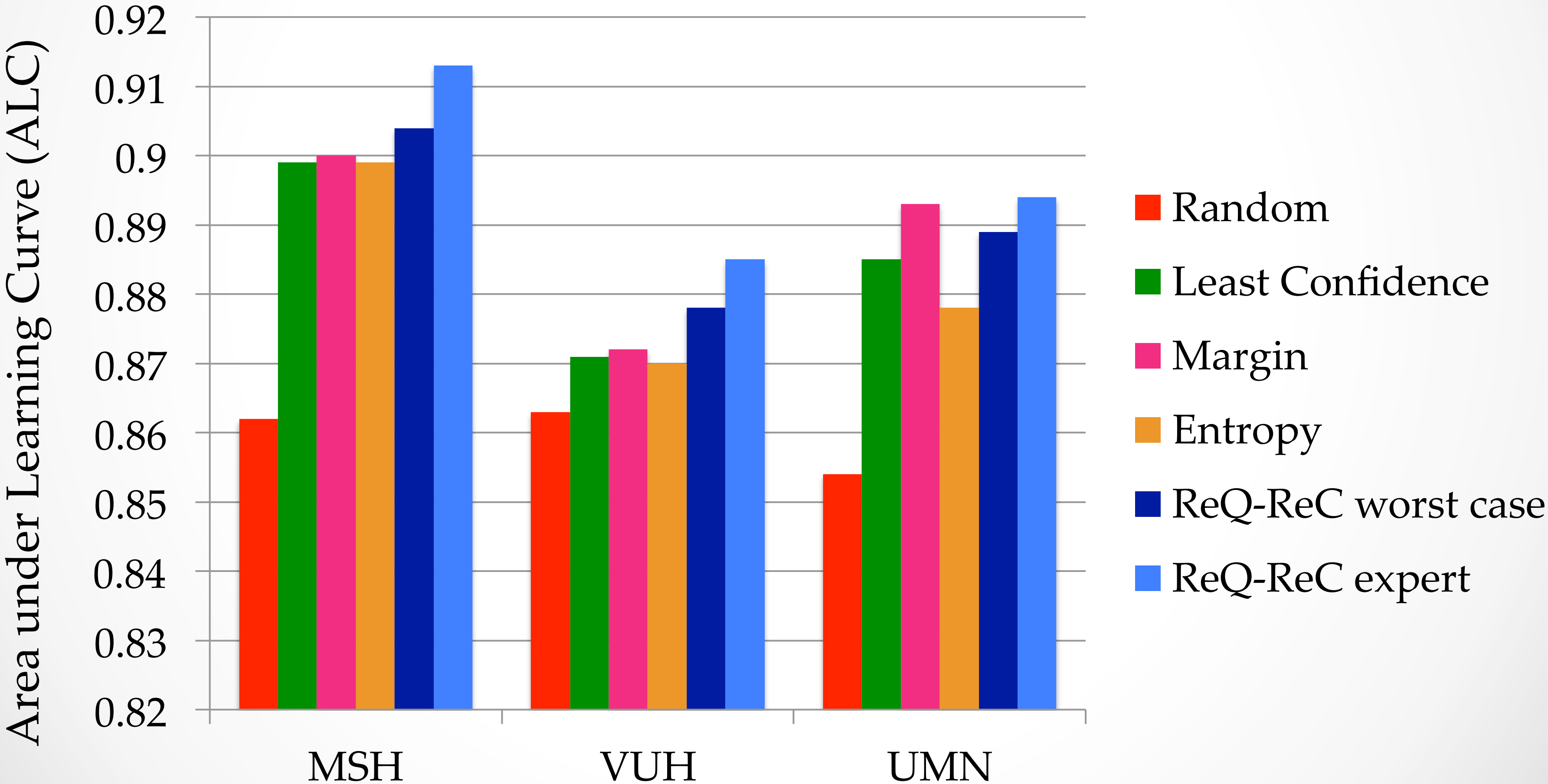


UMN: 75 clinical abbreviations



ReQ-ReC performs best with expert queries

Higher: better



Contents

- Background & Motivation
- A new interactive learning framework
- Evaluation
- Conclusion

Conclusion

- An interactive learning framework for building clinical WSD models
  - ReQuery: domain expert-guided learning
  - ReClassify: statistical model-guided learning

- Future work: evaluate the framework with human experts

Thank you!

raywang@umich.edu

References

- Liu H, Lussier YA, Friedman C. Disambiguating ambiguous biomedical terms in biomedical narrative text: an unsupervised method. Journal of biomedical informatics. 2001 Aug 31;34(4):249-61.
- Chen Y, Cao H, Mei Q, Zheng K, Xu H. Applying active learning to supervised word sense disambiguation in MEDLINE. J Am Med Inform Assoc. 2013 Sep;20(5):1001–6.
- Liu H, Teller V, Friedman C. A multi-aspect comparison study of supervised word sense disambiguation. J Am Med Inform Associ 2004;11:320–31.
- Xu H, Markatou M, Dimova R, Liu H, Friedman C. Machine learning and word sense disambiguation in the biomedical domain: design and evaluation issues. BMC bioinformatics. 2006 Jul 5;7(1):1.
- Liu H, Johnson SB, Friedman C. Automatic resolution of ambiguous terms based on machine learning and conceptual relations in the UMLS. J Am Med Inform Assoc 2002;9:621–36.
- Yu H, Kim W, Hatzivassiloglou V, et al. Using MEDLINE as a knowledge source for disambiguating abbreviations and acronyms in full-text biomedical journal articles. J Biomed Inform 2007;40:150–9.
- Wu Y, Xu J, Zhang Y, Xu H. Clinical abbreviation disambiguation using neural word embeddings. ACL-IJCNLP 2015. 2015 Jul 30:171.
- Xu H, Stetson PD, Friedman C. Combining corpus-derived sense profiles with estimated frequency information to disambiguate clinical abbreviations. In AMIA 2012.