MANCHESTER 1824
The University of Manchester

NaCTeM
The National Centre for Text Mining

Improving the Extraction of Clinical Concepts from Clinical Records

Xiao Fu and Sophia Ananiadou
National Centre for Text Mining
School of Computer Science, University of Manchester

fux@cs.man.ac.uk
sophia.ananiadou@manchester.as.uk

Aims

To develop a **machine learning**-based named entity recognition system to extract **clinical concepts** from electronic medical records **without** the need for any external knowledge resources.

Methods

- **Baseline**
  - NERsuite[1] (the part-of-speech tags, lemmas, and chunk tags)
- **Pre-processing**
  - Truecasing[2]
  - Abbreviation Disambiguation[3]
- **Post-processing**
  - Distributional Similarity
- **Hybrid**
  - Two Combination Schemes (i.e., Sequential and Parallel Schemes)

[1]: http://nersuite.nlplab.org/
[2]: http://argo.nactem.ac.uk/
[3]: http://www.nactem.ac.uk/software/acromine_disambiguation/

Results

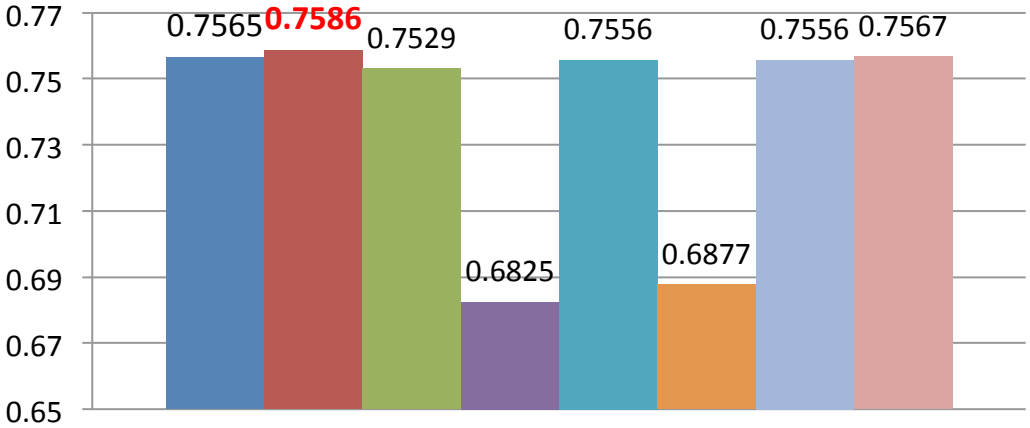

Chart 1. F-scores for Exact Matching

T, truecasing; AD: abbreviation disambiguation;
DS: distributional similarity;
S, sequential scheme; P, parallel scheme.

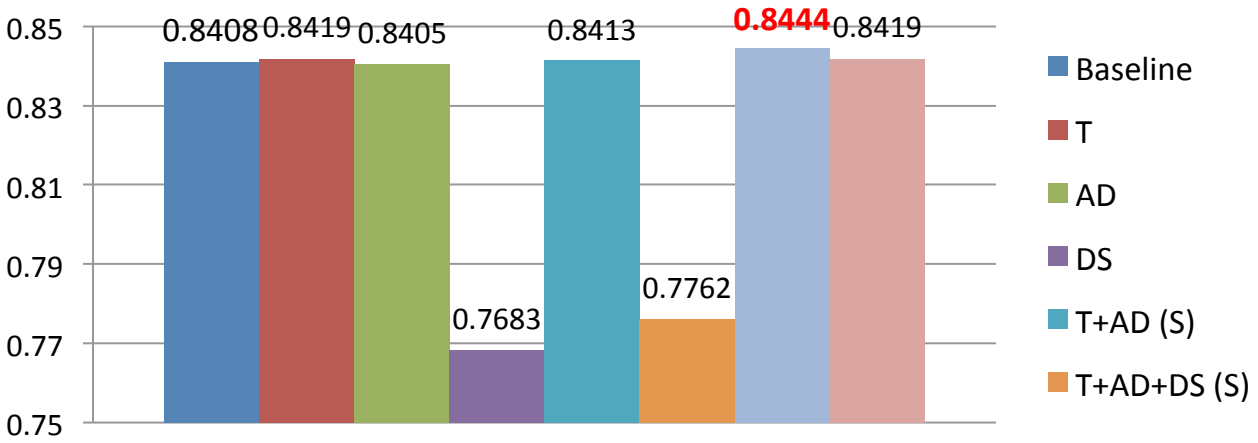

Chart 2. F-scores for Inexact Matching