A.I. in health informatics **lecture 13** reviewing the i2b2 2010 task *and* midterm review

kevin small &
byron wallace

today

- i2b2 2010: what's the state of the science in clinical information extraction?

- midterm review

i2b2 2010: the task

- three aims
  - extraction of medical problems, tests & treatments from discharge summaries and progress notes
  - classification of assertions made on the medical problems
  - classification of *relations* between medical concepts

I2b2 2010: the data

- discharge summaries and progress reports
- participants given 349 annotated, 827 unlabeled
- hold-out (test) set of 477

I2b2 2010

- from free text,
  - (1) find the medical problem,
  - (2) whether it is e.g. ‘present’, ‘absent’, ‘possible’, or ‘hypothetical’ …
  - (3) identify relationships between medical concepts from a closed set (‘treatment is given for the problem’)

features

- token features: ‘Chem-7’ -> ‘Aa-0’; and character *n*-grams
- context features (word bi/tri/quad-grams), wild-cards
- sentence features (grammatical, etc)
- section-level features (e.g., headings)
- document-level features (e.g., doc length)

even more features

- UMLS, cTAKES, ConText
- Brown clustering algorithm
- various home-brew reg-expressions

features, features, features

importantly, our choice of machine learning algorithms allowed us to be versatile in our feature design, and to introduce a large number of features…

task 1

find the medical problem

- approach: semi-markov model
  - semi-markov models allow *sequence* emissions from a given state
- online training (perceptron-style)

task 2

classify assertions

- after step 2 (given that it's a concept), classify assertion
  - as often in practice, use an ensemble
- some hand-tuning

task 3

identify relations

- input: pair of concepts – classify their relationship
- three separate classifiers: treatment-problem; test-problem; problem-problem
  - under-sampling to handle imbalance

results

Table 1 Test set performance for the three tasks

	True positive	False negative	False positive	Recall	Precision	F-score
Task 1: Concepts Task						
System 1.1	37 646	7363	5683	0.8364	0.8688	**0.8523**
System 1.2	36 776	8233	6125	0.8170	0.8572	0.8366
System 1.3	37 663	7346	5787	0.8367	0.8668	0.8515
Task 2: Assertions Task						
System 2.1	17 366	1184	1184	0.9362	0.9362	**0.9362**
System 2.2	17 338	1212	1212	0.9347	0.9347	0.9347
System 2.3	17 197	1353	1353	0.9271	0.9271	0.9271
Task 3: Relations Task						
System 3.1	6296	2809	1965	0.6902	0.7611	0.7239
System 3.2	6269	2801	1896	0.6911	0.7677	0.7274
System 3.3	6288	2782	1838	0.6932	0.7738	**0.7313**

features!

Table 4 Performance for feature accumulations in the Relations Task

Feature set	Recall	Precision	F-score
(a) Baseline	0.646	0.718	0.680
(b) +order/type-sensitive	0.672	0.731	0.700
(c) +rich word features	0.681	0.753	0.715
(d) +domain knowledge	0.694	0.750	0.721
(e) +syntax	0.694	0.763	0.727
(f) +unannotated data	0.693	0.773	0.731

the mid-term

- take-home: you'll have about a week
- likely you'll get it tomorrow
- mostly written; some programming
- please do not collaborate

mid-term topics

- ✓ clinical reasoning
  basic probability, decision theory, &etc
- ✓ computational approaches to above
  symbolic logic, maximum expected value, &etc
- ✓ ontologies
  what they are and why they are useful

mid-term topics

- ✓ machine learning
  classification algorithms, feature representations, structured models
- ✓ nlp
  overlaps with above; structured models, applications to biomedical scenarios
- ✓ information retrieval

questions?

- on either logistics or material?