CliniQA : Highly Reliable Clinical Question Answering System

Ni Yuan, Zhu Hui Jia, Cai Peng, Cao Feng, Zhang Lei, Qiu Zhao Ming

Problem, Goal

- From IT perspective, one of the bottleneck of EBM is the difficulty for physician's to effectively obtain relevant medical research results for a specific clinical question in hand.
- Goal: build a system for physicians that can address clinical questions with highly reliable answers from latest state-of-the-art medical research results. Example:

Q: Is there any evidence for the use of Digoxin to reduce the mortality in patients with heart failure?

A: An RCT from The New England Journal of Medicine concluded that "Digoxin did not reduce overall mortality, but it reduced the rate of hospitalization both overall and for worsening heart failure".

Trial type	randomized, double-blinded, controlled
Followup	37 months
Problem	Patients with heart failure
Intervention	Digoxin
Comparison	Placebo

(a)

Characteristic	Digoxin (N=3397)	Placebo (N=3403)
Age (yr) — mean ±SD	63.4±11.0	63.5±10.8
Ejection fraction — mean ±SD	28.6±8.9	28.4±8.9
Median duration of CHF — mo	17	16
	% of patients	
Female sex	22.2	22.5
Nonwhite race	14.4	14.8
Age >70 yr	26.7	27.4
Method of assessing ejection fraction		
Radionuclide ventriculography	65.0	64.2
Two-dimensional echocardiography	29.5	30.0
Contrast angiography	5.5	5.8
Cardiothoracic ratio >0.55	34.6	34.4
NYHA class		
I	13.7	13.0
II	53.3	54.5
III	30.7	30.5
IV	2.2	1.9

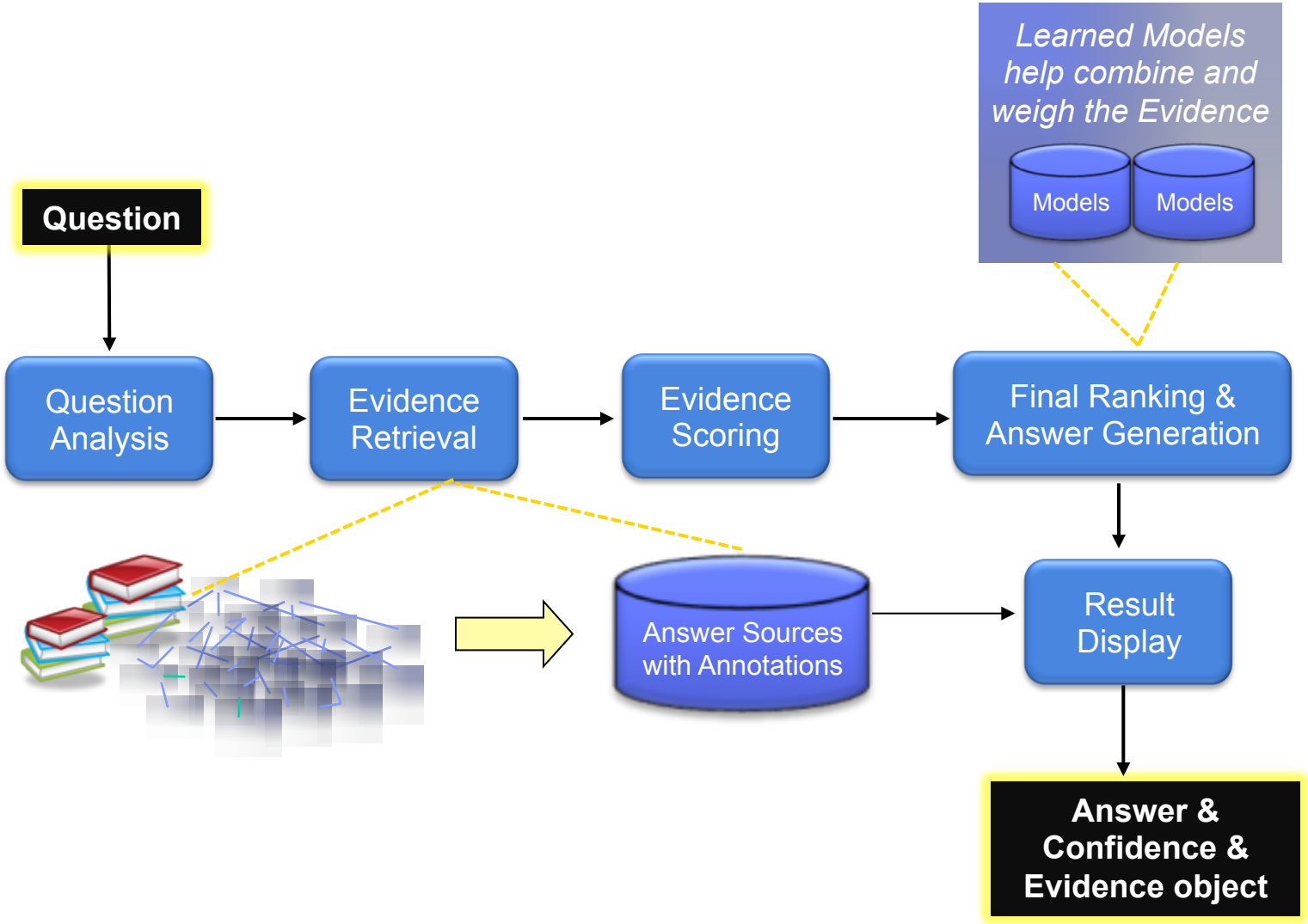
(b)

Table 2. Deaths According to Study Group and Cause.

Cause of Death	Digoxin (N=3397)	Placebo (N=3403)	Absolute Difference*	Risk Ratio (95% CI)†	P Value
	no. of patients (%)		%		
All	1181 (34.8)	1194 (35.1)	−0.4	0.99 (0.91–1.07)	0.80
Cardiovascular	1016 (29.9)	1004 (29.5)	0.4	1.01 (0.93–1.10)	0.78
Worsening heart failure‡	394 (11.6)	449 (13.2)	−1.6	0.88 (0.77–1.01)	0.06
Other cardiac§	508 (15.0)	444 (13.0)	1.9	1.14 (1.01–1.30)	
Other vascular¶	50 (1.5)	45 (1.3)	0.1	1.11 (0.74–1.66)	
Unknown	64 (1.9)	66 (1.9)	−0.1	0.97 (0.69–1.37)	
Noncardiac and nonvascular	165 (4.9)	190 (5.6)	−0.7	0.87 (0.71–1.07)	

(c)

System Architecture – Reuse DeepQA's (Watson) Architecture

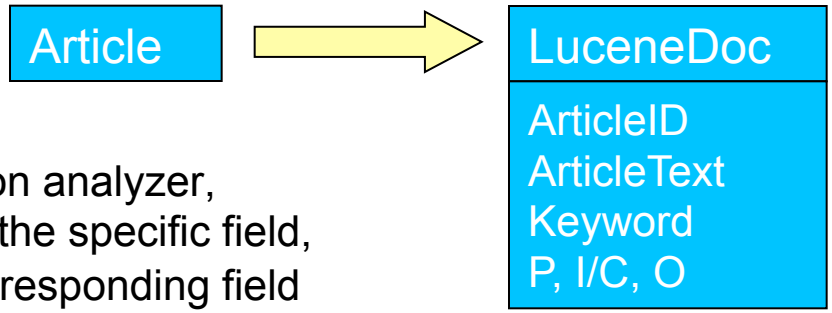


Component Technologies – Question Analyzer

- A set of templates for therapy questions
- The templates are defined in terms of the PICO framework, i.e. Problem (P), Intervention (I), Comparison (C), Outcome (O)
  - <I?, P> asks for the intervention for a given problem, e.g. “how should I treat polymenorrhea in a 14-year-old girl?”
  - ?<I, P> asks for the evidences or the effectiveness of using the Intervention on the Problem, e.g. “what is the evidence for using Metformin in people with type 1 diabetes who are obese and poorly controlled?”
  - <I?> asks for the usage of some intervention, e.g. “Is melatonin good for anything? I do not know anything about melatonin. I need to know the dose”
- Use MetaMap to identify the medical concepts from the questions
- Map the concepts into PICO categories in terms of the types
- Use the SemRap to detect whether a TREAT relationship exists

Component Technologies – Evidence Preprocessor

- Semantic analysis to generate evidence object
  - Identify P, I, C, O elements
    - make use of the linguistic and structure characteristics of the text, for Intervention
    - if the title satisfies the pattern “[XXX] of [YYY] on/for/in [ZZZ]”
    - if there exists a sentence in the objective part of the abstract section, and the sentence satisfies the pattern “the evaluation/effectiveness/impact/effect(s) of [YYY]
    - the most frequently appeared medical concept of semantic type [TREATMENT&DRUG] is also considered as the candidates
  - Identify randomization & followup of the clinical trials
  - Identify the useful tables
- Indexing & Searching
  - Use Lucene to build the indexes
  - Given the search query from the question analyzer, if the question has the requirements on the specific field, the retrieval will be conducted at the corresponding field



Component Technologies – Candidate Scorers

- Bleu Scorer is used to measure the similarity between the question and the matched passages
- Match Coverage Scorers is to measure the percentage of matched concepts in the question
- Patient Gender Scorers is to measure the similarity between the patient genders mentioned in the questions and the patient gender mentioned in the evidences
- Patient Age Scorers is to measure the similarity between the patient ages mentioned in the questions and the patient ages mentioned in the evidences
- Disorder Similarity Scorers is to measure the similarity between the disorders mentioned in the questions and the disorders mentioned in the evidences
- Drug Similarity Scorers is to measure the similarity between the drugs mentioned in the questions and the drugs mentioned in the evidences
- 1-gram, 2-gram, 3-gram
- Search rank

Component Technologies – Answer Generator

- Use machine learning to integrate all scores to get the final ranking

Candidate Evidences	Evidence Feature Scores						
	Blue	Search Rank	N-gram	PatientGen	Patient Age	SimDisorder	SimDrug
$Evidence_1$	0.5	1	0.1	0	0.2	0.4	1
$Evidence_2$	0.4	2	0.1	0	0.4	0.3	1
$Evidence_3$	0.7	3	0.8	1	0.5	0.6	0.9
… …							

Apply the machine learning models

Final Ranking	Confi-dence
$Evidence_{k1}$	0.946
$Evidence_{k2}$	0.152
$Evidence_{k3}$	0.040
… …	

- Extract the Conclusion part of the evidence as the answers

Demo

CliniQA

Ask other questions

Anwser

Your Question:

is Etoricoxib effect for osteoarthritis

The Answer:

Title: Etoricoxib in the treatment of osteoarthritis over 52-weeks: a double-blind, active-comparator controlled trial [NCT00242489] (E411)

conclusion: In this extension study, etoricoxib, at doses ranging from 30 to 90 mg, demonstrated a maintenance of significant clinical efficacy in patients with OA through 52 weeks of treatment. Etoricoxib displayed clinical efficacy similar to diclofenac 150 mg and was generally well tolerated

More

Patients

Age	Gender	Disease	Tables
40 years、62 years	women	osteoarthritis	Baseline Patient Characteristics at Randomization (Patients Entering the Active Comparator Controlled Period - Weeks 6-52) Rates of Adverse Experiences (AEs) of Special Interest per 100 Patient-YearsAll Patients Over The Entire 52 Weeks From The Start of Therapy to the End of the Period

Inteventions

Intevention	Trial	Follow-up	Tables
etoricoxib	RCT	The base study was 14 weeks in duration and consisted of 2 parts; in Part I (6 weeks), patients were allocated to once daily oral etoricoxib 5, 10, 30, 60, 90 mg or placebo. In Part II (8 weeks); the placebo, etoricoxib 5 and 10 mg groups were reallocated to etoricoxib 30, 60, or 90 mg qd or diclofenac 50 mg t.i.d. Treatment was continued for consecutive 12 and 26 week extensions.	

Outcome

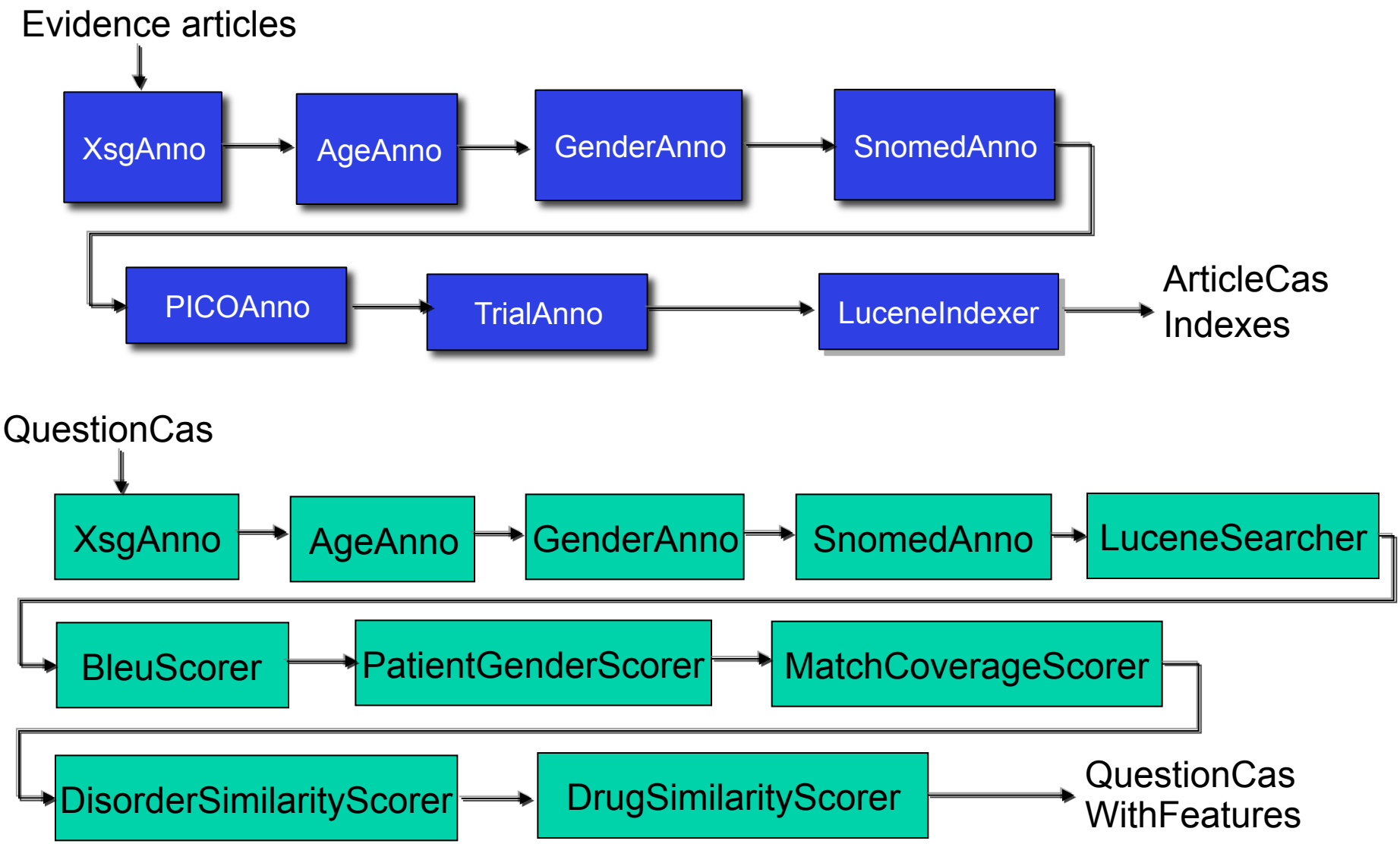
Aspect	Porlarity

IBM.CRL

Prototype Implementation – UIMA Framework

- The UIMA framework is used in the system
- The data pass between components are encapsulate as a CAS (Common Analysis Structure)
- Each component extends the CasAnnotator in UIMA
- Different components are arranged as a pipeline

UIMA Pipeline (1)



Experiment Dataset Preparation (1)

- TripAnswers dataset is leveraged
  - 6265 questions with human generated answers
- Example
  - Question : Are oral B12 supplements in high doses equal effective as B12 intramuscular injections?
  - Answer : *In 2005 Cochrane published the systematic review 'Oral vitamin B12 versus intramuscular vitamin B12 for vitamin B12 deficiency' [1]. This concluded:*
    "The evidence derived from these limited studies suggests that 2000 mcg doses of oral vitamin B12 daily and 1000 mcg doses initially daily and thereafter weekly and then monthly may be as effective as intramuscular administration in obtaining short term haematological and neurological responses in vitamin B12 deficient patients."
    In the full-text of the review, they report:
    "Vitamin B12 replacement has been traditionally administered intramuscularly. However several case control and case series studies have since suggested equal efficacy and safety of the oral route (Chalmers 1958; Ross 1954; Spies 1949). The mechanism for this oral route is most probably that free vitamin B12 can be absorbed both passively (without binding to IF) as well as actively (following binding to IF) in the terminal ileum. Passive diffusion accounts for 1.2% of total absorption with a bioavailability unaffected in patients with pernicious anaemia or gastro-duodenal surgical resection (Berlin 1968; Berlin 1978). High doses of oral vitamin B12 (e.g. 1,000 micrograms daily) may be able to produce adequate absorption of vitamin B12 even in the presence of IF deficiency and therefore be an alternative to the intramuscular route in many patients."

Evaluation Method

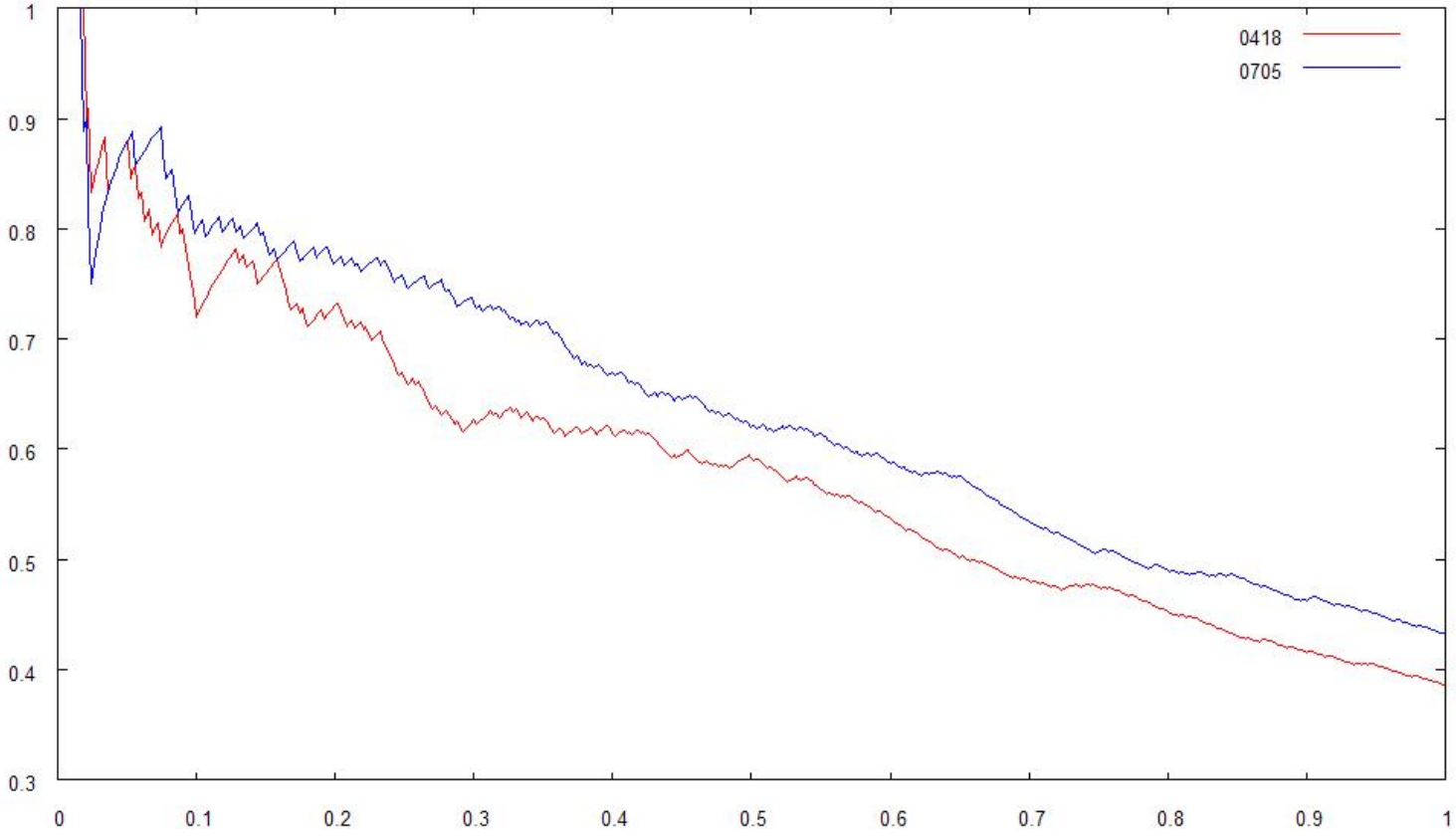
- A simplified explanation of the Put-back experiment:
  - TripAnswers Dataset = { <Q1, A1>, <Q2, A2>, … <Qn, An> }
  - The Test Corpus = { A1, A2, … , An }
  - The Test Question Set is a subset of {Q1, Q2, …, Qn}.
  - For each test question Qi, CliniQA system searches the Test Corpus, and assign answers to the question. We simply check whether answer Ai is assigned in top positions to Qi.

Preliminary Results

- We randomly selected 1000 questions for training and 500 questions for test
- The selected question set has the same distribution of different kinds of questions (i.e. drug, disorder, drug&disorder, other) as the original question set
- Metrics
  - Precision-recall curve
    - Sort the test questions based on their top answer's confidence
    - For each question at position i, suppose from question 1 to question i, the total number of correct answers are c_i, then we get a precision c_i/i at recall i/n
    - A curve is obtained for all precision – recall pairs

Performance Result

- 0418 : first base performance we obtained in April.
- 0705 : with improved annotation and four new scorers, i.e. MatchCoverageScorer, DisorderSimilarityScorer, DrugSimilarityScorer, PersonGenderScorer



Thanks!

Q & A

Preliminary Results

Q1008.xmi

What is the evidence for twice weekly cialis treatment to help with erectile dysfunction post prostatectomy?

Legend

Docume... | Question | Sentence | Snomed... | Token

Click In Text to See Annotation Detail

- Annotations
  - Token
    - Token ("erectile")
      - begin = 68
      - end = 76
  - SnomedConcept
    - SnomedConcept ("erectile dysfunction")
      - begin = 68
      - end = 88
      - name = erectile dysfunction
      - conceptType = disorder
      - code = 398175007
  - Sentence
    - Sentence ("What is the evidence for twice weekly cialis
      - begin = 0
      - end = 107
  - Question
    - Question ("What is the evidence for twice weekly cialis
      - begin = 0
      - end = 108
      - id = Q1008
      - source = TRIP_ANSWERS
      - sourceURL = http://www.tripanswers.org/answer.a
      - metadataItems = null

Select All | Deselect All | Hide Unselected

UIMA Pipeline (2)