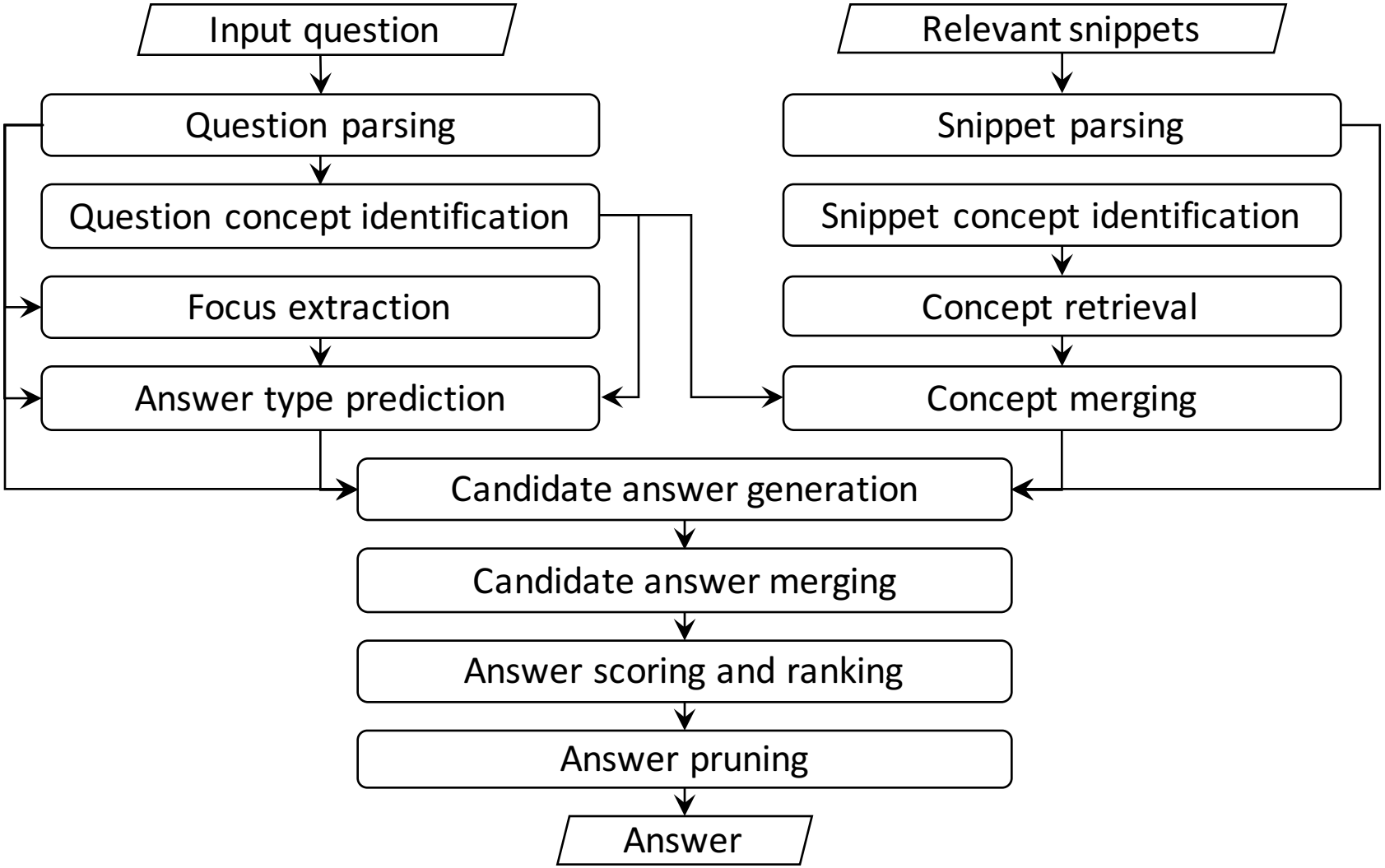
Learning to Answer Biomedical Questions: OAQA at BioASQ 4B

Zi Yang, Yue Zhou, and Eric Nyberg

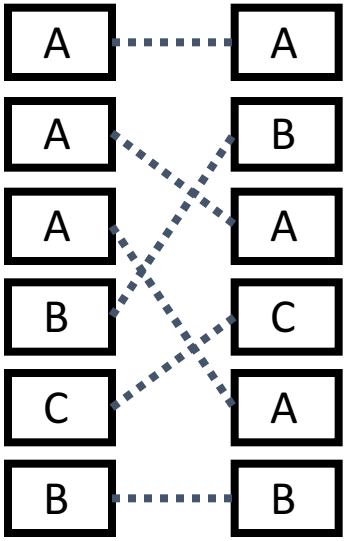
Language Technologies Institute
School of Computer Science
Carnegie Mellon University
{ziy, ehn}@cs.cmu.edu

OAQA@BioASQ-3B (2015) System



Collective Answer Reranking (1/2)

- If a list question asks the names of type A, then ideally the answers of type A are all ranked higher than others.



Common types seen in the top 3 answers	Common types seen in the top 5 answers
2/3	3/5
1/3	1/5
2/3	3/5
0/3	1/5
2/3	3/5
1/3	1/5

Collective Answer Reranking (2/2)

Feature list
Min/max/avg number (and percentage) of **semantic types** that each pair of candidate answers have in **common**
Min/max/avg **Levenshtein edit distance** between each pair of candidate answer variant names • Fludarabine, cytarabine, amsacrine
Min/max/avg **edit distance** between each pair of candidate answer variant names after transformed into their **shape forms** • Oct3/4 -> Aaa0.0, Sox2 -> Aaa0, Klf4 -> Aaa0, and c-Myc -> a.Aaa
Min/max/avg **token distance** between each pair of candidate answer occurrences • Yamanaka factors, Oct3/4, Sox2, Klf4, and c-Myc
The **original score** from the answer scoring prediction

Yes/No Question Answering (1/3)

- Supervised classification problem
- Feature list

Feature list
“Contradictory” concept count in the relevant snippets
Expected answer count in the relevant snippets
Sentiment analysis via positive and negative word count of each relevant snippet
Negation word count of each relevant snippet
Question inversion
Overlapping token count in the relevant snippets

Yes/No Question Answering (2/3)

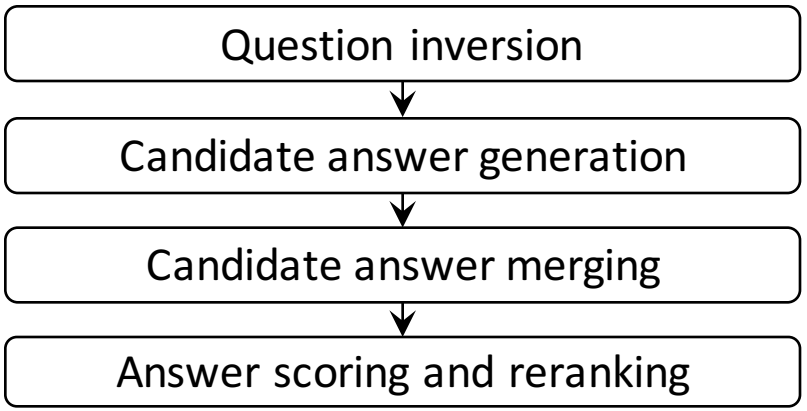
- Clue 1: Negative words
  - *Q: Can NXY-059 be used for treatment of acute ischemic stroke patients?*
  - S1: NXY-059, …, has demonstrated neuroprotection in several animal models of acute ischemic stroke but failed to confirm clinical benefit in the second phase III trial (SAINT-II).
  - S2: NXY-059 is no longer in development following a lack of efficacy found in a Phase III trial in patients with acute ischemic stroke.
  - Sentiment analysis (via sentiment word dictionary) and negation word list

Yes/No Question Answering (2/3)

- Clue 2: “Contradictory” concept
  - *Q: Does HER2 under-expression lead to favorable response to trastuzumab?*
  - S: over-expression of HER2 is reported…
  - *Q: Are chromomethylases present in animal genomes?*
  - S: Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants.
  - Same concept category but different concept
- Clue 3: One special concpet -- expected answer
  - Last concept in the question

Yes/No Question Answering (3/3)

- Clue 4: Question inversion (Kanayama et al., 2012)
  - Leverage a factoid Q/A pipeline to find whether an "inversed" statement is more true than the current statement.
  - *Q: Is pregabalin effective for treatment of patients with restless leg syndrome?*
  - IQ: Patients with what syndrom is pregabalin effective for treatment of?



Misc. Changes (1/2)

Scenario	BioASQ-3B (2015)	BioASQ-4B (2016)
Identify concepts from text snippets	Use MetaMap, LingPipe NER, and noun phrase extractor	Add TmTool (GNormPlus/SR4GN for genes and species, tmChem for chemicals, DNorm for diseases, and tmVar for mutations), and CValue (frequency based term extractor)
Rerank relevant documents, snippets, concepts	Train individual classifiers/learning-to-rank models	A standardized search result reranking interface

Misc. Changes (2/2)

Scenario	BioASQ-3B (2015)	BioASQ-4B (2016)
If an exact answer cannot be identified by a concept search service (e.g. UTS)	Drop it from the answer type prediction training set	Introduce "null" as an answer type
If the GS answer text contains parentheses • *Hydrophilic Interaction Chromatography (HILIC)*	Do nothing	Separate the text inside the parentheses from as a synonym for answer type prediction and answer scoring training • [Hydrophilic Interaction Chromatography, HILIC]
Extract features for answer scoring	Use only numeric features (e.g. count, distance)	Add nominal features (e.g. answer type name, concept type name)

Results – Factoid

- Tested using 3b-5 subset, trained on the rest
  - OF: original feature set
  - AF: additional feature set
  - CR: collective reranking

Method	Len.Ac.	MRR	Str.Ac.
OF + NO	.5000	**.3843**	.3182
OF + CR	.4545	.3791	.3182
AF + CR	**.5455**	.3732	.2727
AF + NO	.5000	.3689	.2727

Results – List

- TP: hard threshold
- RP: relative ratio to the maximum score

Method	F1	Precision	Recall
AF + NO + RP	**.4291**	**.4449**	.4593
AF + CR + RP	.4246	.4045	**.4864**
AF + CR + TP	.3969	.4100	.4267
OF + CR + TP	.3704	.4231	.3645
OF + CR + RP	.3629	.3654	.3874
AF + NO + TP	.3463	.3840	.3677
OF + NO + RP	.3460	.3188	.4431
OF + NO + TP	.1461	.2639	.1183

Results – Yes/No

- CVR: ClassificationViaRegression
- SL: SimpleLogistic
- AY: All Yes
- LR: Logistic regression

Method	Ac.	Neg.Ac.	Pos.Ac.
CVR	**.7143**	**.7778**	.6842
SL	**.7143**	.4444	**.8421**
AY	.6786	.0000	1.0000
LR	.5357	.2222	.6842

Yes/No – False Positive (1/2)

- Concept type mismatch
  - *Q: Are adenylylcyclases always transmembrane proteins?*
  - Key: contradictory concept pair "transmembrane" and "soluble" or "transmembrane adenylyl cyclase (tmAC)" and "soluble AC"
  - transmembrane -> Amino Acid, Peptide, or Protein
  - soluble adenylyl cyclase -> Gene or Genome

Yes/No – False Positive (2/2)

- Contradictory properties or behaviors
  - *Q: Does the 3D structure of the genome remain stable during cell differentiation?*
  - Key: contradictory concept pair "stable" and "reorganization", "alteration", "remodeling", etc.
  - stable -> Qualitative Concept
  - reorganization -> Idea or Concept
  - alteration -> N/A
  - remodelling -> N/A
- *Deep semantic representation for adjectives and verbs only?*

Yes/No – False Negative

- Some snippets contain multiple sentences or clauses, and only one is crucial to answer the question, while others negatively influence the results.
- E.g. "OATP1B1 and OATP1B3-mediated transport of bilirubin was confirmed and inhibition was determined for atazanavir, rifampicin, indinavir, amprenavir, cyclosporine, rifamycin SV and saquinavir."

Future Work

- Open source
- Right place / scenario to plug in deep learning

Thank you!