Question-Answering: Evaluation, Systems, Resources

Ling573
NLP Systems & Applications
April 5, 2011

Roadmap

- Rounding dimensions of QA
  - Evaluation, TREC

- QA systems: Alternate Approaches
  - ISI's Webclopedia
  - LCC's PowerAnswer-2 and Palantir
  - Insight's Patterns

- Resources

Evaluation

- Candidate criteria:
  - Relevance
  - Correctness

Evaluation

- Candidate criteria:
  - Relevance
  - Correctness
  - Conciseness:
    - No extra information

Evaluation

- Candidate criteria:
  - Relevance
  - Correctness
  - Conciseness:
    - No extra information
  - Completeness:
    - Penalize partial answers

Evaluation

- Candidate criteria:
  - Relevance
  - Correctness
  - Conciseness:
    - No extra information
  - Completeness:
    - Penalize partial answers
  - Coherence:
    - Easily readable

Evaluation

- Candidate criteria:
  - Relevance
  - Correctness
  - Conciseness:
    - No extra information
  - Completeness:
    - Penalize partial answers
  - Coherence:
    - Easily readable
  - Justification

Evaluation

- Candidate criteria:
  - Relevance
  - Correctness
  - Conciseness:
    - No extra information
  - Completeness:
    - Penalize partial answers
  - Coherence:
    - Easily readable
  - Justification

- Tension among criteria

Evaluation

- Consistency/repeatability:
  - Are answers scored reliability

Evaluation

- Consistency/repeatability:
  - Are answers scored reliability?

- Automation:
  - Can answers be scored automatically?
  - Required for machine learning tune/test

Evaluation

- Consistency/repeatability:
  - Are answers scored reliability?

- Automation:
  - Can answers be scored automatically?
  - Required for machine learning tune/test
    - Short answer answer keys
      - Litkowski's patterns

Evaluation

- Classical:
  - Return ranked list of answer candidates

Evaluation

- Classical:
  - Return ranked list of answer candidates
  - Idea: Correct answer higher in list => higher score

  - Measure: Mean Reciprocal Rank (MRR)

Evaluation

- Classical:
  - Return ranked list of answer candidates
  - Idea: Correct answer higher in list => higher score

  - Measure: Mean Reciprocal Rank (MRR)
    - For each question,
      - Get reciprocal of rank of first correct answer
        - E.g. correct answer is 4 => ¼
        - None correct => 0
    - Average over all questions

$$MRR = \frac{\sum_{i=1}^{N} \frac{1}{rank_i}}{N}$$

Dimensions of TREC QA

- Applications

Dimensions of TREC QA

- Applications
  - Open-domain free text search
  - Fixed collections
  - News, blogs

Dimensions of TREC QA

- Applications
  - Open-domain free text search
  - Fixed collections
  - News, blogs
- Users
  - Novice
- Question types

Dimensions of TREC QA

- Applications
  - Open-domain free text search
  - Fixed collections
  - News, blogs
- Users
  - Novice
- Question types
  - Factoid -> List, relation, etc
- Answer types

Dimensions of TREC QA

- Applications
  - Open-domain free text search
  - Fixed collections
  - News, blogs
- Users
  - Novice
- Question types
  - Factoid -> List, relation, etc
- Answer types
  - Predominantly extractive, short answer in context
- Evaluation:

Dimensions of TREC QA

- Applications
  - Open-domain free text search
  - Fixed collections
  - News, blogs
- Users
  - Novice
- Question types
  - Factoid -> List, relation, etc
- Answer types
  - Predominantly extractive, short answer in context
- Evaluation:
  - Official: human; proxy: patterns
- Presentation: One interactive track

Webclopedia

- Webclopedia system:
  - Information Sciences Institute (ISI), USC
  - Factoid QA: brief phrasal factual answers

Webclopedia

- Webclopedia system:
  - Information Sciences Institute (ISI), USC
  - Factoid QA: brief phrasal factual answers

- Prior approaches:
  - Form query, retrieve passage, slide window over passages
    - Pick window with highest score

Webclopedia

- Webclopedia system:
  - Information Sciences Institute (ISI), USC
  - Factoid QA: brief phrasal factual answers

- Prior approaches:
  - Form query, retrieve passage, slide window over passages
    - Pick window with highest score
      - E.g. # desirable words: overlap with query content terms

- Issues:

Webclopedia

- Webclopedia system:
  - Information Sciences Institute (ISI), USC
  - Factoid QA: brief phrasal factual answers

- Prior approaches:
  - Form query, retrieve passage, slide window over passages
    - Pick window with highest score
      - E.g. # desirable words: overlap with query content terms

- Issues:
  - Imprecise boundaries

Webclopedia

- Webclopedia system:
  - Information Sciences Institute (ISI), USC
  - Factoid QA: brief phrasal factual answers

- Prior approaches:
  - Form query, retrieve passage, slide window over passages
    - Pick window with highest score
      - E.g. # desirable words: overlap with query content terms

- Issues:
  - Imprecise boundaries: window vs NP/Name
  - Word overlap-based

Webclopedia

- Webclopedia system:
  - Information Sciences Institute (ISI), USC
  - Factoid QA: brief phrasal factual answers

- Prior approaches:
  - Form query, retrieve passage, slide window over passages
    - Pick window with highest score
      - E.g. # desirable words: overlap with query content terms

- Issues:
  - Imprecise boundaries: window vs NP/Name
  - Word overlap-based: synonyms?
  - Single window:

Webclopedia

- Webclopedia system:
  - Information Sciences Institute (ISI), USC
  - Factoid QA: brief phrasal factual answers

- Prior approaches:
  - Form query, retrieve passage, slide window over passages
    - Pick window with highest score
      - E.g. # desirable words: overlap with query content terms

- Issues:
  - Imprecise boundaries: window vs NP/Name
  - Word overlap-based: synonyms?
  - Single window: discontinuous answers?

Webclopedia Improvements

- Syntactic-semantic question analysis

Webclopedia Improvements

- Syntactic-semantic question analysis

- QA pattern matching

Webclopedia Improvements

- Syntactic-semantic question analysis

- QA pattern matching

- Classify QA types to improve answer type ID

- Use robust syntactic-semantic parser for analysis

- Combine word-, syntactic info for answer selection

Webclopedia Architecture

- Query parsing

- Query formulation

- IR

- Segmentation

- Segment ranking

- Segment parsing

- Answering pinpointing & ranking

Webclopedia QA Typology

- Issue: Many ways to express same info need

Webclopedia QA Typology

- Issue: Many ways to express same info need
  - What is the age of the Queen of Holland? How old is the Netherlands' Queen?, …

Webclopedia QA Typology

- Issue: Many ways to express same info need
  - What is the age of the Queen of Holland? How old is the Netherlands' Queen?, …

- Analyzed 17K+ answers.com questions -> 79 nodes
  - Nodes include:
    - Question & answer examples:
      - Q: *Who was Johnny Mathis' high school track coach?*
      - A: *Lou Vasquez, track coach of…and Johnny Mathis*

Webclopedia QA Typology

- Issue: Many ways to express same info need
  - What is the age of the Queen of Holland? How old is the Netherlands' Queen?, ...

- Analyzed 17K+ answers.com questions -> 79 nodes
  - Nodes include:
    - Question & answer examples:
      - Q: *Who was Johnny Mathis' high school track coach?*
      - A: *Lou Vasquez, track coach of...and Johnny Mathis*
    - Question & answer templates
      - Q: *who be <entity>'s <role>, who be <role> of <entity>*
      - A: *<person>, <role> of <entity>*

Webclopedia QA Typology

- Issue: Many ways to express same info need
  - What is the age of the Queen of Holland? How old is the Netherlands' Queen?, ...

- Analyzed 17K+ answers.com questions -> 79 nodes
  - Nodes include:
    - Question & answer examples:
      - Q: *Who was Johnny Mathis' high school track coach?*
      - A: *Lou Vasquez, track coach of...and Johnny Mathis*
    - Question & answer templates
      - Q: *who be <entity>'s <role>, who be <role> of <entity>*
      - A: *<person>, <role> of <entity>*
    - Qtarget: semantic type of answer

Webclopedia QA Typology

VERACITY	YES:NO		
	TRUE:FALSE		
ENTITY	AGENT	NAME	LAST-NAME
			FIRST-NAME
			COMPANY-NAME
			(PROPER-ORGANIZATION)
			ORGANIZATION
			(PROPER-ORGANIZATION)
		GROUP-OF-PEOPLE	
		(PROPER-ORGANIZATION)	
		ANIMAL	
		PERSON	OCCUPATION-PERSON
			GEOGRAPHICAL-PERSON

Question & Answer Parsing

- CONTEX parser:
  - Trained on growing collection of questions

Question & Answer Parsing

- CONTEX parser:
  - Trained on growing collection of questions
    - Original version parsed questions badly

Question & Answer Parsing

- CONTEX parser:
  - Trained on growing collection of questions
    - Original version parsed questions badly
  - Also identifies Qtargets and Qargs:
    - Qtargets:

Question & Answer Parsing

- CONTEX parser:
  - Trained on growing collection of questions
    - Original version parsed questions badly
  - Also identifies Qtargets and Qargs:
    - Qtargets:
      - Parts of speech
      - Semantic roles in parse tree
      - Elements of Typology + additional info

Question & Answer Parsing

- CONTEX parser:
  - Trained on growing collection of questions
    - Original version parsed questions badly
  - Also identifies Qtargets and Qargs:
    - Qtargets:
      - Parts of speech
      - Semantic roles in parse tree
      - Elements of Typology + additional info
      E.g. *Who is Betsy Ross?*
        - Qtarget: WHY-FAMOUS-PERSON; Qargs: "Betsy Ross"

Question & Answer Parsing

- CONTEX parser:
  - Trained on growing collection of questions
    - Original version parsed questions badly
  - Also identifies Qtargets and Qargs:
    - Qtargets:
      - Parts of speech
      - Semantic roles in parse tree
      - Elements of Typology + additional info

      E.g. *Who is Betsy Ross?*
        - Qtarget: WHY-FAMOUS-PERSON; Qargs: "Betsy Ross"
  - Extracted based on 276 hand-written rules
    - 10%: no target

Answer Matching

- Matches:
  - QA patterns in parse tree

  - Qtarget and Qwords in parse tree

  - Words in window

Enhancing
Word-based Match

- Qtarget-specific knowledge: Narrow

Enhancing Word-based Match

- Qtarget-specific knowledge: Narrow
  - Quantities: e.g. population
    - Q: What is the population of New York? – 100K+, M+

Enhancing Word-based Match

- Qtarget-specific knowledge: Narrow
  - Quantities: e.g. population
    - Q: What is the population of New York? – 100K+, M+
    - Biased to typical mean values
  - Abbreviations / Expansions
    - Q: What is NAFTA?

Enhancing Word-based Match

- Qtarget-specific knowledge: Narrow
  - Quantities: e.g. population
    - Q: What is the population of New York? – 100K+, M+
    - Biased to typical mean values
  - Abbreviations / Expansions
    - Q: What is NAFTA?
      - Check that answer includes N,A,F,T, and A
  - Zip code, phone number, etc patterns/NER

Enhancing Word-based Match

- Qtarget-specific knowledge: Narrow
  - Quantities: e.g. population
    - Q: What is the population of New York? – 100K+, M+
    - Biased to typical mean values
  - Abbreviations / Expansions
    - Q: What is NAFTA?
      - Check that answer includes N,A,F,T, and A
  - Zip code, phone number, etc patterns/NER

- Parse information
  - Link discontinuous answer information

QTarget:External Knowledge

- WordNet:

QTarget:External Knowledge

- WordNet:
  - Glosses provide additional world knowledge

QTarget:External Knowledge

- WordNet:
    - Glosses provide additional world knowledge
        - Resolve definition questions
            - *Q: What is the Milky Way?*
            - Candidate 2: the galaxy that contains the Earth
            - Wordnet: Milky Way—the galaxy containing the solar system

QTarget:External Knowledge

- WordNet:
  - Glosses provide additional world knowledge
    - Resolve definition questions
      - *Q: What is the Milky Way?*
      - Candidate 2: the galaxy that contains the Earth
      - Wordnet: Milky Way—the galaxy containing the solar system
    - Provide implicit information
      - *Q1: What is the capital of the United States?*
      -

QTarget:External Knowledge

- WordNet:
  - Glosses provide additional world knowledge
    - Resolve definition questions
      - *Q: What is the Milky Way?*
      - Candidate 2: the galaxy that contains the Earth
      - Wordnet: Milky Way—the galaxy containing the solar system
    - Provide implicit information
      - *Q1: What is the capital of the United States?*
      - S1. Later in the day, the president returned to Washington, the capital of the United States
      -

QTarget:External Knowledge

- WordNet:
  - Glosses provide additional world knowledge
    - Resolve definition questions
      - *Q: What is the Milky Way?*
      - Candidate 2: the galaxy that contains the Earth
      - Wordnet: Milky Way—the galaxy containing the solar system
    - Provide implicit information
      - *Q1: What is the capital of the United States?*
      - S1. Later in the day, the president returned to Washington, the capital of the United States
      - Wordnet: Washington—the capital of the United States

Results

- TREC rank 2 (tied):
  - Contribution: Qtarget, Word window, QA patterns

  - QA patterns too specific (only 4% of answers)

  - Qtarget classification biggest impact