Question-Answering: Systems & Resources

Ling573
NLP Systems & Applications
April 8, 2010
Roadmap

- Two extremes in QA systems:
  - LCC’s PowerAnswer-2
  - Insight’s Patterns...

- Question classification (Li & Roth)
- Resources
PowerAnswer2

- Language Computer Corp.
  - Lots of UT Dallas affiliates

- Tasks: factoid questions

- Major novel components:
  - Web-boosting of results
  - COGEX logic prover
  - Temporal event processing
  - Extended semantic chains

- Results: “Above median”: 53.4% main
Challenges: Co-reference

- Single, basic referent:

<table>
<thead>
<tr>
<th>Target 27 - Jennifer Capriati</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q27.2</td>
</tr>
<tr>
<td>Q27.3</td>
</tr>
</tbody>
</table>
Challenges: Co-reference

- Single, basic referent:

- Multiple possible antecedents:
  - Depends on previous correct answers
Challenges: Events

- Event answers:
  - Not just nominal concepts
Challenges: Events

- Event answers:
  - Not just nominal concepts
  - Nominal events:
    - Preakness 1998
Challenges: Events

- Event answers:
  - Not just nominal concepts
  - Nominal events:
    - Preakness 1998
  - Complex events:
    - Plane clips cable wires in Italian resort
Challenges: Events

- Event answers:
  - Not just nominal concepts
  - Nominal events:
    - Preakness 1998
  - Complex events:
    - Plane clips cable wires in Italian resort

- Establish question context, constraints
PowerAnswer-2

- Factoid QA system:
PowerAnswer-2

- Standard main components:
  - Question analysis, passage retrieval, answer processing
PowerAnswer-2

- Standard main components:
  - Question analysis, passage retrieval, answer processing
- Web-based answer boosting
PowerAnswer-2

- Standard main components:
  - Question analysis, passage retrieval, answer processing
  - Web-based answer boosting

- Complex components:
PowerAnswer-2

- Standard main components:
  - Question analysis, passage retrieval, answer processing
- Web-based answer boosting
- Complex components:
  - COGEX abductive prover
  - Word knowledge, semantics:
    - Extended WordNet, etc
  - Temporal processing
Web-Based Boosting

• Create search engine queries from question
Web-Based Boosting

- Create search engine queries from question
- Extract most redundant answers from search
  - Cf. Dumais et al - AskMSR
Web-Based Boosting

- Create search engine queries from question
- Extract most redundant answers from search
  - Cf. Dumais et al. - AskMSR
- Increase weight on TREC candidates that match
  - Higher weight if higher frequency
Web-Based Boosting

- Create search engine queries from question
- Extract most redundant answers from search
  - Cf. Dumais et al. - AskMSR
- Increase weight on TREC candidates that match
  - Higher weight if higher frequency

Intuition:
- Common terms in search likely to be answer
- QA answer search too focused on query terms
Web-Based Boosting

- Create search engine queries from question
- Extract most redundant answers from search
  - Cf. Dumais et al - AskMSR
- Increase weight on TREC candidates that match
  - Higher weight if higher frequency
- Intuition:
  - Common terms in search likely to be answer
  - QA answer search too focused on query terms
  - Reweighting improves
- Web-boosting improves significantly: 20%
Deep Processing: Query/Answer Formulation

- Preliminary shallow processing:
  - Tokenization, POS tagging, NE recognition, Preprocess
Deep Processing: Query/Answer Formulation

- Preliminary shallow processing:
  - Tokenization, POS tagging, NE recognition, Preprocess

- Parsing creates syntactic representation:
  - Focused on nouns, verbs, and particles
    - Attachment
Deep Processing: Query/Answer Formulation

- Preliminary shallow processing:
  - Tokenization, POS tagging, NE recognition, Preprocess

- Parsing creates syntactic representation:
  - Focused on nouns, verbs, and particles
    - Attachment

- Coreference resolution links entity references
Deep Processing: Query/Answer Formulation

- Preliminary shallow processing:
  - Tokenization, POS tagging, NE recognition, Preprocess

- Parsing creates syntactic representation:
  - Focused on nouns, verbs, and particles
    - Attachment

- Coreference resolution links entity references

- Translate to full logical form
  - As close as possible to syntax
Syntax to Logical Form
Syntax to Logical Form
Syntax to Logical Form
Deep Processing: Answer Selection

- Lexical chains:
  - Bridge gap in lexical choice b/t Q and A
    - Improve retrieval and answer selection
Deep Processing: Answer Selection

- Lexical chains:
  - Bridge gap in lexical choice b/t Q and A
    - Improve retrieval and answer selection
  - Create connections between synsets through topicality
    - Q: When was the internal combustion engine invented?
    - A: The first internal-combustion engine was built in 1867.
    - invent → create_mentally → create → build
Deep Processing: Answer Selection

- Lexical chains:
  - Bridge gap in lexical choice b/t Q and A
    - Improve retrieval and answer selection
  - Create connections between synsets through topicality
    - Q: When was the internal combustion engine invented?
    - A: The first internal-combustion engine was built in 1867.
    - invent → create_mentally → create → build

- Perform abductive reasoning b/t QLF & ALF
  - Tries to justify answer given question
Deep Processing: Answer Selection

- Lexical chains:
  - Bridge gap in lexical choice b/t Q and A
    - Improve retrieval and answer selection
  - Create connections between synsets through topicality
    - Q: *When was the internal combustion engine invented?*
    - A: *The first internal-combustion engine was built in 1867.*
    - invent → create_mentally → create → build

- Perform abductive reasoning b/t QLF & ALF
  - Tries to justify answer given question
  - Yields 10% improvement in accuracy!
Temporal Processing

- 16% of factoid questions include time reference
Temporal Processing

- 16% of factoid questions include time reference
- Index documents by date: absolute, relative
Temporal Processing

- 16% of factoid questions include time reference
- Index documents by date: absolute, relative
- Identify temporal relations b/t events
  - Store as triples of (S, E1, E2)
  - S is temporal relation signal – e.g. during, after
Temporal Processing

- 16% of factoid questions include time reference
- Index documents by date: absolute, relative
- Identify temporal relations b/t events
  - Store as triples of (S, E1, E2)
    - S is temporal relation signal – e.g. during, after
- Answer selection:
  - Prefer passages matching Question temporal constraint
  - Discover events related by temporal signals in Q & As
  - Perform temporal unification; boost good As
Temporal Processing

- 16% of factoid questions include time reference
- Index documents by date: absolute, relative
- Identify temporal relations b/t events
  - Store as triples of (S, E1, E2)
    - S is temporal relation signal – e.g. during, after
- Answer selection:
  - Prefer passages matching Question temporal constraint
  - Discover events related by temporal signals in Q & As
  - Perform temporal unification; boost good As
- Improves only by 2%
  - Mostly captured by surface forms
## Results

<table>
<thead>
<tr>
<th></th>
<th>PowerAnswer-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factoid</td>
<td>0.713</td>
</tr>
<tr>
<td>List</td>
<td>0.468</td>
</tr>
<tr>
<td>Other</td>
<td>0.228</td>
</tr>
<tr>
<td>Overall</td>
<td>0.534</td>
</tr>
</tbody>
</table>

Table 2: Results in the main task.
Overview

- Key sources of improvement:
  - Shallow processing:
    - Web-boosting: +20%
Overview

- Key sources of improvement:
  - Shallow processing:
    - Web-boosting: +20%
  - Deep processing:
    - COGEX logic prover + semantics: 10%
    - Temporal processing: 2%
Overview

- Key sources of improvement:
  - Shallow processing:
    - Web-boosting: +20%
  - Deep processing:
    - COGEX logic prover + semantics: 10%
    - Temporal processing: 2%
  - Relation queries:
    - All relatively shallow:
      - Biggest contributors: Keyword extraction, Topic signatures
Patterns of Potential Answer Expressions...

- “Insight”
- Shallow-pattern-based approach
  - Contrasts with deep processing techniques
Patterns of Potential Answer Expressions...

- “Insight”

- Shallow-pattern-based approach
  - Contrasts with deep processing techniques

- Intuition:
  - Some surface patterns highly correlated to information
Patterns of Potential Answer Expressions...

- “Insight”
- Shallow-pattern-based approach
  - Contrasts with deep processing techniques

Intuition:
- Some surface patterns highly correlated to information
  - E.g. Mozart (1756-1791)
Patterns of Potential Answer Expressions...

- “Insight”

- Shallow-pattern-based approach
  - Contrasts with deep processing techniques

- Intuition:
  - Some surface patterns highly correlated to information
    - E.g. Mozart (1756-1791)
    - Person – birthdate, death date
      - Pattern: Capitalized word; paren, 4 digits; dash; 4 digits; paren
      - Attested 850 times in a corpus
Pattern Library

- Potentially infinite patterns
Pattern Library

- Potentially infinite patterns
- Pattern structure:
  - Fixed components:
    - Words, characters, symbols
Pattern Library

- Potentially infinite patterns

- Pattern structure:
  - Fixed components:
    - Words, characters, symbols
  - Variable components:
    - Usually query terms and answer terms
Pattern Library

- Potentially infinite patterns

- Pattern structure:
  - Fixed components:
    - Words, characters, symbols
  - Variable components:
    - Usually query terms and answer terms
  - List of 51 pattern elements – combined for patterns
    - Ordered or unordered
Pattern Library

- Potentially infinite patterns

- Pattern structure:
  - Fixed components:
    - Words, characters, symbols
  - Variable components:
    - Usually query terms and answer terms
  - List of 51 pattern elements – combined for patterns
    - Ordered or unordered
  - More complex patterns are typically more indicative
Other Examples

- Post questions: Who is the Queen of the Netherlands?
Other Examples

- Post questions: Who is the Queen of the Netherlands?
- Beatrix, Queen of the Netherlands
Other Examples

- Post questions: Who is the Queen of the Netherlands?
- Beatrix, Queen of the Netherlands

Pattern elements:
- Country name
- Post name
- Person name
- Title (optional)
  - In some order
Basic Approach

- Question analysis:
  - Identify detailed question type
Basic Approach

- Question analysis:
  - Identify detailed question type

- Passage retrieval
  - Collect large number of retrieval snippets
    - Possibly with query expansion
Basic Approach

- Question analysis:
  - Identify detailed question type

- Passage retrieval
  - Collect large number of retrieval snippets
    - Possibly with query expansion

- Answer processing:
  - Find matching patterns in candidates
    - 10s of patterns/answer type
Results

- Best result in TREC-10
- MRR (strict) 0.676:
  - Correct: 289; 120 unanswered

- Retrieval based on shallow patterns
  - Bag of patterns, and sequences
  - Still highly effective
Question Classification: Li & Roth
Roadmap

- Motivation:
Why Question Classification?
Why Question Classification?

- Question classification categorizes possible answers
Why Question Classification?

- Question classification categorizes possible answers
- Constrains answers types to help find, verify answer

Q: What Canadian city has the largest population?
- Type?
Why Question Classification?

- Question classification categorizes possible answers
- Constrains answers types to help find, verify answer

Q: What Canadian city has the largest population?
- Type? -> City
- Can ignore all non-city NPs
Why Question Classification?

- Question classification categorizes possible answers
- Constrains answers types to help find, verify answer

Q: What Canadian city has the largest population?
- Type? -> City
- Can ignore all non-city NPs

- Provides information for type-specific answer selection
  - Q: What is a prism?
  - Type? ->
Why Question Classification?

- Question classification categorizes possible answers
  - Constrains answers types to help find, verify answer

  \[Q: \text{What Canadian city has the largest population?}\]
  - Type? -> City
  - Can ignore all non-city NPs

- Provides information for type-specific answer selection
  - \[Q: \text{What is a prism?}\]
  - Type? -> Definition
    - Answer patterns include: ‘A prism is...’
Challenges
Challenges

- Variability:
  - What tourist attractions are there in Reims?
  - What are the names of the tourist attractions in Reims?
  - What is worth seeing in Reims?
    - Type?
Challenges

• Variability:
  • What tourist attractions are there in Reims?
  • What are the names of the tourist attractions in Reims?
  • What is worth seeing in Reims?
    • Type? -> Location
Challenges

• Variability:
  • What tourist attractions are there in Reims?
  • What are the names of the tourist attractions in Reims?
  • What is worth seeing in Reims?
    • Type? -> Location

• Manual rules?
Challenges

- Variability:
  - What tourist attractions are there in Reims?
  - What are the names of the tourist attractions in Reims?
  - What is worth seeing in Reims?
    - Type? -> Location

- Manual rules?
  - Nearly impossible to create sufficient patterns

- Solution?
Challenges

- Variability:
  - What tourist attractions are there in Reims?
  - What are the names of the tourist attractions in Reims?
  - What is worth seeing in Reims?
    - Type? -> Location

- Manual rules?
  - Nearly impossible to create sufficient patterns

- Solution?
  - Machine learning – rich feature set
Approach

- Employ machine learning to categorize by answer type
  - Hierarchical classifier on semantic hierarchy of types
    - Coarse vs fine-grained
      - Up to 50 classes

- Differs from text categorization?
Approach

- Employ machine learning to categorize by answer type
  - Hierarchical classifier on semantic hierarchy of types
    - Coarse vs fine-grained
      - Up to 50 classes
  - Differs from text categorization?
    - Shorter (much!)
    - Less information, but
    - Deep analysis more tractable
Approach

- Exploit syntactic and semantic information
- Diverse semantic resources
Approach

- Exploit syntactic and semantic information
  - Diverse semantic resources
    - Named Entity categories
    - WordNet sense
    - Manually constructed word lists
    - Automatically extracted semantically similar word lists
Approach

- Exploit syntactic and semantic information
  - Diverse semantic resources
    - Named Entity categories
    - WordNet sense
    - Manually constructed word lists
    - Automatically extracted semantically similar word lists

Results:
- Coarse: 92.5%; Fine: 89.3%
- Semantic features reduce error by 28%
<table>
<thead>
<tr>
<th>Class</th>
<th>#</th>
<th>Class</th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ABBREVIATION</strong></td>
<td></td>
<td><strong>LOCATION</strong></td>
<td></td>
</tr>
<tr>
<td>abbreviation</td>
<td>2</td>
<td>city</td>
<td>44</td>
</tr>
<tr>
<td>expression</td>
<td>16</td>
<td>country</td>
<td>21</td>
</tr>
<tr>
<td>definition</td>
<td>126</td>
<td>mountain</td>
<td>5</td>
</tr>
<tr>
<td>description</td>
<td>13</td>
<td>other</td>
<td>114</td>
</tr>
<tr>
<td>manner</td>
<td>7</td>
<td>state</td>
<td>11</td>
</tr>
<tr>
<td>reason</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ENTITY</strong></td>
<td>174</td>
<td></td>
<td></td>
</tr>
<tr>
<td>animal</td>
<td>27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>body</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>color</td>
<td>12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>creative</td>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>currency</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>disease/medicine</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>event</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>food</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>instrument</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lang</td>
<td>3</td>
<td>code</td>
<td>1</td>
</tr>
<tr>
<td>letter</td>
<td>0</td>
<td>count</td>
<td>22</td>
</tr>
<tr>
<td>other</td>
<td>19</td>
<td>date</td>
<td>146</td>
</tr>
<tr>
<td>plant</td>
<td>7</td>
<td>distance</td>
<td>38</td>
</tr>
<tr>
<td>product</td>
<td>9</td>
<td>money</td>
<td>9</td>
</tr>
<tr>
<td>religion</td>
<td>1</td>
<td>order</td>
<td>0</td>
</tr>
<tr>
<td>sport</td>
<td>3</td>
<td>other</td>
<td>24</td>
</tr>
<tr>
<td>substance</td>
<td>20</td>
<td>period</td>
<td>18</td>
</tr>
<tr>
<td>symbol</td>
<td>2</td>
<td>percent</td>
<td>7</td>
</tr>
<tr>
<td>technique</td>
<td>1</td>
<td>speed</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>temp</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>vol.size</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>weight</td>
<td>4</td>
</tr>
<tr>
<td><strong>DESCRIPTION</strong></td>
<td>153</td>
<td></td>
<td></td>
</tr>
<tr>
<td>definition</td>
<td>126</td>
<td></td>
<td></td>
</tr>
<tr>
<td>description</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>manner</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>reason</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>HUMAN</strong></td>
<td>171</td>
<td></td>
<td></td>
</tr>
<tr>
<td>group</td>
<td>24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>individual</td>
<td>140</td>
<td></td>
<td></td>
</tr>
<tr>
<td>title</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>description</td>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Learning a Hierarchical Question Classifier

- Many manual approaches use only:
Learning a Hierarchical Question Classifier

- Many manual approaches use only:
  - Small set of entity types, set of handcrafted rules
Learning a Hierarchical Question Classifier

- Many manual approaches use only:
  - Small set of entity types, set of handcrafted rules
  - Note: Webclopedia’s 96 node taxo w/276 manual rules
Learning a Hierarchical Question Classifier

- Many manual approaches use only:
  - Small set of entity types, set of handcrafted rules
    - Note: Webclopedia’s 96 node taxo w/276 manual rules
- Learning approaches can learn to generalize
  - Train on new taxonomy, but
Learning a Hierarchical Question Classifier

● Many manual approaches use only:
  ● Small set of entity types, set of handcrafted rules
    ● Note: Webclopedia’s 96 node taxo w/276 manual rules

● Learning approaches can learn to generalize
  ● Train on new taxonomy, but
    ● Someone still has to label the data...

● Two step learning: (Winnow)
  ● Same features in both cases
Learning a Hierarchical Question Classifier

- Many manual approaches use only:
  - Small set of entity types, set of handcrafted rules
    - Note: Webclopedia’s 96 node taxo w/276 manual rules

- Learning approaches can learn to generalize
  - Train on new taxonomy, but
    - Someone still has to label the data...

- Two step learning: (Winnow)
  - Same features in both cases
    - First classifier produces (a set of) coarse labels
    - Second classifier selects from fine-grained children of coarse tags generated by the previous stage
    - Select highest density classes above threshold
Features for Question Classification

- Primitive lexical, syntactic, lexical-semantic features
- Automatically derived
- Combined into conjunctive, relational features
- Sparse, binary representation
Features for Question Classification

- Primitive lexical, syntactic, lexical-semantic features
  - Automatically derived
  - Combined into conjunctive, relational features
  - Sparse, binary representation

- Words
  - Combined into ngrams
Features for Question Classification

- Primitive lexical, syntactic, lexical-semantic features
  - Automatically derived
  - Combined into conjunctive, relational features
  - Sparse, binary representation

- Words
  - Combined into ngrams

- Syntactic features:
  - Part-of-speech tags
  - Chunks
  - Head chunks: 1st N, V chunks after Q-word
Syntactic Feature Example

- Q: Who was the first woman killed in the Vietnam War?
Syntactic Feature Example

- Q: Who was the first woman killed in the Vietnam War?

- POS: [Who WP] [was VBD] [the DT] [first JJ] [woman NN] [killed VBN] {in IN} [the DT] [Vietnam NNP] [War NNP] [? .]
Syntactic Feature Example

- Q: Who was the first woman killed in the Vietnam War?

- POS: [Who WP] [was VBD] [the DT] [first JJ] [woman NN] [killed VBN] {in IN} [the DT] [Vietnam NNP] [War NNP] [? .]

- Chunking: [NP Who] [VP was] [NP the first woman] [VP killed] [PP in] [NP the Vietnam War] ?
Syntactic Feature Example

- Q: Who was the first woman killed in the Vietnam War?

- POS: [Who WP] [was VBD] [the DT] [first JJ] [woman NN] [killed VBN] {in IN} [the DT] [Vietnam NNP] [War NNP] [? .]

- Chunking: [NP Who] [VP was] [NP the first woman] [VP killed] [PP in] [NP the Vietnam War] ?

- Head noun chunk: ‘the first woman’
Semantic Features

- Treat analogously to syntax?
Semantic Features

- Treat analogously to syntax?
- Q1: What’s the semantic equivalent of POS tagging?
Semantic Features

- Treat analogously to syntax?
  - Q1: What’s the semantic equivalent of POS tagging?
  - Q2: POS tagging > 97% accurate;
    - Semantics? Semantic ambiguity?
Semantic Features

- Treat analogously to syntax?
  - Q1: What’s the semantic equivalent of POS tagging?
  - Q2: POS tagging > 97% accurate;
    - Semantics? Semantic ambiguity?

- A1: Explore different lexical semantic info sources
  - Differ in granularity, difficulty, and accuracy
Semantic Features

- Treat analogously to syntax?
  - Q1: What’s the semantic equivalent of POS tagging?
  - Q2: POS tagging > 97% accurate;
    - Semantics? Semantic ambiguity?

- A1: Explore different lexical semantic info sources
  - Differ in granularity, difficulty, and accuracy
  - Named Entities
  - WordNet Senses
  - Manual word lists
  - Distributional sense clusters
Tagging & Ambiguity

- Augment each word with semantic category

- What about ambiguity?
  - E.g. ‘water’ as ‘liquid’ or ‘body of water’
Tagging & Ambiguity

- Augment each word with semantic category

What about ambiguity?
- E.g. ‘water’ as ‘liquid’ or ‘body of water’
- Don’t disambiguate
  - Keep all alternatives
  - Let the learning algorithm sort it out
  - Why?
Semantic Categories

- Named Entities
  - Expanded class set: 34 categories
    - E.g. Profession, event, holiday, plant,...
Semantic Categories

- Named Entities
  - Expanded class set: 34 categories
    - E.g. Profession, event, holiday, plant,...

- WordNet: IS-A hierarchy of senses
  - All senses of word + direct hyper/hyponyms
Semantic Categories

- Named Entities
  - Expanded class set: 34 categories
    - E.g. Profession, event, holiday, plant, ...

- WordNet: IS-A hierarchy of senses
  - All senses of word + direct hyper/hyponyms

- Class-specific words
  - Manually derived from 5500 questions
    - E.g. Class: Food
      - {alcoholic, apple, beer, berry, breakfast brew butter candy cereal champagne cook delicious eat fat ..}
    - Class is semantic tag for word in the list
Semantic Types

- Distributional clusters:
  - Based on Pantel and Lin
  - Cluster based on similarity in dependency relations
  - Word lists for 20K English words
Semantic Types

- Distributional clusters:
  - Based on Pantel and Lin
  - Cluster based on similarity in dependency relations
  - Word lists for 20K English words
    - Lists correspond to word senses
    - Water:
      - Sense 1: { oil gas fuel food milk liquid}
      - Sense 2: {air moisture soil heat area rain}
      - Sense 3: {waste sewage pollution runoff}
Semantic Types

- Distributional clusters:
  - Based on Pantel and Lin
  - Cluster based on similarity in dependency relations
  - Word lists for 20K English words
    - Lists correspond to word senses
    - Water:
      - Sense 1: \{ oil, gas, fuel, food, milk, liquid \}
      - Sense 2: \{ air, moisture, soil, heat, area, rain \}
      - Sense 3: \{ waste, sewage, pollution, runoff \}
  - Treat head word as semantic category of words on list
Evaluation

- Assess hierarchical coarse->fine classification
- Assess impact of different semantic features
- Assess training requirements for diff’t feature set
Evaluation

- Assess hierarchical coarse->fine classification
- Assess impact of different semantic features
- Assess training requirements for diff’lt feature set
  - Training:
    - 21.5K questions from TREC 8,9; manual; USC data
  - Test:
    - 1K questions from TREC 10,11
Evaluation

- Assess hierarchical coarse->fine classification
- Assess impact of different semantic features
- Assess training requirements for diff’t feature set

Training:
- 21.5K questions from TREC 8,9; manual; USC data

Test:
- 1K questions from TREC 10,11

Measures: Accuracy and class-specific precision
Results

- Syntactic features only:

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Word</th>
<th>POS</th>
<th>Chunk</th>
<th>Head(SYN)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coarse</td>
<td>85.10</td>
<td>91.80</td>
<td>91.80</td>
<td>92.50</td>
</tr>
<tr>
<td>Fine</td>
<td>82.60</td>
<td>84.90</td>
<td>84.00</td>
<td>85.00</td>
</tr>
</tbody>
</table>

- POS useful; chunks useful to contribute head chunks
- Fine categories more ambiguous
Results

- Syntactic features only:
  - POS useful; chunks useful to contribute head chunks
  - Fine categories more ambiguous

- Semantic features:
  - Best combination: SYN, NE, Manual & Auto word lists
  - Coarse: same; Fine: 89.3% (28.7% error reduction)
Results

- Syntactic features only:
  - POS useful; chunks useful to contribute head chunks
  - Fine categories more ambiguous

- Semantic features:
  - Best combination: SYN, NE, Manual & Auto word lists
    - Coarse: same; Fine: 89.3% (28.7% error reduction)

- Wh-word most common class: 41%
<table>
<thead>
<tr>
<th>Class</th>
<th>#</th>
<th>Precision[c]</th>
<th>Class</th>
<th>#</th>
<th>Precision[c]</th>
</tr>
</thead>
<tbody>
<tr>
<td>abb</td>
<td>2</td>
<td>100%</td>
<td>desc</td>
<td>25</td>
<td>36%</td>
</tr>
<tr>
<td>exp</td>
<td>17</td>
<td>94.11%</td>
<td>manner</td>
<td>8</td>
<td>87.5%</td>
</tr>
<tr>
<td>animal</td>
<td>27</td>
<td>85.18%</td>
<td>reason</td>
<td>7</td>
<td>85.71%</td>
</tr>
<tr>
<td>body</td>
<td>4</td>
<td>100%</td>
<td>gr</td>
<td>19</td>
<td>89.47%</td>
</tr>
<tr>
<td>color</td>
<td>12</td>
<td>100%</td>
<td>ind</td>
<td>154</td>
<td>90.25%</td>
</tr>
<tr>
<td>cremat</td>
<td>13</td>
<td>76.92%</td>
<td>title</td>
<td>4</td>
<td>100%</td>
</tr>
<tr>
<td>currency</td>
<td>6</td>
<td>100%</td>
<td>desc</td>
<td>3</td>
<td>100%</td>
</tr>
<tr>
<td>dismed</td>
<td>4</td>
<td>50%</td>
<td>city</td>
<td>41</td>
<td>97.56%</td>
</tr>
<tr>
<td>event</td>
<td>4</td>
<td>75%</td>
<td>country</td>
<td>21</td>
<td>95.23%</td>
</tr>
<tr>
<td>food</td>
<td>6</td>
<td>100%</td>
<td>mount</td>
<td>2</td>
<td>100%</td>
</tr>
<tr>
<td>instru</td>
<td>1</td>
<td>100%</td>
<td>LOC:other</td>
<td>116</td>
<td>89.65%</td>
</tr>
<tr>
<td>lang</td>
<td>3</td>
<td>100%</td>
<td>state</td>
<td>14</td>
<td>78.57%</td>
</tr>
<tr>
<td>ENTY:other</td>
<td>24</td>
<td>37.5%</td>
<td>count</td>
<td>24</td>
<td>91.66%</td>
</tr>
<tr>
<td>plant</td>
<td>3</td>
<td>100%</td>
<td>date</td>
<td>145</td>
<td>100%</td>
</tr>
<tr>
<td>product</td>
<td>6</td>
<td>66.66%</td>
<td>dist</td>
<td>37</td>
<td>97.29%</td>
</tr>
<tr>
<td>religion</td>
<td>1</td>
<td>100%</td>
<td>money</td>
<td>6</td>
<td>100%</td>
</tr>
<tr>
<td>sport</td>
<td>4</td>
<td>75%</td>
<td>NUM:other</td>
<td>15</td>
<td>93.33%</td>
</tr>
<tr>
<td>substance</td>
<td>21</td>
<td>80.95%</td>
<td>period</td>
<td>20</td>
<td>85%</td>
</tr>
<tr>
<td>symbol</td>
<td>2</td>
<td>100%</td>
<td>perc</td>
<td>9</td>
<td>77.77%</td>
</tr>
<tr>
<td>termeq</td>
<td>22</td>
<td>63.63%</td>
<td>speed</td>
<td>8</td>
<td>100%</td>
</tr>
<tr>
<td>veh</td>
<td>7</td>
<td>71.42%</td>
<td>temp</td>
<td>4</td>
<td>100%</td>
</tr>
<tr>
<td>def</td>
<td>125</td>
<td>97.6%</td>
<td>weight</td>
<td>4</td>
<td>100%</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>1000</td>
<td><strong>89.3%</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Observations

- Effective coarse and fine-grained categorization
  - Mix of information sources and learning
  - Shallow syntactic features effective for coarse
  - Semantic features improve fine-grained
    - Most feature types help
      - WordNet features appear noisy
      - Use of distributional sense clusters dramatically increases feature dimensionality

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NE</td>
<td>0.23</td>
</tr>
<tr>
<td>SemWN</td>
<td>16</td>
</tr>
<tr>
<td>SemCSR</td>
<td>23</td>
</tr>
<tr>
<td>SemSWL</td>
<td>557</td>
</tr>
</tbody>
</table>
Software Resources

- Build on existing tools
  - Focus on QA specific tasks
- General: Machine learning tools
Software Resources

• General: Machine learning tools
  • Mallet:  http://mallet.cs.umass.edu
  • Weka toolkit: www.cs.waikato.ac.nz/ml/weka/
Software Resources

• General: Machine learning tools
  • Mallet: http://mallet.cs.umass.edu
  • Weka toolkit: www.cs.waikato.ac.nz/ml/weka/

• NLP toolkits, collections:
  • GATE: http://gate.ac.uk
  • NLTK: http://www.nltk.org
  • LingPipe: alias-i.com/lingpipe/
  • Stanford NLP tools: http://nlp.stanford.edu/software/
Software Resources: Specific

- Information retrieval:
  - Lucene: [http://lucene.apache.org](http://lucene.apache.org) (on patas)
    - Standard system, tutorials
    - High quality research system
    - Linked to textbook on IR
Software Resources: Cont’d

- POS taggers:
  - Stanford POS tagger
  - Treetagger
  - Maxent POS tagger
  - Brill tagger

- Stemmers: http://snowball.tartarus.org
  - Implementations of Porter stemmer in many langs

- Sentence splitters
  - NIST
Software Resources

• Parsers:
  • Constituency parser
    • Stanford parser
    • Collins/Bikel parser
    • Charniak parser
  • Dependency parsers
    • Minipar

• WSD packages:
  • WordNet::Similarity
Software Resources

- Semantic analyzer:
  - Shalmaneser

- Databases, ontologies:
  - WordNet
  - FrameNet
  - PropBank
Information Resources

- Proxies for world knowledge:
  - WordNet: Synonymy; IS-A hierarchy
Information Resources

- Proxies for world knowledge:
  - WordNet: Synonymy; IS-A hierarchy
  - Wikipedia
Information Resources

- Proxies for world knowledge:
  - WordNet: Synonymy; IS-A hierarchy
  - Wikipedia
  - Web itself
  - ....

- Training resources:
  - Question classification sets (UIUC)
  - Other TREC QA data (Questions, Answers)