Question Classification

Ling573
NLP Systems and Applications
April 25, 2013

Deliverable #3

- Posted: Code & results due May 10
- Focus: Question processing
  - Classification, reformulation, expansion, etc
- Additional: general improvement motivated by D#2

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: 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? -> 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%

Question Hierarchy

Class	#	Class	#
ABBREVIATION	18	term	19
abbreviation	2	vehicle	7
expression	16	word	0
DESCRIPTION	153	**HUMAN**	171
definition	126	group	24
description	13	individual	140
manner	7	title	4
reason	7	description	3
ENTITY	174	**LOCATION**	195
animal	27	city	44
body	5	country	21
color	12	mountain	5
creative	14	other	114
currency	8	state	11
disease/medicine	3	**NUMERIC**	289
event	6	code	1
food	7	count	22
instrument	1	date	146
lang	3	distance	38
letter	0	money	9
other	19	order	0
plant	7	other	24
product	9	period	18
religion	1	percent	7
sport	3	speed	9
substance	20	temp	7
symbol	2	vol.size	4
technique	1	weight	4

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't 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:

Classifier	Word	POS	Chunk	Head(SYN)
Coarse	85.10	91.80	91.80	**92.50**
Fine	82.60	84.90	84.00	**85.00**

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

Results

- Syntactic features only:

Classifier	Word	POS	Chunk	Head(SYN)
Coarse	85.10	91.80	91.80	**92.50**
Fine	82.60	84.90	84.00	**85.00**

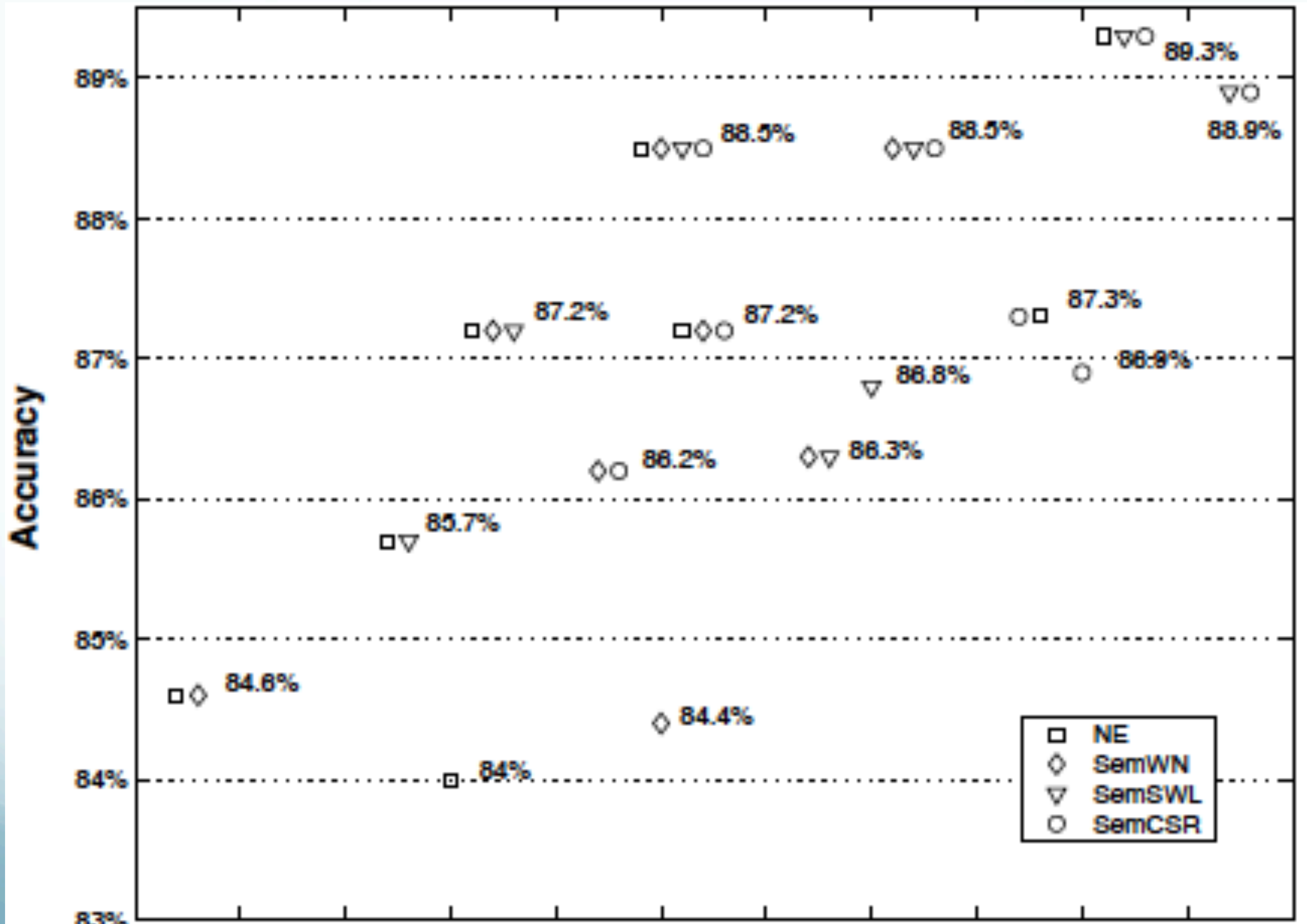
  - POS useful; chunks useful to contribute head chunks
  - Fine categories more ambiguous
- Semantic features:
  - Best combination: SYN, NE, Manual & Auto word lists
    - Coarse: same; Fine: 89.3% (28.7% error reduction)

Results

- Syntactic features only:

Classifier	Word	POS	Chunk	Head(SYN)
Coarse	85.10	91.80	91.80	**92.50**
Fine	82.60	84.90	84.00	**85.00**

  - 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%



Accuracy
89%
88%
87%
86%
85%
84%
83%
89.3%
88.9%
88.5%
88.5%
87.2%
87.2%
87.3%
86.9%
86.8%
86.2%
86.3%
85.7%
84.6%
84.4%
84%
NE
SemWN
SemSWL
SemCSR

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

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

NE	0.23
SemWN	16
SemCSR	23
SemSWL	557