Deliverable #2: Question Classification

Thomas Phan
Jason Shaw
Varden Wang

Focus: Feature Design

- Surface
  - Unigrams
- Syntactic
  - POS tags
- Semantic
  - WSD
  - NER
  - WordNet

- All classification done using Mallet's MaxEnt

Surface Features

- Unigrams only
- No stemming or case neutralizing
- Weak tokenization

- Classification on Li and Roth data:
  - Coarse: 84.6%
  - Fine: 75.8%

Syntactic Features

- Words tagged with POS
- NLTK tagger trained on Penn Treebank

- Classification on Li and Roth data:
  - Coarse: 85.6%
  - Fine: 76.8%

Semantic Features

- Word sense disambiguation
- Unsupervised Lesk disambiguator using WordNet
  - Coarse: 85.4%
  - Fine: 73.8%

- Named entity recognition
- Stanford NER package
- Identified PERSONs, ORGANIZATIONs, and LOCATIONs
  - Coarse: 37.8%
  - Fine: 12.0%

Semantic Features (cont'd)

- Classification sense lists from WordNet
- Map Question Hierarchy to high-level senses
- Test for presence of these senses as hypernyms
- Used in conjunction with other features, performs better than WSD alone
  - Coarse: 75.8%
  - Fine: 64.6%

Class	Sense
ENTITY	entity#1
food	food#1, food#2
plant	plant#1
HUMAN	human#2
individual	individual#1
LOCATION	location#1
city	city#1, city#2

Putting it all together

- Optimal results when using unigrams, POS tags, and classification sense lists

Training data	Test data	Type	Accuracy
Li and Roth	TREC 10	Coarse	88.0%
Li and Roth	TREC 10	Fine	80.2%
Li and Roth, TREC 2004	TREC 10	Coarse	88.0%
Li and Roth, TREC 2004	TREC 2005	Coarse	83.5%

- The addition of NER tagging improved fine classification to 81.6%

Potential improvements

- Finer-grained NER categories
- Head chunking
- Classification sense lists without WSD