Question Classification

Chase Hermsen, Sergei Lushtak,
Joshua Lutes

Approach
- Features
  - bigrams, unigrams, et cetera
- Models
  - BalancedWinnow, SVM, MaxEnt, et cetera

Evaluation
- tables of results
- initial attempts with all of the models
- results with more features on most promising models
- results with cross validation

Feature extraction

Transformations:

- lower case
- Porter stemming

Feature templates:

- ngrams
- hypernyms and hyponyms
- question words
- similar words
- part of speech

Original Machine Choice

- Lots of models available in Mallet
  - Maximum Entropy
  - Decision Tree
  - Winnow
  - Balanced Winnow
- Also, SVM's from libSVM
- How to choose?

Where do babies come from ?

- Text was transformed to lower case
- Text stemmed with Porter stemming algorithm

where do babi come from ?

bigrams: <s>_where, where_do, do_babi, babi_come, come_from, from_?, ?_</s>

where do babi come from ?

Naive Bayes	LOC:other
Maximum Entropy	DESC:desc
Decision Tree	DESC:desc
Winnow	LOC:other
Balanced Winnow	LOC:other

Evaluation

	Training Accuracy	Test Accuracy
Naive Bayes	0.97	0.80
Maximum Entropy	0.99	0.84
Decision Tree	0.51	0.45
Winnow	0.65	0.54
Balanced Winnow	1.00	0.87

Hypernyms and Hyponyms

- WordNet through NLTK
- Add all hypernyms and all hyponyms as features.

Similar Words

- Distributional similarity in Brown corpus
- Extraordinarily slow

Part of Speech Tags

Two different types of features:

POS-Pairs: <s>_<s> N_Joe V_ate Adj_delicious N_sand...
POS-Bigrams: <s>_N N_V V_Adj Adj_N ...
POS Unigrams: <s> N V Adj N ...

where do babi come from ?

unigrams: where, do, babi, come, from, ?

bigrams: <s>_where, where_do, do_babi, babi_come, come_from, from_?, ?_</s>

question_word: qword_where

hypernyms and hyponyms

similar words

part of speech

How to choose our features

In a word, cross-validation!

- Started with single test corpus
  - Strange results.
- Cross validation ensures more consistency of choosing a "good" model for the blind test.
  - 90/10 splits, ten instances
  - Average scores presented
- Training set: Combination of train5452 and TREC-2004

Cross Validation Results

	MaxEnt	BalancedWinnow
Bigrams	0.833	0.829
Bi + Unigrams	**0.873**	0.857
Tri + Bi + Unigrams	0.863	0.850
(1:4)-Grams	0.857	0.854
Bi + Uni + QuestionWords	0.859	0.848
Bi + Uni + Hyps	0.840	0.821
Bi + Uni + POS-Bigrams	0.864	0.862
Bi + Uni + POS-B + POS-U	0.868	0.857
Bi + unigram-POS pairs	0.850	0.840

Final Test Data Results

Run on the blind data set:

0.8175824175824242