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?

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Location/Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naive Bayes</td>
<td>LOC:other</td>
</tr>
<tr>
<td>Maximum Entropy</td>
<td>DESC:desc</td>
</tr>
<tr>
<td>Decision Tree</td>
<td>DESC:desc</td>
</tr>
<tr>
<td>Winnow</td>
<td>LOC:other</td>
</tr>
<tr>
<td>Balanced Winnow</td>
<td>LOC:other</td>
</tr>
</tbody>
</table>
## Evaluation

<table>
<thead>
<tr>
<th>Method</th>
<th>Training Accuracy</th>
<th>Test Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naive Bayes</td>
<td>0.97</td>
<td>0.80</td>
</tr>
<tr>
<td>Maximum Entropy</td>
<td>0.99</td>
<td>0.84</td>
</tr>
<tr>
<td>Decision Tree</td>
<td>0.51</td>
<td>0.45</td>
</tr>
<tr>
<td>Winnow</td>
<td>0.65</td>
<td>0.54</td>
</tr>
<tr>
<td>Balanced Winnow</td>
<td>1.00</td>
<td>0.87</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Feature Set</th>
<th>MaxEnt</th>
<th>BalancedWinnow</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bigrams</td>
<td>0.833</td>
<td>0.829</td>
</tr>
<tr>
<td>Bi + Unigrams</td>
<td>0.873</td>
<td>0.857</td>
</tr>
<tr>
<td>Tri + Bi + Unigrams</td>
<td>0.863</td>
<td>0.850</td>
</tr>
<tr>
<td>(1:4)-Grams</td>
<td>0.857</td>
<td>0.854</td>
</tr>
<tr>
<td>Bi + Uni + QuestionWords</td>
<td>0.859</td>
<td>0.848</td>
</tr>
<tr>
<td>Bi + Uni + Hyps</td>
<td>0.840</td>
<td>0.821</td>
</tr>
<tr>
<td>Bi + Uni + POS-Bigrams</td>
<td>0.864</td>
<td>0.862</td>
</tr>
<tr>
<td>Bi + Uni + POS-B + POS-U</td>
<td>0.868</td>
<td>0.857</td>
</tr>
<tr>
<td>Bi + unigram-POS pairs</td>
<td>0.850</td>
<td>0.840</td>
</tr>
</tbody>
</table>
Final Test Data Results

Run on the blind data set:

0.81758241758242