Automatic Grammaticality Judgements

Joachim Wagner, Jennifer Foster and Josef van Genabith

National Centre for Language Technology
School of Computing
Dublin City University
Ireland

NCLT/CNGL MT Workshop
24th July 2008
Automatic Grammaticality Judgements

**Input:** a sentence, $s$

**Output:** a judgement on the grammaticality of $s$

Judgement can be categorical or graded.

Uses

- Computer-aided language learning or grammar checking: as a first step towards diagnosing an error
- Automatic essay grading
- To decide which grammar to use in a robust parsing setup
- To rank the output of natural language generator or MT systems?
**Automatic Grammaticality Judgements**

**Input:** a sentence, $s$

**Output:** a judgement on the grammaticality of $s$

Judgement can be categorical or graded.

**Uses**

- Computer-aided language learning or grammar checking: as a first step towards diagnosing an error
- Automatic essay grading
- To decide which grammar to use in a robust parsing setup
- To rank the output of natural language generator or MT systems?
Automatic Grammaticality Judgements

**Input:** a sentence, \( s \)

**Output:** a judgement on the grammaticality of \( s \)

Judgement can be categorical or graded.

Uses

- Computer-aided language learning or grammar checking: as a first step towards diagnosing an error
- Automatic essay grading
- To decide which grammar to use in a robust parsing setup
- To rank the output of natural language generator or MT systems?
Approaches to Grammaticality Judging

3 Basic Methods

1. N-gram-based classifier
2. Precision-grammar-based classifier
3. Probabilistic-parsing-based classifier

- All methods use the J48 decision tree machine learning algorithm.
- Training data – large corpus of grammatical sentences and an equally large corpus of ungrammatical sentences.
Approaches to Grammaticality Judging

3 Basic Methods

1. N-gram-based classifier
2. Precision-grammar-based classifier
3. Probabilistic-parsing-based classifier

- All methods use the J48 decision tree machine learning algorithm.
- Training data – large corpus of grammatical sentences and an equally large corpus of ungrammatical sentences.
N-gram-based Classifier

Classifies a sentence as ungrammatical if it contains an unusual part of speech sequence.

- The classifier relies on the frequencies of least frequent bigram, trigram, 4-gram, 5-gram, 6-gram and 7-gram in the sentence.
- Frequencies obtained from a reference corpus of grammatical sentences.
N-gram-based Classifier

Classifies a sentence as ungrammatical if it contains an unusual part of speech sequence.

- The classifier relies on the frequencies of least frequent bigram, trigram, 4-gram, 5-gram, 6-gram and 7-gram in the sentence.
- Frequencies obtained from a reference corpus of grammatical sentences.
Precision-grammar-based Classifier

Classifies a sentence using a parser and a broad-coverage hand-written grammar.

- ParGram English LFG and XLE engine
- The classifier uses the following information:
  1. whether or not the sentence can be parsed (without resorting to special robustness mechanisms)
  2. parsing time
  3. number of parses found
  4. number of optimal constraints used during parsing
  5. number of unoptimal constraints used during parsing
  6. number of words (excluding punctuation)
Precision-grammar-based Classifier

Classifies a sentence using a parser and a broad-coverage hand-written grammar.

- **ParGram English LFG and XLE engine**

  - The classifier uses the following information:
    1. whether or not the sentence can be parsed (without resorting to special robustness mechanisms)
    2. parsing time
    3. number of parses found
    4. number of optimal constraints used during parsing
    5. number of unoptimal constraints used during parsing
    6. number of words (excluding punctuation)
Precision-grammar-based Classifier

Classifies a sentence using a parser and a broad-coverage hand-written grammar.

- ParGram English LFG and XLE engine
- The classifier uses the following information:
  1. whether or not the sentence can be parsed (without resorting to special robustness mechanisms)
  2. parsing time
  3. number of parses found
  4. number of optimal constraints used during parsing
  5. number of unoptimal constraints used during parsing
  6. number of words (excluding punctuation)
Statistical-parsing-based Classifier

Classifies a sentence using output of Charniak and Johnson parser

1. Trained on grammatical Wall Street Journal sentences
2. Trained on ungrammatical versions of the Wall Street Journal sentences
3. Trained on a mixture of above

The classifier uses:

1. the parse probabilities assigned by the three grammars
2. structural properties of the parse trees
Statistical-parsing-based Classifier

Classifies a sentence using output of Charniak and Johnson parser

1. Trained on grammatical Wall Street Journal sentences
2. Trained on ungrammatical versions of the Wall Street Journal sentences
3. Trained on a mixture of above

The classifier uses:

1. the parse probabilities assigned by the three grammars
2. structural properties of the parse trees
Classifier Experiments

Classifiers trained on sentences from the British National Corpus

- 50%: original grammatical sentences
- 50%: artificially created ungrammatical sentences
- 10-fold cross validation
Evaluation Metrics

- Accuracy on ungrammatical data

\[
\text{acc\_ungram} = \frac{\text{\#correctly classified as ungrammatical}}{\text{\#ungrammatical sentences}}
\]

- Accuracy on grammatical data

\[
\text{acc\_gram} = \frac{\text{\#correctly classified as grammatical}}{\text{\#grammatical sentences}}
\]

- Independent of error density of test data
Classifier Results

Accuracy graph

![Accuracy Graph Image]

Classifier Results

accuracy graph
Classifier Results

Region of improvement

![Graph showing the region of improvement with coordinates (acc ungram, acc gram).]
Classifier Results

Region of degradation

![Graph showing the region of degradation with axes labeled as acc_ungram and acc_gram. The graph has a grid with values ranging from 0 to 1 on both axes.]
Classifier Results

Undecided

![Graph showing classifier results with undecided points]
Comparing classifier results on artificial data
Comparing classifier results on artificial data
Comparing classifier results on artificial data

![Graph showing the comparison of classifier results on artificial data. The graph plots Acc_gram against Acc_ungram for different combinations of classifiers: prob, ngram+prob, xle+prob, ngram, xle+ngram, and all 3. The points are marked with diamonds.]
Comparing classifier results on artificial data

- hard to compare methods
- n-gram method inferior to prob method
Train multiple classifiers on different subsets of data

- Each classifier votes whether sentence is grammatical
- Parameter: number of votes required for final decision
- Plot accuracy for all possible parameter values
- Allows us to compare methods
Accuracy Tradeoff with Voting Scheme

![Graph showing accuracy tradeoff with voting scheme]
Applying BNC-trained classifiers to WSJ data
How can this be used in MT?

- Rank a n-best list produced by an MT system
- Add as a feature to a discriminative reranker
- See whether it correlates with MT evaluation metrics