CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

Insights from computational linguistics in teaching and assessing written English

Ted Briscoe

University of Cambridge

100
CAMBRIDGE ENGLISH
CENTENARY 1913–2013

The task: automated writing feedback

Automated writing feedback

Automatically evaluate the quality (CEFR level) of free text and provide grades and feedback

Challenges

- Accurate grades and effective and detailed feedback
- Grade with high correlation to human examiners
- Provide pedagogically useful feedback like human teachers

Applications

1. Self-assessment and self-tutoring
2. Second marking of high stakes exams
3. Computer-adaptive testing / teaching

Advantages

- Immediate and iterated grades and feedback
- Encourage exploratory writing development
- Application of consistent assessment criteria
- Reduced teacher / examiner workload
- Cost-effective approach to teaching / grading

Script-level grading

Text assessment

Overall assessment of someone's proficiency by scoring the text as a whole

1. Assess general linguistic competence
   - Identify textual features which are proxies for intrinsic qualities of writing competence
   - Predict a score using a weighted combination of the features
   - Evaluate predicted scores
2. Provide scoring feedback

Machine Learning

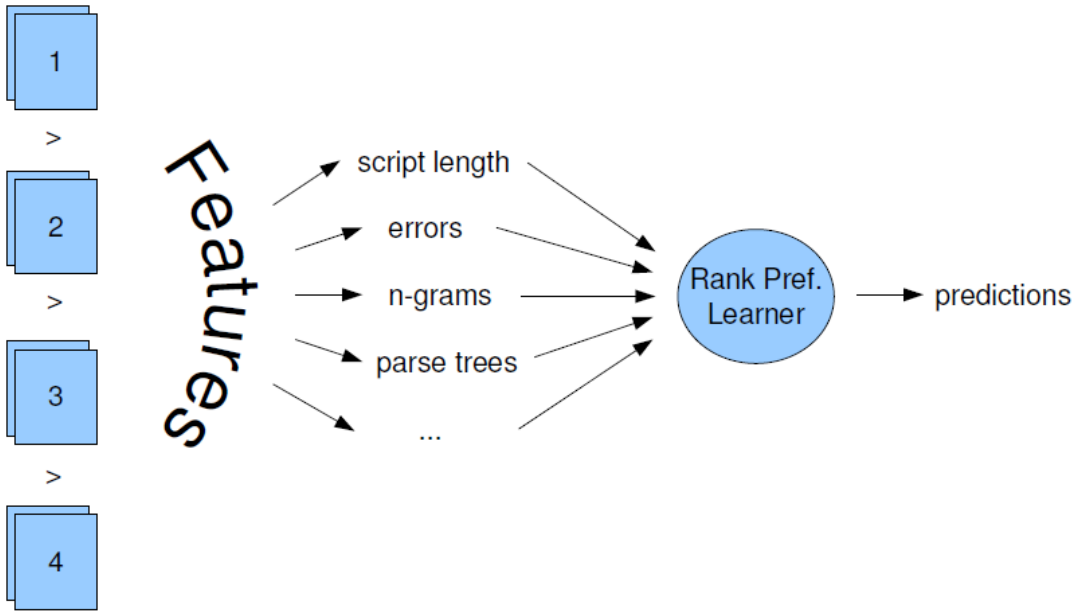
Supervised Machine Learning

- Use data to make inferences
- Use annotated data in order to make inferences about new unannotated data

Approaches

- Classification (labels)
- Regression (real values)
- Rank Preference (ordering)

Rank Preference



Feature Space

1. **Word sequences**
    1. I go
    2. suggest idea
    3. the people is

Feature Space

1. **Word sequences**
    1. I go
    2. suggest idea
    3. the people is
2. **Part-of-speech (PoS) sequences**
    1. VV0 VV0 (e.g., keep develop)
    2. NN2 VVG (e.g., children smiling)

Feature Space

1. **Word sequences**
    1. I go
    2. suggest idea
    3. the people is
2. **Part-of-speech (PoS) sequences**
    1. VV0 VV0 (e.g., keep develop)
    2. NN2 VVG (e.g., children smiling)
3. **Grammatical constructions**
    1. V1/modal bse/+- (e.g., can only travel in July)
    2. S/pp-ap s-r (e.g., for better or worse, he left)
    3. T/txt-frag (e.g., but know Kim knew)

Feature Space

1. **Word sequences**
    1. I go
    2. suggest idea
    3. the people is
2. **Part-of-speech (PoS) sequences**
    1. VV0 VV0 (e.g., keep develop)
    2. NN2 VVG (e.g., children smiling)
3. **Grammatical constructions**
    1. V1/modal bse/+- (e.g., can only travel in July)
    2. S/pp-ap s-r (e.g., for better or worse, he left)
    3. T/txt-frag (e.g., but know Kim knew)
4. **Other features**
    1. Error rate
    2. Script length

Evaluation

Correlation between human scores and the system-predicted scores

Features	Pearson's correlation	Spearman's correlation
word seq	0.601	0.598
+PoS seq	0.682	0.687
+script length	0.692	0.689
+syntax	0.714	0.712
+error rate	0.741	0.773
Upper bound	**0.796**	**0.792**

Regression vs. Rank Preference

Comparison with previous work

Model	Pearson's correlation	Spearman's correlation
Regression	0.697	0.706
Our model	0.741	0.773

CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

Script-level feedback

Overall score

An overall score is assigned on a scale from red for a text that looks like it may be at level B1 or below to green for a text that shows some evidence of being at level B2 or above.

Change answer

Some people learn a foreign language in order to widen their horizons and etc.

Perhaps you prefer to stay on dry land. Can u sea the see from were you live?

Word count: 31

Save

Word-level feedback

Error detection and correction

Ensure high precision and good coverage

1. Corpus-derived rules
   - Error rules from the Cambridge Learner Corpus (CLC)
   - Detect incorrect word sequences
   - At least 90% incorrect occurrences
2. Electronic Dictionary-derived rules

CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

Word-level feedback

Response text

Some people learn a foreign language in order to widen their horizons and etc. Perhaps you prefer to stay on dry land.

Can u sea the see from were you live?

Possible errors

and Insertion: This word may not actually be needed.

etc. Substitution: A different word might work better here. Have you contemplated using 'so on'?

u Register: The word you have chosen might be inappropriate for this writing task. If so, 'you' would be a more conventional choice.

sea Confusion: You may have confused this with a similar-looking word. 'see' seems more likely in this context.

the Insertion: This word may not actually be needed.

see Confusion: You may have confused this with a similar-looking word. 'sea' seems more likely in this context.

Sentence-level grading

Sentence evaluation

Assess and score the quality of individual sentences, independently of their context

Challenges

- Limited linguistic evidence that can be extracted automatically
- Difficulty in acquiring annotated data

Feature Space

1. Main verbs, nouns, adjectives, subordinating conjunctions and adverbs
2. Clausal subjects and modifiers
3. Affixes
4. Features representing syntax
5. Error counts
6. Number of words forming an error

Sentence-level feedback

In the past people didn't have electricity and if they wanted, for example, to read or to cook something they used to light a fire.
You must have a TV because you can learn about what is happening in the world and you can see some places that you haven't been to.
You can enjoy watching a film if you have some free time.
In our daily life, however, we seldom notice how easy a life we've got or, what is more, how difficult our grandparents found it.

In the past the people didn't have electiity and if they wanted for example to read or to cook something they used to do in the fire.
You must have TV because you can liten what it happend in the world and you can watch some places that you didn't go.
You can enjoy you time to watch a film if you have free time.
In our daily life, however, we seldom notice how much convinient life we've got, what is more, how much inconvinient our grandparents had got.

CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

Self-Assessment and Tutoring (SAT) system

Assessed answer

Overall score

An overall score is assigned on a scale from red for a text that looks like it may be at level B1 or below to green for a text that shows some evidence of being at level B2 or above.

Detailed feedback (Help)

Score feedback | Error feedback | Combined

This view combines the information contained in the score feedback and error feedback views. A red box indicates that explanations/corrections are available and can be viewed by hovering over the relevant word.

Some people learn a foreign language in order to widen their horizons and etc. Perhaps you prefer to stay on dry land.

Can u sea the see from were you live?

Change answer

Some people learn a foreign language in order to widen their horizons and etc. Perhaps you prefer to stay on dry land.

Can u sea the see from were you live?

Word count: 31

Trials

2 User Trials

- Over 450 students participated, expected to be at or above the upper-intermediate level
- 3000 submissions total, including revisions
  - Over 600,000 words
  - Average response length: 200 words
- Average number of revisions: 3.2
- Median of number of revisions: 2
- Max number of revisions: 54
- Score given to the last revision is higher than that given to the initial revision in over 80% of cases

User satisfaction

	Trial 1	Trial 2
Using the SAT system helps me to write better in English	3.80	3.92
I find the SAT system useful for understanding my mistakes	3.74	3.96
I think the sentence colouring is useful	3.74	4.15
I think the word-level information [error feedback] is useful	3.86	4.12
The SAT system is easy to use	4.45	4.49
The feedback on my writing is clear	3.80	3.93
If you have used the SAT system before, has it improved since the last time?	—	3.86

Table: Average feedback scores on a scale from 1 (strongly disagree) to 5 (strongly agree)

Conclusions

Results

- Effective feedback / grading at three different levels of granularity (script, sentence, word)
- Visualisation displays information in an intuitive and easily interpretable way
- Usefulness and usability of the tool confirmed through questionnaire-based evaluations

Future Work

- Improve methodologies used for providing error feedback
- Add further functionality
  - L1-specific feedback
  - Discourse organisation feedback
- Gather more and better-annotated data

CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

Thank you!

For Papers:

Search:

’Alta Institute Cambridge’

URL:

http://www.cl.cam.ac.uk/ ejb1/epp.html

100 CAMBRIDGE ENGLISH CENTENARY 1913–2013