NUS at the HOO 2011 Pilot Shared Task

Daniel Dahlmeier[1], Hwee Tou Ng[1,2], and Thanh Phu Tran[2]

[1]NUS Graduate School for Integrative Sciences and Engineering
[2]Department of Computer Science, National University of Singapore

System Architecture

Input ⇒ Spelling correction ⇒ Article correction ⇒ Preposition correction ⇒ Output

Spelling Correction

- Open-source spell checker Aspell.
- In-domain dictionary from ACL-ANTHOLOGY.
- Exclude words which are too short or include hyphens or capitalization.

Language Model Filter

- In each step, only keep corrections that strictly increase the normalized language model score.

Article Correction

- Multi-class linear classifier.
- Normalize *an* to *a* for classification.
- Predict most likely article for each NP

$$\hat{y} = \arg\max_{y \in \{a, the, \epsilon\}} \mathbf{u}_y^T X.$$

- Rule-based heuristic to restore *an*.
- Thresholding: only propose correction if

$$\|\mathbf{u}_{\hat{y}}^T X - \mathbf{u}_{y_{observed}}^T X\| > t.$$

Preposition Correction

- Multi-class linear classifier.
- Predict most likely preposition for each PP.
- Possible classes: *about, among, at, by, for, in, into, of, on, to*, and *with*.
- Thresholding: only propose correction if

$$\|\mathbf{u}_{\hat{y}}^T X - \mathbf{u}_{y_{observed}}^T X\| > t.$$

Experiments

- Split HOO development data into HOO-TUNE and HOO-HELDOUT.
- Train classifiers on well-edited examples from CL-JOURNAL.
- Train language model filters on ACL-ANTHOLOGY (spelling) and WEB 1T 5-GRAM (articles, prepositions).
- Tune threshold parameters on HOO-TUNE when testing on HOO-HELDOUT or on the complete HOO development data when testing on HOO-TEST.

Evaluation

- Overall micro-averaged F_1 detection,recognition, and correction scores as defined in the official HOO overview paper.
- Individual F_1 scores for each error category.

$$P = \frac{\#\{\text{detected / recognized / corrected}\}}{\#\{\text{proposed corrections}\}}$$

$$R = \frac{\#\{\text{detected / recognized / corrected}\}}{\#\{\text{gold corrections}\}}$$

$$F_1 = 2 \times \frac{P \times R}{P + R}$$

- With bonus scores, give credit for missed optional corrections.

HOO Held-Out Development Data

Step	Detection		Recognition		Correction	
	wb	w/o b	wb	w/o b	wb	w/o b
PRE	.2152	.0000	.2152	.0000	.2152	.0000
+SPEL	.2219	.0095	.2190	.0063	.2162	.0031
+ART	.2681	.1093	.2520	.0917	.2455	.0846
+PREP	.2973	.1354	.2763	.1123	.2657	.1008

Table: Overall F_1 scores with (wb) and without bonus (w/o b) on the HOO-HELDOUT data after pre-processing (PRE), spelling (SPEL), article (ART), and preposition correction (PREP).

Step	Detection		Recognition		Correction	
	wb	w/o b	wb	w/o b	wb	w/o b
SPEL	.2667	.2667	.2667	.2667	.2667	.2667
ART	.3455	.3011	.3455	.3011	.3246	.2796
PREP	.2692	.2353	.2308	.1961	.1731	.1373

Table: Individual F_1 scores for each error category with (wb) and without bonus (w/o b) on the HOO-HELDOUT data.

HOO Test Data

Step	Detection		Recognition		Correction	
	wb	w/o b	wb	w/o b	wb	w/o b
PRE	.1553	.0000	.1553	.0000	.1553	.0000
+SPEL	.1663	.0093	.1629	.0093	.1611	.0075
+ART	.2718	.1552	.2545	.1373	.2209	.1014
+PREP	.2840	.1774	.2686	.1615	.2274	.1177

Table: Overall F_1 scores with (wb) and without bonus (w/o b) on the HOO-TEST data.

Step	Detection		Recognition		Correction	
	wb	w/o b	wb	w/o b	wb	w/o b
SPEL	.4706	.4706	.4706	.4706	.4706	.4706
ART	.3591	.3404	.3466	.3277	.2630	.2426
PREP	.3409	.2000	.3409	.2000	.2614	.1200

Table: Individual F_1 scores for each error category with (wb) and without bonus (w/o b) on the HOO-TEST data.

{danielhe, nght, thanhphu}@comp.nus.edu.sg