Noisy Channel for Low Resource Grammatical Error Correction

Simon Flachs, Ophélie Lacroix and Anders Søgaard

sfl@siteimprove.com

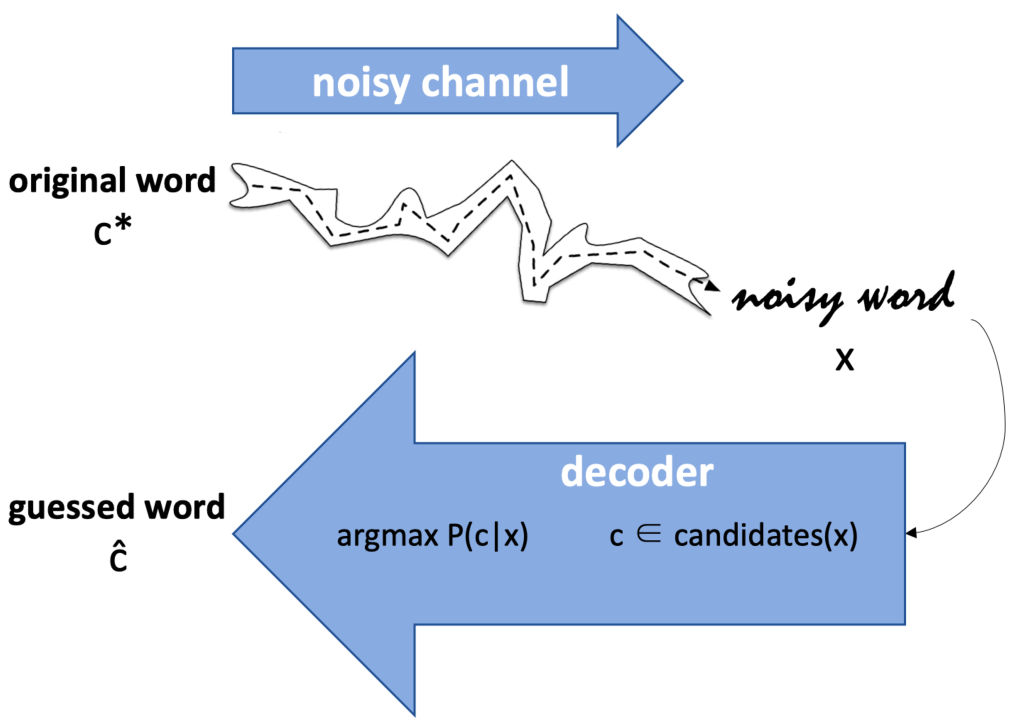
Introduction

- Our contribution to the low-resource track of the BEA 2019 shared task on GEC
- Ranked as the 6th best performing system

In our approach we:

- Formalize GEC in the **noisy channel** framework,
- Generate confusion sets from the Wikipedia edit history
- Estimate a channel model based on edit frequency counts
- Combine existing pre-trained language models
- Use beam search to find the optimal combination of corrections

The Noisy Channel



Intuition

- Each word, $\boldsymbol{x}$, in a sentence has a true underlying word, $\boldsymbol{c^*}$
- $\boldsymbol{c^*}$ has been passed through a noisy communication channel
- The channel has potentially modified $\boldsymbol{c^*}$ into an erroneous surface form

Method

- **Goal**: find the hidden word, $\boldsymbol{c^*}$, that generated $\boldsymbol{x}$.
- Use a confusion set, $\boldsymbol{C}$, of candidates for each $\boldsymbol{x}$
- Choose $\boldsymbol{c} \in \boldsymbol{C}$ that maximizes $\boldsymbol{P(c|x)}$

$$\begin{aligned}\hat{c} &= \arg\max_{c \in C} P(c|x) \\ &= \arg\max_{c \in C} \underbrace{P(x|c)}_{\text{channel model}} * \underbrace{P(c)}_{\text{language model}} \qquad \text{(Bayes' rule)}\end{aligned}$$

System

Channel Model

Non-word errors

- Words not in vocabulary and not named entities
- Use the inverse Levenshtein distance to distribute probability between the candidates

Real-word errors

- Assumption: the probability of $\boldsymbol{x}$ being wrong is 5% ($\boldsymbol{\alpha} = 0.05$)
- $\boldsymbol{\alpha}$ is distributed between candidates based on edit frequency counts from the Wikipedia revision history.

Language Models

BERT

- Conditions on both left and right context.
- We train multiple models specialized on different tasks
  - PoS tag prediction: verb form (VB/VBG/VBN/VBP/VBZ) and noun number (NN/NNS) errors
  - Word prediction: real-word and non-word errors
  - Comma prediction: we remove all commas and let the model predict where to insert commas.

GPT-2

- Only conditions on left context.
- We include the previous sentence when estimating probabilities.

Combination

- Combine components to make the final prediction:

$$\hat{c} = \arg\max_{c \in C} \; P_{Channel}(x|c) * P_{BERT}(c) * P_{GPT-2}(c)$$

- Use beam search to efficiently explore combinations of corrections in order to find the optimal output sentence (beam width = 3).

Confusion Sets

- **Real-word confusions** Gather 348 edit pairs from Wikipedia revision histories
- **Non-word confusions** Use suggestions from the Enchant library
- **Noun number confusions** Use singular/plural nouns derived from Wiktionary
- **Verb form confusions** Use all possible verb inflections derived from the Unimorph project

Results

Error type	#	P	R	$F_{0.5}$
M:Punct	422	80.10	38.15	65.66
R:Adj	24	12.50	4.17	8.93
R:Adv	17	33.33	5.88	17.24
R:Conj	5	2.22	20.00	2.70
R:Det	129	20.48	52.71	23.34
R:Morph	128	46.15	18.75	35.71
R:Noun	70	50.00	8.57	25.42
R:Noun:Infl	19	42.86	31.58	40.00
R:Noun:Num	290	43.79	68.31	47.18
R:Orth	349	10.20	1.43	4.59
R:Other	618	20.43	6.15	13.95
R:Part	15	38.89	46.67	40.23
R:Prep	292	39.49	58.56	42.24
R:Pron	50	34.15	56.00	37.04
R:Spell	321	76.51	75.08	76.22
R:Verb	134	25.00	2.99	10.10
R:Verb:Form	169	47.96	55.62	49.32
R:Verb:Infl	7	100.00	85.71	96.77
R:Verb:SVA	146	74.39	83.56	76.06
R:Verb:Tense	160	42.50	10.62	26.56
U:Punct	118	34.90	88.14	39.69
All error types	4498	44.52	28.88	40.17

Ablation Analysis

	P	R	$F_{0.5}$
Chan + BERT + GPT + beam	44.52	28.88	40.17
- beam	40.29	29.19	37.44 (-2.73)
- GPT	37.03	28.98	35.08 (-5.09)
- BERT	42.31	29.89	39.06 (-1.11)
- Chan	43.50	29.49	39.73 (-0.44)

Conclusion & Future work

- Approached GEC with a **noisy channel** framework
- Explored combinations of different language models, a channel model and beam search
- Each of the components has a positive effect

Future work

- Explore using more advanced channel models (e.g. using phonetic features)
- Adapt to handle insertions and deletions