Spelling Error Correction with Soft-Masked BERT

Authors: Shaohua Zhang, Haoran Huang, Jicong Liu and Hang Li

발표자: 구선민

Introduction

Spelling error correction

- 주어진 문장에서 spelling errors를 감지하고, grammatical fashion으로 자동으로 교정
- 기본적으로 인간 수준의 언어 이해 능력이 필요하므로 중요하지만 어려운 작업

- 기존 BERT는 input sequence의 15%만 Masking 되어 오류 감지를 위한 충분한 양의 학습이 되지 못함
- 따라서 오류 확률을 예측하는 Detection Network와 오류 보정 확률을 예측하는 Correction Network 구조를 가진 Soft-Masked BERT를 제안

Model

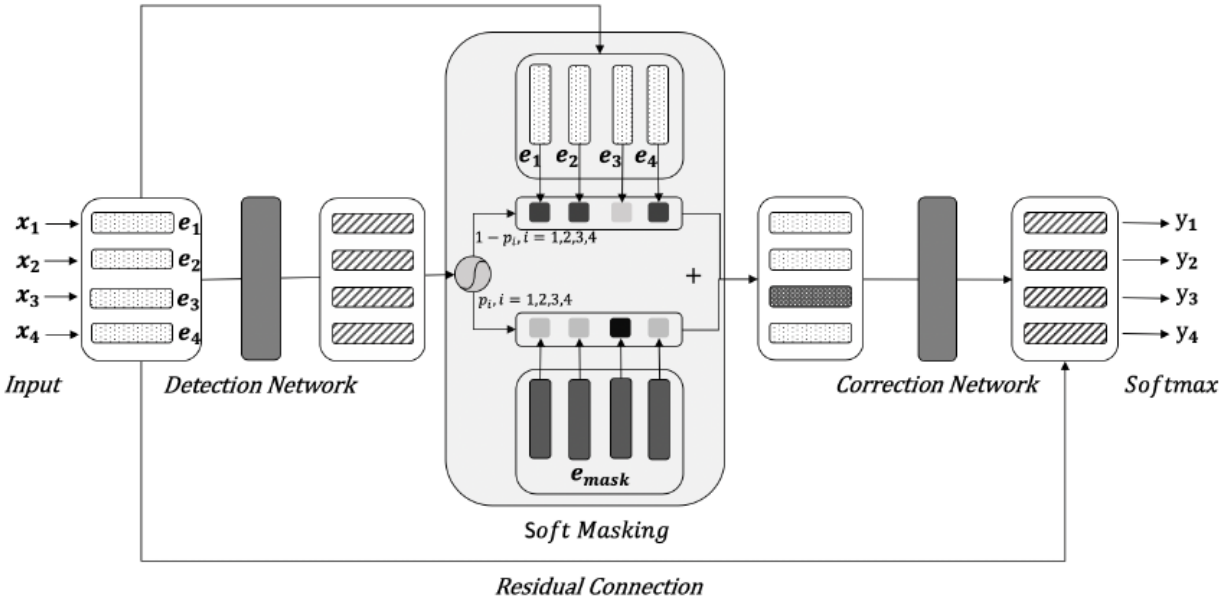

Figure 1: Architecture of Soft-Masked BERT

Detection Network: 오류 확률을 예측

Correction Network: 오류 보정 확률을 예측

- Input sequence의 각 character를 embedding
- detection network에 sequence of embeddings를 input으로 넣어 sequence of characters에 대한 오류 확률 outputs
- 오류 확률에 따라 가중치가 부여된 입력 임베딩과 [MASK] 임베딩의 가중 합계를 계산하여 soft way로 error 마스킹
- Correction network sequence of soft-masked embeddings를 input으로 넣어 오류 교정 확률 outputs

Model

Detection Network

- Sequential binary labeling model

input $E = (e_1, e_2, \cdots, e_n)$: BERT의 3가지 임베딩(Word embedding + Position embedding + segment embedding으로 생성)

output $G = (g_1, g_2, \cdots, g_n)$: 0 – correct, 1-incorrect

$$p_i = P_d(g_i = 1|X) = \sigma(W_d h_i^d + b_d)$$

$$\overrightarrow{h_i^d} = \text{GRU}(\overrightarrow{h}_{i-1}^d, e_i)$$

$$\overleftarrow{h_i^d} = \text{GRU}(\overleftarrow{h}_{i+1}^d, e_i)$$

$$h_i^d = [\overrightarrow{h_i^d}; \overleftarrow{h_i^d}]$$

$$e_i^{'} = p_i \cdot e_{mask} + (1 - p_i) \cdot e_i$$

P : 각 character가 incorrect일 확률

h: GRU의 hidden state
양방향이기 때문에 GRU의 concatenation 결과로 사용

e': soft masked embedding
오류 확률이 높으면 e_{mask}로, 낮으면 e_i로 임베딩

Model

Correction Network

- Sequential multi-class labeling model based on BERT

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1; \cdots, \text{head}_h)W^O$$

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V)$$

$$\text{FFN}(X) = \max(0, XW_1 + b_1)W_2 + b_2$$

$$P_c(y_i = j|X) = \text{softmax}(Wh_i' + b)[j]$$

$$h_i' = h_i^c + e_i$$

P_c : 각 character의 error correction 확률

h_i' : hidden state

h_i^c : final layer의 hidden state

e_i: character x의 input embedding

Model

Learning

- Sequential multi-class labeling model based on BERT

$$\text{as} \ = \{(X_1, Y_1), (X_2, Y_2), \ldots, (X_N, Y_N)\}.$$

$$\mathcal{L}_d = -\sum_{i=1}^{n} \log P_d(g_i|X)$$

$$\mathcal{L}_c = -\sum_{i=1}^{n} \log P_c(y_i|X)$$

$$\mathcal{L} = \lambda \cdot \mathcal{L}_c + (1 - \lambda) \cdot \mathcal{L}_d$$

- Original sequence와 corrected sequence의 pair로 구성
- Learning process는 error detection과 error correction을 optimizing

- L_d: objective for training of the detection network
- L_c: objective for training of the correction network
- 두 개의 함수 선형 결합

Experiments Results

Table 2: Performances of Different Methods on CSC

Test Set	Method	Detection				Correction			
		Acc.	Prec.	Rec.	F1.	Acc.	Prec.	Rec.	F1.
SIGHAN	NTOU (2015)	42.2	42.2	41.8	42.0	39.0	38.1	35.2	36.6
	NCTU-NTUT (2015)	60.1	71.7	33.6	45.7	56.4	66.3	26.1	37.5
	HanSpeller++ (2015)	70.1	**80.3**	53.3	64.0	69.2	**79.7**	51.5	62.5
	Hybird (2018b)	-	56.6	69.4	62.3	-	-	-	57.1
	FASPell (2019)	74.2	67.6	60.0	63.5	73.7	66.6	59.1	62.6
	Confusionset (2019)	-	66.8	73.1	69.8	-	71.5	59.5	64.9
	BERT-Pretrain	6.8	3.6	7.0	4.7	5.2	2.0	3.8	2.6
	BERT-Finetune	80.0	73.0	70.8	71.9	76.6	65.9	64.0	64.9
	Soft-Masked BERT	**80.9**	73.7	**73.2**	**73.5**	**77.4**	66.7	**66.2**	**66.4**
News Title	BERT-Pretrain	7.1	1.3	3.6	1.9	0.6	0.6	1.6	0.8
	BERT-Finetune	80.0	65.0	61.5	63.2	76.8	55.3	52.3	53.8
	Soft-Masked BERT	**80.8**	**65.5**	**64.0**	**64.8**	**77.6**	**55.8**	**54.5**	**55.2**

Table 3: Impact of Different Sizes of Training Data

Train Set	Method	Detection				Correction			
		Acc.	Prec.	Rec.	F1.	Acc.	Prec.	Rec.	F1.
500,000	BERT-Finetune	71.8	49.6	48.2	48.9	67.4	36.5	35.5	36.0
	Soft-Masked BERT	**72.3**	**50.3**	**49.6**	**50.0**	**68.2**	**37.9**	**37.4**	**37.6**
1,000,000	BERT-Finetune	74.2	54.7	51.3	52.9	70.0	41.6	39.0	40.3
	Soft-Masked BERT	**75.3**	**56.3**	**54.2**	**55.2**	**71.1**	**43.6**	**41.9**	**42.7**
2,000,000	BERT-Finetune	77.0	59.7	57.0	58.3	73.1	48.0	45.8	46.9
	Soft-Masked BERT	**77.6**	**60.0**	**58.5**	**59.2**	**73.7**	**48.4**	**47.3**	**47.8**
5,000,000	BERT-Finetune	80.0	65.0	61.5	63.2	76.8	55.3	52.3	53.8
	Soft-Masked BERT	**80.8**	**65.5**	**64.0**	**64.8**	**77.6**	**55.8**	**54.5**	**55.2**

감사합니다