

# BERT for Joint Intent Classification and Slot Filling

## Prerequisites

- Language Model : . Next Word Prediction
- Encoder-Decoder : 가 . 1 RNN ,  
1 RNN (concat)
- Transformer :  
- RNN , Context Vector  
1. 가 2. (Vanishing Gradient)
- BERT : Transformer Encoder

## Joint BERT

- BERT NLP tasks SOTA , NLU (Natural Language Understanding)
- NLU (Intent classification) (Slot filling)가 task , Joint BERT BERT  
가 task (Joint)

## NLU

- 가 ,
- NLP (utterance) 가 (semantic parse)
- Google Home, Amazon Alexa , (Chatbot)

## NLU Tasks

- Intent Classification : ( )
- Slot Filling : (utterance) task (slot label)  
- )  
- intent :  
- entity : ( )  
- slot filling (slot labels) : , , ( , .. )  
- )  
- intent :  
- slot filling : 가 = , =

## Joint Intent Classification and Slot Filling

- Intent Classification :  
- [CLS]  
- BERT 가 [CLS] , intent BERT

$$y^i = \text{softmax}(\mathbf{W}^i h_1 + \mathbf{b}^i)$$

- Slot Filling :

- [CLS] slot filling label
- [CLS] BERT
- WordPiece , 가 sub , 가
- ) pairing --> pair, ##ing pair

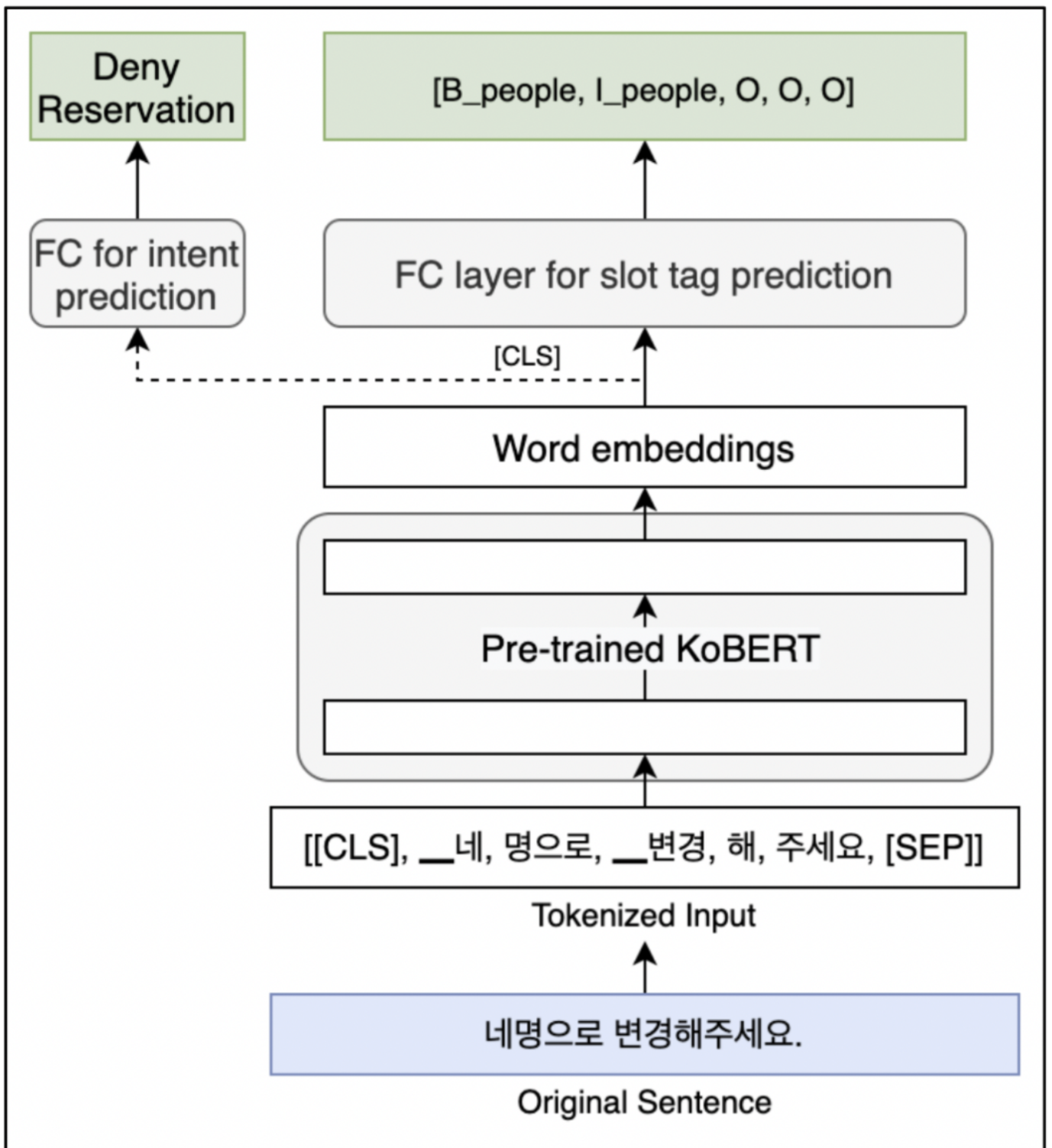
$$y_n^s = \text{softmax}(\mathbf{W}^s \mathbf{h}_n + \mathbf{b}^s), n \in 1 \dots N$$

- Tokenization
  - Token compatibility BERT WordPiece tokenizer
- Objective function
  - Intent classification Slot filling (Conditional Probability)

$$p(y^i, y^s | \mathbf{x}) = p(y^i | \mathbf{x}) \prod_{n=1}^N p(y_n^s | \mathbf{x})$$

- $\mathbf{x}$  = input token sequence
- (Pi product notation)
- $P(X, Y | Z) = P(X | Z) * P(Y | Z)$
- $y^i: 1$  ([CLS] )
- $y^s:$  ( )

- Joint BERT Architecture



## Conditional Random Field (CRF)

- Random Field : The representation of a joint distribution for a given set of random observations
- Conditional : Conditional Probability
- (sentence-level) (network) LSTM ,
- Slot Label Prediction (dependent) CRF

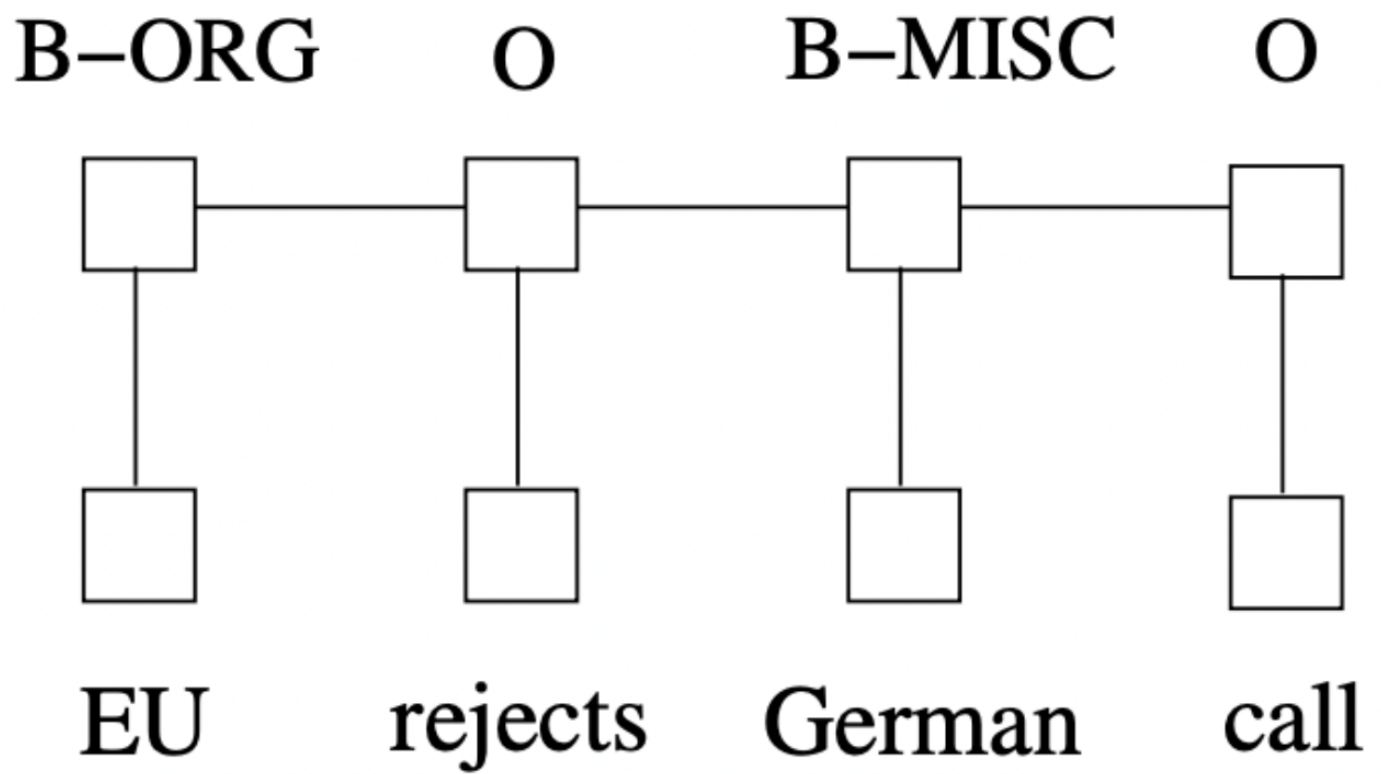
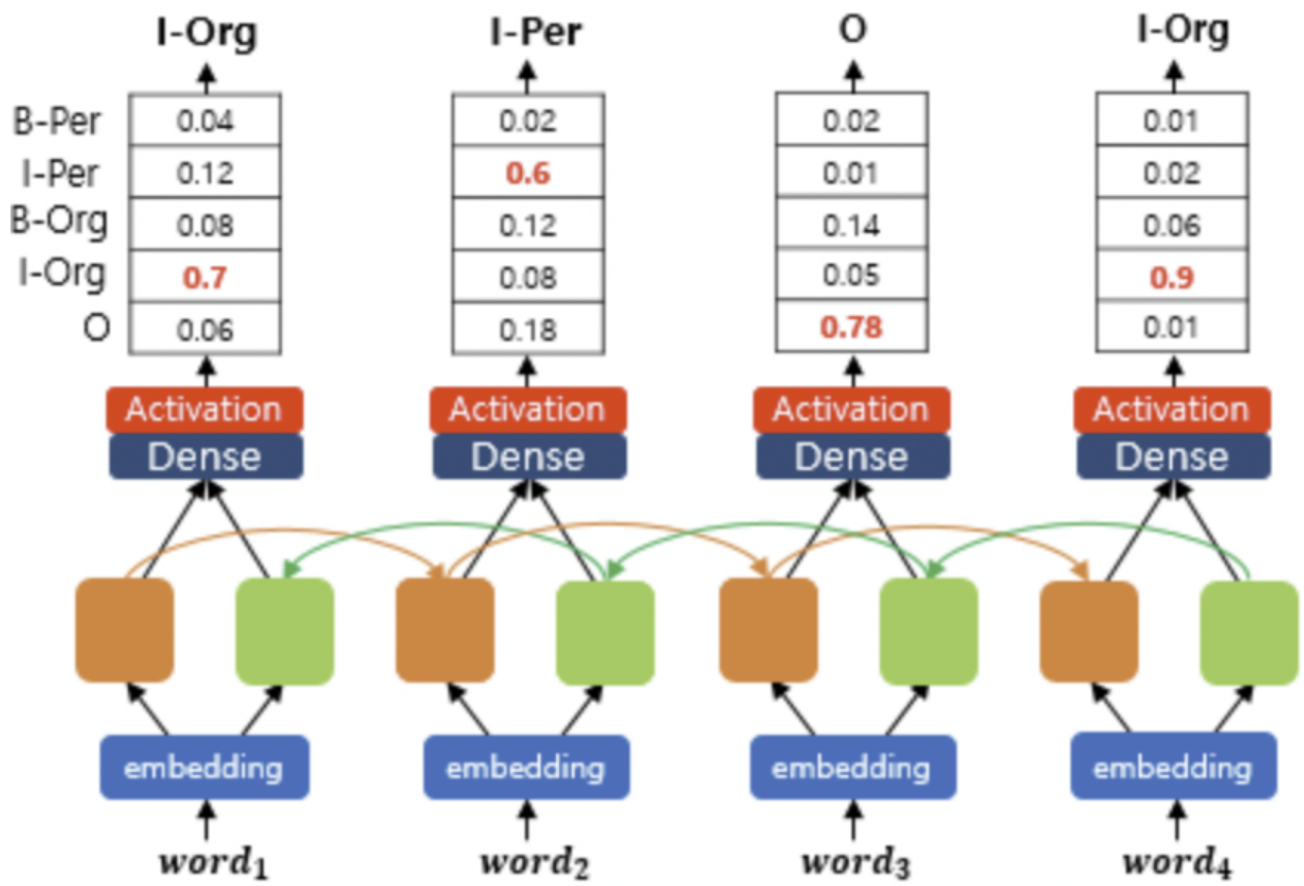


Figure 5: A CRF network.

- BIO representation : B(Begin), I(Inside), O(Outside). (B) (I) (I) (I) (O) (O)가(O) (O)
- “ B-Per ”, “ I-Per ”, “ B-Org ”, “ I-Org ”, “ O ” BIO , I



• CRF

가

가

. CRF





Models	Snips			ATIS		
	Intent	Slot	Sent	Intent	Slot	Sent
RNN-LSTM (Hakkani-Tür et al., 2016)	96.9	87.3	73.2	92.6	94.3	80.7
Atten.-BiRNN (Liu and Lane, 2016)	96.7	87.8	74.1	91.1	94.2	78.9
Slot-Gated (Goo et al., 2018)	97.0	88.8	75.5	94.1	95.2	82.6
Joint BERT	<b>98.6</b>	<b>97.0</b>	<b>92.8</b>	97.5	<b>96.1</b>	88.2
Joint BERT + CRF	98.4	96.7	92.6	<b>97.9</b>	96.0	<b>88.6</b>

} Baseline

### Ablation Analysis

- Joint
- The jointBERT model fine-tuned with only 1 epoch already outperforms the first group of models in Table 2.

Model	Epochs	Intent	Slot
Joint BERT	30	98.6	97.0
No joint	30	98.0	95.8
Joint BERT	40	98.3	96.4
Joint BERT	20	99.0	96.0
Joint BERT	10	98.6	96.5
Joint BERT	5	98.0	95.1
Joint BERT	1	98.0	93.3

Table 3: Ablation Analysis for the Snips dataset.

intent_cls.png	124 KB	2022-03-14
slot_fill.png	118 KB	2022-03-14
obj_func.png	171 KB	2022-03-14
architec.png	2.24 MB	2022-03-14
crf.png	1.35 MB	2022-03-14
crf_network.png	635 KB	2022-03-14
crf2.png	893 KB	2022-03-14
metric_result.png	410 KB	2022-03-14
ablation.png	260 KB	2022-03-14