



label : ["B", "I", "O", "X", "[CLS]", "[SEP]"]

num-labels = label 개수 + 1

argmax로 최댓값 → 각 token에 해당하는 label

softmax

[batch_size, max_seq_size, 7]

BERT_VER

reshape

[batch_size * seq_length, num_labels]

W [num_labels, hidden_size] + b

[batch_size * seq_length, hidden_size]

drop_out, reshape → W와 차원을 맞추기 위해

final_hidden_layer

←----- get-sequence-output [batch 크기, seq_length, hidden_size]

BERT

; return final hidden layer of Encoder embedding_size