

Bi-LSTM with CRF and Attention on Named Entity Recognition

Group 104

Shengbo Qin



1 Data preprocessing

1.1 Word Tokenization

Word tokenization is applied to both training sentence, training label, validation sentence, validation label and test sentence to tokenize into words.

1.2 Case-folding

Case-folding is performed to training, validation, and test sentence by reducing all letters to lower case, since ‘CAR’, ‘Car’ and ‘car’ have same meaning in natural language, but they are different for machine.

1.3 Lemmatization

Lemmatization is applied to both training, validation and test sentence, it reduces inflections or variant forms to base form, to reduce complexity of model but retain the meaning of the words.

2 Input Embedding

2.1 Syntactic Textual Feature Embedding

Dependency Parsing is the syntactic textual feature input embedding for the model. Applying dependency parsing can help the model to avoid Syntactic Ambiguities problem, to get a better understanding of sentence structure. The structure of a sentence is described in terms of a set of binary relations between words in the sentence. Larger constituency is not directly encoded in the dependency analysis. [1]

A pre-trained statistical model for English is imported through Spacy api. Since the data is tokenized and for dependency parsing, we need sentences instead of token, tokens are joined for each sentence and get dependency parse for each word. A Word2Vec

model is trained to get embedding matrix for each dependency parsing and we keep 50 dimensions of word as syntactic textual feature embedding input.

2.2 Semantic Textual Feature Embedding

Word embedding matrix is generated by pre-trained model ‘glove-twitter-100’ with dimensions of 100 for each word.

2.3 Domain Feature Embedding

Term Frequency-Inverse Document Frequency (TFIDF) is selected as domain feature input of model.

$$w_{i,j} = tf_{i,j} \times \log \left(\frac{N}{df_i} \right)$$

$w_{i,j}$ = weight of term i in document j

$tf_{i,j}$ = number of occurrences of term i in document j

N = total number of documents

df_i = number of documents containing term i

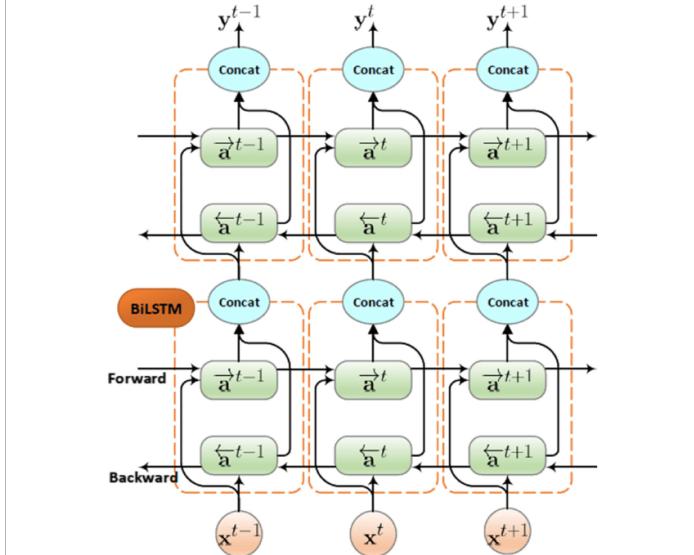
(Figure 1: TF-IDF Calculation Equation)

The term frequency (TF) is calculated by number of occurrences of a word in a given document. The document frequency is calculated by number of documents containing the word. TFIDF is used to calculate the importance of a word to a document in a collection. A high importance keyword should have high TFIDF with high TF and low DF, so that it filters out common terms and only keep important keyword. [2]

3 NER model

3.1 Stacked Seq2Seq model

A Bi-LSTM model with multiple stack layers is implemented in the model. Comparing to RNN model, a LSTM model can solve long-term dependency problem where RNN cannot remember long-term memory, but LSTM can remember with its special structure of ‘gate’. Bi-LSTM takes not only words before selector as input, but also takes words after selector as input. So Bi-LSTM has a better understanding to the word content since it “read” both previous sequence and next sequence. A stacked Bi-LSTM model is implemented in the model and tested for the best number of layers (Figure 2).



(Figure 2: stacked Bi-LSTM model) [3]

3.2 Attention

Attention is added into the model to help the model to extract critical information and improve the performance of the model by assigning different weights to each part of the input. Attention weighs the hidden state of all steps and focus on the more important hidden state information in the entire text. Three types of attention score calculation is added into the model where S_t^T and h_i represent query and key for a hidden state:

3.2.1 Dot-product

$$\text{Score}(S_t, h_t) = S_t^T h_t$$

A dot-product calculation is applied to query and key to get the attention weight of the hidden state. Then SoftMax is used to normalize the calculated weight. Next use dot-product calculation with normalized weight and value to get the attention output. In this model, key and value are the same, that is key = value.

3.2.2 Scaled dot-product

$$\text{Score}(S_t, h_t) = \frac{S_t^T h_t}{\sqrt{n}}$$

Scaled dot-product uses dimension of hidden state as a scaling factor to scale dot-product of query and key to calculate the score of attention. The rest steps are as same as dot product calculation as section 3.2.1.

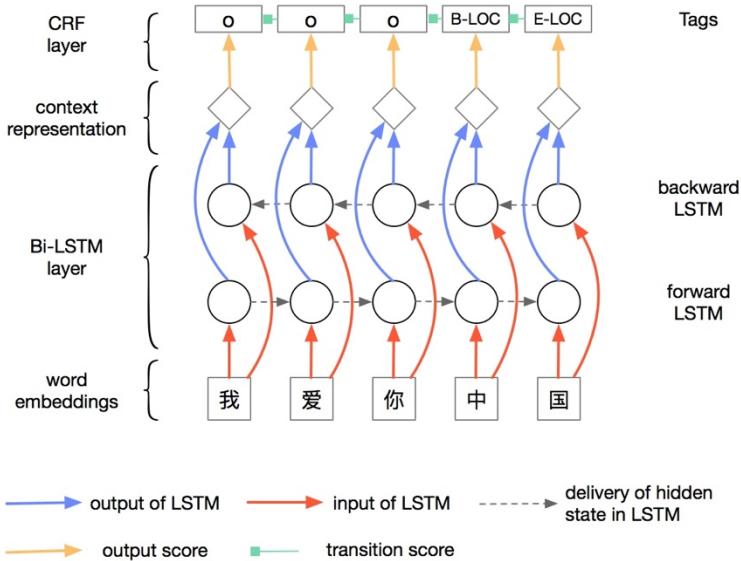
3.2.3 Cosine similarity

$$\text{Score}(S_t, h_t) = \text{cosine}[S_t, h_t]$$

Instead of dot product, this attention calculation calculates cosine similarity between query and key to find out the attention weight based on content similarity. [4]

3.3 CRF Attachment

CRF calculates the probability distribution of possible sequence y under the condition of given x to predict current tags by using neighbor tag information. Hidden Markov Model is probabilistic graphical model to predict hidden variables from observed variables. CRF uses HMM but without hypotheses of the observations are independent so that the model can handle the dependency between predicted entity. Viterbi algorithm is used to find out the best path with path score by using dynamic programming so that the global best sequence is returned (Figure 3).



(Figure 3: Bi-LSTM with CRF model) [5]

There is one bug with the model without CRF when calculating accuracy and generating classification report. We believe the cal_acc function need to be modified to fit the Bi-LSTM output instead of Bi-LSTM with CRF output.

4 Evaluation

4.1 Evaluation setup

4.1.1 Datasets

The main experiments are conducted on part of re3d dataset from Defense Science and Technology Laboratory, U.K, which focusing on entity and relationship extraction relevant to somebody operating in the role of a defense and security intelligence analyst. [6]

4.1.2 Baseline Model

We compare our model with one representative and state-of-the-art models:

Bi-LSTM CRF: Bi-LSTM CRF model is a state-of-the-art NER model without considering attention and stacked layer of Bi-LSTM. Bi-LSTM precisely capture the meaning of each word and relationships between input token where CRF are classifiers to capture the dependency between predicted labels.

4.1.3 Implementation

For hyper-parameter settings, the hidden state size of BiLSTM is 200. The dropout rate is set as 0.2. We use the SGD optimizer with learning rate 0.01. We use glove-twitter-100 pre-trained word2vec model to get semantic embedding. We use 20 epochs to train all models.

4.2 Evaluation result

4.2.1 Performance Comparison

Models	F1 Score
Base Model	0.7069
Baseline Model	0.8083
Ours - Best Model	0.8282

(Table 1: Performance Comparison)

The base model has no attention implemented, no CRF implemented, stacked layer is 1, and has only semantic as input. The model has only one layer of Bi-LSTM so the model is not complex enough to learn features of input data, therefore f1 score is low. Comparing base model with baseline model (Table 1), baseline model has all three different input embedding with CRF implemented, the f1 score is 10% higher than base model. For our best model, we have scale dot product attention implemented, CRF implemented with 2 stacked layer of Bi-LSTM and all three different input embedding,

the f1 score has improved 2% compared to baseline model, with the major difference of attention and stacked layer of Bi-LSTM.

4.2.2 Different input embedding model

Models	F1 Score
Base Model (using semantic)	0.8083
+ syntactic (dependency parsing)	0.7914
+ domain (TF-IDF)	0.8166
+ ALL (semantic, syntactic, domain)	0.8197

(Table 2: Models with different input embedding model)

There is one interesting thing we found during experiment (Table 2). We added syntactic embedding to the input, which is the dependency parsing of input data, our accuracy decreases a bit compare to base model, however, when we add all input embedding to the model, we have the best result comparing other types of input. One potential reason can be syntactic embedding is very sensitive to the sentence length, because as the length increases, the effective of syntactic embedding will decrease so accuracy will decrease. However, TF-IDF finds out the importance of a word to a document collection or corpus, so it added value to each word in terms of importance and frequency, it may increase f1 score combining with syntactic embedding.

4.2.3 Different attention strategy

Models	F1 Score
Base Model	0.8197
+ dot product	0.8195
+ scaled dot product	0.8165
+ cosine similarity	0.7791

(Table 3: Models with different attention strategy)

Attention does not perform well in this model (Table 3). One potential reason can be that in this dataset all sentences and tokens have relative average importance of value and even after applying attention to the model, the attention weights are close for all hidden state so that it does not add much value to the model. The dot product and scaled dot product attention have similar f1 score compare to base model which does not have any attention. However, cosine similarity attention decreases accuracy we consider our model have potential defects for cosine similarity attention calculation, but we couldn't find out.

4.2.4 Different Stacked layer

Models	F1 Score
Base Model (layer =1)	0.8197
layer =2	0.8129
layer =3	0.8093
layer =5	0.7977
layer =8	0.6485

(Table 4: Models with different stacked layer)

Multiple stacked layers of Bi-LSTM model increase complexity of model. In this task, the f1 score does not change much when layer is 1, 2 or 3 (Table 4). However when we increase layer to 5 or 8, the accuracy decreases significantly. After checking output of model for text input for layer number is 8, we found that the model classifies all tokens as O (Figure 4), we believe this is because when layer is 8, the model is having gradient vanishing problem between layers, this is because the model is too complex and our input data is relatively small, so that the gradient of layers close to input state update slow and little, the efficiency of model drops significantly.

	precision	recall	f1-score	support
B-DocumentReference	0.0000	0.0000	0.0000	20
B-Location	0.0000	0.0000	0.0000	186
B-MilitaryPlatform	0.0000	0.0000	0.0000	16
B-Money	0.0000	0.0000	0.0000	5
B-Nationality	0.0000	0.0000	0.0000	8
B-Organisation	0.0000	0.0000	0.0000	280
B-Person	0.0000	0.0000	0.0000	102
B-Quantity	0.0000	0.0000	0.0000	55
B-Temporal	0.0000	0.0000	0.0000	47
B-Weapon	0.0000	0.0000	0.0000	38
I-DocumentReference	0.0000	0.0000	0.0000	83
I-Location	0.0000	0.0000	0.0000	265
I-MilitaryPlatform	0.0000	0.0000	0.0000	16
I-Money	0.0000	0.0000	0.0000	10
I-Nationality	0.0000	0.0000	0.0000	1
I-Organisation	0.0000	0.0000	0.0000	416
I-Person	0.0000	0.0000	0.0000	173
I-Quantity	0.0000	0.0000	0.0000	26
I-Temporal	0.0000	0.0000	0.0000	62
I-Weapon	0.0000	0.0000	0.0000	45
O	0.6485	1.0000	0.7867	3420
accuracy			0.6485	5274
macro avg	0.0309	0.0476	0.0375	5274
weighted avg	0.4205	0.6485	0.5102	5274

(Figure 4: Result of model when number of layers is 8)

4.2.5 With/Without CRF

Models	F1 Score
With CRF	0.8197
Without CRF	0.7069

(Table 5: Models with/without CRF)

The f1 score of model increases significantly when CRF is added into model (Table 5). We believe this is because model without CRF will always find the label with highest score as output in Bi-LSTM but in real word language there are label dependency exist and sometimes the label with highest score is not correct. CRF can help model to learn label dependency such that some label cannot be the first of sentence and label a cannot follow label b.

References

1. Speech and Language Processing, Daniel Jurafsky & James H. Martin, December 2020, [online], available at: <https://web.stanford.edu/~jurafsky/slp3/14.pdf>
2. KNN TFIDF Based Named Entity Recognition, B.Upendraa, Dr. A. Sudheer Babu, December 2016, [online], available at: <https://www.ijtsdr.org/papers/IJSDR1612008.pdf>
3. COMP5046-2021S1-A2, Dr. Caren Han, June 2021, [online], available at: <https://www.kaggle.com/c/2021-comp5046-a2/overview/ner-model>
4. Neural Turing Machines, Alex Graves, Greg Wayne, Ivo Danihelka, December 2014, [online], available at: <https://arxiv.org/pdf/1410.5401v2.pdf>
5. Chinese NER based Bi-LSTM and CRF, November 2017, [online], available at: <http://www.kiwee.me/chinese-ner-based-bi-lstm-and-crf>
6. entity-recognition-datasets, Juan Diego Rodriguez, September 2018, [online], available at: <https://github.com/juand-r/entity-recognition-datasets/tree/master/data/re3d>

Appendix

Running log with different configuration of model:

- CRF = false,
no attention,
layer = 1,
input = semantic(word-emb)
(Base model)

```

Epoch:1, Training loss: 14874.99, train acc: 0.7589, val loss: 4565.16, val acc: 0.7091, time: 88.29s
Epoch:2, Training loss: 8465.77, train acc: 0.7755, val loss: 3408.15, val acc: 0.7351, time: 86.48s
Epoch:3, Training loss: 6367.66, train acc: 0.7754, val loss: 2876.45, val acc: 0.7258, time: 84.28s
Epoch:4, Training loss: 5205.36, train acc: 0.7789, val loss: 2554.84, val acc: 0.7218, time: 84.90s
Epoch:5, Training loss: 4397.43, train acc: 0.7742, val loss: 2407.63, val acc: 0.7133, time: 85.52s
Epoch:6, Training loss: 3773.15, train acc: 0.7725, val loss: 2326.33, val acc: 0.7048, time: 84.73s
Epoch:7, Training loss: 3223.77, train acc: 0.7582, val loss: 2245.07, val acc: 0.6936, time: 87.16s
Epoch:8, Training loss: 2794.31, train acc: 0.7592, val loss: 2220.14, val acc: 0.6856, time: 88.88s
Epoch:9, Training loss: 2443.06, train acc: 0.7589, val loss: 2258.68, val acc: 0.6864, time: 87.21s
Epoch:10, Training loss: 2089.54, train acc: 0.7604, val loss: 2302.32, val acc: 0.6815, time: 86.01s
Epoch:11, Training loss: 1781.10, train acc: 0.7667, val loss: 2386.35, val acc: 0.6847, time: 85.80s
Epoch:12, Training loss: 1539.20, train acc: 0.7851, val loss: 2441.91, val acc: 0.6928, time: 85.08s
Epoch:13, Training loss: 1367.97, train acc: 0.7928, val loss: 2528.08, val acc: 0.6980, time: 84.35s
Epoch:14, Training loss: 1228.03, train acc: 0.7854, val loss: 2603.94, val acc: 0.6873, time: 87.16s
Epoch:15, Training loss: 1054.49, train acc: 0.7928, val loss: 2711.38, val acc: 0.6959, time: 86.74s
Epoch:16, Training loss: 932.01, train acc: 0.8099, val loss: 2829.49, val acc: 0.7107, time: 85.90s
Epoch:17, Training loss: 864.96, train acc: 0.8215, val loss: 2984.32, val acc: 0.7162, time: 84.53s
Epoch:18, Training loss: 834.50, train acc: 0.8181, val loss: 3029.25, val acc: 0.7152, time: 84.21s
Epoch:19, Training loss: 677.77, train acc: 0.8317, val loss: 3207.10, val acc: 0.7203, time: 86.78s
Epoch:20, Training loss: 673.39, train acc: 0.8206, val loss: 3175.33, val acc: 0.7069, time: 86.24s

```

- CRF = true,
no attention,
layer = 1,
input = semantic(word-emb)
(Baseline model)

Epoch:1, Training loss: 14215.47, train acc: 0.7682, val loss: 4399.70, val acc: 0.7165, time: 100.80s
 Epoch:2, Training loss: 8165.73, train acc: 0.8162, val loss: 3358.36, val acc: 0.7622, time: 100.17s
 Epoch:3, Training loss: 6132.31, train acc: 0.8395, val loss: 2848.25, val acc: 0.7763, time: 100.73s
 Epoch:4, Training loss: 5009.88, train acc: 0.8600, val loss: 2513.75, val acc: 0.7897, time: 101.18s
 Epoch:5, Training loss: 4280.94, train acc: 0.8751, val loss: 2362.61, val acc: 0.7958, time: 100.57s
 Epoch:6, Training loss: 3661.65, train acc: 0.8896, val loss: 2249.18, val acc: 0.8043, time: 100.59s
 Epoch:7, Training loss: 3185.98, train acc: 0.9031, val loss: 2167.11, val acc: 0.8125, time: 100.66s
 Epoch:8, Training loss: 2780.73, train acc: 0.9124, val loss: 2172.15, val acc: 0.8132, time: 99.62s
 Epoch:9, Training loss: 2439.79, train acc: 0.9219, val loss: 2197.49, val acc: 0.8104, time: 99.95s
 Epoch:10, Training loss: 2114.55, train acc: 0.9340, val loss: 2222.65, val acc: 0.8123, time: 100.19s
 Epoch:11, Training loss: 1861.09, train acc: 0.9388, val loss: 2307.57, val acc: 0.8127, time: 100.01s
 Epoch:12, Training loss: 1636.92, train acc: 0.9433, val loss: 2424.79, val acc: 0.8138, time: 100.59s
 Epoch:13, Training loss: 1463.31, train acc: 0.9456, val loss: 2530.32, val acc: 0.8091, time: 100.27s
 Epoch:14, Training loss: 1326.13, train acc: 0.9570, val loss: 2601.70, val acc: 0.8089, time: 99.89s
 Epoch:15, Training loss: 1087.80, train acc: 0.9575, val loss: 2613.35, val acc: 0.8064, time: 100.34s
 Epoch:16, Training loss: 995.62, train acc: 0.9666, val loss: 2817.18, val acc: 0.8081, time: 100.34s
 Epoch:17, Training loss: 852.01, train acc: 0.9666, val loss: 2870.97, val acc: 0.8075, time: 99.98s
 Epoch:18, Training loss: 770.69, train acc: 0.9704, val loss: 2967.68, val acc: 0.8134, time: 100.65s
 Epoch:19, Training loss: 706.61, train acc: 0.9709, val loss: 3208.45, val acc: 0.8083, time: 100.65s
 Epoch:20, Training loss: 620.70, train acc: 0.9724, val loss: 3228.11, val acc: 0.8055, time: 100.91s

	precision	recall	f1-score	support
B-DocumentReference	0.3846	0.2500	0.3030	20
B-Location	0.6875	0.5914	0.6358	186
B-MilitaryPlatform	0.5000	0.0625	0.1111	16
B-Money	0.1667	0.2000	0.1818	5
B-Nationality	0.2727	0.3750	0.3158	8
B-Organisation	0.7328	0.6071	0.6641	280
B-Person	0.8469	0.8137	0.8300	102
B-Quantity	0.6596	0.5636	0.6078	55
B-Temporal	0.7391	0.7234	0.7312	47
B-Weapon	0.2903	0.2368	0.2609	38
I-DocumentReference	0.6053	0.2771	0.3802	83
I-Location	0.6364	0.5283	0.5773	265
I-MilitaryPlatform	0.0000	0.0000	0.0000	16
I-Money	0.8000	0.4000	0.5333	10
I-Nationality	0.0000	0.0000	0.0000	1
I-Organisation	0.7609	0.6274	0.6877	416
I-Person	0.8716	0.7457	0.8037	173
I-Quantity	0.3000	0.1154	0.1667	26
I-Temporal	0.8431	0.6935	0.7611	62
I-Weapon	0.2941	0.1111	0.1613	45
0	0.8453	0.9380	0.8893	3420
accuracy			0.8083	5274
macro avg	0.5351	0.4219	0.4572	5274
weighted avg	0.7923	0.8083	0.7950	5274

- CRF = true,
no attention,
layer = 1,
input = semantic(word-emb) + dependency parsing

```

Epoch:1, Training loss: 17657.75, train acc: 0.6999, val loss: 6240.19, val acc: 0.6485, time: 112.56s
Epoch:2, Training loss: 13880.45, train acc: 0.6999, val loss: 5519.64, val acc: 0.6485, time: 112.65s
Epoch:3, Training loss: 12404.71, train acc: 0.7062, val loss: 4808.31, val acc: 0.6593, time: 111.68s
Epoch:4, Training loss: 10285.54, train acc: 0.7406, val loss: 4041.49, val acc: 0.6799, time: 112.80s
Epoch:5, Training loss: 8531.53, train acc: 0.7647, val loss: 3679.44, val acc: 0.6995, time: 115.64s
Epoch:6, Training loss: 7435.15, train acc: 0.7808, val loss: 3437.45, val acc: 0.7107, time: 115.39s
Epoch:7, Training loss: 6600.67, train acc: 0.8074, val loss: 3277.42, val acc: 0.7308, time: 114.32s
Epoch:8, Training loss: 5877.82, train acc: 0.8213, val loss: 3164.92, val acc: 0.7414, time: 113.74s
Epoch:9, Training loss: 5259.18, train acc: 0.8355, val loss: 3052.36, val acc: 0.7539, time: 113.18s
Epoch:10, Training loss: 4715.38, train acc: 0.8511, val loss: 3016.09, val acc: 0.7611, time: 113.44s
Epoch:11, Training loss: 4226.52, train acc: 0.8680, val loss: 2967.57, val acc: 0.7624, time: 112.95s
Epoch:12, Training loss: 3789.89, train acc: 0.8766, val loss: 2961.20, val acc: 0.7664, time: 109.72s
Epoch:13, Training loss: 3427.71, train acc: 0.8886, val loss: 2938.99, val acc: 0.7691, time: 110.68s
Epoch:14, Training loss: 3088.16, train acc: 0.9046, val loss: 2919.19, val acc: 0.7702, time: 110.58s
Epoch:15, Training loss: 2807.89, train acc: 0.9126, val loss: 2824.34, val acc: 0.7812, time: 111.53s
Epoch:16, Training loss: 2525.07, train acc: 0.9184, val loss: 2959.90, val acc: 0.7782, time: 111.44s
Epoch:17, Training loss: 2342.21, train acc: 0.9220, val loss: 2982.86, val acc: 0.7859, time: 111.26s
Epoch:18, Training loss: 2128.48, train acc: 0.9283, val loss: 2977.92, val acc: 0.7846, time: 111.68s
Epoch:19, Training loss: 1959.92, train acc: 0.9384, val loss: 3000.29, val acc: 0.7892, time: 111.45s
Epoch:20, Training loss: 1748.11, train acc: 0.9429, val loss: 3095.14, val acc: 0.7893, time: 112.82s

```

	precision	recall	f1-score	support
B-DocumentReference	0.6667	0.1000	0.1739	20
B-Location	0.6076	0.5161	0.5581	186
B-MilitaryPlatform	1.0000	0.0625	0.1176	16
B-Money	0.0000	0.0000	0.0000	5
B-Nationality	0.0000	0.0000	0.0000	8
B-Organisation	0.7273	0.6286	0.6743	280
B-Person	0.7579	0.7059	0.7310	102
B-Quantity	0.6562	0.3818	0.4828	55
B-Temporal	0.6471	0.4681	0.5432	47
B-Weapon	0.5000	0.0789	0.1364	38
I-DocumentReference	0.4286	0.0723	0.1237	83
I-Location	0.5844	0.5094	0.5444	265
I-MilitaryPlatform	0.0000	0.0000	0.0000	16
I-Money	1.0000	0.2000	0.3333	10
I-Nationality	0.0000	0.0000	0.0000	1
I-Organisation	0.7170	0.6274	0.6692	416
I-Person	0.7586	0.6358	0.6918	173
I-Quantity	0.2857	0.1538	0.2000	26
I-Temporal	0.7593	0.6613	0.7069	62
I-Weapon	0.5000	0.0444	0.0816	45
0	0.8316	0.9415	0.8832	3420
accuracy			0.7914	5274
macro avg	0.5442	0.3232	0.3644	5274
weighted avg	0.7698	0.7914	0.7696	5274

- CRF = true,
no attention,
layer = 1,
input = semantic(word-emb) + TF-IDF

```

Epoch:1, Training loss: 14667.72, train acc: 0.7496, val loss: 4837.56, val acc: 0.6961, time: 102.74s
Epoch:2, Training loss: 9116.33, train acc: 0.8101, val loss: 3603.09, val acc: 0.7575, time: 102.50s
Epoch:3, Training loss: 7090.06, train acc: 0.8323, val loss: 3107.00, val acc: 0.7763, time: 102.68s
Epoch:4, Training loss: 5936.69, train acc: 0.8487, val loss: 2723.93, val acc: 0.7871, time: 103.29s
Epoch:5, Training loss: 5163.32, train acc: 0.8616, val loss: 2473.86, val acc: 0.7984, time: 103.29s
Epoch:6, Training loss: 4463.60, train acc: 0.8752, val loss: 2325.86, val acc: 0.8075, time: 102.84s
Epoch:7, Training loss: 4082.47, train acc: 0.8832, val loss: 2320.80, val acc: 0.8026, time: 103.74s
Epoch:8, Training loss: 3626.50, train acc: 0.8958, val loss: 2234.13, val acc: 0.8104, time: 102.51s
Epoch:9, Training loss: 3351.64, train acc: 0.9113, val loss: 2055.12, val acc: 0.8193, time: 102.95s
Epoch:10, Training loss: 3025.23, train acc: 0.9180, val loss: 2017.58, val acc: 0.8159, time: 102.38s
Epoch:11, Training loss: 2714.53, train acc: 0.9176, val loss: 2142.93, val acc: 0.8132, time: 102.67s
Epoch:12, Training loss: 2557.06, train acc: 0.9254, val loss: 2199.98, val acc: 0.8104, time: 102.67s
Epoch:13, Training loss: 2305.84, train acc: 0.9339, val loss: 2142.97, val acc: 0.8199, time: 102.98s
Epoch:14, Training loss: 2158.34, train acc: 0.9380, val loss: 2173.63, val acc: 0.8132, time: 102.89s
Epoch:15, Training loss: 1956.88, train acc: 0.9488, val loss: 2125.50, val acc: 0.8201, time: 102.56s
Epoch:16, Training loss: 1885.45, train acc: 0.9545, val loss: 2111.99, val acc: 0.8174, time: 102.65s
Epoch:17, Training loss: 1701.02, train acc: 0.9518, val loss: 2312.22, val acc: 0.8117, time: 102.83s
Epoch:18, Training loss: 1658.67, train acc: 0.9568, val loss: 2278.26, val acc: 0.8182, time: 102.98s
Epoch:19, Training loss: 1616.31, train acc: 0.9647, val loss: 2260.28, val acc: 0.8096, time: 103.60s
Epoch:20, Training loss: 1514.58, train acc: 0.9642, val loss: 2322.99, val acc: 0.8161, time: 102.61s

```

	precision	recall	f1-score	support
B-DocumentReference	0.3158	0.3000	0.3077	20
B-Location	0.6818	0.6452	0.6630	186
B-MilitaryPlatform	0.6667	0.1250	0.2105	16
B-Money	0.5000	0.2000	0.2857	5
B-Nationality	0.4000	0.2500	0.3077	8
B-Organisation	0.7198	0.6607	0.6890	280
B-Person	0.8365	0.8529	0.8447	102
B-Quantity	0.7143	0.7273	0.7207	55
B-Temporal	0.7561	0.6596	0.7045	47
B-Weapon	0.2800	0.1842	0.2222	38
I-DocumentReference	0.5000	0.2771	0.3566	83
I-Location	0.7188	0.5208	0.6039	265
I-MilitaryPlatform	0.0000	0.0000	0.0000	16
I-Money	1.0000	0.6000	0.7500	10
I-Nationality	0.0000	0.0000	0.0000	1
I-Organisation	0.7062	0.6587	0.6816	416
I-Person	0.8447	0.7861	0.8144	173
I-Quantity	0.3000	0.1154	0.1667	26
I-Temporal	0.8913	0.6613	0.7593	62
I-Weapon	0.3333	0.1111	0.1667	45
0	0.8598	0.9357	0.8961	3420
accuracy			0.8166	5274
macro avg	0.5726	0.4415	0.4834	5274
weighted avg	0.8006	0.8166	0.8044	5274

- CRF = true,
no attention,
layer = 1,
input = semantic(word-emb) + dependency parsing + TF-IDF

```

Epoch:1, Training loss: 14394.12, train acc: 0.7536, val loss: 4450.40, val acc: 0.6964, time: 104.33s
Epoch:2, Training loss: 8735.51, train acc: 0.8069, val loss: 3360.10, val acc: 0.7588, time: 104.41s
Epoch:3, Training loss: 6672.09, train acc: 0.8329, val loss: 2767.63, val acc: 0.7789, time: 105.15s
Epoch:4, Training loss: 5550.08, train acc: 0.8394, val loss: 2582.60, val acc: 0.7783, time: 104.50s
Epoch:5, Training loss: 4850.08, train acc: 0.8657, val loss: 2245.56, val acc: 0.7994, time: 104.48s
Epoch:6, Training loss: 4324.57, train acc: 0.8796, val loss: 2149.02, val acc: 0.8055, time: 105.32s
Epoch:7, Training loss: 3854.12, train acc: 0.8796, val loss: 2233.13, val acc: 0.7990, time: 104.58s
Epoch:8, Training loss: 3446.60, train acc: 0.8991, val loss: 2061.60, val acc: 0.8096, time: 104.46s
Epoch:9, Training loss: 3173.33, train acc: 0.9098, val loss: 2061.63, val acc: 0.8098, time: 104.86s
Epoch:10, Training loss: 2850.42, train acc: 0.9187, val loss: 2001.19, val acc: 0.8182, time: 104.87s
Epoch:11, Training loss: 2675.81, train acc: 0.9178, val loss: 2098.31, val acc: 0.8125, time: 104.71s
Epoch:12, Training loss: 2498.29, train acc: 0.9254, val loss: 2106.86, val acc: 0.8140, time: 105.20s
Epoch:13, Training loss: 2314.35, train acc: 0.9389, val loss: 2041.03, val acc: 0.8153, time: 105.18s
Epoch:14, Training loss: 2209.72, train acc: 0.9414, val loss: 2043.96, val acc: 0.8208, time: 104.48s
Epoch:15, Training loss: 2018.47, train acc: 0.9454, val loss: 2140.92, val acc: 0.8231, time: 104.85s
Epoch:16, Training loss: 1989.32, train acc: 0.9499, val loss: 2115.83, val acc: 0.8256, time: 104.80s
Epoch:17, Training loss: 1734.00, train acc: 0.9406, val loss: 2423.14, val acc: 0.8125, time: 104.78s
Epoch:18, Training loss: 1704.49, train acc: 0.9506, val loss: 2329.85, val acc: 0.8106, time: 105.28s
Epoch:19, Training loss: 1536.68, train acc: 0.9587, val loss: 2293.03, val acc: 0.8193, time: 104.33s
Epoch:20, Training loss: 1413.91, train acc: 0.9663, val loss: 2239.25, val acc: 0.8208, time: 104.68s

```

	precision	recall	f1-score	support
B-DocumentReference	0.4118	0.3500	0.3784	20
B-Location	0.7126	0.6667	0.6889	186
B-MilitaryPlatform	0.0000	0.0000	0.0000	16
B-Money	0.3333	0.2000	0.2500	5
B-Nationality	0.2222	0.2500	0.2353	8
B-Organisation	0.6929	0.6929	0.6929	280
B-Person	0.8317	0.8235	0.8276	102
B-Quantity	0.6613	0.7455	0.7009	55
B-Temporal	0.7750	0.6596	0.7126	47
B-Weapon	0.3235	0.2895	0.3056	38
I-DocumentReference	0.5106	0.2892	0.3692	83
I-Location	0.7005	0.5472	0.6144	265
I-MilitaryPlatform	0.0000	0.0000	0.0000	16
I-Money	1.0000	0.5000	0.6667	10
I-Nationality	0.0000	0.0000	0.0000	1
I-Organisation	0.7237	0.7115	0.7176	416
I-Person	0.8323	0.7457	0.7866	173
I-Quantity	0.5833	0.2692	0.3684	26
I-Temporal	0.8491	0.7258	0.7826	62
I-Weapon	0.3333	0.1556	0.2121	45
0	0.8699	0.9269	0.8975	3420
accuracy			0.8197	5274
macro avg	0.5413	0.4547	0.4861	5274
weighted avg	0.8057	0.8197	0.8102	5274

- CRF = true,
no attention,
layer = 2,
input = semantic(word-emb) + dependency parsing + TF-IDF

```

Epoch:1, Training loss: 16332.14, train acc: 0.7073, val loss: 5261.77, val acc: 0.6536, time: 115.09s
Epoch:2, Training loss: 10600.63, train acc: 0.7696, val loss: 3983.21, val acc: 0.7148, time: 114.24s
Epoch:3, Training loss: 8002.49, train acc: 0.8053, val loss: 3249.38, val acc: 0.7497, time: 114.86s
Epoch:4, Training loss: 6504.88, train acc: 0.8207, val loss: 2880.50, val acc: 0.7611, time: 115.23s
Epoch:5, Training loss: 5503.59, train acc: 0.8427, val loss: 2551.40, val acc: 0.7835, time: 115.08s
Epoch:6, Training loss: 4866.41, train acc: 0.8579, val loss: 2266.85, val acc: 0.7960, time: 114.75s
Epoch:7, Training loss: 4321.47, train acc: 0.8653, val loss: 2217.10, val acc: 0.8013, time: 115.06s
Epoch:8, Training loss: 3935.55, train acc: 0.8793, val loss: 2097.88, val acc: 0.8081, time: 115.41s
Epoch:9, Training loss: 3576.62, train acc: 0.8881, val loss: 1985.07, val acc: 0.8100, time: 114.94s
Epoch:10, Training loss: 3271.33, train acc: 0.9037, val loss: 1946.39, val acc: 0.8180, time: 115.77s
Epoch:11, Training loss: 2955.26, train acc: 0.9101, val loss: 2007.85, val acc: 0.8189, time: 115.10s
Epoch:12, Training loss: 2684.51, train acc: 0.9208, val loss: 1973.32, val acc: 0.8210, time: 114.66s
Epoch:13, Training loss: 2508.04, train acc: 0.9173, val loss: 2121.28, val acc: 0.8074, time: 115.38s
Epoch:14, Training loss: 2390.86, train acc: 0.9283, val loss: 1996.39, val acc: 0.8161, time: 115.67s
Epoch:15, Training loss: 2179.36, train acc: 0.9340, val loss: 2070.44, val acc: 0.8189, time: 115.58s
Epoch:16, Training loss: 2021.76, train acc: 0.9409, val loss: 2055.19, val acc: 0.8159, time: 115.25s
Epoch:17, Training loss: 1817.55, train acc: 0.9368, val loss: 2168.27, val acc: 0.8032, time: 115.24s
Epoch:18, Training loss: 1641.49, train acc: 0.9581, val loss: 2139.85, val acc: 0.8165, time: 114.38s
Epoch:19, Training loss: 1639.58, train acc: 0.9528, val loss: 2277.53, val acc: 0.8163, time: 114.51s
Epoch:20, Training loss: 1480.57, train acc: 0.9663, val loss: 2176.22, val acc: 0.8168, time: 115.26s

```

	precision	recall	f1-score	support
B-DocumentReference	0.5000	0.3500	0.4118	20
B-Location	0.6826	0.6129	0.6459	186
B-MilitaryPlatform	0.6667	0.2500	0.3636	16
B-Money	0.3333	0.2000	0.2500	5
B-Nationality	0.0000	0.0000	0.0000	8
B-Organisation	0.6596	0.6643	0.6619	280
B-Person	0.8252	0.8333	0.8293	102
B-Quantity	0.6721	0.7455	0.7069	55
B-Temporal	0.8095	0.7234	0.7640	47
B-Weapon	0.2581	0.2105	0.2319	38
I-DocumentReference	0.5581	0.2892	0.3810	83
I-Location	0.7216	0.5283	0.6100	265
I-MilitaryPlatform	0.3750	0.1875	0.2500	16
I-Money	1.0000	0.4000	0.5714	10
I-Nationality	0.0000	0.0000	0.0000	1
I-Organisation	0.6604	0.6779	0.6690	416
I-Person	0.8462	0.7630	0.8024	173
I-Quantity	0.5625	0.3462	0.4286	26
I-Temporal	0.8246	0.7581	0.7899	62
I-Weapon	0.2500	0.1111	0.1538	45
0	0.8691	0.9243	0.8958	3420
accuracy			0.8129	5274
macro avg	0.5750	0.4560	0.4961	5274
weighted avg	0.8015	0.8129	0.8039	5274

- CRF = true,
no attention,
layer = 3,
input = semantic(word-emb) + dependency parsing + TF-IDF

Epoch:1, Training loss: 16215.58, train acc: 0.6999, val loss: 5569.65, val acc: 0.6485, time: 124.57s
 Epoch:2, Training loss: 12318.10, train acc: 0.7246, val loss: 4719.34, val acc: 0.6670, time: 123.98s
 Epoch:3, Training loss: 9897.84, train acc: 0.7590, val loss: 3946.78, val acc: 0.7107, time: 124.53s
 Epoch:4, Training loss: 8117.97, train acc: 0.7857, val loss: 3365.26, val acc: 0.7285, time: 124.99s
 Epoch:5, Training loss: 6896.53, train acc: 0.8033, val loss: 3059.85, val acc: 0.7425, time: 124.79s
 Epoch:6, Training loss: 6166.00, train acc: 0.8165, val loss: 2753.53, val acc: 0.7619, time: 126.68s
 Epoch:7, Training loss: 5464.24, train acc: 0.8426, val loss: 2378.22, val acc: 0.7829, time: 125.52s
 Epoch:8, Training loss: 4881.20, train acc: 0.8434, val loss: 2380.34, val acc: 0.7825, time: 125.55s
 Epoch:9, Training loss: 4431.94, train acc: 0.8648, val loss: 2192.30, val acc: 0.7922, time: 125.65s
 Epoch:10, Training loss: 4158.86, train acc: 0.8694, val loss: 2136.47, val acc: 0.7984, time: 126.55s
 Epoch:11, Training loss: 3746.05, train acc: 0.8810, val loss: 2096.33, val acc: 0.8015, time: 126.60s
 Epoch:12, Training loss: 3466.40, train acc: 0.8881, val loss: 2027.27, val acc: 0.8058, time: 127.65s
 Epoch:13, Training loss: 3238.94, train acc: 0.8958, val loss: 2118.90, val acc: 0.8104, time: 127.89s
 Epoch:14, Training loss: 2923.60, train acc: 0.9089, val loss: 2000.85, val acc: 0.8161, time: 126.89s
 Epoch:15, Training loss: 2741.99, train acc: 0.9116, val loss: 2003.50, val acc: 0.8129, time: 125.50s
 Epoch:16, Training loss: 2560.33, train acc: 0.9182, val loss: 2073.75, val acc: 0.8108, time: 125.60s
 Epoch:17, Training loss: 2370.08, train acc: 0.9251, val loss: 2048.78, val acc: 0.8138, time: 125.25s
 Epoch:18, Training loss: 2221.80, train acc: 0.9298, val loss: 2096.56, val acc: 0.8153, time: 124.65s
 Epoch:19, Training loss: 2086.11, train acc: 0.9292, val loss: 2159.50, val acc: 0.8113, time: 124.94s
 Epoch:20, Training loss: 1973.66, train acc: 0.9336, val loss: 2260.35, val acc: 0.8106, time: 124.55s

	precision	recall	f1-score	support
B-DocumentReference	0.4000	0.1000	0.1600	20
B-Location	0.6805	0.6183	0.6479	186
B-MilitaryPlatform	1.0000	0.0625	0.1176	16
B-Money	0.3333	0.2000	0.2500	5
B-Nationality	0.4000	0.2500	0.3077	8
B-Organisation	0.7202	0.6250	0.6692	280
B-Person	0.8316	0.7745	0.8020	102
B-Quantity	0.6667	0.6909	0.6786	55
B-Temporal	0.7027	0.5532	0.6190	47
B-Weapon	0.4167	0.1316	0.2000	38
I-DocumentReference	0.8750	0.1687	0.2828	83
I-Location	0.7150	0.5396	0.6151	265
I-MilitaryPlatform	1.0000	0.1250	0.2222	16
I-Money	1.0000	0.4000	0.5714	10
I-Nationality	0.0000	0.0000	0.0000	1
I-Organisation	0.7303	0.5337	0.6167	416
I-Person	0.9015	0.6879	0.7803	173
I-Quantity	0.3750	0.1154	0.1765	26
I-Temporal	0.8810	0.5968	0.7115	62
I-Weapon	0.3333	0.0667	0.1111	45
0	0.8338	0.9582	0.8917	3420
accuracy			0.8093	5274
macro avg	0.6570	0.3904	0.4491	5274
weighted avg	0.7976	0.8093	0.7894	5274

- CRF = true,
no attention,
layer = 5,
input = semantic(word-emb) + dependency parsing + TF-IDF

Epoch:1, Training loss: 16253.77, train acc: 0.6999, val loss: 5832.76, val acc: 0.6485, time: 147.96s
 Epoch:2, Training loss: 13397.14, train acc: 0.6999, val loss: 5346.42, val acc: 0.6485, time: 146.31s
 Epoch:3, Training loss: 12653.74, train acc: 0.6999, val loss: 5129.88, val acc: 0.6485, time: 147.07s
 Epoch:4, Training loss: 12237.07, train acc: 0.6999, val loss: 4888.56, val acc: 0.6485, time: 147.14s
 Epoch:5, Training loss: 11824.56, train acc: 0.6999, val loss: 4710.44, val acc: 0.6532, time: 147.36s
 Epoch:6, Training loss: 11172.54, train acc: 0.7072, val loss: 4389.80, val acc: 0.6555, time: 146.89s
 Epoch:7, Training loss: 10364.17, train acc: 0.7272, val loss: 4030.35, val acc: 0.6818, time: 147.20s
 Epoch:8, Training loss: 9250.78, train acc: 0.7529, val loss: 3553.50, val acc: 0.7025, time: 146.89s
 Epoch:9, Training loss: 8185.76, train acc: 0.7639, val loss: 3305.49, val acc: 0.7139, time: 147.02s
 Epoch:10, Training loss: 7446.62, train acc: 0.7760, val loss: 3093.25, val acc: 0.7237, time: 147.07s
 Epoch:11, Training loss: 6899.21, train acc: 0.7932, val loss: 2913.85, val acc: 0.7387, time: 146.86s
 Epoch:12, Training loss: 6432.99, train acc: 0.8068, val loss: 2846.30, val acc: 0.7467, time: 146.77s
 Epoch:13, Training loss: 6064.18, train acc: 0.8253, val loss: 2622.99, val acc: 0.7567, time: 146.96s
 Epoch:14, Training loss: 5565.28, train acc: 0.8263, val loss: 2613.49, val acc: 0.7617, time: 147.53s
 Epoch:15, Training loss: 5205.30, train acc: 0.8357, val loss: 2528.29, val acc: 0.7617, time: 147.53s
 Epoch:16, Training loss: 4863.98, train acc: 0.8432, val loss: 2492.54, val acc: 0.7734, time: 147.56s
 Epoch:17, Training loss: 4527.44, train acc: 0.8587, val loss: 2356.37, val acc: 0.7918, time: 146.93s
 Epoch:18, Training loss: 4298.99, train acc: 0.8623, val loss: 2283.74, val acc: 0.7920, time: 147.45s
 Epoch:19, Training loss: 3920.12, train acc: 0.8732, val loss: 2218.66, val acc: 0.7945, time: 147.23s
 Epoch:20, Training loss: 3705.59, train acc: 0.8718, val loss: 2207.80, val acc: 0.7948, time: 147.50s

	precision	recall	f1-score	support
B-DocumentReference	0.0000	0.0000	0.0000	20
B-Location	0.6333	0.6129	0.6230	186
B-MilitaryPlatform	0.5000	0.0625	0.1111	16
B-Money	0.0000	0.0000	0.0000	5
B-Nationality	0.0000	0.0000	0.0000	8
B-Organisation	0.7162	0.5679	0.6335	280
B-Person	0.8764	0.7647	0.8168	102
B-Quantity	0.8462	0.2000	0.3235	55
B-Temporal	0.7812	0.5319	0.6329	47
B-Weapon	0.3333	0.0263	0.0488	38
I-DocumentReference	0.0000	0.0000	0.0000	83
I-Location	0.7702	0.4679	0.5822	265
I-MilitaryPlatform	0.3333	0.0625	0.1053	16
I-Money	0.0000	0.0000	0.0000	10
I-Nationality	0.0000	0.0000	0.0000	1
I-Organisation	0.7413	0.5096	0.6040	416
I-Person	0.8973	0.7572	0.8213	173
I-Quantity	0.0000	0.0000	0.0000	26
I-Temporal	0.7674	0.5323	0.6286	62
I-Weapon	0.0000	0.0000	0.0000	45
O	0.8106	0.9699	0.8831	3420
accuracy			0.7977	5274
macro avg	0.4289	0.2888	0.3245	5274
weighted avg	0.7593	0.7977	0.7653	5274

- CRF = true,
no attention,
layer = 8,
input = semantic(word-emb) + dependency parsing + TF-IDF

Epoch:1, Training loss: 17706.41, train acc: 0.6999, val loss: 6097.22, val acc: 0.6485, time: 178.41s
Epoch:2, Training loss: 14026.09, train acc: 0.6999, val loss: 5558.51, val acc: 0.6485, time: 177.51s
Epoch:3, Training loss: 13162.79, train acc: 0.6999, val loss: 5282.58, val acc: 0.6485, time: 177.39s
Epoch:4, Training loss: 12720.30, train acc: 0.6999, val loss: 5110.67, val acc: 0.6485, time: 177.67s
Epoch:5, Training loss: 12436.28, train acc: 0.6999, val loss: 4978.73, val acc: 0.6485, time: 177.95s
Epoch:6, Training loss: 12235.67, train acc: 0.6999, val loss: 4888.95, val acc: 0.6485, time: 178.61s
Epoch:7, Training loss: 12083.29, train acc: 0.6999, val loss: 4803.16, val acc: 0.6485, time: 177.01s
Epoch:8, Training loss: 11973.38, train acc: 0.6999, val loss: 4741.09, val acc: 0.6485, time: 178.44s
Epoch:9, Training loss: 11872.80, train acc: 0.6999, val loss: 4687.19, val acc: 0.6485, time: 177.76s
Epoch:10, Training loss: 11786.88, train acc: 0.6999, val loss: 4645.02, val acc: 0.6485, time: 177.66s
Epoch:11, Training loss: 11720.29, train acc: 0.6999, val loss: 4611.90, val acc: 0.6485, time: 177.70s
Epoch:12, Training loss: 11660.85, train acc: 0.6999, val loss: 4578.39, val acc: 0.6485, time: 178.23s
Epoch:13, Training loss: 11609.51, train acc: 0.6999, val loss: 4556.74, val acc: 0.6485, time: 177.47s
Epoch:14, Training loss: 11564.93, train acc: 0.6999, val loss: 4531.97, val acc: 0.6485, time: 178.86s
Epoch:15, Training loss: 11518.66, train acc: 0.6999, val loss: 4513.71, val acc: 0.6485, time: 177.73s
Epoch:16, Training loss: 11475.48, train acc: 0.6999, val loss: 4484.34, val acc: 0.6485, time: 179.93s
Epoch:17, Training loss: 11436.52, train acc: 0.6999, val loss: 4472.09, val acc: 0.6485, time: 180.14s
Epoch:18, Training loss: 11414.74, train acc: 0.6999, val loss: 4466.38, val acc: 0.6485, time: 178.02s
Epoch:19, Training loss: 11381.70, train acc: 0.6999, val loss: 4448.64, val acc: 0.6485, time: 179.88s
Epoch:20, Training loss: 11364.62, train acc: 0.6999, val loss: 4435.17, val acc: 0.6485, time: 178.78s
> precision recall f1-score support
B-DocumentReference 0.0000 0.0000 0.0000 20
B-Location 0.0000 0.0000 0.0000 186
B-MilitaryPlatform 0.0000 0.0000 0.0000 16
B-Money 0.0000 0.0000 0.0000 5
B-Nationality 0.0000 0.0000 0.0000 8
B-Organisation 0.0000 0.0000 0.0000 280
B-Person 0.0000 0.0000 0.0000 102
B-Quantity 0.0000 0.0000 0.0000 55
B-Temporal 0.0000 0.0000 0.0000 47
B-Weapon 0.0000 0.0000 0.0000 38
I-DocumentReference 0.0000 0.0000 0.0000 83
I-Location 0.0000 0.0000 0.0000 265
I-MilitaryPlatform 0.0000 0.0000 0.0000 16
I-Money 0.0000 0.0000 0.0000 10
I-Nationality 0.0000 0.0000 0.0000 1
I-Organisation 0.0000 0.0000 0.0000 416
I-Person 0.0000 0.0000 0.0000 173
I-Quantity 0.0000 0.0000 0.0000 26
I-Temporal 0.0000 0.0000 0.0000 62
I-Weapon 0.0000 0.0000 0.0000 45
0 0.6485 1.0000 0.7867 3420
accuracy 0.6485 5274
macro avg 0.0309 0.0476 0.0375 5274
weighted avg 0.4205 0.6485 0.5102 5274

- CRF = true,
Attention = dot product,
layer = 1,
input = semantic(word-emb) + dependency parsing + TF-IDF

```

Epoch:1, Training loss: 14658.56, train acc: 0.7622, val loss: 4356.93, val acc: 0.7105, time: 105.68s
Epoch:2, Training loss: 8623.45, train acc: 0.8064, val loss: 3355.85, val acc: 0.7548, time: 105.09s
Epoch:3, Training loss: 6611.09, train acc: 0.8314, val loss: 2823.87, val acc: 0.7673, time: 105.18s
Epoch:4, Training loss: 5489.47, train acc: 0.8523, val loss: 2459.17, val acc: 0.7928, time: 105.70s
Epoch:5, Training loss: 4832.11, train acc: 0.8661, val loss: 2248.77, val acc: 0.7979, time: 105.60s
Epoch:6, Training loss: 4293.34, train acc: 0.8845, val loss: 2121.61, val acc: 0.8036, time: 104.95s
Epoch:7, Training loss: 3795.16, train acc: 0.8895, val loss: 2111.13, val acc: 0.8102, time: 105.67s
Epoch:8, Training loss: 3501.36, train acc: 0.8952, val loss: 2140.90, val acc: 0.8117, time: 105.38s
Epoch:9, Training loss: 3122.21, train acc: 0.9031, val loss: 2175.25, val acc: 0.8110, time: 105.55s
Epoch:10, Training loss: 2903.23, train acc: 0.9184, val loss: 2042.54, val acc: 0.8159, time: 105.16s
Epoch:11, Training loss: 2655.55, train acc: 0.9262, val loss: 2037.58, val acc: 0.8161, time: 105.79s
Epoch:12, Training loss: 2484.91, train acc: 0.9305, val loss: 2033.60, val acc: 0.8166, time: 105.37s
Epoch:13, Training loss: 2300.59, train acc: 0.9321, val loss: 2080.88, val acc: 0.8157, time: 105.92s
Epoch:14, Training loss: 2188.95, train acc: 0.9353, val loss: 2131.02, val acc: 0.8189, time: 106.05s
Epoch:15, Training loss: 1969.61, train acc: 0.9431, val loss: 2158.31, val acc: 0.8191, time: 106.13s
Epoch:16, Training loss: 1884.96, train acc: 0.9469, val loss: 2104.55, val acc: 0.8223, time: 105.69s
Epoch:17, Training loss: 1752.28, train acc: 0.9541, val loss: 2100.98, val acc: 0.8242, time: 106.50s
Epoch:18, Training loss: 1695.94, train acc: 0.9581, val loss: 2188.35, val acc: 0.8288, time: 106.32s
Epoch:19, Training loss: 1570.68, train acc: 0.9630, val loss: 2213.40, val acc: 0.8199, time: 106.66s
Epoch:20, Training loss: 1449.10, train acc: 0.9668, val loss: 2267.28, val acc: 0.8193, time: 106.47s

```

	precision	recall	f1-score	support
B-DocumentReference	0.2778	0.2500	0.2632	20
B-Location	0.7006	0.6290	0.6629	186
B-MilitaryPlatform	0.5000	0.0625	0.1111	16
B-Money	0.5000	0.2000	0.2857	5
B-Nationality	0.2500	0.2500	0.2500	8
B-Organisation	0.7054	0.6500	0.6766	280
B-Person	0.8384	0.8137	0.8259	102
B-Quantity	0.6452	0.7273	0.6838	55
B-Temporal	0.8095	0.7234	0.7640	47
B-Weapon	0.3571	0.2632	0.3030	38
I-DocumentReference	0.5094	0.3253	0.3971	83
I-Location	0.7065	0.5358	0.6094	265
I-MilitaryPlatform	0.0000	0.0000	0.0000	16
I-Money	1.0000	0.4000	0.5714	10
I-Nationality	0.0000	0.0000	0.0000	1
I-Organisation	0.7153	0.6947	0.7049	416
I-Person	0.8627	0.7630	0.8098	173
I-Quantity	0.5333	0.3077	0.3902	26
I-Temporal	0.8776	0.6935	0.7748	62
I-Weapon	0.4118	0.1556	0.2258	45
0	0.8656	0.9342	0.8986	3420
accuracy			0.8195	5274
macro avg	0.5746	0.4466	0.4861	5274
weighted avg	0.8062	0.8195	0.8092	5274

- CRF = true,
Attention = scaled dot product,
layer = 1,
input = semantic(word-emb) + dependency parsing + TF-IDF

```

Epoch:1, Training loss: 14602.74, train acc: 0.7664, val loss: 4322.76, val acc: 0.7184, time: 106.60s
Epoch:2, Training loss: 8610.70, train acc: 0.8156, val loss: 3329.81, val acc: 0.7590, time: 106.82s
Epoch:3, Training loss: 6615.57, train acc: 0.8384, val loss: 2799.74, val acc: 0.7793, time: 106.15s
Epoch:4, Training loss: 5444.54, train acc: 0.8553, val loss: 2464.68, val acc: 0.7994, time: 106.41s
Epoch:5, Training loss: 4783.89, train acc: 0.8690, val loss: 2255.34, val acc: 0.7984, time: 106.65s
Epoch:6, Training loss: 4264.70, train acc: 0.8796, val loss: 2140.45, val acc: 0.8019, time: 107.40s
Epoch:7, Training loss: 3733.44, train acc: 0.8884, val loss: 2129.83, val acc: 0.8083, time: 105.79s
Epoch:8, Training loss: 3392.33, train acc: 0.8984, val loss: 2125.09, val acc: 0.8136, time: 105.62s
Epoch:9, Training loss: 3070.04, train acc: 0.9117, val loss: 2073.67, val acc: 0.8119, time: 106.41s
Epoch:10, Training loss: 2816.91, train acc: 0.9209, val loss: 2056.87, val acc: 0.8161, time: 106.24s
Epoch:11, Training loss: 2587.66, train acc: 0.9296, val loss: 2023.02, val acc: 0.8125, time: 106.46s
Epoch:12, Training loss: 2404.13, train acc: 0.9328, val loss: 2093.79, val acc: 0.8170, time: 107.21s
Epoch:13, Training loss: 2176.71, train acc: 0.9278, val loss: 2240.00, val acc: 0.8151, time: 107.41s
Epoch:14, Training loss: 2059.89, train acc: 0.9393, val loss: 2164.74, val acc: 0.8197, time: 106.47s
Epoch:15, Training loss: 1844.97, train acc: 0.9440, val loss: 2309.79, val acc: 0.8117, time: 105.85s
Epoch:16, Training loss: 1765.06, train acc: 0.9549, val loss: 2215.06, val acc: 0.8257, time: 105.92s
Epoch:17, Training loss: 1620.98, train acc: 0.9632, val loss: 2176.31, val acc: 0.8242, time: 105.63s
Epoch:18, Training loss: 1614.94, train acc: 0.9626, val loss: 2273.90, val acc: 0.8261, time: 105.83s
Epoch:19, Training loss: 1435.78, train acc: 0.9690, val loss: 2228.91, val acc: 0.8244, time: 106.38s
Epoch:20, Training loss: 1318.74, train acc: 0.9668, val loss: 2403.82, val acc: 0.8165, time: 106.30s

```

	precision	recall	f1-score	support
B-DocumentReference	0.3333	0.1500	0.2069	20
B-Location	0.6784	0.6237	0.6499	186
B-MilitaryPlatform	0.5000	0.0625	0.1111	16
B-Money	0.2500	0.2000	0.2222	5
B-Nationality	0.2857	0.2500	0.2667	8
B-Organisation	0.7056	0.6250	0.6629	280
B-Person	0.8077	0.8235	0.8155	102
B-Quantity	0.6349	0.7273	0.6780	55
B-Temporal	0.7857	0.7021	0.7416	47
B-Weapon	0.3750	0.2368	0.2903	38
I-DocumentReference	0.6667	0.2651	0.3793	83
I-Location	0.7259	0.5396	0.6190	265
I-MilitaryPlatform	0.0000	0.0000	0.0000	16
I-Money	1.0000	0.5000	0.6667	10
I-Nationality	0.0000	0.0000	0.0000	1
I-Organisation	0.7135	0.6346	0.6718	416
I-Person	0.8086	0.7572	0.7821	173
I-Quantity	0.5385	0.2692	0.3590	26
I-Temporal	0.9722	0.5645	0.7143	62
I-Weapon	0.4286	0.1333	0.2034	45
O	0.8567	0.9442	0.8983	3420
accuracy			0.8165	5274
macro avg	0.5746	0.4290	0.4733	5274
weighted avg	0.8017	0.8165	0.8028	5274

- CRF = true,
attention = cosine similarity,
layer = 3,
input = semantic(word-emb) + dependency parsing + TF-IDF

```

Epoch:1, Training loss: 110716.00, train acc: 0.6355, val loss: 10871.57, val acc: 0.5988, time: 112.84s
Epoch:2, Training loss: 22081.70, train acc: 0.6815, val loss: 5854.73, val acc: 0.6183, time: 114.19s
Epoch:3, Training loss: 13060.41, train acc: 0.7060, val loss: 5518.80, val acc: 0.6494, time: 113.47s
Epoch:4, Training loss: 12710.90, train acc: 0.7274, val loss: 4915.12, val acc: 0.6712, time: 113.44s
Epoch:5, Training loss: 11731.90, train acc: 0.7459, val loss: 4797.11, val acc: 0.6799, time: 114.49s
Epoch:6, Training loss: 10450.94, train acc: 0.7586, val loss: 4631.74, val acc: 0.6890, time: 117.70s
Epoch:7, Training loss: 9333.61, train acc: 0.7717, val loss: 4247.90, val acc: 0.7048, time: 115.80s
Epoch:8, Training loss: 7977.70, train acc: 0.7894, val loss: 4671.04, val acc: 0.7218, time: 116.84s
Epoch:9, Training loss: 7774.95, train acc: 0.7755, val loss: 4160.72, val acc: 0.6936, time: 116.20s
Epoch:10, Training loss: 7478.86, train acc: 0.7922, val loss: 3870.66, val acc: 0.7126, time: 115.67s
Epoch:11, Training loss: 6984.84, train acc: 0.7922, val loss: 3563.35, val acc: 0.7050, time: 115.68s
Epoch:12, Training loss: 6415.22, train acc: 0.8113, val loss: 3297.46, val acc: 0.7372, time: 115.50s
Epoch:13, Training loss: 6096.24, train acc: 0.8185, val loss: 3533.37, val acc: 0.7383, time: 116.96s
Epoch:14, Training loss: 5579.98, train acc: 0.8205, val loss: 3485.39, val acc: 0.7317, time: 116.31s
Epoch:15, Training loss: 5338.80, train acc: 0.8336, val loss: 3483.74, val acc: 0.7537, time: 115.42s
Epoch:16, Training loss: 5160.11, train acc: 0.8371, val loss: 3207.05, val acc: 0.7457, time: 116.00s
Epoch:17, Training loss: 4998.54, train acc: 0.8443, val loss: 3257.67, val acc: 0.7529, time: 116.16s
Epoch:18, Training loss: 4616.84, train acc: 0.8438, val loss: 3057.23, val acc: 0.7473, time: 115.37s
Epoch:19, Training loss: 4125.74, train acc: 0.8766, val loss: 2998.58, val acc: 0.7711, time: 115.69s
Epoch:20, Training loss: 4051.88, train acc: 0.8800, val loss: 2913.84, val acc: 0.7791, time: 115.33s

```

- CRF = true,
attention = scaled dot product,
layer = 2,
input = semantic(word-emb) + dependency parsing + TF-IDF
(best model)

```

Epoch:1, Training loss: 15775.15, train acc: 0.7188, val loss: 5071.52, val acc: 0.6670, time: 122.11s
Epoch:2, Training loss: 10649.92, train acc: 0.7746, val loss: 3895.02, val acc: 0.7294, time: 121.03s
Epoch:3, Training loss: 8068.72, train acc: 0.8084, val loss: 3187.46, val acc: 0.7582, time: 120.64s
Epoch:4, Training loss: 6557.02, train acc: 0.8251, val loss: 2752.28, val acc: 0.7704, time: 119.90s
Epoch:5, Training loss: 5524.34, train acc: 0.8443, val loss: 2422.52, val acc: 0.7871, time: 120.47s
Epoch:6, Training loss: 4918.44, train acc: 0.8505, val loss: 2260.65, val acc: 0.7969, time: 120.11s
Epoch:7, Training loss: 4347.20, train acc: 0.8680, val loss: 2100.38, val acc: 0.8038, time: 120.48s
Epoch:8, Training loss: 3978.31, train acc: 0.8834, val loss: 2030.43, val acc: 0.8093, time: 120.93s
Epoch:9, Training loss: 3498.39, train acc: 0.8895, val loss: 2052.07, val acc: 0.8030, time: 121.17s
Epoch:10, Training loss: 3252.48, train acc: 0.9024, val loss: 1963.91, val acc: 0.8138, time: 120.44s
Epoch:11, Training loss: 2895.81, train acc: 0.9088, val loss: 1978.44, val acc: 0.8155, time: 122.87s
Epoch:12, Training loss: 2713.80, train acc: 0.9182, val loss: 1995.10, val acc: 0.8106, time: 121.09s
Epoch:13, Training loss: 2393.25, train acc: 0.9321, val loss: 1961.81, val acc: 0.8180, time: 120.07s
Epoch:14, Training loss: 2267.05, train acc: 0.9380, val loss: 2001.62, val acc: 0.8206, time: 120.17s
Epoch:15, Training loss: 2052.23, train acc: 0.9444, val loss: 1934.64, val acc: 0.8146, time: 120.67s
Epoch:16, Training loss: 1871.79, train acc: 0.9465, val loss: 2080.09, val acc: 0.8161, time: 121.44s
Epoch:17, Training loss: 1717.32, train acc: 0.9508, val loss: 2144.29, val acc: 0.8235, time: 120.59s
Epoch:18, Training loss: 1641.54, train acc: 0.9620, val loss: 2091.52, val acc: 0.8233, time: 121.08s
Epoch:19, Training loss: 1473.61, train acc: 0.9608, val loss: 2251.54, val acc: 0.8176, time: 121.25s
Epoch:20, Training loss: 1375.45, train acc: 0.9630, val loss: 2146.09, val acc: 0.8267, time: 120.76s

```

	precision	recall	f1-score	support
B-DocumentReference	0.5000	0.2500	0.3333	20
B-Location	0.6442	0.7204	0.6802	186
B-MilitaryPlatform	0.0000	0.0000	0.0000	16
B-Money	0.5000	0.2000	0.2857	5
B-Nationality	0.3333	0.1250	0.1818	8
B-Organisation	0.7122	0.7071	0.7097	280
B-Person	0.7768	0.8529	0.8131	102
B-Quantity	0.6667	0.6909	0.6786	55
B-Temporal	0.7727	0.7234	0.7473	47
B-Weapon	0.3226	0.2632	0.2899	38
I-DocumentReference	0.8235	0.3373	0.4786	83
I-Location	0.6391	0.6415	0.6403	265
I-MilitaryPlatform	0.0000	0.0000	0.0000	16
I-Money	1.0000	0.4000	0.5714	10
I-Nationality	0.0000	0.0000	0.0000	1
I-Organisation	0.7446	0.7428	0.7437	416
I-Person	0.7460	0.8150	0.7790	173
I-Quantity	0.6250	0.3846	0.4762	26
I-Temporal	0.7541	0.7419	0.7480	62
I-Weapon	0.4000	0.1778	0.2462	45
O	0.8924	0.9193	0.9057	3420
accuracy			0.8282	5274
macro avg	0.5644	0.4616	0.4909	5274
weighted avg	0.8188	0.8282	0.8206	5274