LIAT Team's Wikipedia Classifier at NTCIR-15 SHINRA2020-ML: Classification Task

Kouta Nakayama
RIKEN AIP, Japan
kouta.nakayama@riken.jp

Satoshi Sekine
RIKEN AIP, Japan
satoshi.sekine@riken.jp

ABSTRACT

This paper reports the document classification system that our team LIAT submitted to the classification task in NTCIR-15 SHINRA2020-ML[2]. We used the outputs of BERT[1] as document embeddings to deal with the longer sentences of Wikipedia. We used the Transformer[3] encoder to classify the document embeddings into each class. Our system was not better than other submission results, but we hope that our results will also be used as a resource.

TEAM NAME

LIAT

SUBTASKS

SHINRA2020-ML: Classification Task

1 INTRODUCTION

SHINRA2020-ML[2] is a shared task to classify Wikipedia in 30 languages into Extended Named Entity (ENE) Hierarchy (ENEH). This task employs version 8.0 of ENEH and classification into 221 classes. We participated in all 30 languages targeted in this task. In this paper, we describe in detail the system we used for classification.

In recent years, pre-trained language models, such as BERT[1], have been utilized for document classification. However, BERT and other transformer-based models can generally only handle around 500 tokens at once due to memory limitations. Therefore, we exploit the outputs of BERT as document embeddings and classify the embeddings into each class using the encoder of the Transformer. This approach allows us to handle input tokens longer than the limit of BERT.

2 MODEL

2.1 BERT for Document Embedding

We fine-tune BERT on the classification task to obtain task-specific document embedding. Specifically, the documents to be classified are divided into a number of tokens that BERT can handle, and they are classified using BERT. In general, when classifying with BERT, the special token [CLS] is combined with the input, and the output for the special token is the classification result, as shown in Figure 1. Here, we handle the intermediate output $T_{[CLS]}$ of fine-tuned BERT for [CLS] as a task-specific document embedding.

2.2 Transformer Encoder for Document Classification

We classify the document embeddings obtained by BERT using the encoder of Transformer, as shown in Figure 2. During training, document embeddings are fixed. The encoder, like BERT, uses the output corresponding to [CLS] to classify the document. The embedding assigned to [CLS] is generated using a dedicated embedding layer and is updated during training.

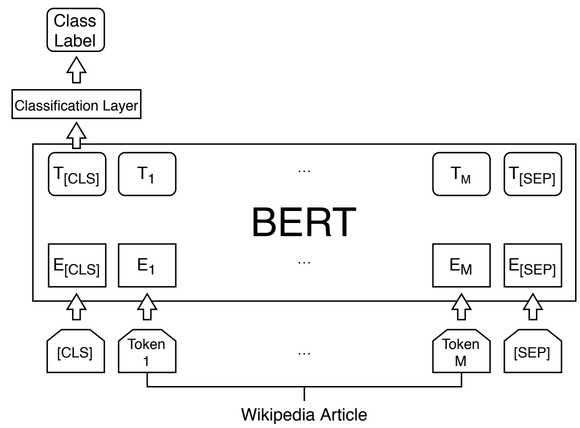

Figure 1: BERT for Document Embedding

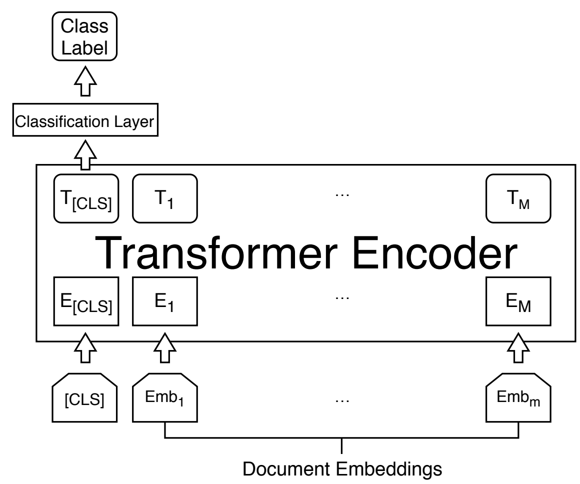

Figure 2: Transformer Encoder for Document Classification

3 EXPERIMENTS

3.1 Model Details

We use pre-trained BERT-Base in 104 languages.[1] Also, we use the Transformers library[4] to build our models. The hyperparameters

[1]We used the cased model. https://github.com/google-research/bert/blob/master/multilingual.md

Hyperparameter	
Epoch	*5*
Batch size	*512*
Gradient accumulation steps	*1*
Sequence length	*256*
Hidden layer dropout	*0.1*
Attention dropout	*0.1*
Learning rate	*5e-5*
Adam $\beta 1$	*0.9*
Adam $\beta 2$	*0.999*
Adam ϵ	*1e-6*
Weight decay	*0.05*

Table 1: Hyperparameters for training BERT.

Hyperparameter	
Epoch	*5*
Batch size	*128*
Gradient accumulation steps	*4*
Sequence length	*63*
Hidden layer dropout	*0.1*
Attention dropout	*0.1*
Learning rate	*5e-5*
Adam $\beta 1$	*0.9*
Adam $\beta 2$	*0.999*
Adam ϵ	*1e-6*
Weight decay	*0.05*

Table 2: Hyperparameters for training Transformer encoder.

we used to train BERT and Transformer encoder are shown in Table 1 and Table 2, respectively. We used the same values as in Table 1 for hyperparameters not mentioned in Table 2.

3.2 Submission Results

We show the official results of the SHINRA2020-ML in Table 3. All scores are macro average F1 measure. *Late submission means reference result submitted after the deadline. The results of our system seem to be inferior in all languages to the results of the best system, such as FPTAI and uomfj. The difference between our system and the best system is shown in Table 4. We seem to have a very low score in hi for our system. Since we did not conduct any hyperparameter search, we consider the training of the model to be converging to a local minimum. In future research, we will monitor the development data score during training to prevent learning failure, such as this one. Our system seems to score particularly poorly in minor languages. We may need to conduct a hyperparameter search, as the learning accuracy depends more heavily on the hyperparameters the less data we have. In future research, we will be searching for hyperparameters of learning as far as our computational resources will enable.*

4 CONCLUSIONS

This paper describes the our system submitted to SHINRA2020-ML. We did not achieve a higher score than other systems. However, the purpose of SHINRA2020-ML is collaborative resource construction, and our results will also be used for ensembles and other purposes. In future work we will adjust our training in more detail, such as the exploring of hyperparameters.

REFERENCES

[1] *Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In* Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Association for Computational Linguistics, Minneapolis, Minnesota, 4171–4186. https://doi.org/10.18653/v1/N19-1423*

[2] *Satoshi Sekine, Masako Nomoto, Kouta Nakayama, Asuka Sumida, Koji Matsuda, and Maya Ando. 2020. Overview of SHINRA2020-ML Task. In* In Proceedings of the NTCIR-15 Conference.

[3] *Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In* Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.). Curran Associates, Inc., 5998–6008. http://papers.nips.cc/paper/7181-attention-is-all-you-need.pdf*

[4] *Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. 2019. HuggingFace's Transformers: State-of-the-art Natural Language Processing.* ArXiv *abs/1910.03771 (2019).*

Group ID		CMVS	FPTAI	LIAT	PribL	PribL	RH312	TKUIM	ousia	uomfj	uomfj	uomfj	vlp	FPTAI	HUKB	HUKB	HUKB	LIAT	ousia
Method		BOW_AT T	BERT	ML-BERT	BERTGR U	BERTLIN CONCAT	RnnGnnXl mr	bert	RoBERTa +wiki2vec +wikidata	jointrep	jointrepPo stprocess	jointrepUn ionPostpr ocess	mlr	BERT	AB	ABC	AC	ML-BERT	RoBERTa +wiki2vec +wikidata
Late Submission														✓	✓	✓	✓	✓	✓
ar	Arabic	5.26	73.25	63.16	**76.27**	75.45	-	4.97	70.52	64.55	64.55	64.55	-	73.25	30.98	30.98	13.51	-	70.52
bg	Bulgarian	-	**83.77**	75.20	-	-	82.13	3.78	-	83.07	83.07	83.07	-	83.28	60.86	61.06	28.09	-	-
ca	Catalan, Valencian	-	52.55	76.28	-	-	-	3.37	-	**79.82**	**79.82**	**79.82**	-	**81.10**	42.34	42.54	16.26	-	80.63
cs	Czech	-	**84.47**	79.46	-	81.19	-	3.37	-	81.29	81.29	81.29	-	83.74	52.61	52.61	18.86	-	-
da	Danish	-	**82.30**	74.80	-	-	-	3.67	-	80.56	80.56	80.56	-	81.74	49.01	49.01	13.99	-	-
de	German	-	22.62	79.49	80.24	79.83	-	3.15	**81.86**	81.03	81.03	81.03	-	81.26	53.72	53.82	26.81	-	-
el	Greek, Modern (1453-)	-	**84.40**	72.43	-	-	-	2.47	-	-	-	-	-	84.10	7.51	7.51	7.51	-	-
en	English	-	82.23	78.56	81.27	80.12	-	3.58	-	**82.73**	82.57	82.68	-	81.96	45.11	45.11	11.92	-	-
es	Spanish, Castilian	-	80.60	77.73	80.30	80.72	-	2.38	80.94	**81.39**	**81.39**	**81.39**	-	80.60	49.21	49.11	19.50	-	80.94
fa	Persian	-	**81.70**	75.42	-	-	-	3.07	-	80.38	80.38	80.38	-	81.52	45.59	45.59	15.66	-	-
fi	Finnish	-	**83.62**	79.13	-	-	-	3.37	-	80.91	80.91	80.91	-	83.36	53.15	53.45	17.06	-	-
fr	French	-	21.59	76.88	77.93	78.52	80.31	2.88	**81.01**	78.21	78.21	78.21	-	80.68	43.84	43.74	11.23	-	**81.01**
he	Hebrew	-	**83.79**	79.11	-	-	-	3.37	-	81.09	81.09	81.09	-	**84.21**	59.95	60.05	15.78	-	-
hi	Hindi	-	**76.43**	16.49	-	-	71.70	3.65	69.75	66.67	66.67	66.67	-	75.65	39.70	39.51	22.02	-	69.75
hu	Hungarian	-	**85.46**	78.93	-	-	-	1.98	-	85.02	85.02	85.02	-	84.78	69.15	69.44	26.09	-	-
id	Indonesian	-	**81.93**	72.45	-	-	77.56	4.37	-	78.51	78.51	78.51	-	81.65	44.07	44.47	16.28	-	-
it	Italian	-	26.55	81.36	81.92	81.89	-	2.57	81.21	**82.02**	**82.02**	**82.02**	-	**82.81**	45.55	45.55	12.06	-	81.21
ko	Korean	-	**83.67**	80.38	81.51	81.04	-	3.49	-	82.51	82.51	82.51	-	**83.77**	63.68	63.98	13.95	-	82.64
nl	Dutch, Flemish	-	**83.29**	79.86	80.95	81.26	-	2.38	-	81.64	81.64	81.64	-	83.17	42.36	42.45	17.12	-	-
no	Norwegian	-	**80.53**	76.50	-	78.39	-	3.58	-	78.79	78.79	78.79	-	80.17	34.58	34.58	11.33	-	-
pl	Polish	-	**84.53**	80.60	82.73	83.46	-	2.59	-	84.52	84.52	84.52	-	84.07	62.72	63.51	32.55	-	-
pt	Portuguese	-	**83.23**	78.49	82.36	81.88	-	3.28	81.40	80.87	80.87	80.87	-	82.70	42.32	42.62	16.10	-	81.40
ro	Romanian, Moldavian, Moldovan	-	**84.60**	76.17	-	-	-	3.57	-	80.83	80.83	80.83	-	84.60	57.60	57.70	28.50	-	-
ru	Russian	-	**84.08**	79.09	82.60	83.07	-	2.78	-	82.90	82.90	82.90	-	83.44	42.04	42.24	11.30	-	-
sv	Swedish	-	**83.18**	71.63	-	-	-	2.49	-	-	-	-	-	**83.44**	50.32	50.62	21.98	79.58	-
th	Thai	-	**81.26**	49.58	-	-	76.77	4.45	76.36	65.02	65.02	65.02	-	81.16	39.98	40.38	24.05	-	76.36
tr	Turkish	-	**86.50**	77.19	84.36	83.23	83.28	3.56	-	84.85	84.85	84.85	-	86.03	61.88	62.48	16.73	-	-
uk	Ukrainian	-	**83.12**	78.71	-	-	-	2.38	-	81.61	81.61	81.61	-	82.61	60.29	60.19	22.51	-	-
vi	Vietnamese	-	**80.34**	75.24	-	-	-	2.98	-	77.06	77.06	77.06	3.08	**80.42**	60.38	60.48	22.14	-	78.42
zh	Chinese	-	**81.25**	77.97	78.38	79.37	-	-	79.76	78.58	78.58	78.58	-	80.60	21.22	21.42	17.57	-	79.76

Table 3: Official results of SHINRA2020-ML

Group ID		LIAT		
Method		ML-BERT	Max	Diff
Late Submission				
it	Italian	81.36	82.81	-1.45
de	German	79.49	81.86	-2.37
zh	Chinese	77.97	81.25	-3.28
ko	Korean	80.38	83.77	-3.39
nl	Dutch, Flemish	79.86	83.29	-3.42
es	Spanish, Castilian	77.73	81.39	-3.66
pl	Polish	80.60	84.53	-3.94
no	Norwegian	76.50	80.53	-4.03
fr	French	76.88	81.01	-4.12
en	English	78.56	82.73	-4.17
uk	Ukrainian	78.71	83.12	-4.41
fi	Finnish	79.13	83.62	-4.50
pt	Portuguese	78.49	83.23	-4.74
ca	Catalan, Valencian	76.28	81.10	-4.83
ru	Russian	79.09	84.08	-4.99
cs	Czech	79.46	84.47	-5.01
he	Hebrew	79.11	84.21	-5.10
vi	Vietnamese	75.24	80.42	-5.18
fa	Persian	75.42	81.70	-6.28
hu	Hungarian	78.93	85.46	-6.53
da	Danish	74.80	82.30	-7.50
ro	Romanian, Moldavian, Moldovan	76.17	84.60	-8.44
bg	Bulgarian	75.20	83.77	-8.57
tr	Turkish	77.19	86.50	-9.32
id	Indonesian	72.45	81.93	-9.49
sv	Swedish	71.63	83.44	-11.80
el	Greek, Modern (1453-)	72.43	84.40	-11.97
ar	Arabic	63.16	76.27	-13.11
th	Thai	49.58	81.26	-31.68
hi	Hindi	16.49	76.43	-59.94

Table 4: Difference from the best system