

Punctuation Restoration with Ensemble of Neural Network Classifiers and Pre-trained Transformers

Michał Marcińczuk

michal.marcinczuk@samurailabs.ai

Samurai Labs / Wrocław University of Science and Technology

October 2021

0 Outline

| 1

① Introduction

② Our approach

③ Evaluation & summary

1 Outline

| 2

① Introduction

② Our approach

③ Evaluation & summary

Input

w środę 23 marca w godzinach rannych silne trzęsienie ziemi dotknęło północno wschodnie wybrzeże japonii

Expected

w środę, 23 marca, w godzinach rannych, silne trzęsienie ziemi dotknęło północno- wschodnie wybrzeże japonii.

2 Outline

| 4

① Introduction

② Our approach

③ Evaluation & summary

2 Task representation

Text

w środę, 23 marca, w
godzinach rannych, silne
trzęsienie ziemi dotknęło
północno-wschodnie
wybrzeże Japonii.

2 Task representation

Text

w środę, 23 marca, w
godzinach rannych, silne
trzęsienie ziemi dotknęło
północno-wschodnie
wybrzeże japonii.

Representation

w	NONE
środe	COMMA
23	NONE
marca	COMMA
w	NONE
godzinach	NONE
rannych	COMMA
silne	NONE
trzęsienie	NONE
ziemi	NONE
dotknęło	NONE
północno	HYPHEN
wschodnie	NONE
wybrzeże	NONE
japonii	FULLSTOP

2 Neural Network Architecture

Input text: w budynku c 13 wrocławskiej politechniki od godziny 12 00

Tokens: w budynku C 13 wrocławskiej politechniki od godziny 12 00

Subtokens: w budynku C 13 wrocla wskiej poli tech niki od godziny 12 00

Embeddings: transformed-based language model (BERT/RobERTa)

Classifier: dropout + linear layer + argmax

Labels: O O DASH O O O O O COMMA O

Output text: w budynku c-13 wrocławskiej politechniki od godziny 12,00

Parameter	Value
sequence length	512
learning rate	0.00006
optimizer	Adam with weight decay
scheduler	constant
batch size	32
language models	Polish RoBERTa, HerBERT
language models variants	base, large
dropout	0.2
gradient accumulation steps	4
epochs	5

Table: Training parameters

3 Outline

| 8

① Introduction

② Our approach

③ Evaluation & summary

3 Test-A

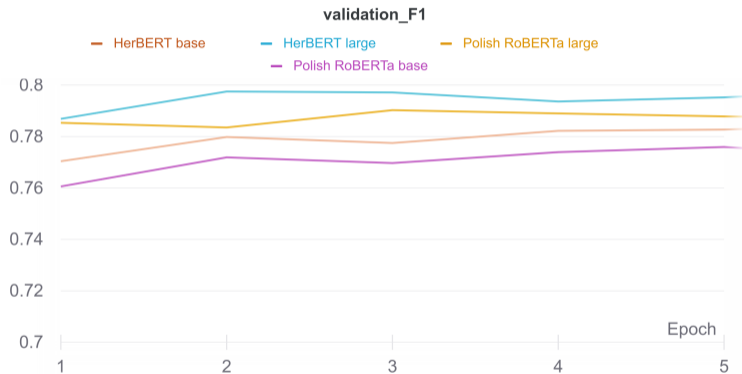


Figure: F1 scores on the test-A dataset for various language models and epochs

	M1	M2	M3	Ensemble
full stop	87.51	88.03	85.49	88.47
comma	77.63	74.86	74.20	77.08
question mark	77.22	79.85	75.10	79.84
exclamation mark	20.00	20.00	36.36	20.00
hyphen	64.43	60.48	63.41	66.30
colon	76.43	79.83	74.94	79.73
ellipses	0.00	0.00	0.00	0.00
all	80.53	79.78	78.17	81.23

Table: Evaluation of the models on the test-D dataset.

Submitter	test-D F1
Krzysztof Wróbel	81.29
Norbert Ropiak	81.25
Michał Marcińczuk	81.23
Dmytro Zhylko	80.34
Tomasz Ziętkiewicz	78.37
Mateusz Piotrowski	76.65

Table: Task leaderboard

3 Conclusions

- ▶ training complexity:
 - > a single model requires 3-4 hours on a single GPU,
- ▶ processing complexity:
 - > 1 minute to process the test-D dataset (244 docs) on a single GPU,
 - > requires "only" 1.5 GB memory on GPU,
 - > ensemble of three models increased the performance by 0.7pp but it nearly tripled the processing time and memory requirements,

