Multilingual-BERT: Classification Experiment

May 13, 2020

1   Data description

Russian dataset was taken from RuTweetCorp and consists of 114911 positive and 111923 negative tweets. English dataset was taken from Kaggle: Twitter sentiment analysis and consists of 56457 positive and 43532 negative tweets. Corresponding datasets' shapes can be viewed in Table 1.

Data	Split	Size
RuTweetCorp	train	181467
	val	22683
	test	22684
Twitter sentiment analysis	train	99989

Table 1: Dataset sizes

2   Experiments

Pre-Trained MBERT: model without fine-tuning, evaluated on the test dataset.

MBERT-Ru: model fine-tuned on Russian data for 1 epoch, evaluated on the Russian test dataset

MBERT-En: model fine-tuned on English data for 6 epochs, evaluated on the Russian test dataset

MBERT-En-Ru: model fine-tuned on English data for 6 epochs and Russian data for 1 epoch, evaluated on the test dataset.

Pre-Trained RuBERT: model without fine-tuning, evaluated on the test dataset

RuBERT: model fine-tuned on Russian data for 1 epoch, evaluated on the Russian test dataset

Comparison table: Table 2

Data can be found here: GoogleDrive

Model	Accuracy	Precision	Recall	F1
Pre-Trained MBERT	49.5%	49.5%	99.9%	66.2%
MBERT-Ru	99.9%	99.9%	99.9%	99.9%
MBERT-En	75.7%	73.4%	79.8%	76.5%
MBERT-En-Ru	99.9%	100%	99.9%	99.9%
Pre-Trained RuBERT	50.0%	70.8%	1%	3%
RuBERT	99.9%	100%	99.8%	99.9%

Table 2: Results obtained on Test dataset