

Проект Transfer-learning for Mail.ru Group dataset

План работы.

1. EDA
2. Подготовить фиксированный трайн-тест
3. Baseline tf-idf with 3 layer perceptron
4. Transfer learning models - создать хорошую модель классификации для использования после пред обученных моделей:
 - a. **Bert-base-multilingual-uncased**
 - b. **Bert-base-multilingual-cased**
 - c. Xlm-mlm-xnli15-1024
 - d. Xlm-mlm-tlm-xnli15-1024
 - e. **Xlm-mlm-17-1280**
 - f. Xlm-mlm-100-1280
 - g. *ulmfit
 - h. **ELMO**
5. Расширить данные по категориям за счет использования ответов к соответствующим вопросам
6. Прогнать Baseline и модели
7. Сравнить результаты

Model	Add info	Clea r	Nor mali ze	Answers use	Embeddi ngs	Resu lt F1
Perceptr on	Weight init, 3 layers linear, dropout - 0.2	+	+	-	Tf-idf, unigram	59.2
lstm_att _cnn	<pre>model_with_att((emb_layer): Embedding(221008, 300) (GRu): LSTM(300, 256, num_layers=3, dropout=0.4, bidirectional=True) (att_soft): Softmax(dim=2) (q_proj): Linear(in_features=512, out_features=512, bias=True) (k_proj): Linear(in_features=512,</pre>	-	-	+ (exempl es = 1,6m)	fasttext	62.8

	<pre> out_features=512, bias=True) (v_proj): Linear(in_features=512, out_features=512, bias=True) (relu): ReLU() (cnn_3gr): Conv1d(1024, 256, kernel_size=(3,), stride=(1,)) (cnn_4gr): Conv1d(1024, 256, kernel_size=(4,), stride=(1,)) (cnn_5gr): Conv1d(1024, 256, kernel_size=(5,), stride=(1,)) (lin1): Linear(in_features=768, out_features=512, bias=True) (lin2): Linear(in_features=512, out_features=28, bias=True) (linear_1): Linear(in_features=768, out_features=512, bias=True) (non_linear_function): ReLU() (drop_out): Dropout(p=0.2, inplace=False) (linear_2): Linear(in_features=512, out_features=28, bias=True) (bn1): BatchNorm1d(768, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True) (bn2): BatchNorm1d(512, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)) </pre>					
lstm_att _cnn	same	+	-	-	fasttext	60.9
lstm_att _cnn	same	-	-	-	fasttext	61.8
lstm_att _cnn	Same+ emb(dropout(0.3)) (12 epoch)	-	-	-	fasttext	63.1
lstm_att _cnn	Same+ emb(dropout(0.3)) обучиться на ответах к каждому из которых канкатенирован рандомно вопрос из той же категории. Полученную модель дообучить на чистых вопросах	-	-	+(EXA MPLES = ALL)	fasttext	63.33
lstm_att	model_with_att(-	-	-	fasttext	63.30

_cnn	<pre> (emb_layer): Embedding(221008, 300) (drop_emb): Dropout(p=0.3, inplace=False) (LSTM): LSTM(300, 315, num_layers=2, batch_first=True, dropout=0.4, bidirectional=True) (q_proj): Linear(in_features=630, out_features=630, bias=True) (k_proj): Linear(in_features=630, out_features=630, bias=True) (v_proj): Linear(in_features=630, out_features=630, bias=True) (att_soft): Softmax(dim=2) (cnn_3gr): Conv1d(1260, 315, kernel_size=(3,), stride=(1,)) (cnn_4gr): Conv1d(1260, 315, kernel_size=(4,), stride=(1,)) (cnn_5gr): Conv1d(1260, 315, kernel_size=(5,), stride=(1,)) (bn1): BatchNorm1d(945, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True) (linear_1): Linear(in_features=945, out_features=630, bias=True) (relu): ReLU() (drop_out): Dropout(p=0.2, inplace=False) (bn2): BatchNorm1d(630, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True) (linear_2): Linear(in_features=630, out_features=28, bias=True)) + Пад сек </pre>					
lstm_att _cnn_pa d		-	-	-	Uncases bert	56,7