presenation

December 6, 2022

[13]:
```
from transformers import AutoTokenizer, AutoModelForSequenceClassification
import torch
import pandas as pd
tokenizer = AutoTokenizer.from_pretrained("nlptown/
 ↪bert-base-multilingual-uncased-sentiment")
model = AutoModelForSequenceClassification.from_pretrained("nlptown/
 ↪bert-base-multilingual-uncased-sentiment")
df = pd.read_csv('../data/train-balanced-sarcasm.csv', encoding='utf8',␣
 ↪on_bad_lines='skip')
```

[34]:
```
summary1 = pd.read_csv('1iter_summary.csv')
summary2 = pd.read_csv('2iter_summary.csv')
```

1 When Sentiment Analysis fails

[14]:
```
comments = [
    "This is a lovely book. I would recommend it to anyone",
    "This book is just a list of numbers. Useless.",
    "OMG, what a suspenseful read",
    "More interesting than the Bible",
    "This is, without a doubt, a more touching story than twilight",
    "A great read. Captivating. I couldn't put it down anymore, when I have␣
 ↪found out that 0.629 is there",
]

tokens = [tokenizer.encode(comment, return_tensors="pt") for comment in␣
 ↪comments]
results = [model(token) for token in tokens]
classes = [torch.argmax(result.logits) for result in results]
classes
```

[14]:
```
[tensor(4), tensor(0), tensor(4), tensor(4), tensor(4), tensor(4)]
```

2 Sarcasm detection is needed

[ ]:
```
from simpletransformers.classification import ClassificationModel
import os
os.environ["TOKENIZERS_PARALLELISM"] = "false"

model = ClassificationModel(
    "roberta", "../models/robertatwitter_reddit/checkpoint-06-12-22-14-57",␣
 ↪use_cuda=False, keepdims=False
)
```

[38]:
```
pred, _ = model.predict(comments)
pred
```

```
  0%|          | 0/6 [00:00<?, ?it/s]

  0%|          | 0/1 [00:00<?, ?it/s]

/home/reka/anaconda3/envs/irony/lib/python3.10/site-
packages/simpletransformers/classification/classification_model.py:2253:
FutureWarning: Unlike other reduction functions (e.g. `skew`, `kurtosis`), the
default behavior of `mode` typically preserves the axis it acts along. In SciPy
1.11.0, this behavior will change: the default value of `keepdims` will become
False, the `axis` over which the statistic is taken will be eliminated, and the
value None will no longer be accepted. Set `keepdims` to True or False to avoid
this warning.
  mode_pred, counts = mode(pred_row)
```

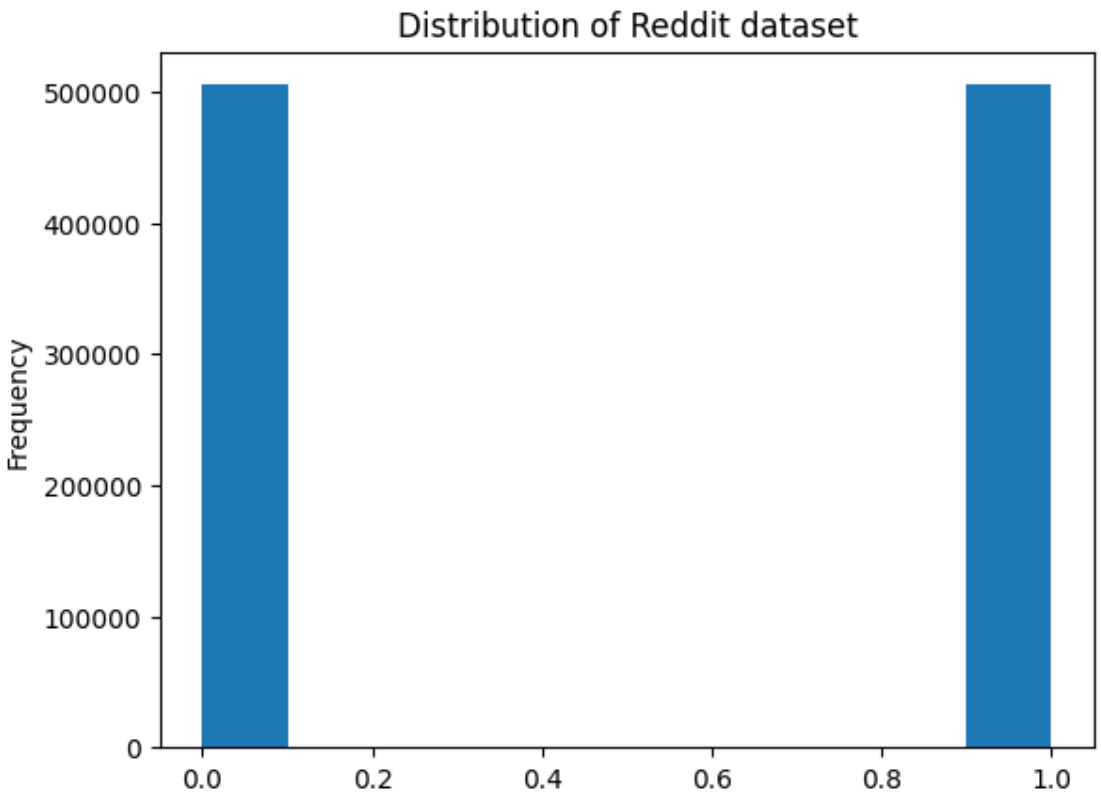
[38]:
```
[0, 0, 1, 1, 0, 0]
```

2.0.1 Dataset

- Twitter (60K data, unbalanced)
- MUSTARD (small data available)
- Reddit (1M data, balanced)

[10]:
```
df['label'].plot(kind='hist', title='Distribution of Reddit dataset')
```

[10]:
```
<AxesSubplot: title={'center': 'Distribution of Reddit dataset'},
ylabel='Frequency'>
```



2.0.2 Model

- pretrained RoBERTa transformer model
- pretrained models used:
  1. https://huggingface.co/cardiffnlp/twitter-roberta-base-irony
  2. https://huggingface.co/jkhan447/sarcasm-detection-RoBerta-base-POS

2.0.3 Hardware Spec

2.0.4 Evaluation

First Iteration

[35]:
```
summary1
```

[35]:

	Pretrained Models	Dataset	Precision	Recall	F1-Score	MCC
0	NoName	MUSTARD	0.675000	0.683544	0.679245	0.243415
1	twitter	MUSTARD	0.862069	0.316456	0.462963	0.301948

Second Iteration

[36]:
```
summary2
```

[36]:

	Pretrained Models	Dataset	Precision	Recall	F1-Score	MCC
0	twitter	twitter	0.550207	0.331998	0.414116	0.173506
1	twitter	reddit	0.677682	0.749624	0.711840	0.505483

2.1 References

- S. K. Bharti, R. K. Gupta, P. K. Shukla, W. A. Hatamleh, H. Tarazi, S. J. Nuagah: Multimodal Sarcasm Detection: A Deep Learning Approach, Wireless Communications and Mobile Computing, Article ID 1653696 (2022).
- S. Oprea, W. Magdy: iSarcasm: A Dataset of Intended Sarcasm, arXiv:1911.03123 (2020).
- H. Yaghoobian, H. R. Arabnia, K. Rasheed: Sarcasm Detection: A Comparative Study, arXiv:2107.02276 (2021).
- A. Kumar, V. Anand: Transformers on Sarcasm Detection with Context, In Proceedings of the Second Workshop on Figurative Language Processing, pages 88–92, Online. Association for Computational Linguistics (2020).
- https://www.projectpro.io/article/bert-nlp-model-explained/558

Special thanks for the Romanian Orthodox Church for (unknowingly) providing computational resources

http://museikon.ro