2306aml133-kritika-assignment13

August 27, 2023

[1]:
```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from tensorflow.keras.preprocessing.text import Tokenizer
from tensorflow.keras.preprocessing.sequence import pad_sequences
from tensorflow.keras.layers import Embedding, LSTM, Dense, Dropout
from tensorflow.keras.models import Sequential
from sklearn.metrics import classification_report

url = "https://raw.githubusercontent.com/Kavitha-Kothandaraman/
 ↪Sarcasm-Detection-NLP/master/Sarcasm_Headlines_Dataset.json"
data = pd.read_json(url, lines=True)

train_data, test_data = train_test_split(data, test_size=0.2, random_state=42)

tokenizer = Tokenizer(oov_token="<OOV>")
tokenizer.fit_on_texts(train_data['headline'])

vocab_size = len(tokenizer.word_index) + 1

train_sequences = tokenizer.texts_to_sequences(train_data['headline'])
train_padded = pad_sequences(train_sequences, maxlen=20, padding='post',␣
 ↪truncating='post')

test_sequences = tokenizer.texts_to_sequences(test_data['headline'])
test_padded = pad_sequences(test_sequences, maxlen=20, padding='post',␣
 ↪truncating='post')

model = Sequential()
model.add(Embedding(vocab_size, 32, input_length=20))
model.add(LSTM(64, dropout=0.2))
model.add(Dense(1, activation='sigmoid'))

model.compile(optimizer='adam', loss='binary_crossentropy',␣
 ↪metrics=['accuracy'])
```

```
model.fit(train_padded, train_data['is_sarcastic'], epochs=5, batch_size=64,␣
 ↪validation_split=0.1)

test_loss, test_acc = model.evaluate(test_padded, test_data['is_sarcastic'])
print("Test accuracy:", test_acc)

predictions = (model.predict(test_padded) > 0.5).astype(int)
print(classification_report(test_data['is_sarcastic'], predictions))
```

```
Epoch 1/5
301/301 [==============================] - 15s 35ms/step - loss: 0.4420 -
accuracy: 0.7729 - val_loss: 0.3214 - val_accuracy: 0.8671
Epoch 2/5
301/301 [==============================] - 10s 33ms/step - loss: 0.2018 -
accuracy: 0.9212 - val_loss: 0.3367 - val_accuracy: 0.8657
Epoch 3/5
301/301 [==============================] - 10s 33ms/step - loss: 0.1035 -
accuracy: 0.9636 - val_loss: 0.4374 - val_accuracy: 0.8559
Epoch 4/5
301/301 [==============================] - 9s 30ms/step - loss: 0.0645 -
accuracy: 0.9789 - val_loss: 0.5143 - val_accuracy: 0.8507
Epoch 5/5
301/301 [==============================] - 10s 33ms/step - loss: 0.0421 -
accuracy: 0.9869 - val_loss: 0.5242 - val_accuracy: 0.8376
167/167 [==============================] - 1s 6ms/step - loss: 0.5403 -
accuracy: 0.8302
Test accuracy: 0.8302134275436401
167/167 [==============================] - 1s 5ms/step
              precision    recall  f1-score   support

           0       0.82      0.90      0.86      2996
           1       0.85      0.74      0.79      2346

    accuracy                           0.83      5342
   macro avg       0.83      0.82      0.82      5342
weighted avg       0.83      0.83      0.83      5342
```