Biswarup Saha
23CS60R45

Report on COVID-19 Fake News Detection using Deep Learning Models

All trained models link: DL_W2_allpretrainedmodels
All Datafile created: DLW2_AllData

Executive Summary

This report outlines the process and findings from a project aimed at detecting COVID-19-related fake news on social media platforms using deep learning models. The project utilized a dataset from the shared task of Constraint@AAAI-2021, comprising 10,600 samples of social media posts labeled as either 'real' or 'fake'. Three types of deep learning models were explored: Deep Neural Network (DNN), 1D Convolutional Neural Network (1D-CNN), and AutoModelForSequenceClassification based on BERT architectures. The report covers dataset preparation, text preprocessing, model training with hyperparameter tuning, and the evaluation of model performance.

Introduction

The proliferation of fake news, especially related to COVID-19, poses significant risks to public health and safety. Automatic detection of such misinformation can help mitigate these risks. This project aimed to develop and evaluate machine learning models capable of classifying social media posts as real or fake news.

Task-1: Dataset Preparation

The dataset was split into training, validation, and test sets using the scikit-learn train_test_split function, adhering to an 80/10/10 distribution. This division ensured a balanced representation of classes in each dataset segment, essential for training and evaluating the models effectively.

Task-2: Preprocessing Social Media Post

Given the nature of social media texts, which include emojis, URLs, and hashtags, careful preprocessing was vital. The preprocessing strategy was designed to retain the semantic essence of the posts while removing noise. Techniques such as emoji translation to text, URL removal, and hashtag processing were applied, maintaining the context and sentiment of the original posts.

Task-3: Obtaining Representations using Bert-based Model

Five BERT-based models were considered for generating text representations:

- bert-base-uncased
- bert-base-cased
- covid-twitter-bert
- twibert-base
- socbert

Each model offers unique perspectives on text representation, leveraging the pre-training on diverse corpora, including COVID-19-specific datasets. The representations obtained were foundational for the subsequent classification tasks.

Task-4: Training Classifiers with Hyperparameter Tuning

Two deep learning models were trained:

> Deep Neural Network (DNN): A custom architecture was designed, focusing on depth and activation functions to capture complex patterns in the data.
> 1D-CNN: Leveraging convolutional layers to interpret the sequential nature of text data, optimizing for local patterns indicative of fake or real news.

Hyperparameter tuning was conducted using a combination of grid search and manual adjustments, optimizing for accuracy while preventing overfitting.

Results:

```
Epoch 1/10, Loss: 0.3128855820815518
Epoch 2/10, Loss: 0.23167701917436886
Epoch 3/10, Loss: 0.2072315534330764
Epoch 4/10, Loss: 0.1955767827073358
Epoch 5/10, Loss: 0.1796591930653689
Epoch 6/10, Loss: 0.1736230719693989
Epoch 7/10, Loss: 0.156926356403614
Epoch 8/10, Loss: 0.14501229123967999
Epoch 9/10, Loss: 0.14341624464909986
Epoch 10/10, Loss: 0.136681812028137
Model saved to 'DNN_bert-base-uncased.pth'.
              precision    recall  f1-score   support

           0       0.96      0.88      0.92       508
           1       0.90      0.97      0.93       552

    accuracy                           0.93      1060
   macro avg       0.93      0.92      0.92      1060
weighted avg       0.93      0.93      0.93      1060
```

```
Epoch 1/10, Loss: 0.3864815536535011
Epoch 2/10, Loss: 0.274706025534081
Epoch 3/10, Loss: 0.2425257718506849
Epoch 4/10, Loss: 0.22329122360866024
Epoch 5/10, Loss: 0.20701516610833834
Epoch 6/10, Loss: 0.1944708049156756
Epoch 7/10, Loss: 0.17849599637512892
Epoch 8/10, Loss: 0.1710144705929846
Epoch 9/10, Loss: 0.1636687963568377
Epoch 10/10, Loss: 0.14593784656586511
Model saved to '1D-CNN_bert-base-uncased.pth'.
              precision    recall  f1-score   support

           0       0.95      0.88      0.91       508
           1       0.89      0.96      0.93       552

    accuracy                           0.92      1060
   macro avg       0.92      0.92      0.92      1060
weighted avg       0.92      0.92      0.92      1060
```

```
Epoch 1/10, Loss: 0.3489933947347245
Epoch 2/10, Loss: 0.2663985440472387
Epoch 3/10, Loss: 0.24664515635033823
Epoch 4/10, Loss: 0.22393263772411168
Epoch 5/10, Loss: 0.21344410815047768
Epoch 6/10, Loss: 0.19962451361682054
Epoch 7/10, Loss: 0.18910393901872186
Epoch 8/10, Loss: 0.18207767536898828
Epoch 9/10, Loss: 0.17008942355805973
Epoch 10/10, Loss: 0.1601618499196363
Model saved to 'DNN_bert-base-cased.pth'.
              precision    recall  f1-score   support

           0       0.95      0.84      0.89       508
           1       0.87      0.96      0.91       552

    accuracy                           0.90      1060
   macro avg       0.91      0.90      0.90      1060
weighted avg       0.91      0.90      0.90      1060
```

```
Epoch 1/10, Loss: 0.42465432460577984
Epoch 2/10, Loss: 0.3166292613688505
Epoch 3/10, Loss: 0.28510955386566667
Epoch 4/10, Loss: 0.26840411866048597
Epoch 5/10, Loss: 0.24681371950878286
Epoch 6/10, Loss: 0.23505422903119394
Epoch 7/10, Loss: 0.2225588561228986
Epoch 8/10, Loss: 0.2157163921532766
Epoch 9/10, Loss: 0.19699655743140096
Epoch 10/10, Loss: 0.18674616952830891
Model saved to '1D-CNN_bert-base-cased.pth'.
              precision    recall  f1-score   support

           0       0.91      0.88      0.90       508
           1       0.90      0.92      0.91       552

    accuracy                           0.90      1060
   macro avg       0.91      0.90      0.90      1060
weighted avg       0.91      0.90      0.90      1060
```

```
weighted avg       0.91      0.90      0.90      1060

Epoch 1/10, Loss: 0.2881992588854I054
Epoch 2/10, Loss: 0.13862429964928976
Epoch 3/10, Loss: 0.10741372968966387
Epoch 4/10, Loss: 0.08794087725815 2923
Epoch 5/10, Loss: 0.06954278785448334
Epoch 6/10, Loss: 0.06378822749767691
Epoch 7/10, Loss: 0.051834678513791814
Epoch 8/10, Loss: 0.05836743936636389
Epoch 9/10, Loss: 0.03284373152488811
Epoch 10/10, Loss: 0.03981984961228423
Model saved to 'DNN_covid-twitter-bert.pth'.
              precision    recall  f1-score   support

           0       0.96      0.95      0.96       508
           1       0.95      0.97      0.96       552

    accuracy                           0.96      1060
   macro avg       0.96      0.96      0.96      1060
weighted avg       0.96      0.96      0.96      1060

Epoch 1/10, Loss: 0.34133222226826665
Epoch 2/10, Loss: 0.21011376964314 93
Epoch 3/10, Loss: 0.17557448055854898
Epoch 4/10, Loss: 0.15442101464880115
Epoch 5/10, Loss: 0.14219678330393332
Epoch 6/10, Loss: 0.12879548779565772
Epoch 7/10, Loss: 0.11583783235494641
Epoch 8/10, Loss: 0.11046861423495806
Epoch 9/10, Loss: 0.10172192791810971
Epoch 10/10, Loss: 0.08960112163375 29
Model saved to '1D-CNN_covid-twitter-bert.pth'.
              precision    recall  f1-score   support

           0       0.96      0.90      0.93       508
           1       0.92      0.97      0.94       552

    accuracy                           0.94      1060
   macro avg       0.94      0.94      0.94      1060
weighted avg       0.94      0.94      0.94      1060

Epoch 1/10, Loss: 0.33133785156788845
Epoch 2/10, Loss: 0.25175147659 65489
Epoch 3/10, Loss: 0.22821553575438275
Epoch 4/10, Loss: 0.21827198221368142
Epoch 5/10, Loss: 0.19652472155274087
Epoch 6/10, Loss: 0.19377865984 48528
Epoch 7/10, Loss: 0.19084499976224495
Epoch 8/10, Loss: 0.17418805334 31881
Epoch 9/10, Loss: 0.17566305971 8894
Epoch 10/10, Loss: 0.17824381823298887
Model saved to 'DNN_twhin-bert-base.pth'.
              precision    recall  f1-score   support

           0       0.94      0.92      0.93       508
           1       0.92      0.95      0.93       552

    accuracy                           0.93      1060
   macro avg       0.93      0.93      0.93      1060
weighted avg       0.93      0.93      0.93      1060
```

```
Epoch 1/10, Loss: 0.3810957713509506
Epoch 2/10, Loss: 0.28298863748896796
Epoch 3/10, Loss: 0.254808642701158
Epoch 4/10, Loss: 0.2327345314172079
Epoch 5/10, Loss: 0.21952158007419334
Epoch 6/10, Loss: 0.2006913848221302
Epoch 7/10, Loss: 0.18759790064310128
Epoch 8/10, Loss: 0.18228084915932619
Epoch 9/10, Loss: 0.16982189416885377
Epoch 10/10, Loss: 0.1587109910499937
Model saved to '1D-CNN_twhin-bert-base.pth'.
              precision    recall  f1-score   support

           0       0.93      0.90      0.91       508
           1       0.91      0.94      0.92       552

    accuracy                           0.92      1060
   macro avg       0.92      0.92      0.92      1060
weighted avg       0.92      0.92      0.92      1060

Epoch 1/10, Loss: 0.3493044173942422
Epoch 2/10, Loss: 0.25907997249994635
Epoch 3/10, Loss: 0.2295660007253008
Epoch 4/10, Loss: 0.2095177099671004
Epoch 5/10, Loss: 0.19343723351100706
Epoch 6/10, Loss: 0.17802356969354288
Epoch 7/10, Loss: 0.15807144788076294
Epoch 8/10, Loss: 0.14961904250085353
Epoch 9/10, Loss: 0.14016745954330237
Epoch 10/10, Loss: 0.12807641566010577
Model saved to 'DNN_socbert.pth'.
              precision    recall  f1-score   support

           0       0.93      0.89      0.91       508
           1       0.91      0.94      0.92       552

    accuracy                           0.92      1060
   macro avg       0.92      0.92      0.92      1060
weighted avg       0.92      0.92      0.92      1060
```

```
Epoch 1/10, Loss: 0.40834840184675075
Epoch 2/10, Loss: 0.29490787763640564
Epoch 3/10, Loss: 0.2592725672811832
Epoch 4/10, Loss: 0.23851568233573212
Epoch 5/10, Loss: 0.21445425852149164
Epoch 6/10, Loss: 0.1992255045557922
Epoch 7/10, Loss: 0.18033977406064294
Epoch 8/10, Loss: 0.1742834393427057
Epoch 9/10, Loss: 0.1564453111258599
Epoch 10/10, Loss: 0.14511221629339008
Model saved to '1D-CNN_socbert.pth'.
              precision    recall  f1-score   support

           0       0.93      0.88      0.90       508
           1       0.89      0.94      0.92       552

    accuracy                           0.91      1060
   macro avg       0.91      0.91      0.91      1060
weighted avg       0.91      0.91      0.91      1060
```

Task-5: Evaluating Machine Learning Models

The models were evaluated based on their performance on the test set, using metrics such as confusion matrix, accuracy, F1-score, precision, and recall (both micro and macro). Custom routines were developed for this evaluation to deepen understanding, alongside the use of scikit-learn's classification_report.

Findings

- The 1D-CNN model demonstrated remarkable precision in detecting fake news, benefiting from its ability to capture local textual features.
- The DNN model showed a balanced performance across both classes, with a slightly higher recall in identifying real news.
- Among the BERT-based models, covid-twitter-bert and socbert provided superior representations for this specific task, likely due to their training on socially relevant datasets.

Conclusion

The project successfully demonstrated the feasibility of using deep learning models to detect COVID-19-related fake news on social media. While each model had its strengths, combining their predictions through ensemble methods could potentially improve overall accuracy and robustness.

Future Work

Further exploration into model ensembles, more extensive hyperparameter tuning, and incorporating additional data sources could enhance the models' predictive capabilities. Additionally, real-time application development for live social media monitoring would be a practical next step.