Comparison between LSTM and GRU on English - France dataset:

|  |  |  |
| --- | --- | --- |
|  | LSTM | GRU |
| Loss | 0.4851 | 0.3802 |
| Accuracy | 0.8562 | 0.8852 |
| F1 score | 0.8774 | 0.9006 |
| Training time | 8m 22.5s | 6m 28.6s |
| Validation loss | 0.5743 | 0.4808 |
| Avalidation accuracy | 0.8311 | 0.8576 |
| Validation F1 score | 0.8551 | 0.8773 |

GRU outperformed LSTM in term of efficency. Not only it has higher accuracy, but F1 score also suggests better overall precision and recall balance in a shorter time.

Comparison of GRU on English – French and English – Vietnamese dataset:

|  |  |  |
| --- | --- | --- |
|  | eng-fra | eng-vie |
| Loss | 0.3802 | 0.3584 |
| Accuracy | 0.8852 | 0.9021 |
| F1 score | 0.9006 | 0.9232 |
| Training time | 6m 28.6s | 18m 52.1s |
| Validation loss | 0.4808 | 0.6553 |
| Avalidation accuracy | 0.8576 | 0.8239 |
| Validation F1 score | 0.8773 | 0.8501 |
| Number of tokens | 98 | 163 |

This model works better on English – Vietnamese dataset, as it has higher accuracy and F1 score. But please note that the time taken to train is also longer due to wider token range.

\*This report based on last epochs, if you want to see full training logs and graphs for further information, read notebooks