

Progress Report

Daniel Biro

December 12, 2020

Contents

1	Experiments - workflow version: 1.0.0; experiments revision: YYY	2
1.1	Methods	2
1.2	Results	2

1 Experiments - workflow version: 1.0.0; experiments revision: YYY

1.1 Methods

To facilitate combining the results of several models for both recoring performance and graphing, the workflow has been updated with a new script (run_experiment.py) and all previous experiments have been re-run using it, increasing the number of training epochs to 1000 and the batch size to 250.

The previous experiments performed were trying out some basic architecture types (GRU, LSTM, RNN, Perceptron), along with a couple of variations on GRU, including different combinations of adding some dropout and making the network bidirectional. These experiments were all performed with 1 hidden layer, 32 units in the hidden layer and a learning rate of 0.001 (results of previous tuning experiments).

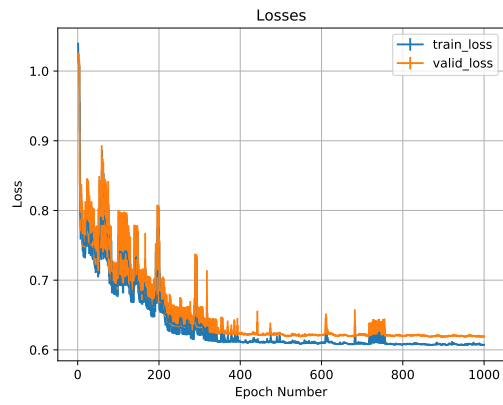
A new experiment performed is using a Transformer network where the recurrent layer is replaced by a transformer encoder layer with layer size (corresponding to features) set to 4 and number of attention heads also set to 4. 10% dropout was applied in this layer and in an additional dropout layer afterwards. The result is flattened into a batch_size x 4 * 81 tensor (where 4 is the number of channels and 81 the sequence length) and then a linear layer and softmax is applied as with the other models.

1.2 Results

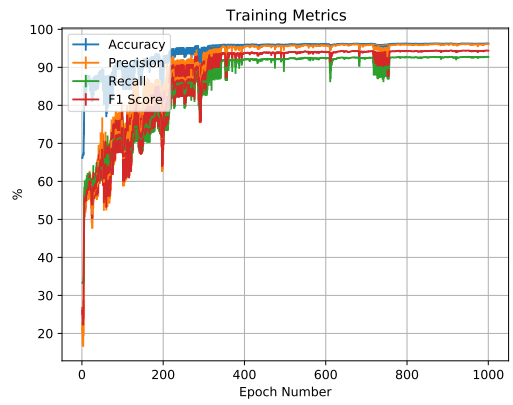
The Perceptron network still performs the best. From the recurrent networks GRU did the best with the results improving the most when the network is made bidirectional and 10% dropout is applied. The Transformer network managed to perform a bit better than the best GRU variant, however model training is significantly longer (approx 10hrs vs approx 30mins). Due to the longer training time, in this report the Transformer results shown are over 3 runs, while the other model results shown are over 10 runs. This and possibly the per epoch model saving implemented during training led to something going wrong for the early epochs, so for this model the graphs become faulty and are not included. Training is under way and the results in the repository will be updated. The rest of the results can be found in the repository.

MODEL	ACCURACY	PRECISION	RECALL	F1 SCORE
GRU	95.47% \pm 0.06	95.14% \pm 0.10	92.33% \pm 0.09	93.72% \pm 0.09
BI-GRU + 10% DROPOUT	95.64% \pm 0.07	95.27% \pm 0.11	92.72% \pm 0.13	93.98% \pm 0.11
TRANSFORMER + 10% DROPOUT	95.65% \pm 0.06	95.83% \pm 0.07	92.85% \pm 0.06	94.32% \pm 00.4
PERCEPTRON	96.24% \pm 0	96.3% \pm 0	93.68% \pm 0	94.97% \pm 0

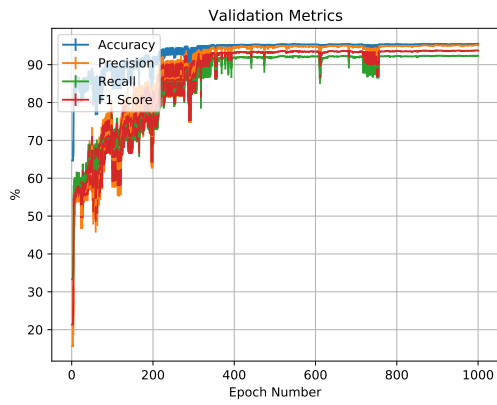
Table 1: Performance of best models on the validation set



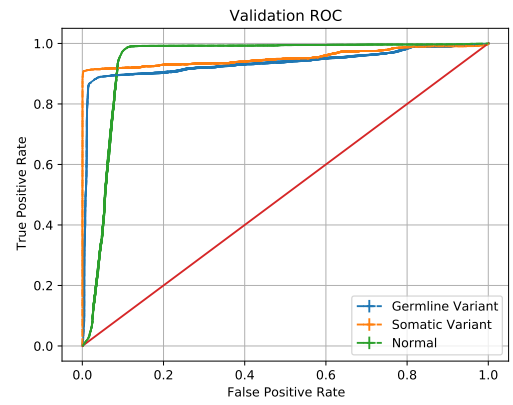
(a)



(b)

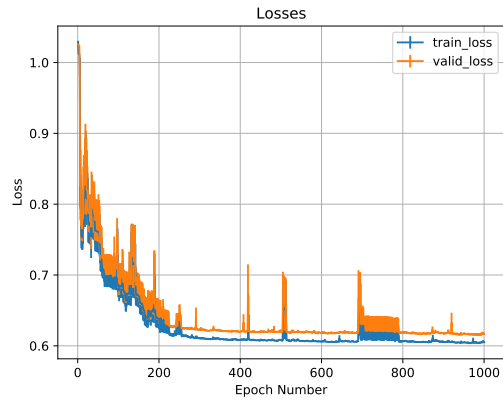


(c)

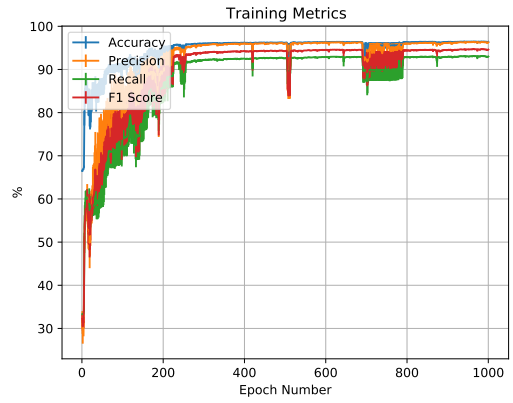


(d)

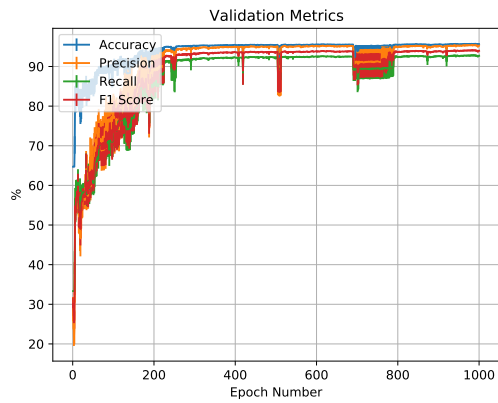
Figure 1: Figures for GRU model



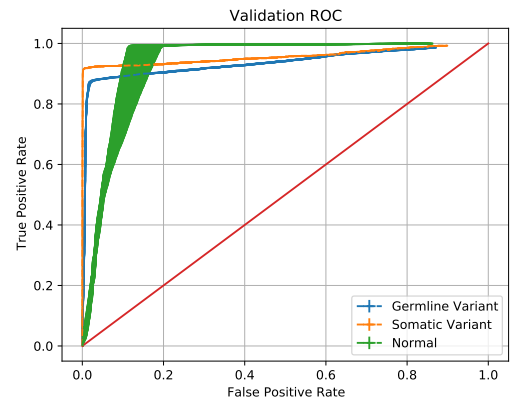
(a)



(b)

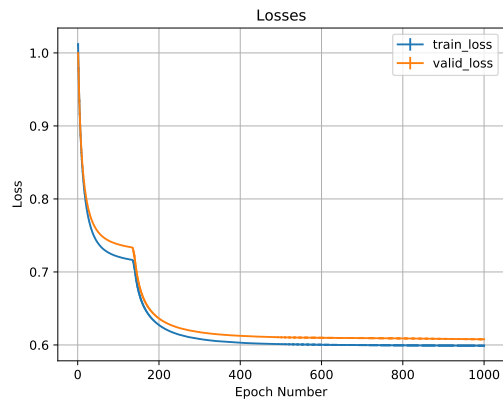


(c)

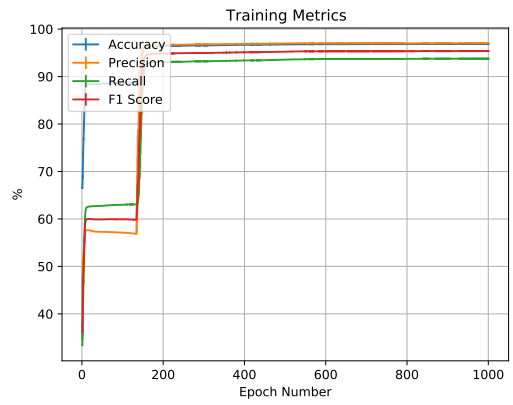


(d)

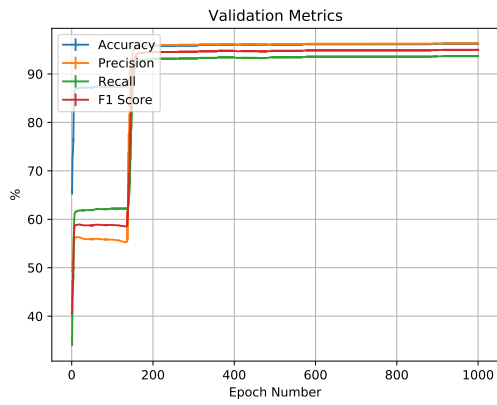
Figure 2: Figures for Bidirectional GRU with Dropout model



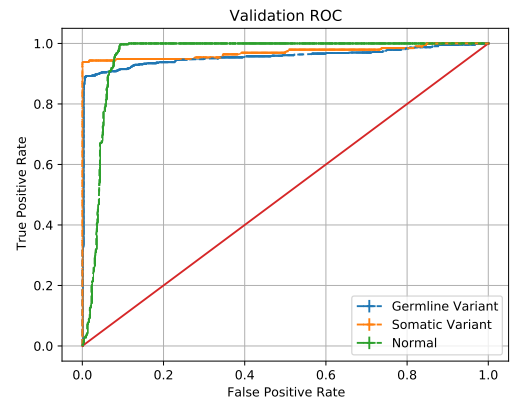
(a)



(b)



(c)



(d)

Figure 3: Figures for Perceptron model