

Assignment 3

Livia Meinhardt

Deep Learning - FGV EMAp

1 Código

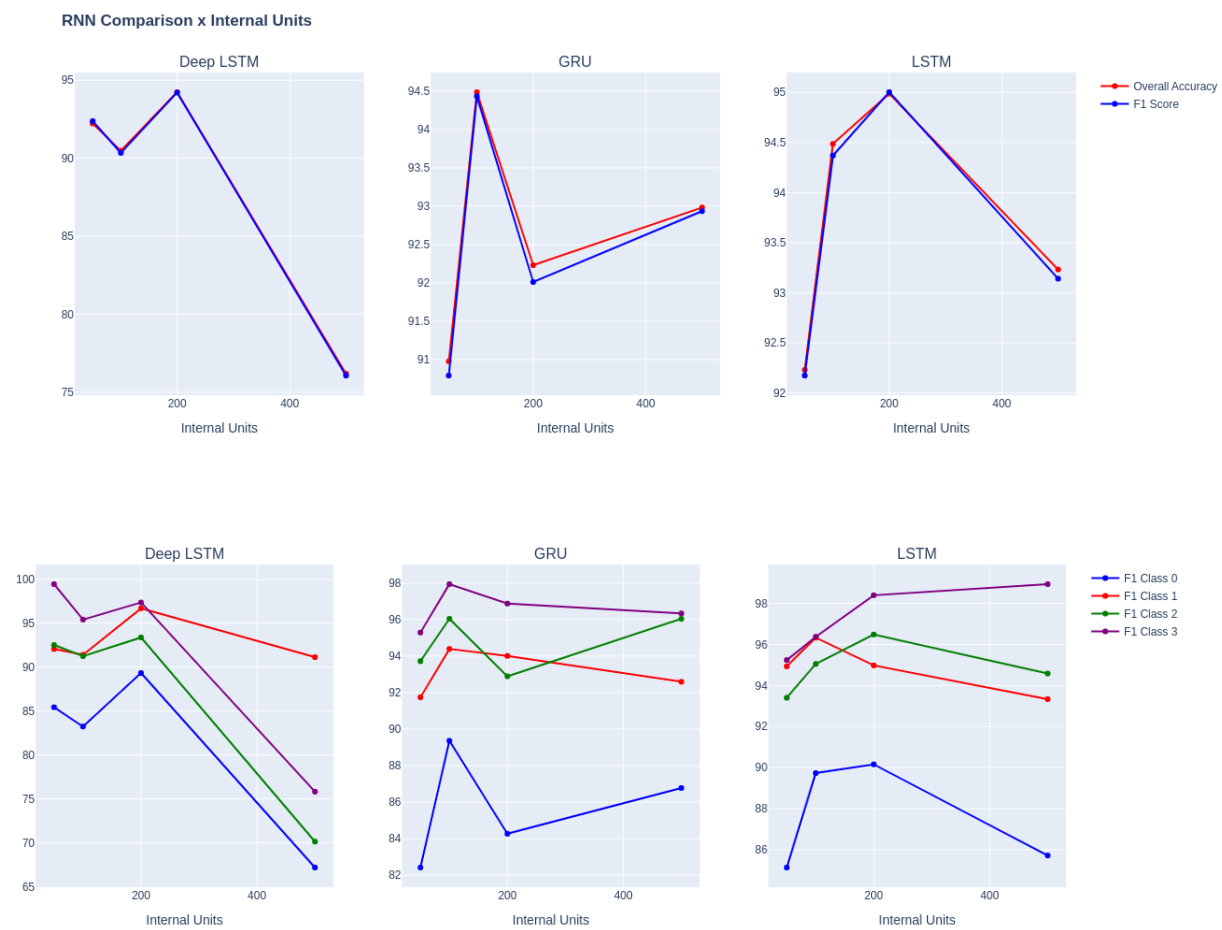
O notebook está disponível no link: <https://colab.research.google.com/gist/liviameinhardt/99182bcd2506173373231310bbebfba591/assign-3-action-recognition-from-ucf101.ipynb>

2 Resultados

A tabela com os resultados, arredondados para duas casas decimais, está apresentada abaixo.

RNN network		Internal Units					
LSTM	50	92.230576	85.128205	94.930876	93.401015	95.238095	92.174548
	100	94.486216	89.72973	96.330275	95.049505	96.373057	94.370642
	200	94.987469	90.15544	94.977169	96.482412	98.395722	95.002686
	500	93.233083	85.714286	93.333333	94.581281	98.93617	93.141268
GRU	50	90.977444	82.417582	91.743119	93.719807	95.287958	90.792117
	100	94.486216	89.361702	94.392523	96.039604	97.938144	94.432993
	200	92.230576	84.269663	94.009217	92.890995	96.875	92.011219
	500	92.982456	86.772487	92.592593	96.039604	96.335079	92.93494
Deep LSTM	50	92.230576	85.436893	92.079208	92.537313	99.470899	92.381079
	100	90.47619	83.243243	91.428571	91.262136	95.431472	90.341356
	200	94.235589	89.340102	96.713615	93.401015	97.382199	94.209233
	500	76.190476	67.169811	91.150442	70.12987	75.816993	76.066779

Para comparar as informações de maneira mais práticas, podemos visualizar os dados graficamente:



2.1 Deep LSTM

Pelos resultados, podemos perceber que a adição da camada feita por esse modelo começa a prejudicar quando possuímos mais internal units. Nesse caso, o modelo se torna menos generalizável, possuindo métricas bem menores que os outros testes para todas as classes (possivelmente terá um overfit com mais unidades internas).

As demais redes também diminuem a acurácia e f1 score médio quando temos 500 unidades internas, porém para esta rede o decrescimento é bem maior (chegando a 76%). Isso pode ser devido ao fato de que esse modelo é o mais flexível (possui mais parâmetros), assim tende a ter um overfit mais rápido do

que os demais, pois se ajusta tão bem ao conjunto de treino que não possui uma performance tão boa nos testes.

No geral a adição da camada não auxilia no modelo, já que comparando com as métricas do LSTM o aumento não é significativo ou são até menores.

2.2 LSTM

Assim como o Deep LSTM, ao utilizar 500 unidades internas o modelo perde muita performance, provavelmente, também, causado pelo aumento da flexibilidade (bias and variance tradeoff). Porém, nesse caso quando usamos somente 50 unidades internas o modelo não tem tanta capacidade de previsão quanto o de 100 ou 200, por exemplo - mas ainda é um bom modelo com 92% de acurácia.

2.3 GRU

Utilizando o GRU o melhor modelo é o que utiliza 100 unidades internas, com 94.5% de acurácia, enquanto para os demais o melhor modelo é o com 200 unidades internas. Podemos perceber que nesse caso esse decréscimo é causado principalmente pelas classes 0 (Basketball) e 2 (GolfSwing), enquanto as demais (Diving e Skiing) decrescem mais lentamente conforme aumentamos as unidades internas.

Intuitivamente, é mais fácil classificar atividades de mergulho e ski nesse universo, uma vez que ambas possuem características mais marcantes que basquete e golf: mergulho sempre será realizado em piscinas, mares, lagos e ski sempre na neve, enquanto basquete e golf podem ser realizados em uma diversidade maior de ambientes (quadras de madeira, concreto, gramado, campo artificial de gold, etc). Nesse mesmo sentido, a classe ski é sempre a que possui um maior F1 score e basquete a menor.