

EP2: RNNs Bidirecionais, Overfitting, Underfitting

Wesley Seidel Carvalho - N. USP 6544342

Belém do Pará, 30 de Outubro de 2020

1 Introdução

Neste documento apresentamos os resultados dos experimentos de análise de comportamento do sobre ajuste de diferentes configurações de redes e hiperparâmetros aplicados na tarefa de classificação de 5 classes (review) obtidas na base de dados b2W.

2 Configurações dos experimentos

Os valores utilizados foram:

- Tamanho do vocabulário: 20000;
- Tamanho máximo da sentença: 40;
- Épocas máxima de treinamento: 100.

Realizamos os experimentos considerando em momentos diferentes as seguintes camadas de embeddings:

- Camada Embedding "vazia" do TensorFlow;
- Camada Embedding adaptada do Word2Vec do NILC.

Para camada de LSTM, utilizamos:

- Camada LSTM com 32 unidades e ativação tangente;
- Camada LSTM Bidirecional com 32 unidades e ativação tangente;.

Em todos os experimentos separamos os dados da seguinte forma:

- Treino: 75
- Teste: 15
- Validação: 10

Na etapa de pré-processamento do texto, ainda antes de usar a camada de vetorização textual, foi realizado apenas a tokenização e a transformação dos caracteres para minúsculo.

Na tabela 1 está o resumo das configurações utilizadas nos experimentos. Em todos os casos, utilizamos sempre o lote (batchsize) de tamanho 32.

Experimento	Tipo Embedding	Tipo LSTM	Dropout
LSTM-0.0	Zerado	LSTM 32 units	0.0
LSTM-0.25	Zerado	LSTM 32 units	0.25
LSTM-0.5	Zerado	LSTM 32 units	0.5
Bidirec-0.0	Zerado	Bidirecional LSTM 32 units	0.0
Bidirec-0.25	Zerado	Bidirecional LSTM 32 units	0.25
Bidirec-0.50	Zerado	Bidirecional LSTM 32 units	0.5
NILC-LSTM-0.0	Word2Vec Nilc	LSTM 32 units	0.0
NILC-LSTM-0.25	Word2Vec Nilc	LSTM 32 units	0.25
NILC-LSTM-0.5	Word2Vec Nilc	LSTM 32 units	0.5
NILC-Bidirec-0.0	Word2Vec Nilc	Bidirecional LSTM 32 units	0.0
NILC-Bidirec-0.25	Word2Vec Nilc	Bidirecional LSTM 32 units	0.25
NILC-Bidirec-0.50	Word2Vec Nilc	Bidirecional LSTM 32 units	0.5

Tabela 1: Experimentos Realizados

2.1 Resultados utilizando o Embedding Vazio (ou Zerado)

A seguir é possível avaliar cada resultado graficamente. Repare que abaixo de cada gráfico existe um pequeno resumo da configuração da rede para o gráfico em questão.

2.1.1 LSTM-0.0

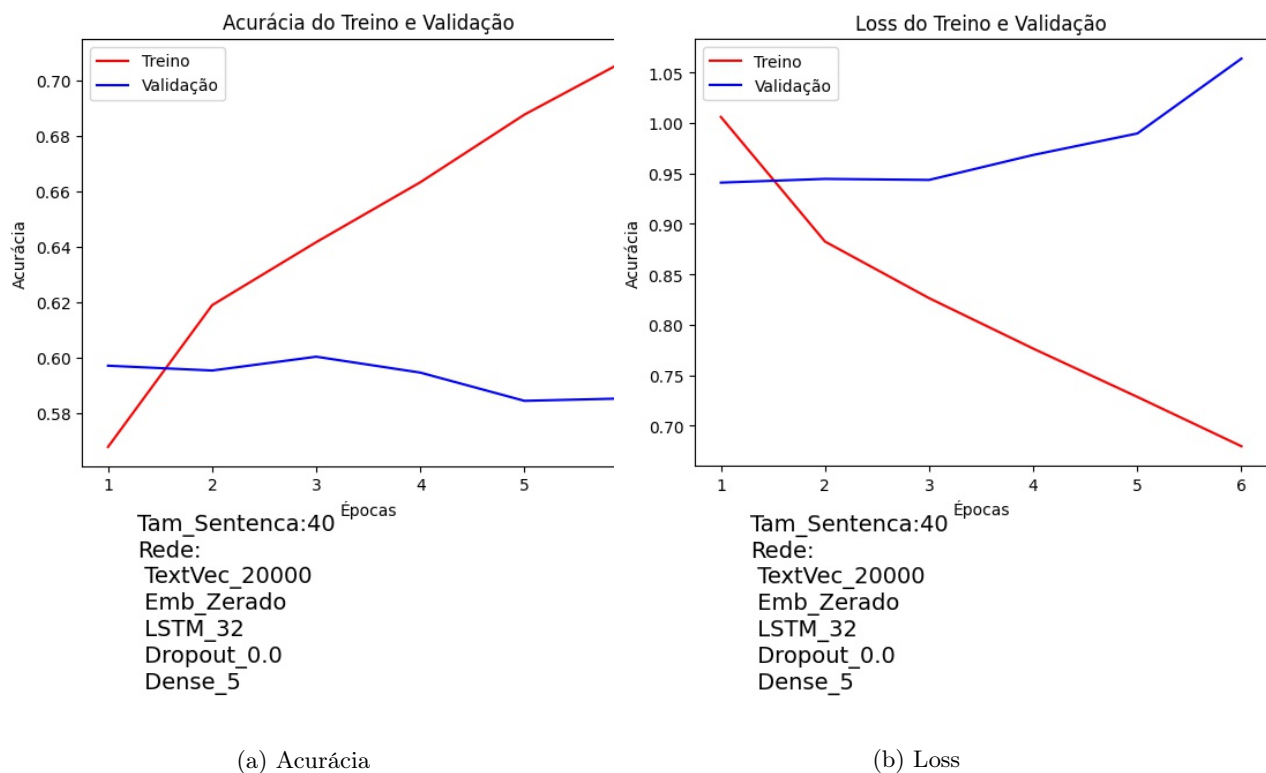
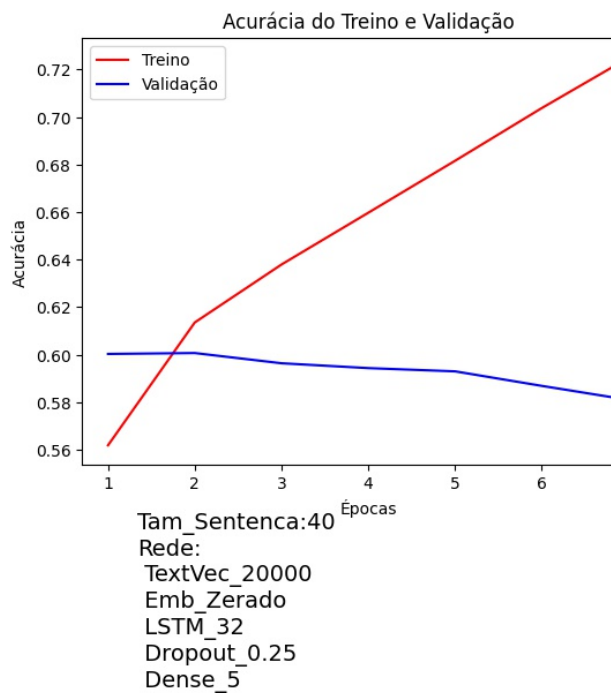
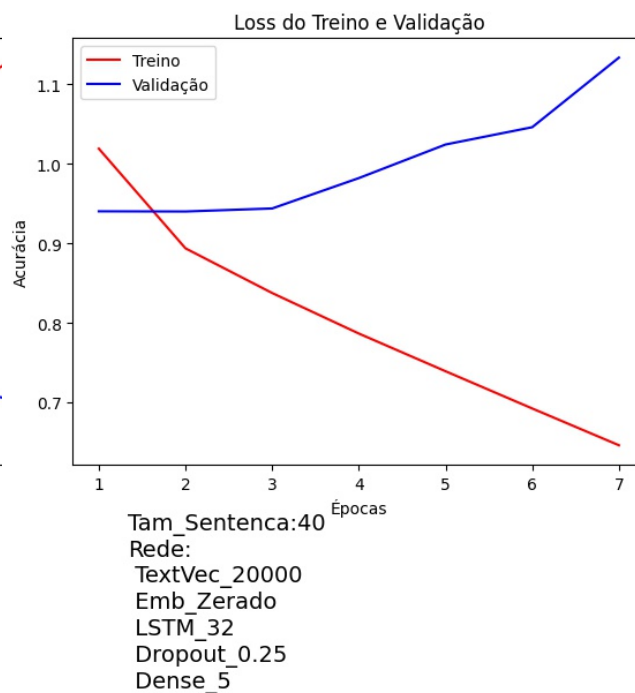


Figura 1: Resultados da rede LSTM-0.0

2.1.2 LSTM-0.25



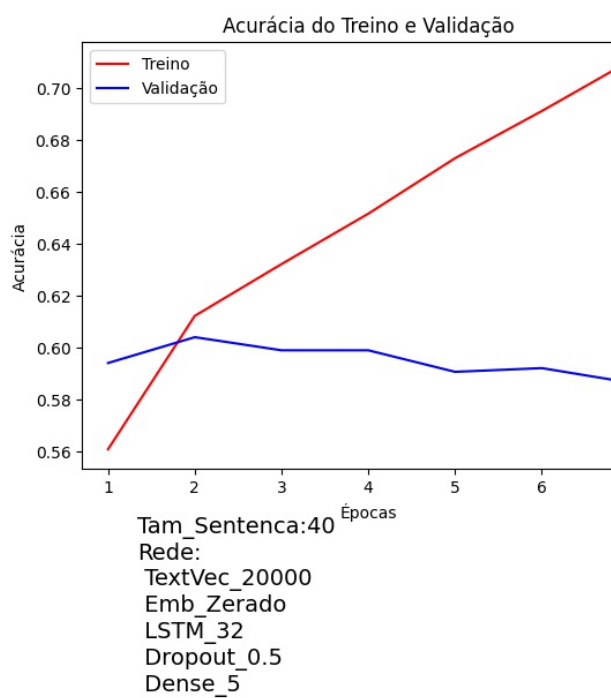
(a) Acurácia



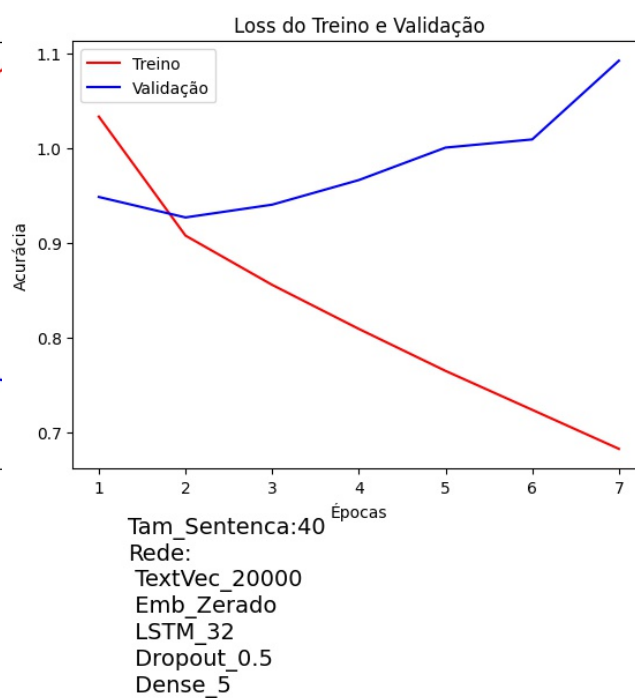
(b) Loss

Figura 2: Resultados da rede LSTM-0.25

2.1.3 LSTM-0.50



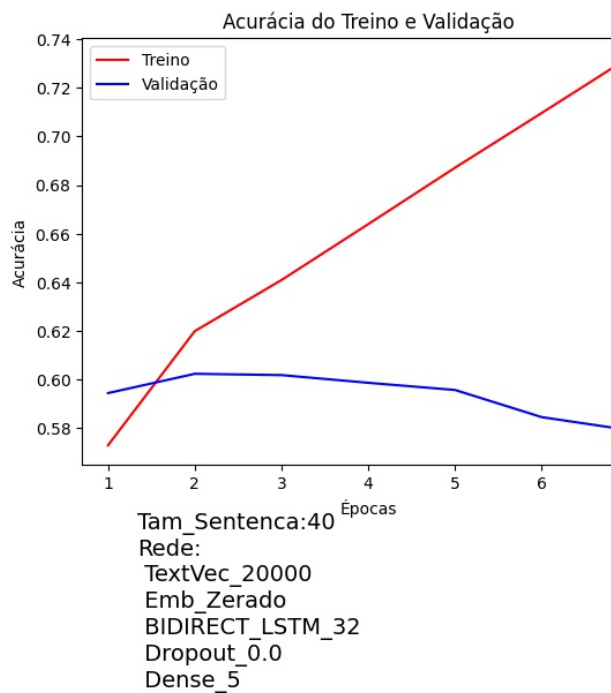
(a) Acurácia



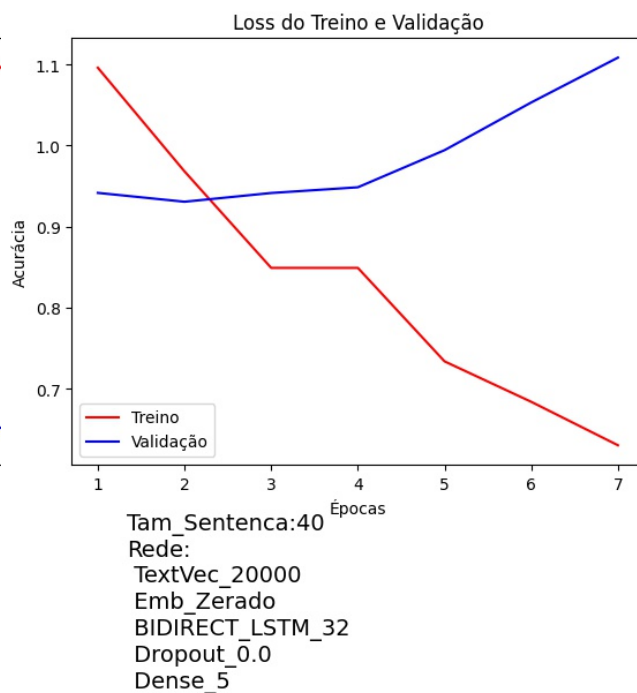
(b) Loss

Figura 3: Resultados da rede LSTM-0.50

2.1.4 Bidirec-0.0



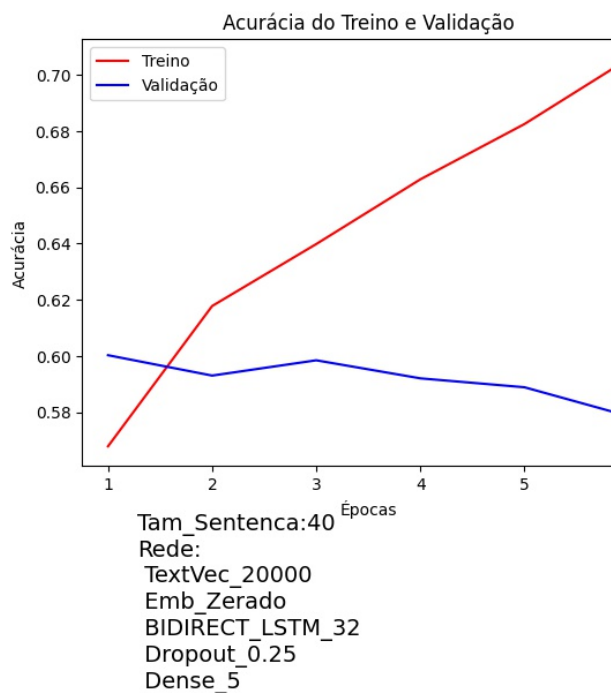
(a) Acurácia



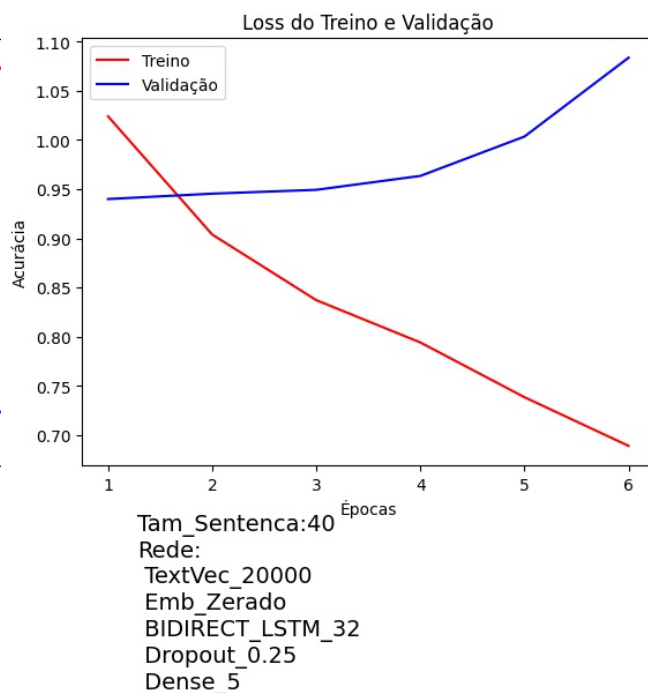
(b) Loss

Figura 4: Resultados da rede Bidirec-0.0

2.1.5 Bidirec-0.25



(a) Acurácia



(b) Loss

Figura 5: Resultados da rede Bidirec-0.25

2.1.6 Bidirec-0.50

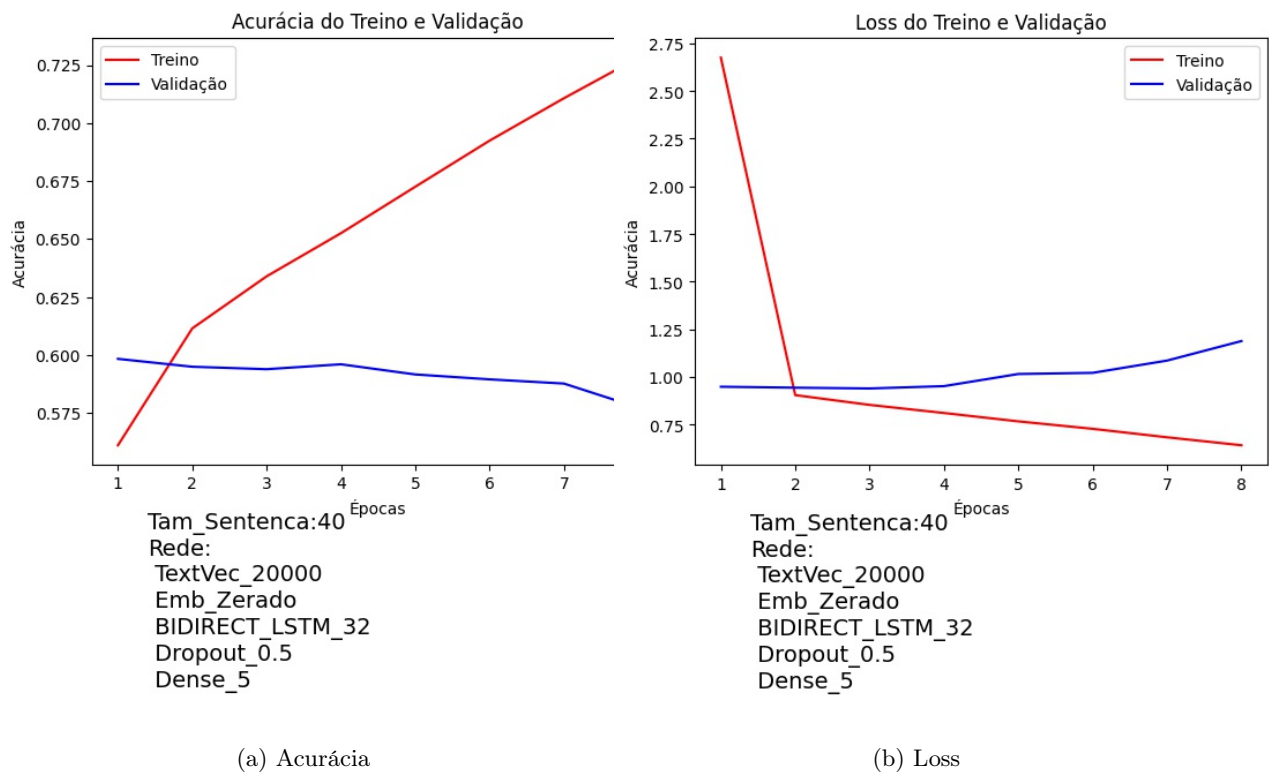


Figura 6: Resultados da rede Bidirec-0.50

2.2 Resultados utilizando o Embedding word2vec do NILC

2.2.1 NILC-LSTM-0.0

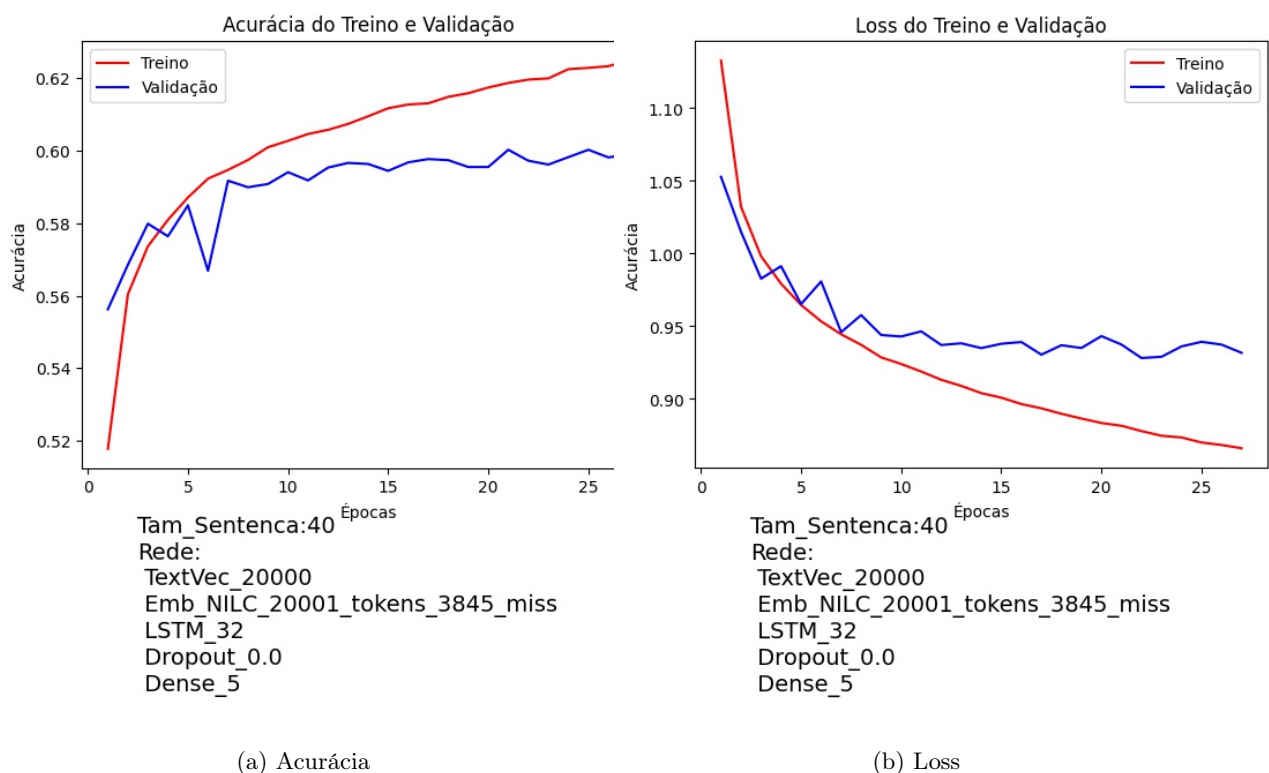


Figura 7: Resultados da rede NILC-LSTM-0.0

2.2.2 NILC-LSTM-0.25

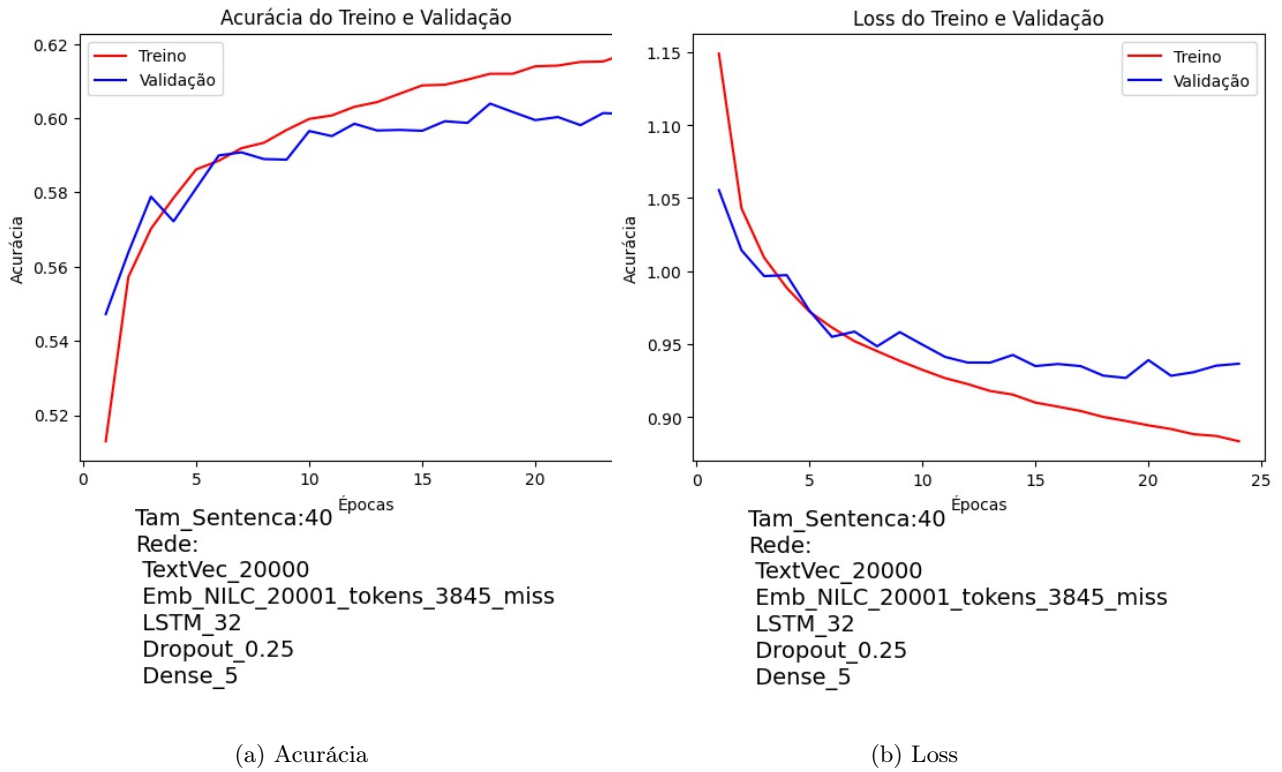


Figura 8: Resultados da rede NILC-LSTM-0.25

2.2.3 NILC-LSTM-0.50

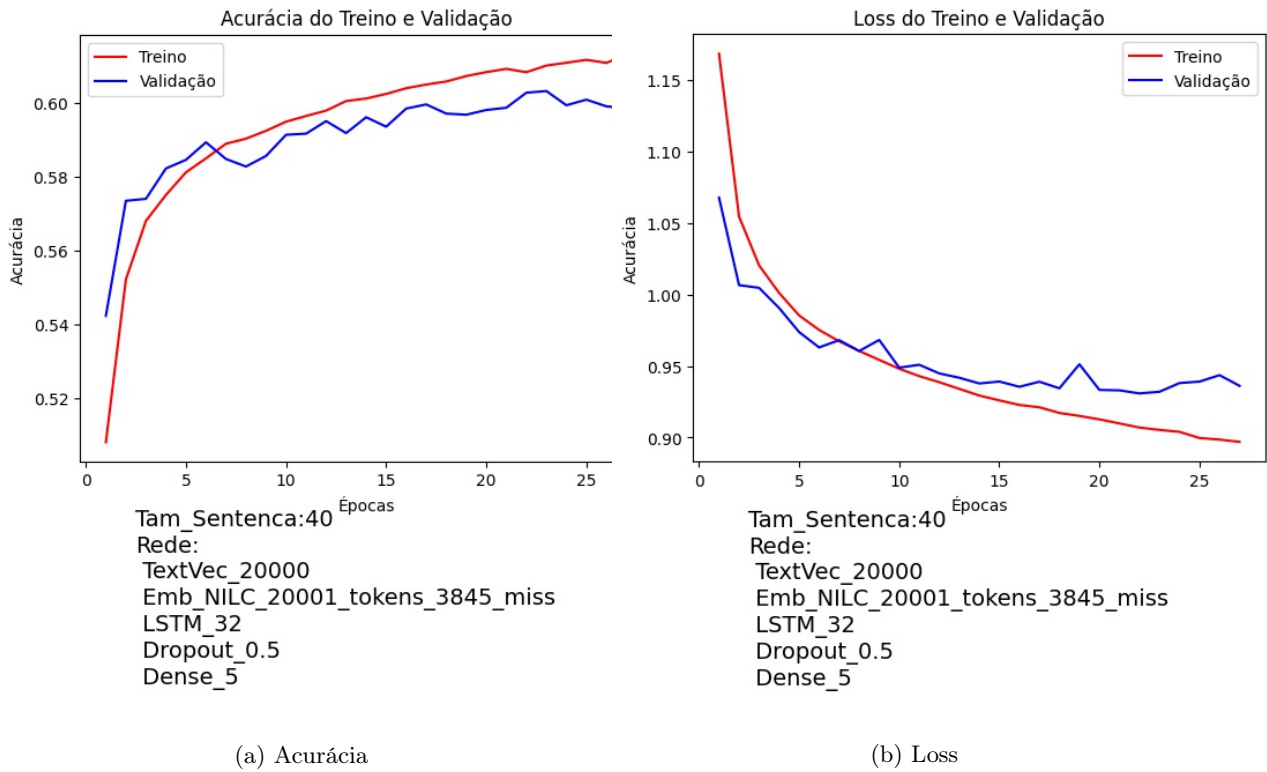


Figura 9: Resultados da rede NILC-LSTM-0.50

2.2.4 NILC-Bidirec-0.0

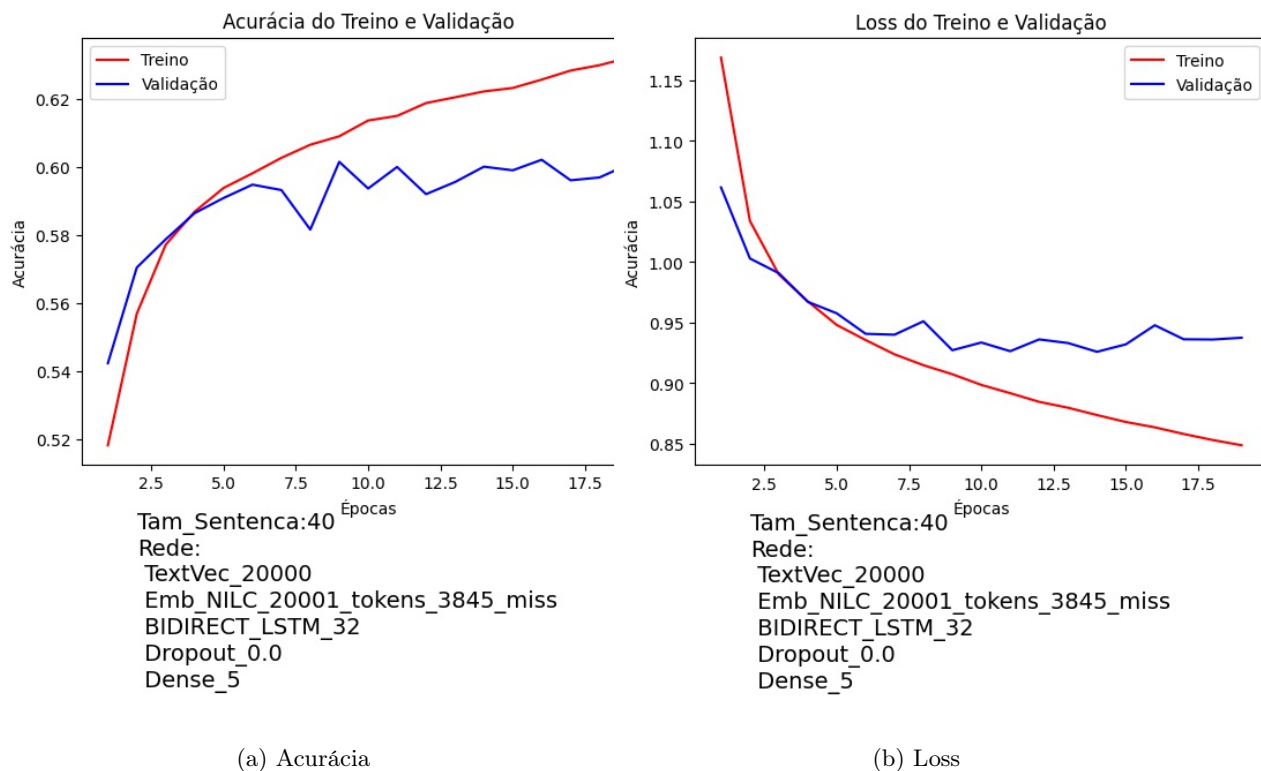


Figura 10: Resultados da rede Bidirec-0.0

2.2.5 NILC-Bidirec-0.25

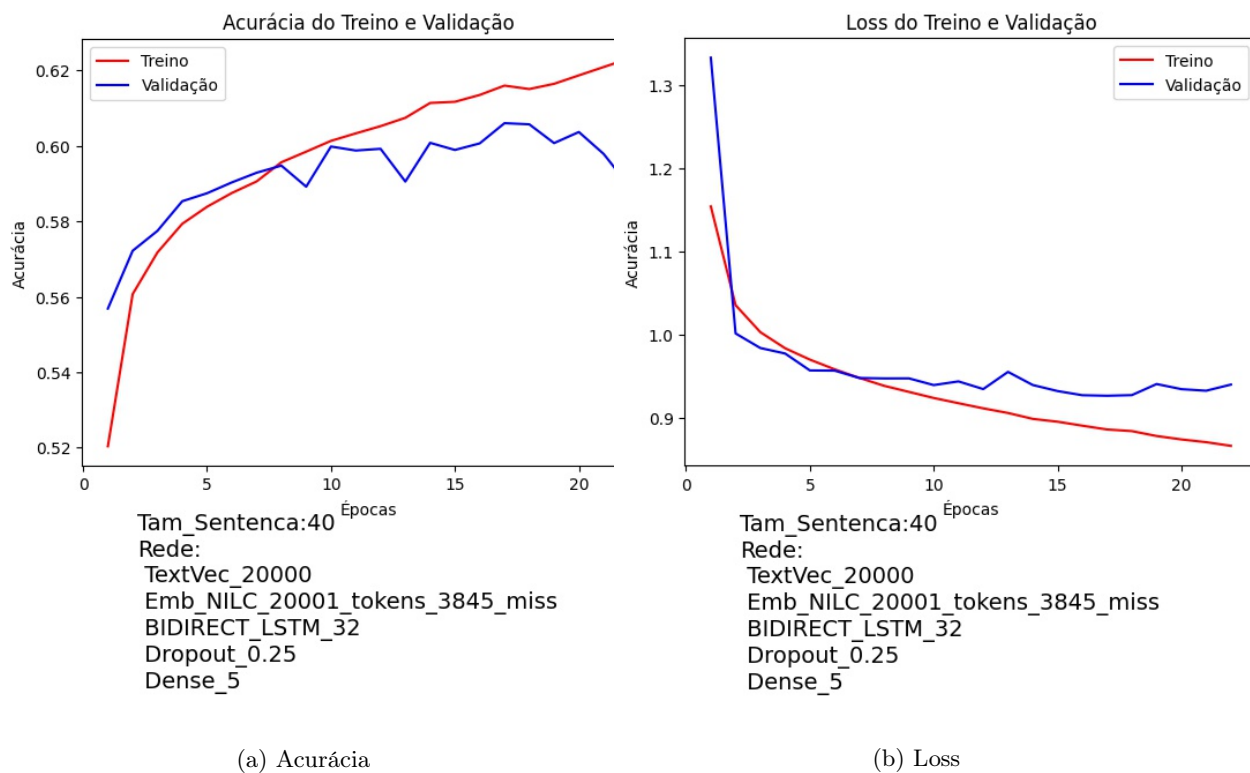


Figura 11: Resultados da rede Bidirec-0.25

2.2.6 NILC-Bidirec-0.50

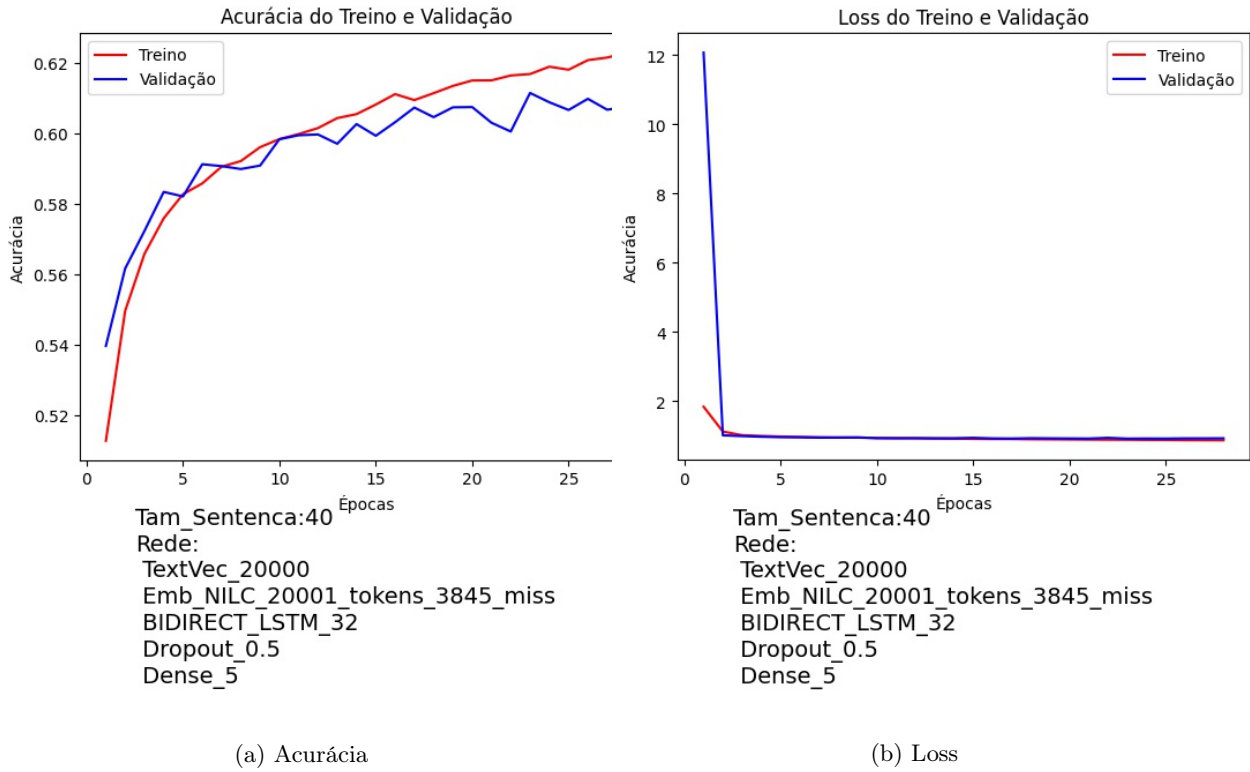


Figura 12: Resultados da rede Bidirec-0.50

2.3 Resultados e Comentários

Na tabela 2 apresentamos os resultados obtidos por experimento realizado. No caso, em cada experimento utilizamos o melhor modelo gerado sem sobreajuste. Neste caso, confiamos nos parâmetros da API do Tensorflow. Para isso foi configurado para que o treinamento pare após 5 épocas sem melhora na medida de Loss entre as épocas. E uma vez que o treinamento é interrompido, a API resgata o melhor modelo da última época.

Experimento	Loss no teste	Acurácia no teste
LSTM-0.0	0.9379	0.5989
LSTM-0.25	0.9423	0.6022
LSTM-0.5	0.9213	0.6042
Bidirec-0.0	0.9292	0.6060
Bidirec-0.25	0.9389	0.5985
Bidirec-0.50	0.9351	0.6003
NILC-LSTM-0.0	0.9371	0.5939
NILC-LSTM-0.25	0.9304	0.5984
NILC-LSTM-0.5	0.9329	0.5970
NILC-Bidirec-0.0	0.9243	0.5991
NILC-Bidirec-0.25	0.9275	0.5984
NILC-Bidirec-0.50	0.9309	0.6007

Tabela 2: Resultados dos experimentos realizados sobre os dados de testes

Aqui neste documento apresentamos apenas os resultados dos experimentos descritos na introdução. No entanto, diversas outras configurações, tanto dos hiperparâmetros, quanto de pré-processamento foram realizadas. Realizamos também o experimento para diferentes tamanhos máximos de sentença e tamanho do vocabulário. Quando tentamos com o vocabulário de 1000 ou 2000 tokens, o sobre-ajuste se apresentou bem mais rápido. Às vezes ainda nas primeiras 3 épocas o sobre-ajuste já ocorria.

Apesar da proposta inicial do trabalho ser de escolher alguma camada de Embedding. Acharmos por bem, trazer as comparações entre duas variações de embedding, no caso, a embedding vazia e outra adaptada

utilizando os pesos do word2vec fornecidos pelo NILC.

Sobre os resultados aqui apresentados, podemos ver que o que mais demorou a sobre-ajustar, foi o modelo "NILC-Bidirec-0.50", que apeans na 25a época teve seu treinamento encerrado. E com esse modelo obtivemos uma acurácia de 0.6007 nos dados de testes.