

# MAC5725

## Linguística Computacional

### Relatório - EP2

Aluno: Felipe de Lima Peressim  
NUSP: 11823558

30 de outubro de 2020

- (a) **Descrição das configurações e dos dos hiper-parâmetros utilizados**

A tabela a seguir descreve as configurações das 3 redes LSTM unidirecionais com os hiper-parâmetros utilizados. Em resumo, a mesma configuração foi utilizada nas três redes unidirecionais, com exceção da probabilidade de *Dropout* que variou entre 0%, 25% e 50% em cada um dos modelos respectivamente.

LSTM – Unidirecional	
Tamanho da entrada	50
Camadas ocultas	1
Camada de embedding	NILC - 200000x50
Qtd células	64
Qtd Neurônios Camada de saída	5
Função de ativação camada de saída	Softmax
Função de ativação camada oculta	Tanh
Função de ativação Recorrente	Sigmoid
Regularização	Dropout - 0%, 25%, e 50%
Otimizador	Adam
Loss	Sparse Categorical Crossentropy
Batch Size	32
Épocas	100

Tabela 1: Configuração dos modelos LSTM Unidirecionais.

A tabela 1 descreve as configurações dos modelos LSTM unidirecionais. Os três modelos em questão foram configurados com os mesmos parâmetros, com exceção da probabilidade de *Dropout* que variou entre 0%, 25% e 50% conforme destacado anteriormente. As sentenças contidas no conjunto de dados foram limitadas a um máximo de 50 caracteres, assim, sentenças com menos de 50 caracteres foram preenchidas e as com mais de 50 foram truncadas. Deste modo, definiu-se o formato da matriz que a camada de entrada da rede espera como sendo  $n \times 50$ , onde  $n$  é

a quantidade de amostras do conjunto de dados. Antes de passar pela camada de entrada cada palavra individual das sentenças foram convertidas para índices que são utilizados pela camada de embedding para recuperar as matrizes com os pesos de cada palavra.

A entrada da rede passa por uma camada de *embedding* treinada com o modelo *Cbow* de 50 dimensões do Centro Interinstitucional de Linguística Computacional (NILC). Para este EP em específico a camada de *embedding* foi construída com os pesos das duzentas mil palavras mais frequentes.

Os modelos em questão possuem apenas uma camada oculta com 64 células. As funções de ativação na camada oculta das redes LSTM não foram alteradas, ou seja, foram mantidas no padrão do artigo original *tanh* e *sigmoid*. Uma camada com *Dropout* foi configurada entre a camada oculta e a camada de saída com as probabilidades supracitadas. Na saída dos modelos foi configurado uma camada densa com 5 neurônios e função de ativação *softmax*.

A função *loss* utilizada foi a *Sparse Categorical Crossentropy*, e o algoritmo *Adam* foi utilizado como função de otimização. Os parâmetros do *Adam* não foram alterados, manteu-se os padrões do Keras que segue o artigo original. O tamanho de cada lote (*batch size*) utilizado durante a etapa de treinamento foi definido como 32 amostras por lote. Por último um total de 100 épocas foram estipuladas.

LSTM – Bidirecional	
Tamanho da entrada	50
Camadas ocultas	1
Foward Layer	LSTM
Backward Layer	LSTM
Camada de embedding	NILC - 200000x50
Qtd células	64
Qtd Neurônios Camada de saída	5
Função de ativação camada de saída	Softmax
Função de ativação camada oculta	Tanh
Função de ativação Recorrente	Sigmoid
Regularização	Dropout - 0%, 25%, e 50%
Otimizador	Adam
Loss	Sparse Categorical Crossentropy
Batch Size	32
Épocas	100

Tabela 2: Configuração dos modelos Bidirecionais.

A tabela 2 exibe a configuração utilizada nas redes Bidirecionais. De maneira análoga aos modelos unidirecionais, a configuração dos três modelos bidirecionais é a mesma, com exceção da variação na probabilidade do *Dropout*. As camadas *forward* e *backward* possuem a mesma configuração das redes LSTM unidirecionais exibida na tabela 1.

Para monitorar o desempenho dos modelos, dois *callbacks* foram utilizados: *Early Stopping* e *Model Check Point*. Ambos configurados para monitorar a acu-

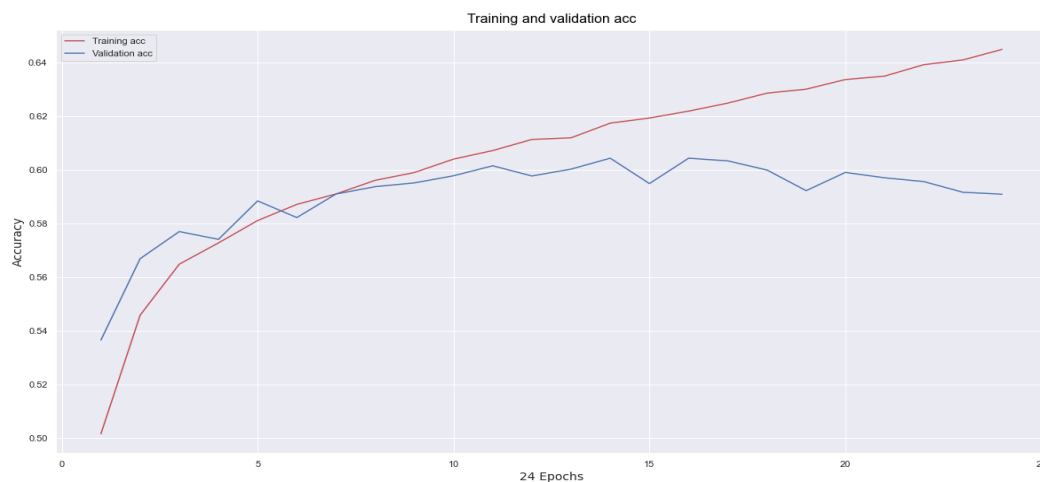
r cia do conjunto de valida  o. Especificamente, o *Early Stopping* foi configurado para interromper o treinamento do modelo quando n o houver melhora de desempenho dentre de 10  pocas de treinamento. O *Model Checkpoint* salva o modelo na  poca em que este obteve a melhor acur cia no conjunto de valida  o.

Por  ltimo, os dados utilizados para treinar, validar e testar o modelo, foram particionados conforme indicado no enunciado, ou seja, 65% para o conjunto de treinamento, 10% para o conjunto de valida  o e 25% para o conjunto de testes. Assim, os modelos contaram com um total de 78501 amostras no conjunto de treino; 12078 amostras no conjunto de valida  o e 30193 amostras no conjunto de testes.

- (b) Os seis gr ficos de valida  o gerados, apontando o n mero de  pocas usados para gerar o modelo sem sobreajuste.

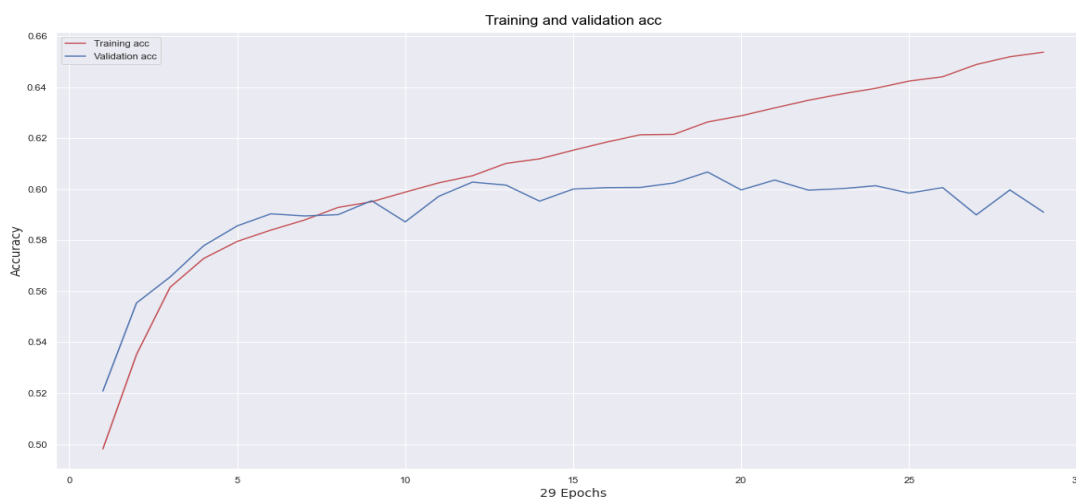
Os gr ficos com as curvas das acur cias obtida pelos modelos nos conjuntos de treinamento e de valida  o s o exibidas as seguir.

Figura 1: Modelo LSTM Unidirecional com 0% de *Dropout*.



O modelo LSTM unidirecional com *Dropout* de 0% (Figura 1) obteve a melhor acur cia durante a  poca de n mero 16, com um total de 60,44% de acur cia no conjunto de valida  o. A partir da  poca 16 o modelo passa a apresentar sobreajuste, ou seja, a acur cia obtida no conjunto de treinamento come a a aumentar de maneira indefinida e a acur cia de valida  o passa a diminuir at  se estabilizar entre 58% e 59%. O treinamento   interrompido pelo callback *Early Stopping* durante a  poca de n mero 24.

Figura 2: Modelo LSTM Unidirecional com 25% de *Dropout*.



O modelo LSTM unidirecional com *Dropout* de 25% (Figura 2) obteve a melhor acurácia durante a época de número 19, com um total de 60,67% de acurácia no conjunto de validação. O modelo começa a apresentar sinais de sobreajuste a partir da época 10, mas é a partir da época 20 que as curvas começam a se distanciar uma da outra mais rapidamente. A acurácia no conjunto de validação apresenta picos de diminuição e aumento mas se estabilizar entre 59% e 60%. O treinamento do modelo é interrompido pelo callback durante a época 29.

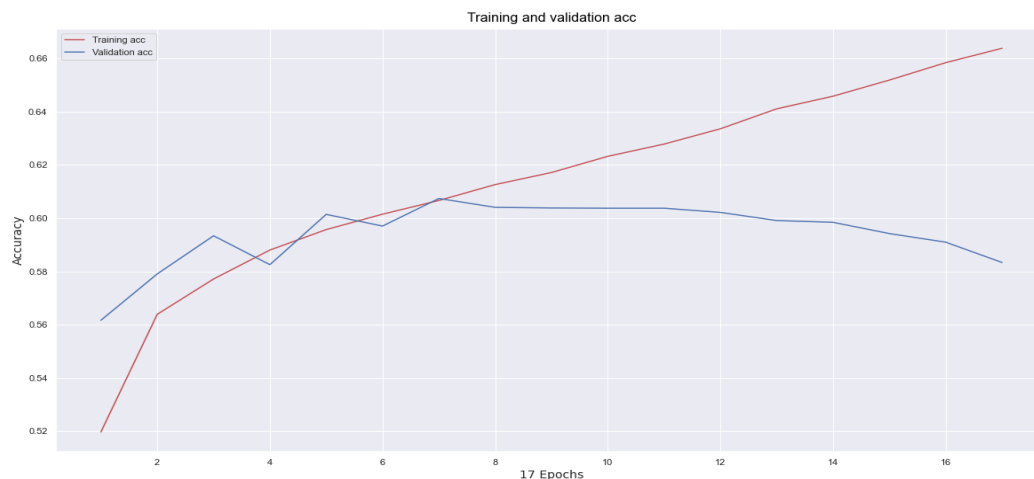
Figura 3: Modelo LSTM Unidirecional com 50% de *Dropout*.



O modelo LSTM unidirecional com *Dropout* de 50% (Figura 3) obteve a melhor acurácia de durante a época de número 17, com um total de 60,47% de acurácia no conjunto de validação. O modelo começa a apresentar sinais de sobreajuste a partir da época 18, mas é a partir da época 20 que as curvas ficam mais distantes

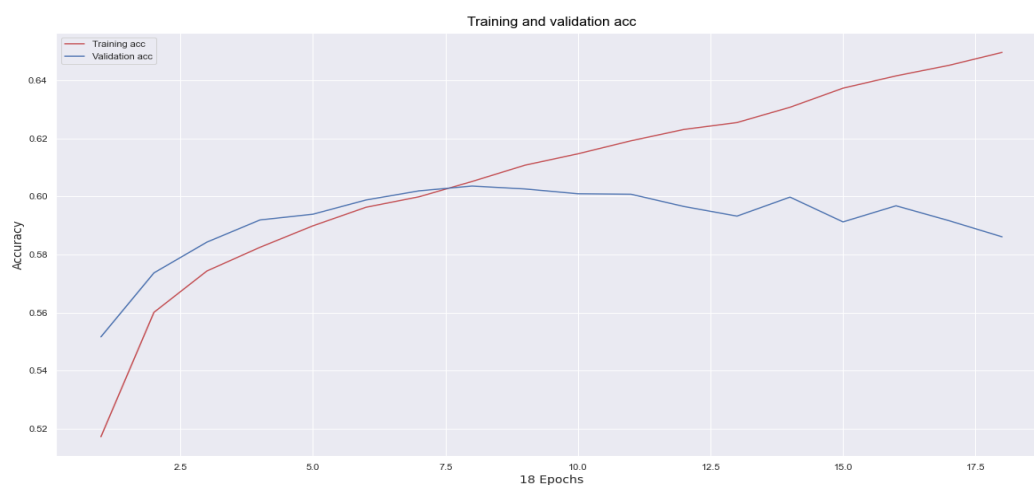
uma da outra mais rapidamente. A acurácia no conjunto de validação se estabiliza em 60%, no entanto a acurácia do conjunto de treinamento continua crescendo mas em um ritmo menor comparado com as curvas das figuras 1 e 2. O treinamento do modelo é interrompido pelo callback durante a época 27.

Figura 4: Modelo LSTM Bidirecional com 0% de *Dropout*.



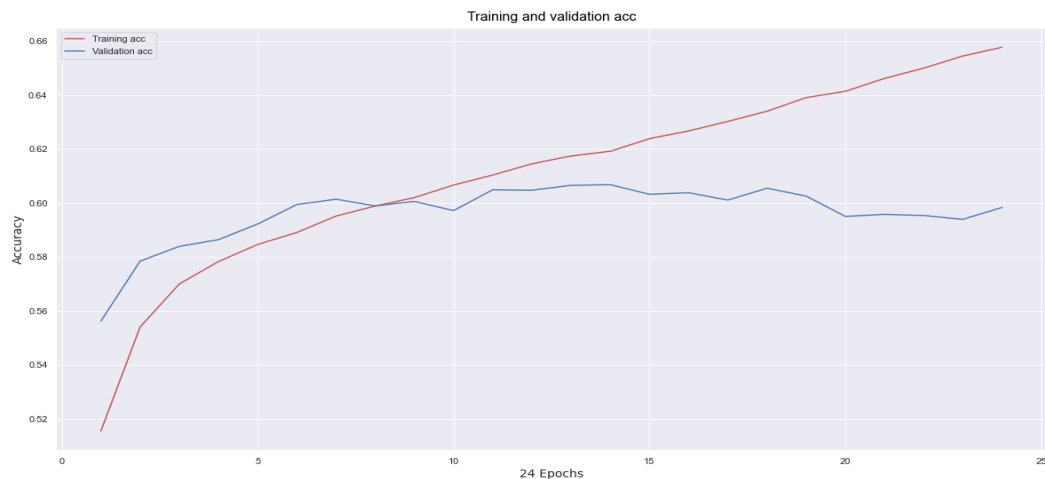
O modelo LSTM bidirecional com *Dropout* de 0% (Figura 4) obteve a melhor acurácia de validação durante a época de número 7, com um total de 60,74% de acurácia no conjunto de validação. A partir da época 8 o modelo passa a apresentar sobreajuste, ou seja, a acurácia obtida no conjunto de treinamento começa a aumentar de maneira indefinida e a acurácia de validação passa a diminuir até se estabilizar em 58%. O treinamento é interrompido pelo callback *Early Stopping* durante a época de número 17.

Figura 5: Modelo LSTM Bidirecional com 25% de *Dropout*.



O modelo LSTM bidirecional com *Dropout* de 25% (Figura 5) obteve a melhor acurácia de validação durante a época de número 10, com um total de 60,26% de acurácia no conjunto de validação. A partir da época 11 o modelo passa a apresentar sobreajuste, ou seja, a acurácia obtida no conjunto de treinamento começa a aumentar de maneira indefinida e a acurácia de validação passa a diminuir até se estabilizar em 59%. O treinamento é interrompido pelo callback *Early Stopping* durante a época de número 18.

Figura 6: Modelo LSTM Bidirecional com 50% de *Dropout*.



O modelo LSTM bidirecional com *Dropout* de 50% (Figura 6) obteve a melhor acurácia de validação durante a época de número 14, com um total de 60,68% de acurácia no conjunto de validação. A partir da época 15 o modelo passa a apresentar sobreajuste, ou seja, a acurácia obtida no conjunto de treinamento começa a aumentar de maneira indefinida e a acurácia de validação passa a diminuir até se estabilizar em 59%. O treinamento é interrompido pelo callback *Early Stopping* durante a época de número 24.

- (c) Uma tabela com as seis acurácias de teste.

Modelo	Acurácia - Conjunto de testes
LSTM Unidirecional Dropout 0%	59,68%
LSTM Unidirecional Dropout 25%	59,30%
LSTM Unidirecional Dropout 50%	59,33%
LSTM Bidirecional Dropout 0%	59,65%
LSTM Bidirecional Dropout 25%	59,60%
LSTM Bidirecional Dropout 50%	59,52%

Tabela 3: Acurácias dos modelos no conjunto de testes.

- (d) Uma discussão dos seus resultados.

As acurácias apresentadas na tabela 3 foram obtidas com os pesos dos modelos que foram armazenados no momento em que estes apresentaram o melhor desempenho, conforme descrito na discussão do item (c). No entanto, conforme se pode observar na tabela 3, todos os modelos apresentaram a mesma acurácia durante a fase de testes, que ficou em torno de 59%.

Uma possível hipótese para a baixa acurácia obtida se concentra na quantidade classes em que se tentou prever, assim como se realizou a classificação de 5 categorias, a quantidade de dados disponíveis não foi o suficiente, e portanto, o modelo teve sua capacidade preditiva limitada.

Deve-se salientar ainda que, como os dados estão desbalanceados, as acurácias obtidas só foram possíveis devido a estratificação realizada no momento de se separar o conjunto de dados em conjuntos de treinamento, validação e testes, visto a desproporcionalidade entre as categorias presentes nas amostras, ou seja, os dados foram separados mantando-se as categorias com a mesma proporção nos três conjuntos. Sem realizar esse procedimento, o modelo apresentava subajuste e não passava de 30% de acurácia.

Embora todos os modelos tenham apresentado a mesma acurácia no conjunto de testes, porém, conforme já discutido anteriormente no item (c), conforme se aumentava a probabilidade de *Dropout*, menor foi o sobreajuste. Isso, fica bastante evidente ao se observar que a distância entre as curvas de acurácia dos modelos diminuem conforme se aumentou a porcentagem da regularização *Dropout*.