

Multilayer Perceptron para Classificação de Sentimento em Textos utilizando a biblioteca TensorFlow

Jessica Imlau Dagostini, Marcos Vinicius de Moura Lima, Neilor Tonin

Universidade Regional Integrada do Alto Uruguai e das Missões (URI - Erechim)

{051137, 045122}@aluno.uricer.edu.br, nat@uricer.edu.br

Abstract. *Artificial Neural Networks are being employed in different classification process. These are mathematical models that aim to simulate the biologic neural networks behavior, with propose to learn some patterns. Once that these networks can be used with a diversity of ends, their applicability makes this approach an ally for the solution of many of problems. Text classification is one of them. The present work aims to present the use of Multilayer Perceptron Neural Networks to classify texts based on their sentiment. Will be analyzed the model itself, as its accuracy according to the characteristics applied to the training.*

Resumo. *Redes Neurais Artificiais vem sendo empregadas para diferentes processos de classificação. Estas são modelos matemáticos que procuram simular redes neurais biológicas, a fim de aprender certos comportamentos. Podendo ser utilizadas para diversos fins, a aplicabilidade de RNAs faz delas uma ótima aliada para solução de problemas. A classificação de textos pode ser enquadrada dentro das aplicações de redes neurais. O presente trabalho apresenta a aplicação de Redes Neurais do tipo Multilayer Perceptron para classificar textos de acordo com o sentimento que os mesmos expressam. Será analisado o modelo em si, bem como sua acurácia perante as características de treinamento aplicadas.*

1. Introdução

Redes Neurais Artificiais (RNAs) vem sendo empregadas para diferentes processos de classificação. Estas são modelos matemáticos que procuram simular redes neurais biológicas, a fim de aprender certos comportamentos [da Silva 2009]. Compostas por neurônios artificiais, denominados *perceptrons*, elas se parecem com o cérebro humano em dois aspectos: o conhecimento é adquirido através de um processo de aprendizagem e as conexões entre os neurônios (chamadas de pesos sinápticos) são utilizadas para guardar o conhecimento adquirido [Haykin 2007].

Uma rede neural pode se enquadrar em uma de diversos tipos de arquiteturas existentes. As redes perceptron de múltiplas camadas (*multilayer perceptron*) são bastante populares e empregadas em diversas aplicações, principalmente na utilização de RNAs para classificação de dados [de Pádua Braga et al. 2007].

O presente trabalho visa apresentar o emprego de uma rede neural *multilayer perceptron* para identificação de sentimento em textos. Para isso, foi utilizado o *data set Sentiment Labelled Sentences* [Kotzias et al. 2015], que contém 3000 avaliações de usuários para os sites *Amazon*, *IMDb* e *Yelp*. Estas avaliações foram classificadas em positivas e

negativas, e a rede tem por objetivo aprender a identificar esses sentimentos a partir desta base de dados.

As próximas seções apresentam o desenvolvimento do mesmo. A Seção 2 apresenta a fundamentação teórica do trabalho, dando especial atenção a contextualização das técnicas utilizadas para essa classificação. A Seção 3 detalha o modelo de rede neural aplicado. A Seção 4 apresenta os resultados das execuções do modelo descrito previamente e discute os mesmos. Por fim, a Seção 5 apresenta as conclusões do presente trabalho e os trabalhos futuros relacionados ao mesmo.

2. Fundamentação Teórica

O cérebro humano é um computador muito complexo que tem a capacidade de organizar seus neurônios para realizar um processamento de forma extremamente rápida e eficiente [Haykin 2007]. Por toda sua capacidade, o ser humano busca entender o seu funcionamento para poder replicar seu processamento.

Redes Neurais Artificiais (RNAs) buscam “emular” o funcionamento conhecido do cérebro humano computacionalmente, a fim de propiciar aprendizagem para máquinas [de Pádua Braga et al. 2007]. Elas são sistemas paralelos distribuídos, constituídos de unidades de processamento simples (os *perceptrons*), que são as unidades que aprendem. Para a aprendizagem acontecer, é necessário “treinar” a rede. Nesse treinamento, há a melhoria dos pesos sinápticos, que é onde a aprendizagem fica de certa forma “armazenada”. [Haykin 2007]

As redes do tipo *multilayer perceptron*, perceptron de multicamadas, vem obtendo sucesso na construção de soluções para problemas de categorização e reconhecimento de padrões [Silva et al. 2012]. Nessa arquitetura, do tipo *perceptron*, a rede é composta por uma camada de entrada dos dados, uma ou mais camadas ocultas (*hidden layers*) e uma camada de saída. Elas são fortemente conectadas, onde todos os neurônios de qualquer camada estão conectados a todos os neurônios da camada anterior [Silva et al. 2012] [de Pádua Braga et al. 2007].

O treinamento de uma RNA *multilayer perceptron* acontece através do algoritmo *backpropagation*, que encontra as derivadas da função de erro em relação aos pesos e *bias* da RNA [Silva et al. 2012]. Esses pesos sinápticos são o que a rede de fato irá aprender. Constituído por duas etapas, o *backpropagation* primeiro propaga as informações da camada de entrada por toda a rede. Essa propagação é a multiplicação do valor de entrada pelo peso, somado a um *bias*. Todo esse resultado passa por uma função de ativação, que definirá se o valor atribuído àquele neurônio será repassado para a próxima camada ou não. Esse processamento é uma simulação das ligações sinápticas que ocorrem no cérebro humano. [Haykin 2007]

Na segunda etapa, a *backward*, é onde de fato acontece o aprendizado da rede. Ela primeiro calcula o erro em relação a camada de saída e retro-propaga esse erro por todas as camadas anteriores, fazendo um reajuste dos pesos sinápticos. Dentre suas variações, o algoritmo de *backpropagation* também leva em consideração uma taxa de aprendizagem para fazer esse reajuste dos pesos. Para encontrar esse erro, a rede calcula a diferença entre o resultado da camada de saída com o resultado esperado para os dados. Esse tipo de aprendizado é denominado aprendizagem supervisionada. [da Silva 2009]

Há diversas aplicações de RNAs para diversos fins. Uma vez que seja necessário uma grande quantidade de processamento de informações, e que nenhum modelo “analítico” de estrutura de dados seja capaz de resolver o problema, a aplicação de redes neurais artificiais é indicada. Uma dessas aplicações é a de Mineração de Textos (MT). “Um dos objetivos da Mineração de Textos é o de descobrir conhecimento útil nos textos e classificar estas informações automaticamente em categorias”[Silva et al. 2004]. A aplicação de MT pode ser utilizada em muitos contextos e é composta por três etapas: coleta dos dados, pré-processamento e classificação. [Silva et al. 2004]

Através do seu conteúdo é possível identificar uma série de categorias a qual o texto em questão pode ser enquadrado. Uma das possíveis aplicações de RNAs para mineração de textos é na identificação de sentimentos dos mesmos. Assim, a partir dessa análise, é possível obter métricas da qualidade do serviço e da felicidade dos usuários quanto ao mesmo.

Para a construção de redes neurais, existem diversas bibliotecas que permitem a criação de modelos sem a necessidade de codificar os cálculos básicos, já popularmente conhecidos. Uma dessas bibliotecas existentes é a *TensorFlow* [Abadi et al. 2016]. O *TensorFlow* utiliza um grafo de fluxo de dados unificado para representar a computação em um algoritmo e o estado em que esse algoritmo opera, incluindo operações matemáticas individuais, seus parâmetros e suas regras de atualização. “Tensores são *arrays* multidimensionais, que vão fluindo pelos nós de um grafo” [Mesquita 2017]. Essa biblioteca se diferencia de outras pois o modelo suporta múltiplas execuções e vértices individuais pode ter estados mutáveis e podem compartilhar diferentes execuções no grafo. [Abadi et al. 2016]

3. O Modelo

Baseado em [Mesquita 2017], a configuração da rede utilizada neste trabalho é graficamente apresentada pela Figura 1. A rede neural é constituída por 4 camadas: a camada de entrada, que recebe os textos de entradas tratados; a camada oculta 1 e a camada oculta 2, ambas ativadas pela função de ativação ELU [Clevert et al. 2015]; e uma camada de saída.

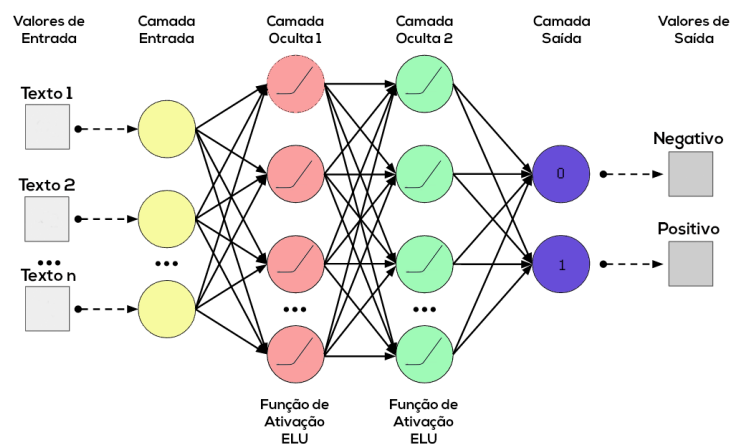


Figura 1. Arquitetura da Rede Neural

A camada de entrada possui tantos neurônios quanto a quantidade de palavras identificadas dentro dos textos. Para serem processados pela *multilayer perceptron*, os textos passam por um pré-processamento. Primeiramente, todas as palavras (tanto do *data set* utilizada para treinamento quanto o de testes) são lidas e armazenadas em um vetor *vocabulário*. Este vetor, além de armazenar as palavras, as armazena de acordo com sua recorrência. Em seguida, esse vocabulário é transformado em índices, onde cada palavra assume um sua respectiva posição no vocabulário.

Como há um número muito grande de informações, tanto de palavras quanto de textos para treinamento, a execução da rede se dá por *batches*. Neste trabalho, cada *batch* assumiu um tamanho de 250 textos. Para cada texto, é montado um vetor de ocorrência de cada palavra. Se uma palavra do vocabulário é identificada no texto, seu valor de entrada será 1. Todo o *batch* é armazenado em uma matriz, onde cada linha identifica um texto e cada coluna uma palavra do vocabulário.

Feito isso, os pesos e *bias* das camadas 1, 2 e de saída são gerados. Eles são números decimais aleatórios, gerados a cada execução da rede. Ambas camadas ocultas são formadas por 100 *perceptrons* cada. Estas são ativadas pela função *ELU*. Proposta em 2015, ela tem se mostrado bastante eficaz. Ela é definida por:

$$ELU(x, \alpha) = \begin{cases} x, & x \geq 0 \\ \alpha(\exp(x) - 1), & c.c. \end{cases} \quad (1)$$

Uma desvantagem técnica dessa ativação é que computá-la é uma operação ineficiente, deixando a rede um pouco mais lenta. Todavia, sua performance é mais alta que as outras funções de ativação conhecidas atualmente. [Clevert et al. 2015]

A camada de saída é composta por dois neurônios. Neste modelo, será utilizada a abordagem *one-hot*, onde somente um dos neurônios ativa para cada classificação. Por exemplo, se o texto for classificado como negativo, somente o primeiro neurônio será ativado, resultado num retorno 1, 0.

Também deve-se definir a taxa de aprendizagem. No caso do presente trabalho, será gerado um vetor de taxas de aprendizado, variando de 0.01 a 0.09 e de 0.1 a 0.9, a fim da proposta dos testes. A rede será executada por *n* épocas, também definidas na aplicação. Como referenciado na Seção 2, em *multilayer perceptrons* o algoritmo de aprendizagem utilizado é o *backpropagation*, formado por 2 etapas. A primeira, *forward*, é a execução das multiplicações dos pesos sinápticos pelos valores dos neurônios, sendo adicionado ao seu *bias* e ativados pela respectiva função de ativação escolhida. Já no *backward*, o erro relativo e a readequação dos pesos acontece. Para essa readequação dos pesos, no modelo será aplicado a variação *AdamOptimizer*. Esta variação leva em conta o valor da taxa de aprendizagem aplicada, fazendo com que ela não diminua agressivamente. Ele adiciona um momento e suaviza ruídos do gradiente. [Guedes 2016]

Por fim é calculado a acuracidade da rede. Para isso, o algoritmo utiliza o *batch* de testes, conferindo qual a porcentagem de acerto ele obtém. No caso do presente trabalho, será montada uma matriz de acurácia para cada uma das 16 diferentes taxa de aprendizagem executadas.

4. Resultados e Discussões

O modelo foi executado diversas vezes a fim de se observar a porcentagem de acurácia do seu aprendizado. Como já citado anteriormente, o *data set Sentiment Labelled Sentences* contém 3000 avaliações de usuários para os sites *Amazon*, *IMDb* e *Yelp*. Nas execuções da rede deste trabalho, o *data set Amazon* foi utilizado como objeto de treinamento e o *Yelp* é o *data set* de testes da mesma.

Com o objetivo de observar o comportamento da rede conforme a mudança de características da mesma, optou-se por executá-la com diferentes taxas de aprendizagem. O modelo foi executado com taxas variando de 0.01 a 0.09 e de 0.1 a 0.9, 5 vezes para cada taxa de ambos intervalos.

A taxa de aprendizado é uma constante de proporcionalidade no intervalo $[0,1]$. Assim quanto maior for essa constante, maior será a mudança nos pesos, aumentando a velocidade do aprendizado, o que pode levar à uma oscilação do modelo na superfície de erro [de Leon F. de Carvalho 2009].

A Figura 2 demonstra o gráfico das médias de acurácia obtidas para ambos intervalos.

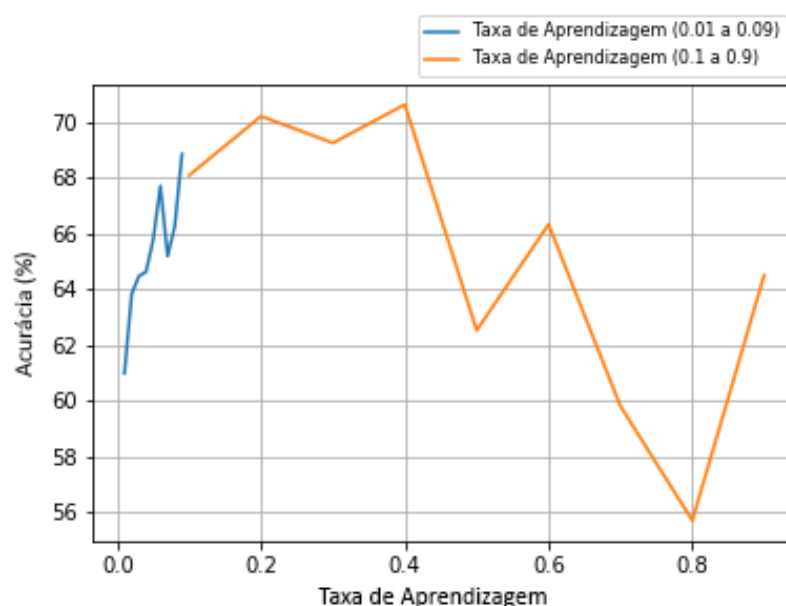
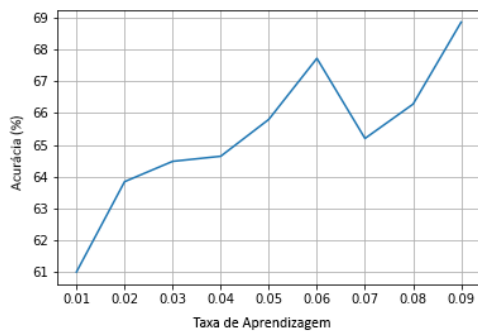
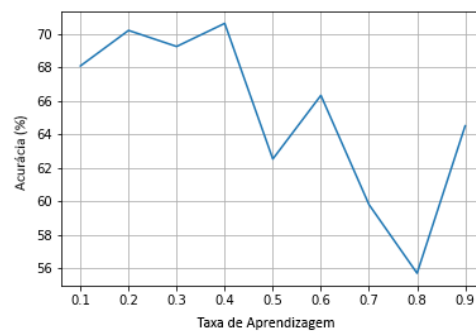


Figura 2. Média de Acurácia para Ambos Intervalos

A partir de sua análise, é possível perceber que a rede obteve um melhor desempenho com os valores de aprendizagem entre 0.2 e 0.4, com a porcentagem de acurácia variando entre 70% e 71%. O intervalo de aprendizagem 0.5 e 0.8 apresentou as piores porcentagens de todos os testes, com uma acurácia de 56%. Pode-se perceber que o modelo trouxe resultados satisfatórios para o *data set* escolhido.



(a) Intervalo 0.01 a 0.09



(b) Intervalo 0.1 a 0.9

Figura 3. Média de Acurácia dos Intervalos

A Figura 3 demonstra o gráfico dos intervalos separados. Nessa separação, é possível notar que a acurácia do intervalo 0.01 e 0.09 (mostrado pela Figura 3(a)) não chegou a 70%. O intervalo também mostrou que, conforme a taxa de aprendizagem foi aumentando, a acurácia também foi melhorada. Já na Figura 3(b) é possível perceber que o modelo atingiu sua máxima otimização com a taxa de aprendizagem em 0.4. Taxas maiores que essa levaram o modelo a convergir de forma errada e se afastar da mais otimizada predição. Com isso, notou-se que para o modelo proposto a taxa de aprendizagem, para apresentar melhor convergência, deve ser um valor intermediário, não sendo demasiado pequeno e nem muito próximo ao máximo.

5. Conclusão

O presente trabalho apresentou uma rede neural *multilayer perceptron*, construída através da biblioteca *TensorFlow*, para a classificação de textos de acordo com o sentimento expresso por eles. Os textos utilizados foram retirados do *data set Sentiment Labelled Sentences*, que possui avaliações de usuários nos sites *Amazon*, *IMDb* e *Yelp*.

Redes da arquitetura perceptron de multicamadas são compostas por uma camada de entrada dos dados, uma ou mais camadas ocultas (*hidden layers*) e uma camada de saída. Elas vem obtendo sucesso na construção de soluções para problemas de categorização e reconhecimento de padrões. Através da execução do modelo, notou-se que o mesmo obteve um desempenho satisfatório nos testes. Em todos eles o modelo apresentou acurácia acima de 50%. Para este modelo de rede que é, de certa forma, mais simplório, os resultados foram bastante eficazes.

O desenvolvimento do presente trabalho propiciou uma maior aprendizagem em relação ao funcionamento e a construção de redes neurais artificiais. A partir dele, será possível desenvolver outros modelos para aplicação desta técnica de classificação, a fim de resolver problemas reais.

Referências

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al. (2016). Tensorflow: a system for large-scale machine learning. In *OSDI*, volume 16, pages 265–283.

- Clevert, D.-A., Unterthiner, T., and Hochreiter, S. (2015). Fast and accurate deep network learning by exponential linear units (elus). *arXiv preprint arXiv:1511.07289*.
- da Silva, A. M. (2009). *Utilização de Redes Neurais Artificiais para a Classificação de SPAM*. PhD thesis, Dissertação (mestrado). CEFET-MG.
- de Leon F. de Carvalho, A. P. (2009). Perceptron multi-camadas (mlp).
- de Pádua Braga, A., de Leon Ferreira, A. C. P., and Ludermir, T. B. (2007). *Redes neurais artificiais: teoria e aplicações*. LTC Editora.
- Guedes, A. B. S. (2016). Reconhecimento de gestos usando redes neurais convolucionadas. *Monografia em Engenharia de Software, Universidade de Brasília*.
- Haykin, S. (2007). *Redes neurais: princípios e prática*. Bookman Editora.
- Kotzias, D., Denil, M., de Freitas, N., and Smyth, P. (2015). From group to individual labels using deep features. In *ACM SIGKDD*.
- Mesquita, D. (2017). Classificando textos com redes neurais e tensorflow.
- Silva, C., Vieira, R., and Osorio, F. (2004). Uso de informações lingüísticas em categorização de textos utilizando redes neurais artificiais. *VIII Simpósio Brasileiro de Redes Neurais,(2004a)*, pages 1–6.
- Silva, R. M., Almeida, T. A., and Yamakami, A. (2012). Análise de desempenho de redes neurais artificiais para classificação automática de web spam. *Revista Brasileira de Computação Aplicada*, 4(2):42–57.