

Redução de Dimensionalidade via Amostragem Compressiva Aplicada a Redes Neurais

Micael Espínola Fonseca Tomaz e Edmar Candeia Gurjão

Resumo— As Redes Neurais Artificiais (RNA) são um tema recorrente na tecnologia, comprovadas por suas diversas aplicações. No treinamento da RNA é necessário usar um grande conjunto de dados para representar as possibilidades entrada e saída de um problema que está sendo modelado, o que leva a uma grande demanda de processamento computacional. Neste trabalho, para reduzir o custo computacional, foi utilizada a Amostragem Compressiva como técnica de redução de dimensionalidade no pré-processamento de uma Rede Neural, os resultados obtidos mostram que é possível treinar a rede e que não há perda significativa no desempenho do treinamento quando se utilizam os dados comprimidos.

Palavras-Chave— Treinamento, Redes Neurais, Redução de Dimensionalidade, Amostragem Compressiva.

Abstract— Artificial Neural Networks (ANN) are a current and recurrent theme in any technological scope and their applications confirm such fact. In machine learning using ANN it is necessary to use a large data set to represent the input and output possibilities of a problem that is being modeled, which leads to a great demand for computational processing. The project aimed to decrease the adversities of computational cost by exploring Compressive Sampling as a dimensional reduction technique in the preprocessing of a Neural Network.

Keywords— Training, Neural Networks, Dimensionality Reduction, Compressed Sensing.

I. INTRODUÇÃO

As Redes Neurais Artificiais já são reconhecidas como o principal método de aprendizagem de máquina da atualidade, podemos ver isso quando notamos as suas mais variadas aplicações em músicas, imagens, segurança cibernética e até em previsões na bolsa de valores. Além disso, com o surgimento da aprendizagem profunda (Deep Learning), não só as utilidades aumentaram, mas também os desafios de sua aplicação.

Definido o modelo da rede neural inicia-se a etapa de treinamento, que consiste em apresentar um conjunto de entradas e saídas conhecidas para que o modelo se ajuste para explicar esses dados [1]. Essa etapa normalmente necessita de grande processamento computacional. Uma forma de contornar esse problema é tratar os dados de entrada, por exemplo reduzindo a dimensão das informações por convolução.

A Amostragem Compressiva consiste em tomar um vetor \mathbf{x} de tamanho $M \times 1$ e explorar sua esparsidade para utilizar uma matriz \mathbf{A} de dimensões $M \times N$ ($N < M$) e obter um vetor $\mathbf{y} = \mathbf{A}\mathbf{x}$ de tamanho $N \times 1$ que é uma versão

comprimida de \mathbf{x} . Essa operação preserva a distância entre os vetores após a compressão [2]. Neste trabalho foi avaliada a utilização da Amostragem Compressiva como método de pré-processamento de uma Rede Neural para reduzir o custo computacional do treinamento, e foi mostrado que esse pré-processamento mantém as características dos dados e que o desempenho da rede não sofreu alterações significativas.

II. METODOLOGIA E DESENVOLVIMENTO

A metodologia da pesquisa apresentou dois pontos de maior relevância. O primeiro deles consistiu em uma introdução às ferramentas e conceitos que seriam empregados ao longo do projeto. Nessa etapa, estabeleceu-se familiaridade com a biblioteca TensorFlow para a construção de modelos neurais, bem como com o conceito de amostragem compressiva por meio do programa 'l1magic' [3], o qual realizava a amostragem compressiva no ambiente MATLAB. Em uma segunda fase, os esforços foram concentrados no pré-processamento dos conjuntos de dados, na criação das redes neurais e em seus subsequentes treinamentos.

A. Sobre o pré-processamento e o treinamento da rede

A dinâmica utilizada no pré-processamento ocorreu da seguinte forma: Inicialmente, no MATLAB, criou-se uma “matriz de amostragem compressiva” com dimensões $M \times N$ ($N < M$), sendo M o tamanho do vetor coluna que será comprimido, enquanto N é o tamanho do novo vetor resultante. Em seguida, essa matriz foi exportada para um arquivo do tipo CSV.

O mencionado arquivo CSV e o banco de dados de interesse foram importados no ambiente virtual do Google Colab. No conjunto de dados, as imagens são representadas matrizes de pixels. Cada pixel é representado por um valor entre 0 e 255 (escala de cinza). Para otimizar o processo, realizamos uma normalização das matrizes, dividindo cada elemento por 255. Na sequência, remodelamos as matrizes em vetores coluna e reduzimos suas dimensões através da amostragem compressiva por meio de multiplicação matricial.

Por último foram realizados treinamentos. Um deles utilizou o banco de dados original, enquanto os outros utilizaram dados pré-processados via Amostragem Compressiva. Considerando a rede treinada com a base original (sem compressão) como referência, foi obtida a classificação para uma entrada nessa rede, bem como na rede treinada com os dados pré-processados. Caso a classificação não coincida, considera-se que ocorreu um erro. A taxa de erros foi calculada como a razão entre a quantidade de erros e o total de imagens testadas.

B. Sobre os modelos neurais

A rede neural desenvolvida segue uma arquitetura sequencial, compreendendo camadas de entrada, uma camada densa intermediária e uma camada de saída. Essa arquitetura foi empregada para ambos os conjuntos de dados abordados. O modelo é configurado com uma camada densa, contendo 128 neurônios ativados pela função “Relu”, enquanto a camada de saída é composta por 10 neurônios ativados pela função “Softmax”. Foi utilizado o otimizador “Adam”, e a função de perda adotada é a “Sparse softmax”.

III. VISÃO GERAL DOS BANCOS DE DADOS

Foram utilizados dois bancos de dados, sendo o primeiro o MNIST, que consiste em um conjunto de 70.000 imagens de dígitos manuscritos, divididos em partições de treinamento e teste para a rede neural [4]. Cada imagem é uma matriz de 28 por 28 pixels e tem um rótulo associado que mostra qual dígito está sendo representado.

Em seguida foi utilizado o Fashion MNIST, que consiste em um conjunto de 60.000 imagens de artigos de roupas e acessórios, divididos em partições de treinamento e teste para a rede neural. Cada imagem é uma matriz de 28 por 28 pixels, associada a um rótulo de 10 classes distintas. Tanto o MNIST quanto o Fashion MNIST possuem imagens em uma escala de cinza, facilitando a aplicação da amostragem compressiva.

IV. RESULTADOS E DISCUSSÃO

No conjunto de dados MNIST e Fashion MNIST originais, cada imagem é representada por um vetor coluna com 784 elementos. Sabendo disso, foi possível reduzir esse vetor para um tamanho único usando a técnica de Amostragem Compressiva durante o pré-processamento. Contudo, percebeu-se que essa redução resulta em um aumento de erros na Rede Neural. Ao aplicar técnicas no conjunto MNIST, foi alcançada uma redução de 85% no tamanho do vetor, com um aumento da probabilidade de erro de apenas 2,1%.

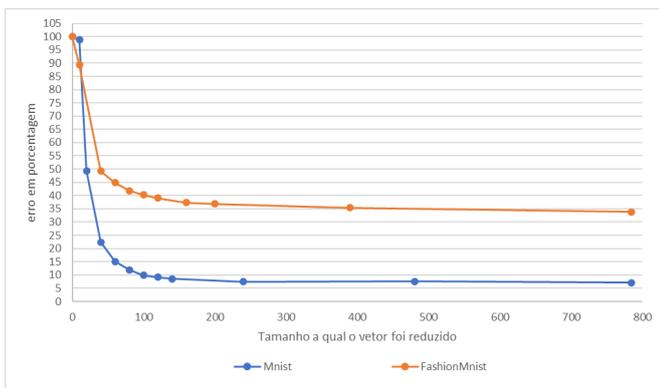


Fig. 1. Probabilidades de erro da rede neural com a redução dimensional do vetor imagem no banco de dados MNIST e FashionMnist.

Para a base Fashion MNIST foi alcançada uma redução de 75% no tamanho do vetor de características das imagens com um aumento na taxa de erro da rede neural de 3%, mostrando a efetividade dessa abordagem proposta, pois houve redução do

custo computacional sem prejuízo significativo do desempenho da rede. Esses resultados são ilustrados na Figura 1.

Os gráficos da Figura 1 mostram a taxa de erro em função do tamanho dos vetores de entrada após a compressão. Observa-se que há um ponto crítico onde a redução dos dados piora excessivamente o desempenho da rede, e portanto não há vantagem em reduzir o custo computacional a partir deste ponto.

A redução mais vantajosa no Fashion MNIST, acontece ao reduzir tamanho da camada de entrada para 200. Neste caso, há uma compressão total de parâmetros analisados de 101.770 para 27.018 e um aumento de apenas 3% na taxa de erro ao final de 10 épocas de treinamento. Na Figura-2 está o desempenho da neural com o banco de dados original e modificado ao decorrer das épocas de treinamento.

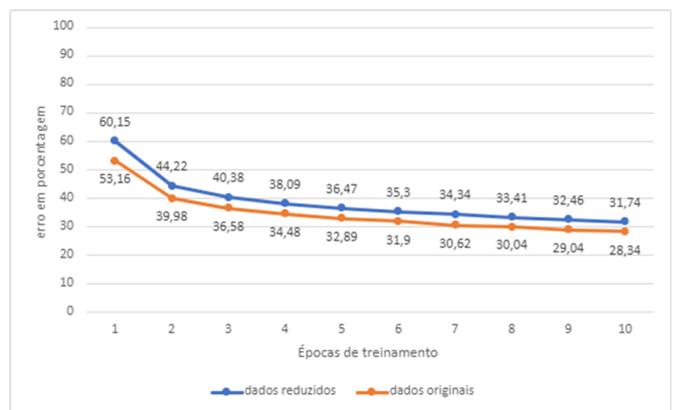


Fig. 2. Desempenho do treinamento do mesmo modelo neural usando o banco de dados Fashion MNIST original e pré-processado via amostragem compressiva.

CONCLUSÕES

Neste trabalho foi explorado o efeito da Amostragem Compressiva no pré-processamento de uma Rede Neural. Os resultados estabeleceram viabilidade no uso dessa técnica como forma de reduzir os gastos computacionais no pré-processamento, e que há espaço para novas investigações sobre a aplicação da Amostragem Compressiva no contexto proposto.

AGRADECIMENTOS

Ao Programa de Iniciação Científica da CNPq (PIBIC-CNPq) da Universidade Federal de Campina Grande (UFCG) pelo apoio financeiro.

REFERÊNCIAS

- [1] S.Haykin, *Neural Networks and Learning Machine*. Pearson, 3rd Edition 2008.
- [2] KUTYNIOK, Y. C. E. *Compressed Sensing: Theory and Applications*. 1. ed. Cambridge, UK: Cambridge University Press, 2012.
- [3] 11-magic, <https://candes.su.domains/software/11magic/>. Acessado em 13 de agosto de 2023.
- [4] MNIST dataset, <https://www.tensorflow.org/datasets/catalog/mnist>. Acessado em 13 de agosto de 2023.