

## 3.04.99 - Engenharia Elétrica

**Estudo e Aplicação de Redes Neurais Generativas**Thiago Soares Laitz<sup>1\*</sup>, Levy Boccato<sup>2</sup>

1. Estudante da Faculdade de Engenharia Elétrica e de Computação (FEEC) da Unicamp

2. Professor/pesquisador da FEEC-Unicamp – orientador

**Resumo**

Modelos generativos lidam com o desafio de sintetizar novas amostras que apresentem características fidedignas em relação às amostras reais e que, portanto, sejam compatíveis com a distribuição dos dados. Tais modelos são úteis para a expansão de bancos de dados (DA, do inglês *data augmentation*) e podem aperfeiçoar a generalização de sistemas, especialmente em cenários de escassez de dados.

Este trabalho foi dedicado ao estudo e aplicação de três redes neurais generativas – (1) rede generativa adversária (GAN), (2) rede auto-codificadora variacional (VAE) e (3) rede auto-codificadora adversária (AAE) – para a síntese de sinais de eletroencefalografia (EEG) no contexto de interfaces cérebro-computador baseadas em potenciais visualmente evocados, visando aumentar a quantidade de amostras para treinamento e, finalmente, aprimorar o desempenho do sistema. Os experimentos realizados revelaram aumentos de acurácia de até 10% para o domínio da frequência, e de 7,25% para o domínio do tempo.

**Autorização legal:** Comitê de Ética da Unicamp (CAAE 0617.0.146.000-10)

**Palavras-chave:** Redes Generativas Adversárias; Autoencoders; Interface Cérebro-Computador

**Apoio financeiro:** PIBIC/CNPq - Unicamp

**Trabalho selecionado para a JNIC:** PIBIC - Unicamp

**Introdução**

Nos últimos anos, várias tarefas desafiadoras do mundo real, como o reconhecimento de padrões em imagens e o processamento de linguagem natural, experimentaram avanços significativos graças ao desenvolvimento de modelos baseados em aprendizado profundo (DL, do inglês *deep learning*). O notório sucesso de DL se deu muito por conta de vários aprimoramentos na estrutura e nas técnicas de treinamento destes modelos, mas também pela maior capacidade computacional atual e, também, pela disponibilidade de quantidades expressivas de dados [1].

Contudo, ainda existem áreas nas quais a coleta de dados é um processo muito custoso, o que faz com o volume de amostras seja reduzido. Nestes casos, é fundamental que os projetistas disponham de estratégias de DA para evitar o sobreajuste (*overfitting*). Um ramo de pesquisa que ainda lida com a escassez de dados envolve o projeto de interfaces cérebro-computador (BCIs, do inglês *brain-computer interfaces*), que são dispositivos que buscam estabelecer um canal de comunicação não-muscular entre o cérebro e um computador [2]. A partir do processamento dos sinais cerebrais, capturados, por exemplo, via EEG, a BCI busca identificar o comando selecionado pelo usuário e, assim, executar a ação correspondente. Neste trabalho, foi considerada uma BCI baseada em potenciais visualmente evocados em regime estacionário (SSVEP, do inglês *steady-state visually evoked potentials*), na qual o indivíduo observa estímulos visuais oscilando em frequências pré-determinadas, sendo a BCI responsável por identificar para qual frequência o indivíduo está olhando [4].

O objetivo deste trabalho foi o estudo e a aplicação de três modelos neurais generativos – (1) GAN, (2) VAE e (3) AAE – para a expansão de bases de dados com amostras sintéticas válidas. Primeiramente, consideramos a geração de dígitos manuscritos para o banco de dados MNIST, bem como a síntese de imagens de quartos de hotéis com a base LSUN-bedrooms. Posteriormente, abordamos o desafio de sintetizar novos sinais de EEG, tanto no domínio temporal quanto no domínio da frequência, para um sistema BCI-SSVEP. Neste caso, avaliamos as taxas de acerto da BCI treinada sem e com os dados sintéticos nos dois domínios (tempo e frequência) considerando 10 voluntários de uma base de dados local [3].

**Metodologia**

A primeira etapa do estudo avaliou o desempenho dos três modelos generativos escolhidos (VAE, AAE e GAN) no desafio de sintetizar dígitos manuscritos da base MNIST, composta por 60 mil imagens 28x28 para treinamento e 10 mil amostras de validação.

Para o VAE, o treinamento foi realizado durante 1000 épocas, com a entropia cruzada como função custo e otimizador *RMSprop*. Além disso, foi adicionado um termo à função custo que simboliza a divergência de Kullback-Leibler entre a distribuição gerada e a distribuição alvo no espaço latente [5]. A estrutura foi composta de camadas densas que se reduziam até um vetor de 2 variáveis no espaço latente, o qual, depois, era expandido até reconstruir a imagem.

Para o AAE, o treinamento durou 5 épocas, com entrada normalizada em  $[-1,1]$ , com estrutura composta

por camadas convolucionais para o codificador e convoluções transpostas para o decodificador, enquanto o discriminador tinha duas camadas densas de 1000 neurônios [6].

Para a GAN, o treinamento foi executado por 100 épocas, com uma estrutura semelhante ao AAE baseada em camadas convolucionais transpostas intercaladas por camadas de *batch normalization*. Na saída da rede discriminadora utilizou-se uma camada densa de 2 neurônios e função *sigmoid*.

Como segunda etapa, foi avaliado o comportamento das duas redes com melhor desempenho no MNIST (AAE e GAN) na síntese de imagens coloridas da base LSUN-bedrooms. Para o treinamento foi utilizado 5% do *dataset* original amostrado aleatoriamente, totalizando 154.841 imagens redimensionadas para 64x64 pixels. As estruturas utilizadas foram semelhantes àquelas da primeira etapa, com pequenas modificações devido à nova dimensão e à inserção de canais adicionais (imagens coloridas). No caso da GAN também foram incorporadas outras técnicas capazes de otimizar o seu desempenho, como o *minibatch discrimination* [7].

Por fim, a etapa final consistiu na aplicação da rede com o melhor desempenho no LSUN-bedroom, a GAN, na geração de sinais de EEG para o projeto de uma BCI-SSVEP. O banco de dados disponível no grupo de pesquisa (CAAE 0617.0.146.000-10) consiste de amostras de 32 pessoas coletadas ao longo de 8 sessões e avaliando 4 frequências de excitação visual (6 Hz, 10 Hz, 12 Hz e 15 Hz) por 12s, com uma taxa de amostragem de 256 Hz. Nos experimentos, consideramos somente os registros do eletrodo Oz para os dez primeiros indivíduos nas frequências de 12 Hz e 15 Hz. Os dados foram divididos em janelas de 3s, com sobreposição de 2s, totalizando 60 padrões para treinamento e 20 de validação por frequência e por indivíduo.

Os testes consideraram como métrica a acurácia dos classificadores de BCI escolhidos: SVM e uma CNN com camadas convolucionais 1d [8]. Os experimentos consideraram o uso dos dados sintéticos em dois domínios: (1) temporal, com classificação e síntese no tempo e (2) frequência, com classificação e geração a partir da transformada de Fourier do sinal.

## Resultados e Discussão

A Figura 1 exibe amostras sintetizadas pelos três modelos após o treinamento no banco de dados MNIST. É possível constatar que a GAN conseguiu gerar traços com maior definição, porém um pouco distorcidos, enquanto as redes baseadas em modelos auto-codificadores geraram dígitos mais ruidosos, embora mais fiéis aos contornos originais.

**Figura 1** – Comparação de modelos: MNIST



Fonte: compilação do autor.

A segunda etapa de análise avaliou os modelos no LSUN-bedroom, cujo resultado é resumido na Figura 2, que apresenta amostras geradas pelos modelos. É fácil ver que a GAN obteve um desempenho muito superior ao AAE, que não foi capaz de aprender as características dos quartos de hotéis. Além disso, pode-se concluir que a GAN obteve um bom desempenho quando comparado a [9], no qual foi utilizado o banco de dados completo. É válido enfatizar que diferentemente de [9], também foi utilizado na estrutura da GAN o *minibatch discrimination* [7], cuja ideia busca retardar o colapso de modo do gerador. Para isso, é criada uma projeção matricial de vários atributos extraídos, de forma que a rede discriminante analise várias amostras de uma só vez ao invés de uma amostra separada.

**Figura 2** - Comparação de modelos: LSUN - bedroom



Fonte: compilação do autor.

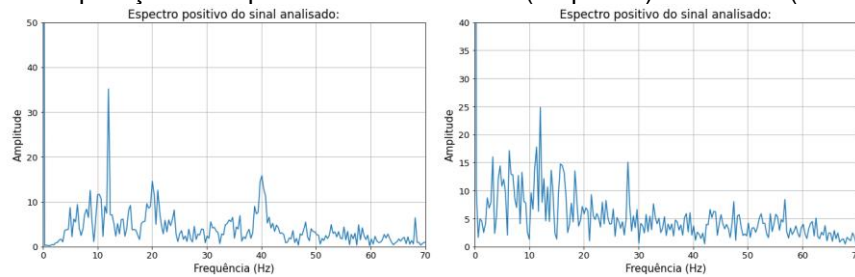
Por fim, a pesquisa realizou experimentos com a GAN nos dois domínios – tempo e frequência – e avaliou o desempenho dos classificadores. A Tabela 1 traz a taxa de acerto dos classificadores quando treinados com e sem a presença de dados sintéticos, adicionando 50, 100 e 150 amostras por frequência. O maior destaque foi a CNN com DA, que apresentou um aumento de precisão em 9 dos 10 casos testados. Os melhores desempenhos para cada classificador estão destacados em negrito.

**Tabela 1: Desempenho da BCI no tempo**

Suj.	SVM Poly	SVM Linear	SVM Gau.	CNN	Poly + 50	Poly + 100	Poly + 150	Lin. + 50	Lin. + 100	Lin. + 150	Gau. + 50	Gau. + 100	Gau. + 150	CNN + 50	CNN + 100	CNN + 150
0	0.6000	0.5000	0.5750	0.600	0.6000	0.6000	0.6250	0.4750	0.4750	0.4500	0.6000	0.6250	0.6000	0.550	<b>0.650</b>	0.625
1	0.6500	0.6000	0.6250	0.675	0.6500	0.7000	0.7000	0.6750	0.6750	0.6500	0.6250	0.6250	0.6250	0.675	<b>0.725</b>	0.700
2	0.6000	0.4500	0.5500	0.575	0.6000	0.5750	0.6000	0.4500	0.4500	0.4750	0.6000	0.6000	0.6000	0.575	<b>0.650</b>	0.650
3	0.6750	0.6000	0.6750	0.600	0.6500	0.6500	0.5250	0.6250	0.6750	0.7000	0.6750	0.6500	0.6750	0.575	<b>0.725</b>	<b>0.725</b>
4	0.6000	0.5000	0.5000	0.625	0.6000	0.5750	0.5750	0.6000	0.6000	0.5500	0.6000	0.6000	0.5750	<b>0.675</b>	0.625	0.575
5	0.7250	0.6500	0.7250	0.700	0.7000	0.7000	0.7000	0.6500	0.6500	0.6500	0.6750	0.6500	0.6500	0.600	0.625	<b>0.750</b>
6	0.7000	0.7250	0.6000	0.700	0.7000	0.6750	0.6500	0.7750	0.7500	0.7500	0.6000	0.5250	0.5500	0.750	<b>0.775</b>	0.750
7	0.9750	0.9750	0.9250	<b>0.975</b>	0.9750	0.9750	0.9750	0.9750	0.9500	0.9500	0.9500	0.9750	0.9250	0.975	0.975	0.975
8	0.5500	0.5000	0.4500	0.525	<b>0.6500</b>	0.6000	0.6250	0.4750	0.4750	0.4500	0.5000	0.5250	0.5000	0.500	0.600	0.525
9	0.5750	0.4500	0.6000	0.525	0.6250	0.6250	0.6250	0.4250	0.4500	0.4500	0.6000	0.6250	<b>0.6500</b>	0.525	0.550	0.550

Fonte: compilado do autor.

Por sua vez, a Figura 3 mostra uma comparação do espectro de uma amostra real com um dado fictício gerado pela GAN treinada no tempo para 12 Hz. É possível observar que existe um pico de energia em torno da frequência de 12 Hz no espectro do dado sintético e que, de certo modo, ele segue a tendência exibida no espectro real, ainda que o dado sintético apresente maior ruído.

**Figura 3: Comparação do espectro de um sinal real (esquerda) e sintético (direita) no tempo**

Fonte: compilado do autor.

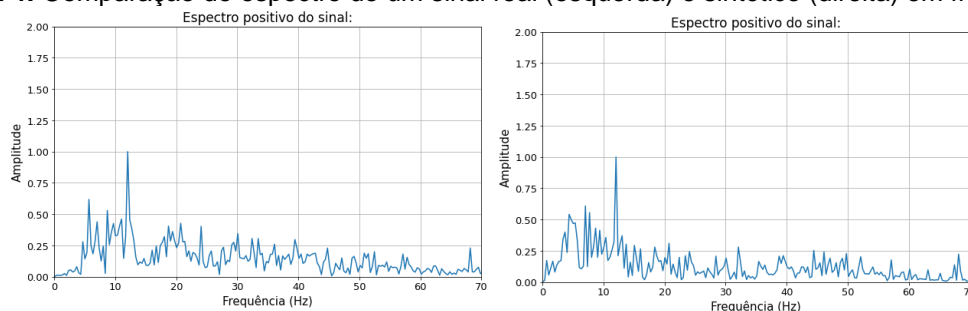
Na Tabela 2 apresentamos os resultados com DA em frequência, i.e, com a GAN e os classificadores treinados a partir da transformada de Fourier dos sinais. Pode-se observar que o desempenho dos classificadores novamente foi superior com a inserção dos dados sintéticos, com ênfase para a CNN com melhor aproveitamento. Além disso, é possível observar que o treinamento da BCI e da GAN em frequência trouxe ganho superiores do que aqueles obtidos no domínio do tempo. É pertinente mencionar que o trabalho [8] já havia realizado experimentos que mostraram um bom aproveitamento dos modelos generativos no domínio do tempo. Não obstante, esta pesquisa conclui que o apelo natural do problema para explorar informação no domínio da frequência é confirmado pelo melhor desempenho da GAN e da BCI-SSVEP quando os espectros são diretamente sintetizados e usados como entrada do classificador.

**Tabela 2: Desempenho da BCI em frequência**

Suj.	SVM Poly	SVM Linear	SVM Gau.	CNN	Poly + 50	Poly + 100	Poly + 150	Lin. + 50	Lin. + 100	Lin. + 150	Gau. + 50	Gau. + 100	Gau. + 150	CNN + 50	CNN + 100	CNN + 150
0	0.9000	0.9000	0.8750	1.0000	0.8500	0.8500	0.8250	0.9000	0.8750	0.8500	0.8500	0.7750	0.7000	<b>1.0000</b>	1.0000	1.0000
1	0.5750	0.6250	0.5250	0.6750	0.5750	0.5750	0.5000	0.5750	0.5500	0.5250	0.5250	0.5250	0.5250	0.6500	<b>0.7250</b>	0.7000
2	0.6250	0.6500	0.6250	0.7750	0.6250	0.6000	0.6250	0.6500	0.6250	0.6000	0.6000	0.6000	0.6000	<b>0.8750</b>	0.7750	0.7750
3	0.5750	0.5750	0.5500	<b>0.6500</b>	0.5000	0.5250	0.5500	0.5000	0.4500	0.4500	0.5250	0.5250	0.5250	0.5250	0.6500	0.6250
4	0.5250	0.5000	0.4750	0.5750	0.5750	0.5000	0.5250	0.5000	0.4500	0.4750	0.5750	0.5250	0.5250	0.5000	0.5750	<b>0.6500</b>
5	0.7750	0.9250	0.8500	0.9750	0.8250	0.8000	0.8250	0.9750	0.9500	0.9500	0.8500	0.8250	0.7500	<b>1.0000</b>	1.0000	1.0000
6	0.6250	0.7750	0.6250	<b>0.8250</b>	0.6750	0.6250	0.7250	0.7000	0.6750	0.7000	0.5750	0.6500	0.5750	0.7750	0.8000	0.8000
7	0.8750	0.9250	0.8750	0.9500	0.8750	0.8500	0.8500	0.9250	0.9250	0.9250	0.8250	0.8000	0.8250	0.9500	0.9500	<b>1.0000</b>
8	0.5500	0.4750	0.5000	<b>0.6000</b>	0.5250	0.5250	0.5250	0.5000	0.5250	0.5250	0.5250	0.5250	0.5000	0.5250	0.5750	0.5750
9	0.9750	0.9750	0.9750	1.0000	0.9750	0.9750	0.9500	0.9750	0.9750	0.9750	0.9750	0.9500	0.9250	<b>1.0000</b>	1.0000	1.0000

Fonte: compilado do autor.

Por fim, a Figura 4 ilustra uma comparação do espectro de um dado real e um dado sintético com a GAN treinada em frequência, na qual é possível observar uma grande redução dos ruídos apresentados na versão treinada no tempo.

**Figura 4: Comparação do espectro de um sinal real (esquerda) e sintético (direita) em frequência**

Fonte: compilado do autor.

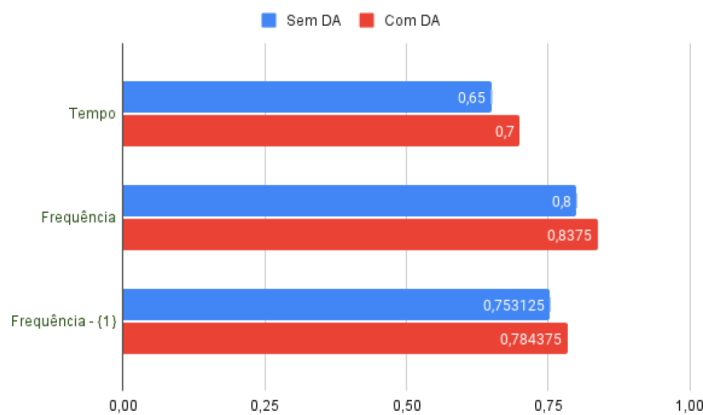
## Conclusões

Através dos experimentos realizados neste trabalho, i.e, a aplicação de três modelos neurais generativos na tarefa de síntese de dados para DA, foi possível investigar a capacidade de cada estrutura de aprender a reproduzir as características dos dados de treinamento em novas amostras válidas. Para o banco de dados MNIST, todos os modelos conseguiram um desempenho satisfatório, todavia, com o aumento da complexidade dos dados de treinamento, ao por exemplo, adicionar mais canais de entrada para imagens coloridas (base LSUN-bedroom) ou ao utilizar sinais de EEG, apenas a GAN extraiu as características adequadamente.

Diante disso, conclui-se que as redes generativas se destacam como uma forma de DA que foi capaz de aumentar a precisão dos classificadores no sistema BCI-SSVEP em até 10% para o domínio da frequência, e de 7,25% para o domínio do tempo. Um resumo dos principais resultados se encontra na Figura 5, que mostra o desempenho médio da BCI-SSVEP obtido com o melhor classificador e para dados em ambos os domínios, além da acurácia média (para a frequência) excluindo os casos em que a acurácia já era 100% antes do DA.

É possível notar que DA via GAN trouxe benefícios para o sistema BCI-SSVEP operando em ambos os domínios (tempo e frequência). Além disso, vemos que embora no treinamento de redes neurais seja amplamente difundido o uso de dados brutos, ao utilizar o espectro dos sinais de EEG em vez das janelas temporais, a qualidade dos dados sintéticos e a precisão dos classificadores foram superiores (83,75% vs. 70%), evidenciando que o pré-processamento baseado no apelo natural do problema é, de fato, útil para o sistema.

**Figura 5:** Média da precisão de classificação para as frequências de 12 Hz e 15 Hz.



Fonte: compilado do autor.

## Referências bibliográficas

- [1] A. Géron, Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow, 2nd ed. O'Reilly Media, 2019
- [2] E. Lashgari, D. Liang, and U. Maoz, "Data augmentation for deep-learning-based electroencephalography," *Journal of Neuroscience Methods*, p. 108885, 2020.
- [3] S. N. d. C. LEITE, "Contribuições ao desenvolvimento de interfaces cérebro-computador baseadas em potenciais evocados visualmente em regime estacionário," Ph.D. dissertation, Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação, 2016.
- [4] B. Graimann, G. Pfurtscheller, and B. Allison, *Brain-Computer Interfaces: Revolutionizing Human-Computer Interaction*. Springer, 2010.
- [5] C. Doersch, "Tutorial on variational autoencoders," arXiv preprint arXiv:1606.05908, 2016.
- [6] A. Makhzani, J. Shlens, N. Jaitly, and I. J. Goodfellow, "Adversarial autoencoders," *CoRR*, vol. abs/1511.05644, 2015. [Online]. Available: <http://arxiv.org/abs/1511.05644>
- [7] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training gans," arXiv preprint arXiv:1606.03498, 2016.
- [8] N. K. N. Aznan, A. Atapour-Abarghouei, S. Bonner, J. D. Connolly, N. Al Moubayed, and T. P. Breckon, "Simulating brain signals: Creating synthetic EEG data via neural-based generative models for improved SSVEP classification," in 2019 International Joint Conference on Neural Networks (IJCNN). IEEE, 2019, pp. 1–8.
- [9] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," arXiv preprint arXiv:1511.06434, 2015.