

Título em Português: Desenvolvendo Redes Neurais Artificiais com estrutura de Redes Complexas

Título em Inglês: Developing Artificial Neural Networks with Complex Networks structure

Autor: Paulo Ricardo Sturion

Instituição: Universidade de São Paulo

Unidade: Instituto de Física de São Carlos

Orientador: Odemir Martinez Bruno

Área de Pesquisa / SubÁrea: Metodologia e Técnicas da Computação

Agência Financiadora: Sem fomento

Desenvolvendo Redes Neurais Artificiais com Estrutura de Redes Complexas

Autor: Paulo Ricardo Sturion

Orientador: Odemir Martinez Bruno

Instituto de Física de São Carlos - Universidade de São Paulo

prsturion@usp.br

Objetivos

Este trabalho investiga a aplicação de estruturas de conexão inspiradas no modelo Watts–Strogatz (WS) [2] de redes complexas pequeno-mundo em Redes Neurais MLP como uma aprimoração e extensão do trabalho [3]. Experimentos controlados tiveram como objetivos: (i) comparar o desempenho da arquitetura proposta com a MLP tradicional e (ii) analisar a influência de seus hiperparâmetros. A principal contribuição é a proposição de técnicas que aumentam a eficiência e a plausibilidade biológica [2] das ANNs, fomentando avanços em inteligência artificial e computação neural.

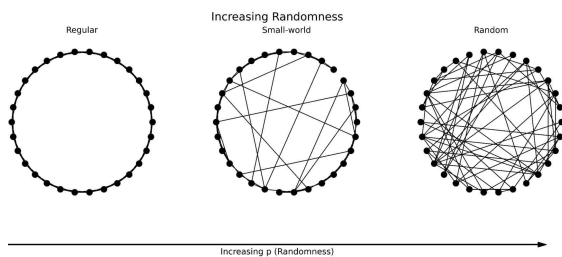


Figura 1: Modelo de rede Watts-Strogatz.

Métodos e Procedimentos

• **Preparação dos Dados:** foram utilizados os *benchmarks* de visão computacional MNIST e Fashion-MNIST, escolhidos por sua ampla

adoção e pela capacidade de capturar a não-linearidade das ANNs.

• **Desenvolvimento teórico e implementação:** Adaptou-se o modelo WS para MLPs de forma compatível com *backpropagation*. No WS, conexões de uma rede regular são remanejadas com probabilidade p [2], criando atalhos que reduzem o caminho mínimo médio entre nós (fig. 1). O efeito pequeno-mundo surge para p intermediário. Supôs-se que esse efeito poderia melhorar a performance da rede. O modelo foi, portanto, ajustado para permitir conexões entre camadas não-adjacentes, mantendo cada neurônio ligado apenas a $(2k + 1)$ vizinhos da camada seguinte, com k configurável (fig. 2).

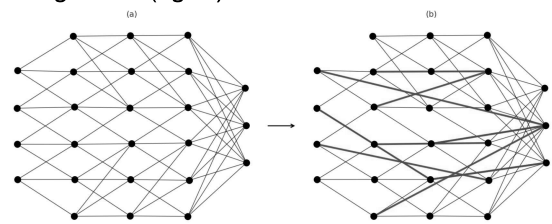


Figura 2: Modelo de rede proposto (a) antes do *rewiring*; (b) pós *rewiring*.

Para incluir as conexões entre camadas não-adjacentes, o *forwardpass* foi modificado segundo:

$$X_n = \phi \left[\sum_{i=0}^{n-1} X_i (A_{i,n}^T \odot W_{i,n}^T) + B_n \right] \quad (1)$$

onde A é a máscara que indica as conexões ativas entre camadas, filtrando os pesos de W pelo produto de Hadamard.

• **Realização dos experimentos:** Foram realizados experimentos de treinamento e teste com um *ensemble* de 100 *seeds* aleatórias, a fim de obter dados que relacionam: (i) estrutura de conexão e desempenho da rede; (ii) parâmetros do modelo e sua eficácia.

Resultados

Na figura 3 e na tabela 1, estão os resultados obtidos e suas respectivas descrições na legenda. Gráficos similares foram obtidos para Fashion-MNIST.

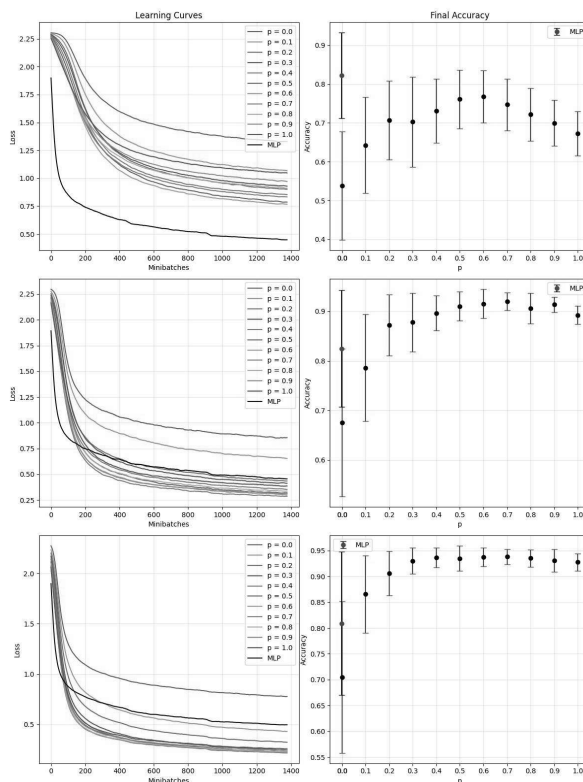


Figura 3: Curva de aprendizado e acurácia final no *dataset* MNIST com diferentes valores de p para $k = 1, 3, 5$, respectivamente.

Tabela 1: Acurácia final máxima do modelo WS em comparação com MLP para diferentes profundidades de redes ($k = 3$).

	MNIST		Fashion-MNIST	
<i>hidden</i>	WS (%)	MLP (%)	WS (%)	MLP (%)
64	90 ± 5	91 ± 7	82 ± 3	81 ± 7
64,32	91 ± 3	86 ± 11	83.0 ± 1.8	79 ± 8
64,32,32	92 ± 2	82 ± 12	83.2 ± 1.2	74 ± 12
64,32,32,32	91.8 ± 1.5	75 ± 15	83.5 ± 1.2	70 ± 12

Conclusões

Na figura, o efeito pequeno-mundo melhora a acurácia da rede em valores intermediários de p (cf. $k = 1, k = 3$); para k maiores, o efeito satura (cf. $k = 5$). Assim, recomenda-se ajuste fino de k e p para maximizar desempenho com menos conexões.

A tabela mostra que cada modelo se sai melhor em uma profundidade de rede distinta. A rede WS mais precisa superou a MLP de melhor desempenho, usando apenas ~5,4% das conexões, em ambos os *datasets*. Além disso, apresentou menor desvio padrão no *ensemble*, revelando maior robustez à inicialização.

Conclui-se que o modelo WS supera as MLPs tradicionais, oferecendo melhor acurácia e confiabilidade com menor número de conexões.

O autor declara não haver conflito de interesses.

Referências

[1] LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *Nature*, Londres, v. 521, n. 7553, p. 436–444, 28 maio 2015. DOI: 10.1038/nature14539.

[2] WATTS, D. J.; STROGATZ, S. H. Collective dynamics of ‘small-world’ networks. *Nature*, Londres, v. 393, n. 6684, p. 440–442, 1998. DOI: 10.1038/30918.

[3] SIMARD, D.; NADEAU, L.; KRÖGER, H. Fastest learning in small-world neural networks. *Physics Letters A*, v. 336, n. 1, p. 8–15, 2005.

Developing Artificial Neural Networks with Complex Network Structure

Author: Paulo Ricardo Sturion

Supervisor: Odemir Martinez Bruno

Instituto de Física de São Carlos - Universidade de São Paulo

prsturion@usp.br

Objectives

This work investigates the application of connection structures inspired by the Watts–Strogatz (WS) [2] small-world complex network model in MLP Neural Networks as an improvement and extension of the work [3]. Controlled experiments aimed to: (i) compare the performance of the proposed architecture with the traditional MLP and (ii) analyze the influence of its hyperparameters. The main contribution is the proposition of techniques that increase the efficiency and biological plausibility [2] of ANNs, fostering advances in artificial intelligence and neural computing.

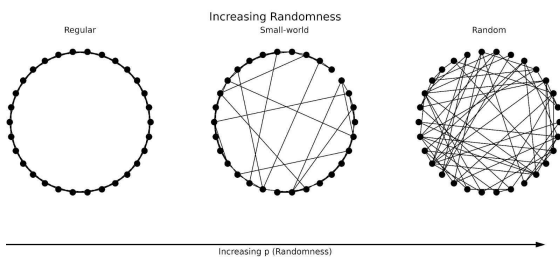


Figure 1: Watts-Strogatz network model.

Methods

- **Data Preparation:** the MNIST and Fashion-MNIST computer vision benchmarks were used, chosen for their wide adoption and their ability to capture the non-linearity of ANNs.

- **Theoretical development and implementation:** The WS model was adapted for MLPs in a way compatible with backpropagation. In WS, connections of a regular network are rewired with probability p [2], creating shortcuts that reduce the average shortest path between nodes (Fig. 1). The small-world effect arises for intermediate p . It was assumed that this effect could improve the network's performance. The model was, therefore, adjusted to allow connections between non-adjacent layers, while keeping each neuron connected only to $(2k + 1)$ neighbors of the next layer, with k configurable (Fig. 2).

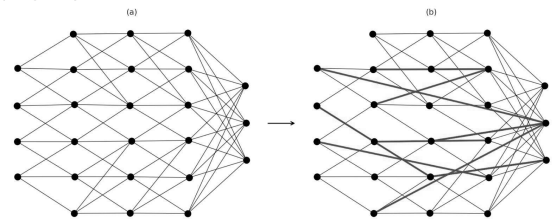


Figure 2: Proposed network model (a) before rewiring; (b) after rewiring.

To include the connections between non-adjacent layers, the forward pass was modified according to:

$$X_n = \phi \left[\sum_{i=0}^{n-1} X_i (A_{i,n}^T \odot W_{i,n}^T) + B_n \right] \quad (1)$$

where A is the mask that indicates the active connections between layers, filtering the weights of W through the Hadamard product.

• **Execution of experiments:** Training and testing experiments were carried out with an ensemble of 100 random seeds, in order to obtain data relating: (i) connection structure and network performance; (ii) model parameters and their effectiveness.

Results

In Figure 3 and Table 1, the results obtained and their respective descriptions in the caption are presented. Similar plots were obtained for Fashion-MNIST.

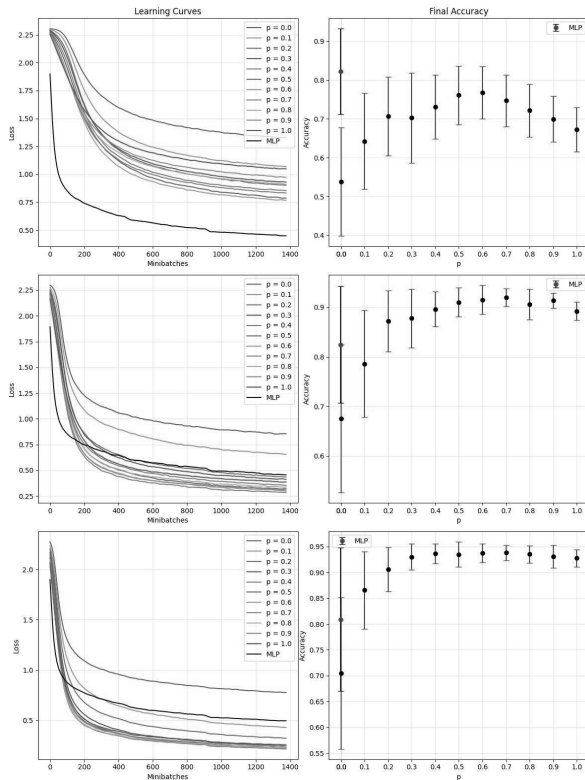


Figure 3: Learning curve and final accuracy on the MNIST dataset with different values of p for $k = 1, 3, 5$, respectively.

Table 1: Maximum final accuracy of WS compared with MLP for different network depths ($k = 3$).

	MNIST		Fashion-MNIST	
<i>hidden</i>	WS (%)	MLP (%)	WS (%)	MLP (%)
64	90 ± 5	91 ± 7	82 ± 3	81 ± 7
64,32	91 ± 3	86 ± 11	83.0 ± 1.8	79 ± 8
64,32,32	92 ± 2	82 ± 12	83.2 ± 1.2	74 ± 12
64,32,32,32	91.8 ± 1.5	75 ± 15	83.5 ± 1.2	70 ± 12

Conclusions

In the figure, the small-world effect improves network accuracy at intermediate values of p (cf. $k = 1$, $k = 3$); for larger k , the effect saturates (cf. $k = 5$). Thus, fine-tuning of k and p is recommended to maximize performance with fewer connections.

The table shows that each model performs better at a different network depth. The most accurate WS network outperformed the best-performing MLP while using only ~5.4% of the connections, in both datasets. In addition, it presented a lower standard deviation in the ensemble, revealing greater robustness to initialization.

It is concluded that the WS model outperforms traditional MLPs, offering better accuracy and reliability with a smaller number of connections.

The author declares no conflict of interest.

Referências

[1] LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. Nature, Londres, v. 521, n. 7553, p. 436–444, 28 maio 2015. DOI: 10.1038/nature14539.

[2] WATTS, D. J.; STROGATZ, S. H. Collective dynamics of ‘small-world’ networks. Nature, Londres, v. 393, n. 6684, p. 440–442, 1998. DOI: 10.1038/30918.

[3] SIMARD, D.; NADEAU, L.; KRÖGER, H. Fastest learning in small-world neural networks. Physics Letters A, v. 336, n. 1, p. 8–15, 2005.