

**Modalidade do trabalho:** Relatório técnico-científico  
**Evento:** XXIII Seminário de Iniciação Científica

## **IMPLEMENTAÇÃO PARALELA DE UMA ARQUITETURA DE REDE NEURAL ARTIFICIAL DISTRIBUÍDA E REALIMENTADA<sup>1</sup>**

**Jessica De Almeida Berlezi<sup>2</sup>, Eldair Fabricio Dorneles<sup>3</sup>, Edson Luiz Padoin<sup>4</sup>, Rogério S. M. Martins<sup>5</sup>.**

<sup>1</sup> Trabalho realizado no contexto do projeto Sistema de Decisão Adaptativo Baseado em Redes Neurais para Uma Arquitetura Aplicada a Agricultura de Precisão.

<sup>2</sup> Bolsista PIBIC/Unijuí, jessica.berlezi@gmail.com

<sup>3</sup> Bolsista PROBIC/FAPERGS, fabriciud@gmail.com

<sup>4</sup> Professor Orientador, padoin@unijui.edu.br

<sup>5</sup> Professor Pesquisador, rogerio.martins@unijui.edu.br

### **1. Introdução**

A maior parte das funções do nosso corpo são controladas pelo sistema nervoso central (SNC). Para fazer isso ele recebe informações do ambiente, seja com os receptores visuais nos olhos, receptores auditivos nos ouvidores, ou receptores de outros tipos, e informações do próprio corpo, pelos receptores táteis na superfície da pele. Após recebê-las, o SNC transmite, armazena, altera e utiliza estas informações (GUYTON, HALL, 2011).

Este processamento é regido por elementos processadores biológicos que operam em paralelo tendo como objetivo a produção de ações apropriadas para cada uma de suas funcionalidades (DA SILVA, SPATTI, FLAUZINO, 2010). A ideia de construir sistemas considerados inteligentes e as potencialidades de tais sistemas tornasse atrativa ao observar que computadores possuem muitas características em comum com o sistema nervoso dos seres vivos (GUYTON, HALL, 2011). Entre estes sistemas as Redes Neurais Artificiais (RNA), que permitem a resolução de problemas advindos de diferentes áreas do conhecimento.

Por mais poderoso que possa ser o último processador produzido, e da razão de crescimento do desempenho computacional promovido pela indústria, algumas aplicações demandam de maiores recursos computacionais. Dentre elas, destaca-se a área de inteligência artificial, incluindo redes neurais, robótica e reconhecimento de padrões (YAMIN, 2006). Nesse contexto, para reduzir o tempo de execução, a técnica de paralelização tem sido empregada.

O objetivo deste trabalho é desenvolver uma arquitetura de RNA do tipo Recorrente (RNAR), ou realimentada, na forma paralela. A escolha desta arquitetura se deve a sua característica de realimentação. Na realimentação, as saídas dos neurônios são realimentadas como sinais de entrada para outros neurônios (DA SILVA, SPATTI, FLAUZINO, 2010), em nosso caso, para toda a rede.

**Modalidade do trabalho:** Relatório técnico-científico  
**Evento:** XXIII Seminário de Iniciação Científica

A realimentação garante variação temporal do sistema e posterior inserção de características de aprendizagem, como o uso de Algoritmo Genético (escolha do melhor adaptado), por exemplo.

## 2. Metodologia

A realização desta pesquisa está dividida em três etapas. A primeira consiste na implementação da arquitetura da RNA Recorrente com processamento sequencial, onde é proposto um modelo de neurônio artificial (nodo) similar em estrutura e funcionamento, ao neurônio biológico. A segunda trata da paralelização da implementação da RNA. Por fim, na terceira etapa analisa-se o ganho obtido com a paralelização da RNA.

As RNAs possuem a capacidade de aquisição e manutenção do conhecimento (baseado em informações). Elas podem ser definidas como um conjunto de unidades de processamento, caracterizadas por neurônios artificiais, que são interligados por um grande número de interconexões (DA SILVA, SPATTI, FLAUZINO, 2010). Esta estrutura foi construída por vetores e matrizes de pesos sinápticos desenvolvidos em linguagem de programação C. O sistema operacional utilizado foi Ubuntu 14.04.2 LTS com Kernel versão 3.13.0-53-generic. O equipamento possui processador Intel x64 Pentium T4200 Dual Core com tecnologia HT, o qual possibilita a execução de 4 threads utilizando os 4GB de memória RAM.

Para a paralelização do algoritmo RNAR sequencial foi utilizada a tecnologia openMP. OpenMP fornece diretivas, funções de biblioteca e variáveis de ambiente que permitem criar e controlar a execução de programas paralelos (CHAPMAN, JOST, VAN DER PAS, 2008).

## 3. Resultados e Discussões

Nesta seção apresenta-se a implementação, diagramação e formulação da RNA recorrente, bem como a técnica de paralelização utilizada para o desenvolvimento do algoritmo paralelo. Também são apresentados os resultados alcançados com o estudo sobre o SNC biológico.

Muitas atividades do sistema nervoso se iniciam pelas experiências sensoriais que excitam os receptores sensoriais. Estas informações são processadas e geram respostas mentais e motoras apropriadas. Mais de 99% de toda a informação sensorial é descartada pelo cérebro como irrelevante e sem importância (GUYTON, HALL, 2011).

O cérebro humano, uma rede neural, é composto de neurônios. Os neurônios biológicos são divididos em três seções inter-relacionadas: corpo da célula, dendritos e axônio. Os dendritos recebem os impulsos nervosos (informações) de outros neurônios, transmitindo-os até o corpo da célula. Em seguida, a informação é transformada em novo impulso, o qual é transmitido a outros neurônios através do axônio. A conexão feita entre o axônio de um neurônio e o dendrito de outro neurônio é denominada "sinapse". As sinapses formam as ligações entre os neurônios, compondo

**Modalidade do trabalho:** Relatório técnico-científico  
**Evento:** XXIII Seminário de Iniciação Científica

assim as redes neurais e funcionando como válvulas capazes de controlar a transmissão de impulsos (fluxo de informação) entre os neurônios na rede neural. O efeito da sinapse sobre o sinal que trafega através dela é ajustado conforme a experiência, implementando a capacidade de adaptação de toda rede neural (DE MELO, 2006).

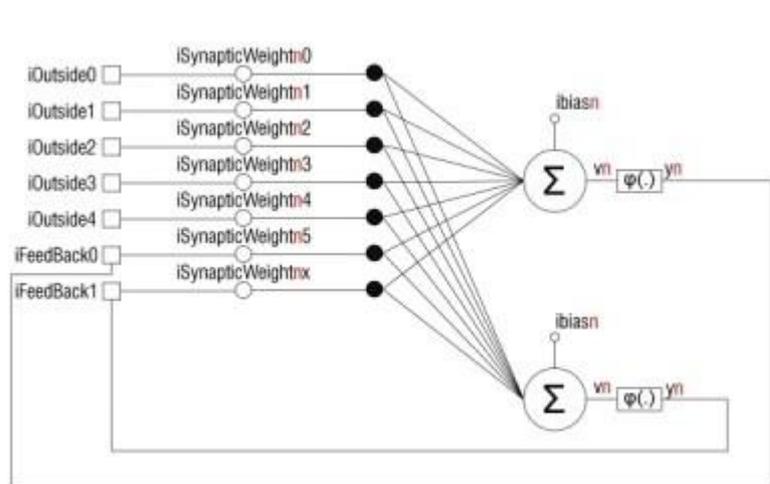


Figura 1. Rede Neural Artificial Recorrente com dois Neurônios. Fonte: Próprio Autor.

Os neurônios artificiais, analogamente ao que ocorre no cérebro, são interconectados, formando a rede neural artificial, como representado na Figura 1. Assim, é possível criar estruturas para generalização de padrões através de variadas entradas, produzindo uma ou mais saídas que podem representar uma ação ou objeto do mundo real (padrão) como resposta aos diferentes dados apresentados na entrada.

A representação matemática do modelo representado na Figura 1, é apresentada na equação:

$$v_n = \sum_{i=0}^x iSynapticWeight_{ni} * iOutside_i + iBias_n$$

onde n representa qual neurônio está sendo processado e x a iteração atual sobre este neurônio dada as suas respectivas entradas. Isto em uma dada mutação m, ou seja, o número de variações temporais da RNA.

Conforme observação do esquema da Figura 1, RNAs com arquitetura recorrente possuem suas saídas realimentadas como sinais de entrada para outros neurônios. A característica da

**Modalidade do trabalho:** Relatório técnico-científico  
**Evento:** XXIII Seminário de Iniciação Científica

realimentação qualifica tais redes para processamento dinâmico de informações, isto é, estas podem ser utilizadas em sistemas variantes em relação ao tempo (DA SILVA, SPATTI, FLAUZINO, 2010).

Assim, a saída produzida atual leva também em consideração as saídas anteriores. Este comportamento pode ser melhor visto na saída produzida na junção aditiva, vn, onde as entradas de retorno da rede (iFeedBacki), são estendidas a fórmula como as demais entradas, sendo apenas representativas.

Em (DA SILVA, SPATTI, FLAUZINO, 2010) são listadas as características da realimentação em uma RNAR:

- Comportamento tipicamente dinâmico;
- Capacidade de memorizar relacionamentos;
- Possibilidade de armazenamento de informações; e
- Facilidade de implementação em hardware análogo.

Os fenômenos naturais são inerentemente paralelos. Escrever programas sequenciais implica impor uma ordem às ações que são potencialmente independentes e que poderiam executar concorrentemente (YAMIN, 2006). Como na programação sequencial é inevitável arbitrar uma ordem particular na qual as ações serão colocadas, é necessário encontrar trechos de código em que a ordem de execução praticamente não importe para o entendimento e funcionamento do problema real (YAMIN, 2006).

Deste modo, cada neurônio possui seu processamento independente dos demais em uma mutação. Posteriormente seus resultados são unidos e uma nova mutação pode ser iniciada, considerando os resultados anteriores. Logo, o processamento individual de cada neurônio em uma mutação, são exemplos onde a ordem da execução praticamente não importa e as técnicas de paralelização podem ser empregadas.

Pesos Sinápticos	Neurônios	Tempo Sequencial (ms)	Tempo openMP(2th) (ms)	Speed-up (%)
300	150	3,73	2,12	56,8
600	300	14,90	8,97	60,2
1200	600	58,50	34,77	59,4
2400	1200	230,94	155,17	67,2
4800	2400	935,58	483,13	51,6
9600	4800	3683,55	2005,05	54,4
19200	9600	14938,93	7613,07	51,0
38400	19200	66077,01	38063,63	57,6

**Modalidade do trabalho:** Relatório técnico-científico  
**Evento:** XXIII Seminário de Iniciação Científica

Na Tabela 1 são apresentados os tempos de execução mensurados na execução dos algoritmos sequencial e paralelo com a API openMP. Com a execução do algoritmo paralelo obteve-se uma redução significativa do tempo de execução em todos os testes. O Speed-up representa o ganho de desempenho alcançado com a paralelização da versão sequencial. Ele é calculado pela razão entre o melhor tempo sequencial e o tempo da versão paralela (DE ROSE, 2006). Para todos pesos sinápticos e neurônios testados o ganho foi acima de 50% com a utilização dos cores disponíveis no equipamento.

#### 4. Conclusões e trabalhos futuros

A partir dos resultados alcançados com os testes experimentais, observa-se que dobrando-se o número de neurônios da RNA, e dobrando-se os pesos sinápticos, obtêm-se um aumento de 4 vezes no tempo de execução. A Figura 2, relaciona o tamanho do problema com o tempo de execução (log) mensurado nos testes realizados. A partir destes dados percebe-se que o problema apresenta um crescimento linear, para as condições utilizadas, na RNA.

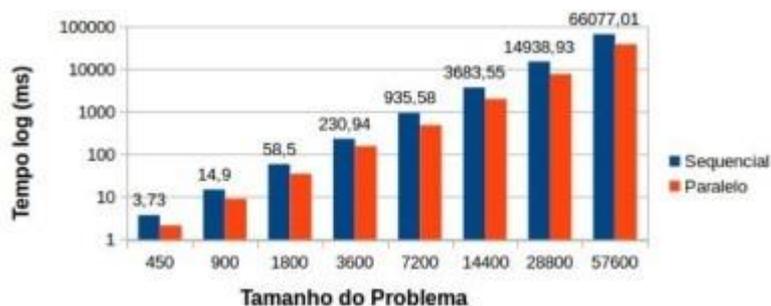


Figura 2. Ordem do Tamanho do Problema. Fonte: Próprio Autor.

Como trabalhos futuros pretende-se utilizar novas técnicas de paralelização do algoritmo e realizar novos testes utilizando placas de vídeo da NVIDIA com a tecnologia CUDA. Concluída esta etapa, pretende-se utilizar dados reais do projeto “Sistema de Decisão Adaptativo Baseado em Redes Neurais para Uma Arquitetura Aplicada a Agricultura de Precisão”, nos testes para validação do modelo.

5. Palavras-chave: Paralelismo; Inteligência Artificial; Sistemas Inteligentes; openMP; Linguagem de Programação C.

6. Agradecimentos

**Modalidade do trabalho:** Relatório técnico-científico

**Evento:** XXIII Seminário de Iniciação Científica

Agradeço a UNIJUI pela concessão da bolsa através do Programa Institucional de Bolsas de Iniciação Científica (PIBIC) que permitiu o desenvolvimento desta pesquisa.

## 7. Referências Bibliográficas

CHAPMAN, Barbara; JOST, Gabriele; VAN DER PAS, Ruud. Using OpenMP: portable shared memory parallel programming. MIT press, 2008.

DA SILVA, Ivan Nunes; SPATTI, Danilo Hernane; FLAUZINO, Rogério Andrade. Redes Neurais Artificiais para engenharia e ciências aplicadas curso prático. Artliber, 2010.

DE MELO, Francisco Ramos et al. Generalizador Neural de Espaços de Aprendizagem em Sistemas Tutores Inteligentes. In: Anais do Simpósio Brasileiro de Informática na Educação. 2006. p. 141-150.

DE ROSE, César A. F. Fundamentos de Processamento de Alto Desempenho. Ijuí: ERAD Escola Regional de Alto Desempenho, 2006.

GUYTON, Arthur Clifton; HALL, John E. Guyton & Hall - Tratado de fisiologia médica. Elsevier Brasil, 2011.

YAMIN, Adenauer Corrêa. Reflexões sobre aspectos que promovem a adoção do paralelismo. Ijuí: ERAD- Escola Regional de Alto Desempenho, 2006.