

USO DE REDES GSN^{ff} PARA PONDERAÇÃO DE VARIÁVEIS DE ENTRADA EM CLASSIFICADORES

Weber Martins^{1,2}, Luciano de Souza Miguel¹
weber@eee.ufg.br, lucianoz@uol.com.br

¹ Universidade Federal de Goiás, Brasil
Escola de Eng. Elétrica e de Computação – Grupo PIRENEUS
UFG – EEEF – Praça Universitária s/n – Goiânia – Goiás – Brazil

² Universidade Católica de Goiás, Brasil
Departamento de Psicologia
Av. Universitária, 1440, Setor Universitário – Goiânia – Goiás – Brazil

Abstract

This paper presents the novel use of networks of Goal Seeking Neurons (GSN) on the evaluation of the relative importance of dimensions of multidimensional data. Therefore, it is possible to measure the role of each input variable in the classification of patterns defined by a sample. The best performances have been observed at smaller topologies. This fact, combined with the one-shot training algorithm (where each pattern is employed just once), leads to an efficient system. Compared with the standard classifier, the levels of correct answers and saturation have been improved. The decreasing of saturation level creates more flexibility to learn secondary rules (more space to new internal mappings). In order to evaluate the proposed system more thoroughly, we have chosen the prediction of the Brazilian Northeast pluviometric time series as a real test situation. We wanted to predict if the next month would present more or less rain than the previous month. Results were promising.

Resumo

Este artigo apresenta o uso inédito de redes de Goal Seeking Neurons (GSN) na avaliação da importância relativa de dados multidimensionais. Assim, é possível medir o papel de cada variável de entrada na classificação de padrões definidos por uma amostra. Os melhores desempenhos foram observados nas menores topologias. Este fato, combinado com o algoritmo de treinamento one-shot (onde cada padrão é empregado uma única vez), conduz a um sistema eficiente. Comparado com o classificador padrão, os níveis de respostas corretas e de saturação foram melhorados. O decréscimo do nível de saturação cria mais flexibilidade para aprender regras secundárias (mais espaço para novos mapeamentos internos). No sentido de avaliar o sistema proposto mais detalhadamente, escolhemos a predição da série temporal pluviométrica do Nordeste Brasileiro como uma situação real de teste. O objetivo da tarefa foi prever se o próximo mês apresentaria mais ou menos chuva que o mês anterior. Os resultados foram promissores.

Keywords: redes GSN, classificadores, ponderação de variáveis.

1. Introdução

Um dos propósitos do uso de técnicas de Inteligência Artificial é o reconhecimento de padrões [1], a análise de um conjunto de dados e as relações entre as dimensões que os definem. Tal análise é substancialmente mais difícil quando o volume de dados a ser analisado é grande e complexo.

O presente estudo propõe uma técnicas utilizando redes neurais GSN [2-5], derivadas de Redes Booleanas [6-9], para valorar, dentro de um grupo de variáveis, as importâncias relativas.

Redes neurais artificiais [10-12] se popularizam mais a cada dia, desempenhando um papel importante nas aplicações que requerem a capacidade de reconhecimento de padrões, imagens e sons, e na avaliação do comportamento ou tendência de uma série conhecida de dados.

No mundo real, informações são geradas a cada segundo. Essas informações muitas vezes carregam, em si, dados que expressam o comportamento do processo que as gerou, identificam um período de tempo do qual esse processo fez parte, e até mesmo podem ajudar a prever resultados baseados no histórico delas.

Freqüentemente, o volume de dados a ser manipulado é muito grande, dificultando a localização das informações que realmente interessam para a análise requerida. Nesse ponto, as redes neurais mostram-se como técnicas promissoras.

A otimização das redes neurais – visando melhores desempenhos com aprendizados mais flexíveis e contribuindo para melhorias da generalização – torna-se uma tarefa consideravelmente importante diante da necessidade de maiores recursos computacionais e da urgência quanto ao tempo de resposta.

As redes neurais avaliam as informações fornecidas conjunta e integralmente, utilizando tais dados para o processamento e definição do resultado.

Na avaliação de um conjunto de dados, pode-se verificar que freqüentemente nem todas as informações possuem a mesma importância relativa. Algumas

informações influenciam fortemente, estão intrínseca e fortemente relacionadas ou, até mesmo, definem diretamente o resultado avaliado, enquanto outras contribuem bem menos.

Selecionar, em um conjunto de informações, as mais importantes (do ponto de vista de que elas influenciam o resultado positiva ou negativamente) pode fazer com que a rede neural não desperdice recursos disponíveis de processamento com dados de pouca importância, generalize mais (melhorando o desempenho) e utilize os recursos disponíveis de processamento mais eficientemente. Em redes GSN, o uso de posições internas de memória é crítico. O adiamento de altos níveis de saturação, problema inerente ao modelo GSN, é uma das grandes metas.

O problema central é, portanto, a determinação da importância relativa das variáveis componentes, numa tarefa de reconhecimento de padrões, a partir de um conjunto de dados disponível. Como problema secundário, temos a diminuição do nível de saturação de redes GSN.

A hipótese principal conjectura que o uso do sistema proposto permite mensurar a importância relativa das variáveis de entrada.

As hipóteses secundárias são:

- O nível de saturação do sistema proposto é menor quando comparado ao sistema padrão, uma vez que as variáveis mais importantes são escolhidas mais frequentemente.
- A taxa de acerto do sistema proposto é maior quando comparado ao sistema padrão, uma vez que maior atenção é dispensada para variáveis mais importantes.

2. Goal Seeking Neurons

A rede GSN (do Inglês, Goal Seeking Neuron) [2] é um modelo de rede booleana [6] que tenta melhorar algumas limitações de outras redes booleanas antecessoras [7], como a utilização de memória e a recuperação de padrões [9]. Ao contrário de outras redes neurais tradicionais [10-12], totalmente interconectadas e dependentes de cálculos matemáticos com funções contínuas – que contribuem para uma dificuldade maior de processamento – a rede GSN possui arquitetura de árvore (muitas vezes, binária) e seu processamento neuronal é mais simples (geralmente, um circuito lógico, sem grandezas contínuas). O modelo GSN, bem como toda a família de redes neurais artificiais sem pesos, tem motivado estudos frequentes na literatura desde a década de 1960 [13-24]. Vale ressaltar a adequação natural deste modelo a implementações em hardware, tais como FPGA [25], onde o leitor interessado pode resgatar um pouco dessa interessante sub-área.

Ao contrário de outros modelos neurais booleanos, que empregam escolhas probabilísticas para definir saídas neuronais, resultando em comportamento não determinístico em termos globais [26], as redes GSN sempre respondem o mesmo valor para a mesma estimulação de entrada.

Foram propostas redes GSN [2] para aprendizado supervisionado e não supervisionado. Interessamos

apenas o modelo GSNff, onde o aprendizado supervisionado é utilizado. Assim, no sentido de abreviar nossa exposição e detalhar o que de fato é o foco deste trabalho, apenas tal modelo será apresentado e descrito. Quando escrevemos “redes GSN”, subentende-se, portanto, que estamos nos referindo a “redes GSNff”. Vários melhoramentos foram propostos para este tipo de arquitetura [3,4].

Um neurônio GSN (ver Figura 1) é formado por:

- Duas (ou mais) entradas onde são aplicados os valores vindos de outros neurônios, no caso do neurônio pertencer a um nível intermediário, ou valores do padrão corrente (componente de entrada do dado amostral focalizado) quando o neurônio estiver na camada de entrada.
- Uma saída, cujo valor é aplicado na entrada de outro neurônio ou contém a saída da rede quando o neurônio estiver na camada de saída.
- Um conjunto endereçável que disponibiliza posições internas de memória interna no sentido de implementar objetivos (mapeamentos) assumidos. O conjunto endereçável de um neurônio (com duas entradas) é composto, portanto, de quatro posições internas de memórias, inicialmente em estado indefinido, contendo o valor indefinido “u”. Sendo n o número de entradas do neurônio, o conjunto endereçável é composto de 2^n posições internas de memória.

Um conjunto endereçado que conterá os valores internos do neurônio, que foram endereçados pelas entradas correntes (atuais). Quando todas as entradas possuem valor definido (“0” ou “1”), o conjunto endereçado tem apenas uma única posição endereçada. Quando uma única entrada tem valor indefinido, duas posições são endereçadas e assim por diante. Sendo k o número de entradas com valor indefinido para um certo padrão de entrada, o conjunto endereçado compõe-se de 2^k posições internas endereçadas.

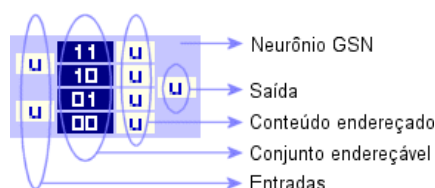
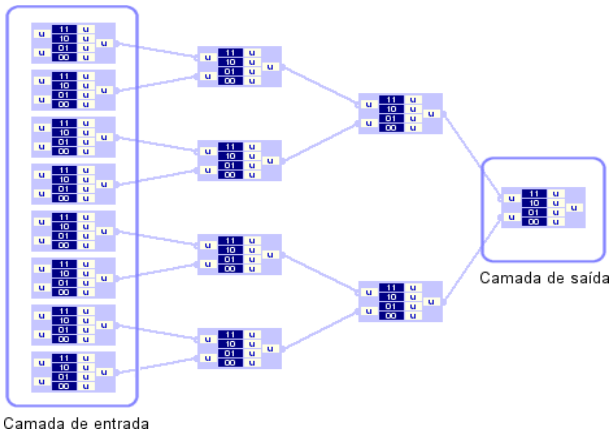


Figura 1: Estrutura interna de um Goal Seeking Neuron.

Uma rede GSN possui a topologia de uma árvore. A camada de entrada está associada à base da árvore, enquanto o ápice da árvore (ponto de convergência de toda estrutura) define a unidade de saída. Normalmente, são utilizadas estruturas homogêneas e, mais, árvores binárias. A redução do número de entradas diminui o número de parâmetros livres (ajustáveis no treinamento) dentro do neurônio, forçando-o a maiores generalizações.

O objetivo da rede GSN é construir internamente mapeamentos que reproduzam ao máximo o conjunto de padrões de treinamento. O objetivo de um neurônio intermediário é suprir, em cada situação, o valor adequado ao neurônio que capta sua saída particular no

sentido de fazer o neurônio de saída responder adequadamente.



Camada de entrada

Figura 2: Estrutura interna de uma Rede GSNff.

O treinamento de redes GSN, onde cada padrão é apresentado uma única vez, conduz seus neurônios a dois estados, validação e aprendizagem. Inicialmente, todos os conteúdos endereçados de todos os neurônios contêm o valor indefinido, “u”. Na validação, o neurônio mostra todos os mapeamentos possíveis de serem implementados, enquanto, no estado de aprendizagem, um mapeamento local concordante com o mapeamento coletivo desejado é executado na forma do armazenamento de um valor lógico não indefinido, “0” ou “1”.

3. Sistema Proposto

O presente estudo desenvolve um sistema formado por redes GSN capaz de identificar, para fins de classificação, a importância relativa de cada variável componente da representação de dados amostrais. Tal conhecimento provê informação sobre a sustentação do resultado apresentado pelo sistema classificador, possibilitando, ainda, a construção de classificadores com menos neurônios. Topologias menores possuem, geralmente, melhor capacidade de generalização, sendo mais rápidas (tanto em software quanto em hardware) e mais fáceis de serem implementadas fisicamente.

O sistema trabalha com padrões binários e, para este estudo e sem perda de generalidade, apenas duas classes foram analisadas. A rigor, redes GSN são adequadas à situação de falta de conhecimento de variáveis de entrada (quando o valor indefinido “u” pode ser utilizado). Em outras palavras, o sistema está preparado para aceitar não apenas os valores lógicos definidos, “0” e “1”, mas também o valor indefinido “u”.

Com as variáveis de entrada valoradas, definimos classificadores onde as variáveis são escolhidas para cada rede componente através de uma busca genética (onde o grau de adaptação é a importância da variável). Em comparação ao sistema originalmente proposto, que escolhe as variáveis de entrada aleatoriamente, nasce a hipótese de que o sistema proposto apresenta melhor desempenho devido à valoração correta da importância das entradas e da escolha baseada neste viés.

Outra hipótese razoável é a ocorrência de menor índice de saturação, devido ao fato da rede utilizar mais vezes as mesmas variáveis e estas estarem mais bem relacionadas ao mapeamento pretendido.

A idéia básica para valoração das variáveis de entrada é considerar seu papel em redes GSN durante o treinamento. Variáveis mais importantes [7] conduzem a melhores desempenhos: mais exemplos aprendidos e menos saturação.

Operacionalmente, com base no número de exemplos aprendidos e nível de saturação (adotados como primeiro e segundo critério de ordenação respectivamente), construímos um ranking do conjunto de redes analisadas. A posição de uma rede específica, sua colocação no ranking, é tomada como uma nota (uma medida parcial de importância) para todas as variáveis usadas em sua composição. Para a definição das redes constituintes do conjunto a ser ordenado, cabe lembrar que as variáveis de entrada são escolhidas através de uma distribuição aleatória uniforme, ou seja, todas variáveis apresentam a mesma chance.

Matematicamente, a definição de importância é dada por:

$$V_x = \sum_{j=1}^n R_j^x$$

Onde

V_x é a importância relativa da variável x ,

R_j^x é a importância da variável x para a rede j ,

e n é o total de redes do conjunto analisado.

A importância da variável x para a rede j , por sua vez, é definida por:

$$R_j^x = \begin{cases} 0 & x \notin \text{rede}_j \\ P_j & x \in \text{rede}_j \end{cases}$$

Onde

P_j é a posição da rede no ranking das n redes analisadas.

Para forçar frequências de utilização semelhantes, o mecanismo que sorteia aleatoriamente as variáveis de entrada é impedido de escolher novamente uma variável específica enquanto todas as outras ainda não foram escolhidas. Com exceção da escolha da primeira variável de entrada da primeira rede, somente quando todas variáveis foram usadas, torna-se possível escolher de todo o conjunto de variáveis de entrada. Garantimos, conseqüentemente, que não apareça duas vezes na mesma rede.

Avaliadas as importâncias das variáveis de entrada, tais valores definem a distribuição probabilística necessária para guiar a escolha das variáveis de entrada das redes que compõem o sistema classificador. A distribuição uniforme (assumida quando todas as variáveis são consideradas de mesma importância) é, portanto, substituída pela distribuição obtida pelo cálculo das importâncias relativas.

Variáveis relativamente mais importantes são escolhidas mais freqüentemente no sistema proposto. O modelo padrão (proposto pelo criador da arquitetura GSN) não utiliza este viés, tornando as escolhas das variáveis equiprováveis.

Tabela 1: Resultado médio de P1 (100 avaliações, 3 redes de 2 entradas)

Mês/Ano	DESEMPENHO		SATURAÇÃO	
	PADRÃO %	PROPOSTO %	PADRÃO %	PROPOSTO %
01/1901	53,80	55,90	100,00	97,85
02/1902	55,77	57,23	100,00	96,50
03/1903	57,60	66,87	100,00	97,37
04/1904	56,57	56,87	100,00	97,52
05/1905	66,77	77,97	100,00	96,58
06/1906	56,60	56,87	100,00	98,07
07/1907	53,80	54,83	100,00	97,37
08/1908	68,47	73,63	100,00	94,77
09/1909	86,70	88,60	100,00	97,93
10/1910	58,03	59,93	100,00	97,27
11/1911	53,73	54,57	100,00	98,03
12/1912	52,07	53,70	100,00	98,53

A definição da importância relativa das variáveis que modelam um fenômeno é bastante útil para sua compreensão e, no caso de um sistema de auxílio à decisão, também para utilização de seus resultados. Quando um meteorologista avisa que estamos diante de uma estação chuvosa, gostaríamos pelo menos de saber quais variáveis foram consideradas e em que medida.

Fenômenos descritos por um número elevado de variáveis são mais bem entendidos quando tal número é reduzido sem perda de informação. Em nosso caso, obteremos classificadores baseados em redes menores e, conseqüentemente, capazes de maior generalização, mais rápidas e, por fim, mais fáceis de implementar tanto em software quanto em hardware.

Por fim, a diminuição do índice de saturação em redes GSN propicia a ampliação da capacidade dinâmica de aprendizagem: a rede pode aprender mais, pois mais posições indefinidas são mantidas no treinamento.

4. Resultados

Primeiramente, cabe registrar o sucesso da arquitetura proposta quando utilizamos dados sintéticos. Em poucas palavras, definimos duas classes a partir de dois modelos básicos antagônicos (2 matrizes 16x16, uma com todos os bits "1" e outra com todos os bits "0") e introduzimos faixas de ruído com intensidade controlada. A importância da variável (pixel isolado), conseqüentemente, é inversamente proporcional ao nível de ruído a que foi submetida. Tal fato pôde ser observado com clareza e consistência durante as simulações. O que passamos a demonstrar, entretanto, é a utilização do sistema em tarefa realística e historicamente complexa.

Foram utilizados dados que representam o índice mensal de precipitação pluviométrica no Nordeste Brasileiro (cidades de Fortaleza e Quixeramobim), do ano de 1849 a 2002 [27].

Os dados foram submetidos a dois processos de avaliação com diferentes parâmetros. Um dos processos (nomeado P1) consistindo de 100 avaliações de três redes, com cada rede contendo duas entradas (um neurônio), e outro processo (nomeado P2) com três avaliações de 100 redes de oito entradas cada uma (sete neurônios). O resultado médio está demonstrado nas tabelas 1 e 2.

O conjunto de avaliação contém o mês de janeiro do ano de 1901, o mês seguinte do próximo ano e assim sucessivamente, compondo doze meses de anos consecutivos. Deste modo, estamos focalizando o desempenho no conjunto de teste.

Os valores obtidos com os dois processos de avaliação mostram claramente que a quantidade e a topologia das redes afetam o resultado final quanto ao desempenho. Comparado ao sistema padrão, o sistema proposto conduz a menores saturações.

Tabela 2: Resultado médio de P2 (3 avaliações, 100 redes de 8 entradas)

Mês/Ano	DESEMPENHO		SATURAÇÃO	
	PADRÃO %	PROPOSTO %	PADRÃO %	PROPOSTO %
01/1901	43,37	42,42	88,08	85,41
02/1902	55,95	62,15	89,32	85,43
03/1903	61,22	69,35	89,66	87,12
04/1904	38,84	29,29	90,74	88,16
05/1905	54,95	70,72	89,03	86,45
06/1906	51,49	38,10	87,12	84,15
07/1907	43,28	36,89	89,41	86,78
08/1908	63,14	80,99	89,08	83,76
09/1909	66,52	54,17	83,71	82,84
10/1910	39,39	39,32	84,58	83,38
11/1911	55,12	56,02	89,64	87,09
12/1912	50,48	52,98	87,58	86,58

Um maior número de processos de pequenas redes com poucas entradas realmente utilizam melhor as variáveis identificadas como mais importantes do conjunto de variáveis de entrada, como observado nos problemas artificiais e nas figuras 7 e 8 (ver final).

A Figura 9 mostra as variáveis identificadas na amostra segundo a sua importância, tendo como base a situação P1 (100 avaliações, 3 redes de 2 entradas). A análise da Figura 9 sugere o efeito da estabilidade na avaliação da importância relativa das variáveis

5. Conclusão

Este trabalho apresentou um novo conceito na área de redes neurais: o uso da fase de treinamento como fornecedora de resultados intermediários importantes não ligados diretamente às tarefas de classificação, previsão e otimização. No caso, o

resultado intermediário (nem por isso de menor importância) é a ponderação das variáveis componentes de padrões multidimensionais. Tal ponderação pôde auxiliar um classificador convencional baseado em redes GSN a escolher melhores topologias para suas redes componentes.

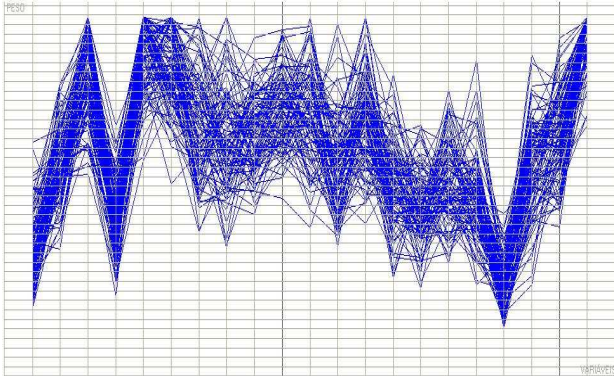


Figura 9: Visão sobreposta da valorização de importância das variáveis de entrada em P1.

No cenário de padrões booleanos, a transformação realizada, no estudo do comportamento histórico das chuvas no nordeste brasileiro, de uma tarefa de previsão para ser tratada como classificação abre perspectivas para novas aplicações.

A diminuição do nível de saturação em classificadores baseados em redes GSN surge como importante ganho, possibilitando maior aprendizagem e flexibilidade do sistema.

De modo conclusivo, podemos dizer que o sistema proposto conseguiu avaliar a importância relativa das variáveis de entrada e forneceu sustentação para escolha de melhores topologias iniciais para classificadores baseados em GSN no aspecto de saturação interna.

Quando situações mais simples (dados sintéticos) foram utilizadas, o sistema se comportou muito bem. Entretanto, a dificuldade da tarefa enfrentada, apesar de inibir melhores resultados, serviu-nos como teste mais exigente, mais rigoroso, sugerindo o emprego futuro de outros classificadores.

Referências

- [1] Russel, S.J. & Norvig, P., *Inteligência Artificial*, Editora Campus, 2ª edição, 2004.
- [2] Filho, E. Investigation of Boolean neural network based on a novel Goal-Seeking-Neuron, Tese de Doutorado, Electronic Engineering Laboratories, University of Kent, UK, 1990.
- [3] Martins, W. 1994. "The Improvement of Classifiers Based On Weightless Neural Networks", Tese de Doutorado, Department of Electronics, University of York, UK.
- [4] Martins, W. & Pinheiro Jr, C.B., On-Line Expansion of Goal Seeking Neuron Networks, Anais da IEEE-INNS-ENNS International Joint Conference on Neural Networks (IJCNN'00), Como, Italia, Volume 4, pp. 4523-4528, 2000.
- [5] Martins, W. & Silva, J.C., Uso de redes GSN em tarefas de ordenação (ranking) aplicadas à medida de adimplência de clientes, Anais do V Congresso Brasileiro de Redes Neurais, Rio de Janeiro, vol. 1, pp 529-534, 2001.
- [6] Aleksander, I & Albrow, R. C. Adaptive logic circuits. *Computer Journal*, pp. 11-65, 1968.
- [7] Aleksander, I. *Neural Computing Architectures*, North Oxford Academic Press, 1989.
- [8] Austin, J., "A Review of RAM based Neural Networks", Anais do Fourth International Conference on Microelectronics for, pp. 58-66, IEEE Computer Society Press, 1994.
- [9] Lurdermir, T.B. et al, Weightless neural models: a review of current and past Works, *Neural Computing Surveys*, vol 2, pp 41-46, 1998.
- [10] McCulloch, W. S. & Pitts, W. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 1943.
- [11] Kohonen T., *Self-Organizing Maps*. Berlin. Springer-Verlag, 2ª edição, 1977.
- [12] Rumelhart, D. E., Hinton, G. E. & Williams, R. J., Learning representations by back-propagating errors, *Nature*, pp 533-536, 1986.
- [13] Aleksander, I & Mandani, E. H. Microcircuit learning net improved recognition by means of a pattern feedback. *Electronics Letters*. Vol. 4, pp 425-426, 1968.
- [14] Bisset, D. L., Filho, E and Fairhurst, M. C., A comparative study of neural network structures for practical application in a pattern recognition environment. Proceedings of the IEE International Conference on Artificial Neural Networks, pp. 378-382. Londres, 1989.
- [15] Adeodato P. J. L. & Taylor J. G., Analysis of the storage capacity of RAM-based neural networks Anais da Weightless Neural Networks Workshop 1995, pp. 103 - 110, D. L. Bisset (ed.), Canterbury, UK, 1995.
- [16] Gustin, V., Realization of a fuzzy Boolean neural network classifier, *Microprocessing and Microprogramming*, vol 40, 1, pp 23-31, 1994.
- [17] Austin, J., Uncertain Reasoning with RAM based Neural Networks, *Journal of Intelligent Systems*, vol. 2, pp. 121-154, Freund Publishing (UK), 1992.
- [18] Bhide, S., John, N. & Kabula, M.R., A Boolean Neural Network Approach for the Traveling Salesman Problem, *IEEE Transactions on Computers*, vol 42, 10, pp 1271-1278, 1993.
- [19] Martins, W. & Allinson, N.M., Imposing cooperation and avoiding false examples by fast control of presentation order, Anais da Intl Conf on ANN, ICANN'95, Paris, vol 1, pp 180-185, Springer, Berlin, 1995.
- [20] Austin, J. *RAM-based Neural Networks*, Progress in Neural Processing 9, Singapore: World Scientific Pub. Co., 1998.

- [21] Linneberg, C. & Jørgensen, T.M., Discretization Methods for Encoding of Continuous Input Variables for Boolean Neural Networks, Anais da IJCNN'99, Washington, 1999.
- [22] Lindblad, T. et al, RAM-based neural networks for data mining applications, pp 3390-52, Anais da SPIE Vol. 3390, 1998.
- [23] Howells, W.G.J. et al, An exploration of a new paradigm for weightless RAM-based neural networks, *Connection Science*, Vol 12, 2000.
- [24] Geoffrey, G., Aleksander, I., Predicting the behaviour of G-RAM networks, *Neural Networks*, Elsevier Science Ltd, vol 16, 1, pp 91-100, 2003.
- [25] Simões, E.V., Uebel, L.F., Barone, D.C., Fast Prototyping of Artificial Neural Networks: GSN Digital Implementation. IV Intl Conf on Microelectronics for Neural Networks and Fuzzy Systems, Itália, 1994.
- [26] Kan, W K and Aleksander I. 1987. A probabilistic logic neuron network for associative learning, Anais do IEEE ICNN, vol II, 541-548, San Diego, USA.
- [27] Dados Pluviométricos do Nordeste Brasileiro, [http://tao.atmos.washington.edu/data_sets/brazil/#values]



Figura 7: Desempenho e saturação médios em P1 (100 avaliações, 3 redes de 2 entradas)

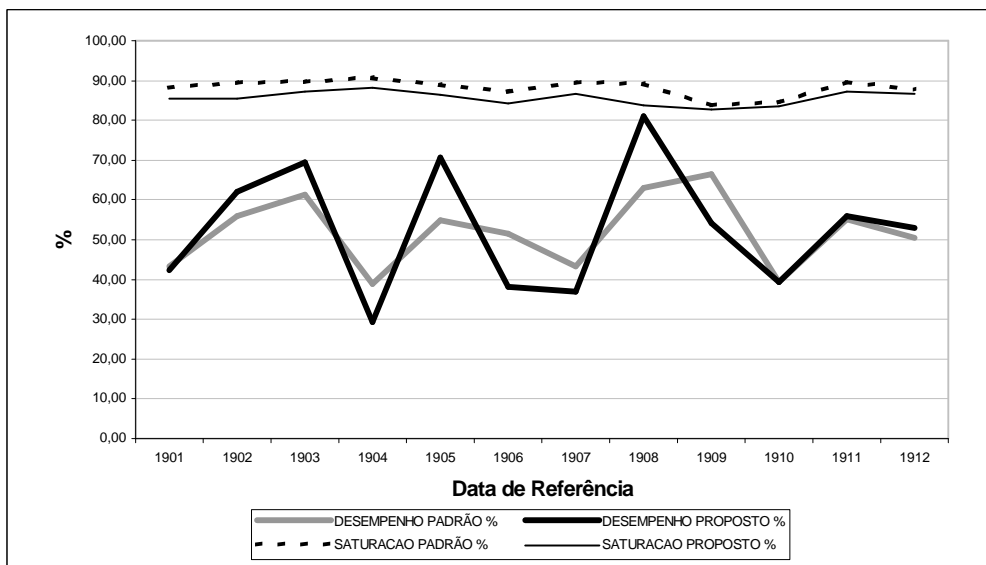


Figura 8: Desempenho e saturação médios em P2 (3 avaliações, 100 redes de 8 entradas)