

REDUÇÃO DE NÚMERO DE NEURÔNIOS DE RAIOS VARIÁVEL EM REDES TIPO ART

Luiz P. Caloba¹, Max S. Dutra², Alvaro D. Orjuela^{1,3}, Ivanovich Lache²

caloba@lps.ufrj.br, max@mecanica.coppe.ufrj.br, dorjuela@ieee.org, ilache@ufrj.br

¹ Laboratório de Processamento de Sinais - LPS – Programa de Engenharia Elétrica COPPE/UFRJ

² Laboratório de Projeto de Máquinas e Robótica – LABROB – Programa de Engenharia Mecânica COPPE/UFRJ

³ Programa de Ingeniería Electrónica y Biomédica - Universidad Antonio Nariño – Bogotá – Colombia

Abstract

The neural network's relevance for clustering different kind of data has encouraged diverse works in order to optimize their implementation. Between the areas of research, exist ones that seek to reduce the number of neurons needed to define a specific class. In this work the authors present an algorithm that allows the reduction of number of neurons with variable radius in ART's networks, the procedure used is listed as semi-supervised, but has features that allow its application in conjunction with not supervised techniques. Several tests were performed to assess the effectiveness of this method, two of the most important and informative examples are presented.

Resumo.

A importância das redes neurais para a clusterização de diferentes tipos de dados tem impulsionado diversos trabalhos para conseguir técnicas que otimizem a aplicação das mesmas. Entre as áreas de pesquisa estão aquelas que procuram a redução do número de neurônios necessários para definir uma classe. Neste trabalho os autores apresentam um algoritmo que permite a redução de número de neurônios de raios variáveis numa rede tipo ART, o procedimento utilizado é catalogado como semi-supervisionado, mas com características que permitem sua aplicação em conjunto com técnicas não supervisionadas. Diversos testes foram feitos avaliando a eficiência deste método, dois dos mais importantes e esclarecedores da técnica são apresentados.

Palavra chave: ART Network, Semi-supervised training, Non-supervised reduction

Introdução

A clusterização por meio de redes neurais é uma técnica que tem demonstrado ser útil em diversos tipos de aplicações, usando técnicas de treinamento supervisionado ou não supervisionado, mantendo sempre o objetivo geral de obter redes que possam identificar padrões semelhantes entre diversos dados [1][2].

Da mesma forma, existem diversos tipos de redes neurais que podem levar a cabo dito procedimento, entre as mais conhecidas se encontram [3].

- Camada de Kohonen
 - Supervisionada
 - Não supervisionada
- Rede ART
 - Supervisionada
 - Não supervisionada
- Rede Counterpropagation.

Todos os tipos de redes apresentadas anteriormente tem uma topologia em comum, a qual se pode observar na Figura (1), onde se acham os 3 tipos de redes apresentados anteriormente. O bloco denominado de 1, representa a camada de Kohonen e as redes ART, a variável X representa a entrada, w^i e b_i são o centro e o bias de cada neurônio, finalmente, Y é a saída da rede neural [4][5].

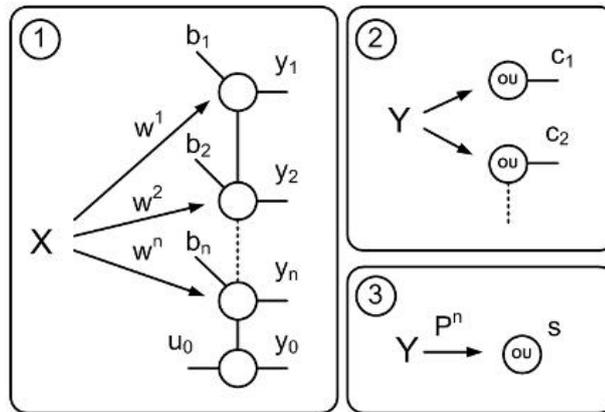


Figura 1. Topologia de redes.

O bloco 1 unido com o bloco 2, permite a criação de redes capazes de definir, depois de um treinamento ART, que neurônios representam uma determinada classe. Finalmente a união do bloco 1 com o bloco 3 permite criar a rede Counterpropagation onde os termos P^n representam os padrões que representam cada neurônio.

A variável denominada de *bias* (b_i), permite-lhe a um neurônio i ter uma área de cobertura maior o menor dependendo do seu valor, a relação do b_i com o raio de cobertura do neurônio (r_i) está dada pela equação (1)

$$b_i = -\left(1 - \frac{r_i^2}{2}\right) \quad (1)$$

Definido b_i , é possível pensar em redes que tenham neurônios de diferentes raios de cobertura [5], raios que permitiriam a criação de clusters bem definidos com uma quantidade mínima de neurônios, este tipo de topologia se pode observar na Figura (2).

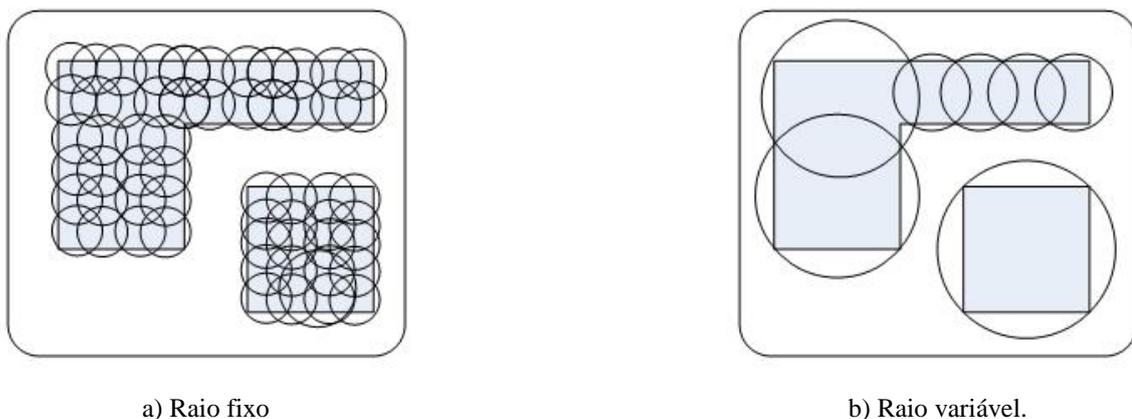


Figura 2. Importância do raio variável.

O grande dilema é o procedimento que tem que se levar a cabo para poder gerar este tipo de redes, como treiná-las e como ajusta-las, este procedimento será descrito na seguinte seção.

1. Desenvolvimento da rede.

O primeiro passo é saber qual é o tipo de restrições que se querem aplicar às redes de raio variável, entre as opções estudadas se tem:

- Redução de Neurônios: Reduzir ao máximo a quantidade de neurônios, identificando os distintos clusteres, a única restrição é não invadir clusters alheios.
- Manter limites do cluster: Reduzir neurônios mantendo os limites do cluster o mais fiel possível à realidade.

O primeiro destes itens, Redução de Neurônios, é trabalhado neste documento, e como foi apresentado anteriormente, a meta é reduzir a quantidade de neurônios necessária para representar corretamente uma classe, os passos necessários para levar a cabo este procedimento foram.

1. Carregar dados X
2. Normalizar dados X/I .
3. Achar raio mínimo dos neurônios r_0 .
4. Treinamento ART.
5. Definir Cluster (não supervisionado).
6. Modificar a rede.
 - a. Definir que entradas ativam cada neurônio (*Acti*).
 - b. Definir a distâncias entre os neurônios (*Dista*).
 - c. Selecionar um neurônio (*i*) e achar seus vizinhos.
 - d. Selecionar o neurônio vizinho (*j*) mais perto de *i*.
 - e. Definir as entradas que são representadas por *i* e $i(X_{ji})$.
 - f. Definir um neurônio que seja a media ponderada de (X_{ji}) com raio r_n .
 - g. Se o neurônio W_n com r_n é ativo com todos os dados (X_{ji}) pode passar a (*h*), se não, tem que acrescentar o raio, e voltar a (*g*).
 - h. Avaliar se o novo W_n representa algum dado que não é de seu cluster, se representa voltar a (*c*) com um outro neurônio *i* senão ir para (*j*).
 - i. Modificar o neurônio *i*, $W_i=W_n$ e $r_i=r_n$, W_j e r_j são zero.
 - j. Voltar para (*c*) até avaliar todos os neurônios.
7. Eliminar os neurônios não usados, $W_x=0$.
8. Voltar para (6) até que não sejam eliminados mais neurônios em (7).

2. Resultados

O primeiro passo é avaliar o comportamento da rede neural quando é usada para classificar diferentes tipos de formas de clusteres e diferentes tipos de classes. Para conseguir este objetivo foram gerados 2 tipos de dados diferentes que são apresentados à rede neural, esta segue os passos definidos na seção 2 tentando diminuir o numero inicial de neurônios da rede mantendo as classes achadas na primeira iteração.

2.1. Exemplo de redução.

O grupo de dados apresentados na Figura (3.a) foram escolhidos para ser analisados pela rede, estes dados devem ser normalizados para poder ser classificados corretamente, a normalização dos dados gera uma terceira componente que permite que todos os dados se localizem sobre uma esfera de raio unitário, em outras palavras o módulo de X_n para todo n é igual a 1. Os dados normalizados se podem observar na Figura (3.b).

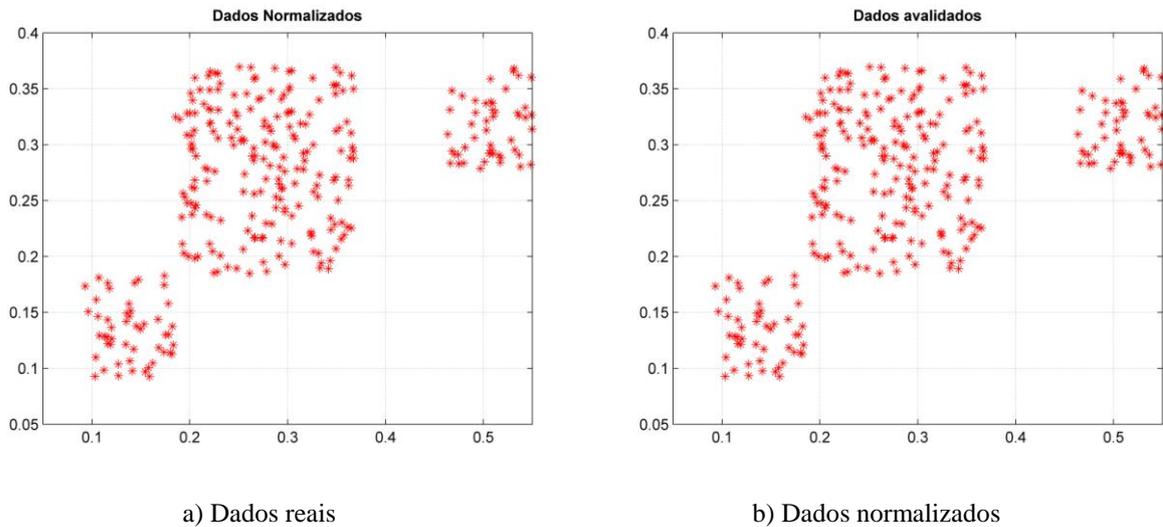


Figura 3. Dados do experimento.

O seguinte passo é achar a distância entre cada neurônio e encontrar a distribuição para mencionado valor, o valor d_{ij} é a distância entre o dado i e o dado j e esta representado na Equação (2) abaixo.

$$d_{ij}^2 = |x_i - x_j|^2 \quad (2)$$

Obtendo os valores d_{ij}^2 se acha a distribuição que para os dados apresentados na Figura (3) é a desenhada na Figura (4).

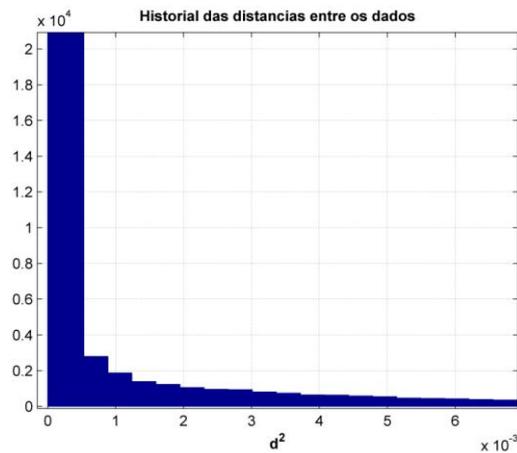


Figura 4. Histograma de d_{ij}^2 .

Com estes valores é possível achar o valor do raio inicial r_0 segundo a Equação (3) abaixo descrita.

$$f_0^2 = 1.5m_{ij}^2 \quad (3)$$

Logo de ter obtido o valor para r_0 é iniciado o treinamento tipo ART para achar os neurônios que representem aos dados apresentados. A primeira rede achada é apresentada na Figura (5.a), posteriormente e logo de fazer as análises de vizinhança são achados os clusters que são apresentados na Figura (5.b).

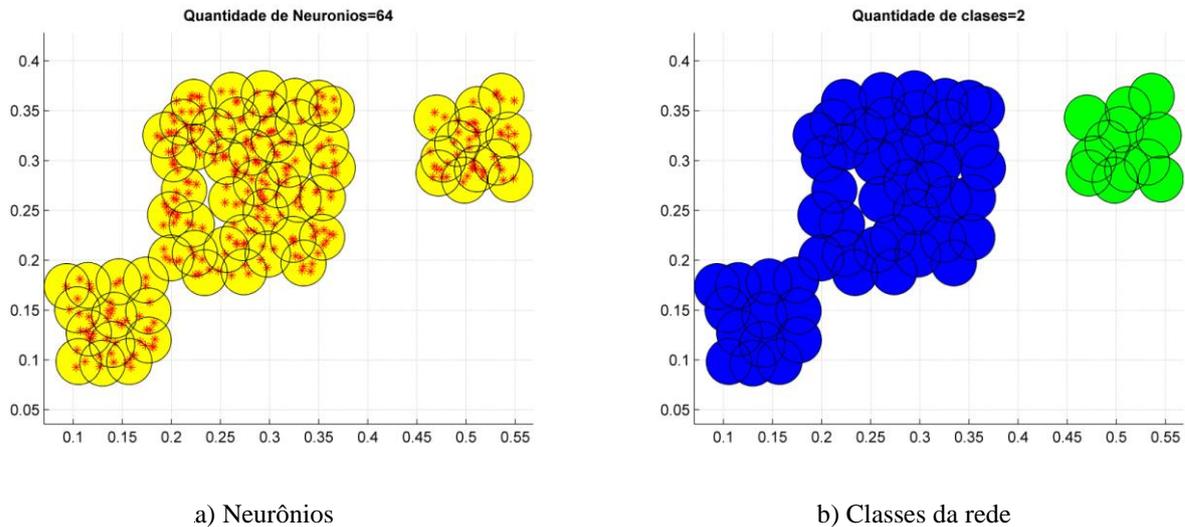


Figura 5. Nova topologia da rede.

A verdadeira projeção, devido a que os neurônios ao igual que os dados se encontram numa esfera de raio unitário, é mostrada na Figura (6).

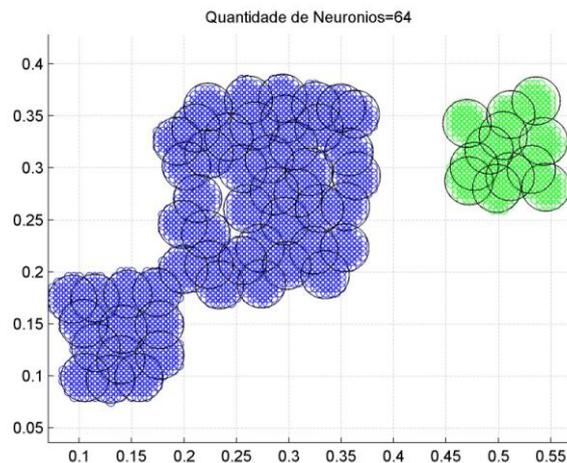
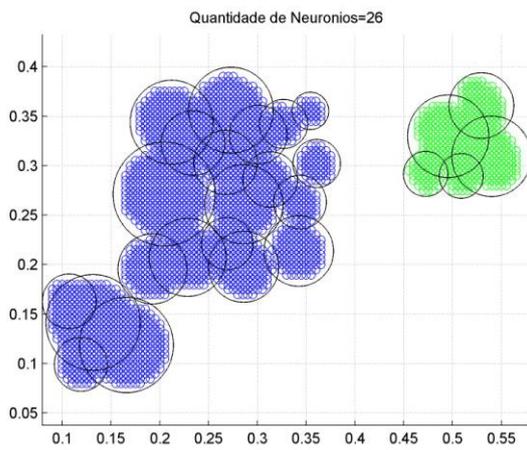


Figura 6. Domínio real da clusterização, 1 iteração.

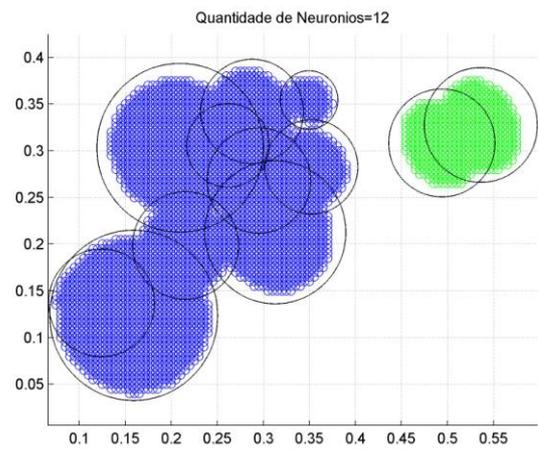
O seguinte procedimento consistiu em aplicar a rotina de modificação da rede até que não houvesse diminuição da mesma, os resultados e a topologia das redes obtidas se podem observar nas figuras contidas na Tabela 1.

Tabela 1. Tabela de figuras

| Aplicação Rotina | Figura do domino da rede |
|------------------|--------------------------|
| 1 | 7.a |
| 2 | 7.b |
| 3 | 8.a |
| 4 | 8.b |

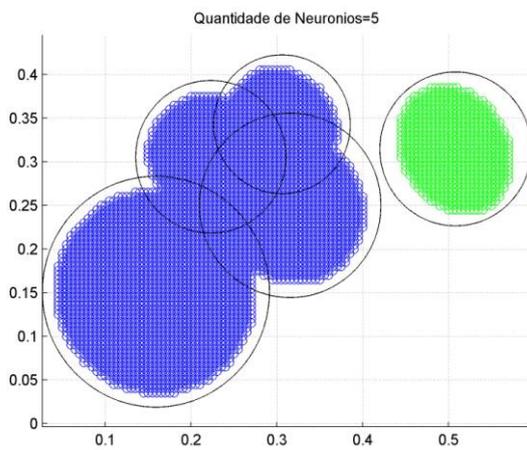


a) 2 Iterações.

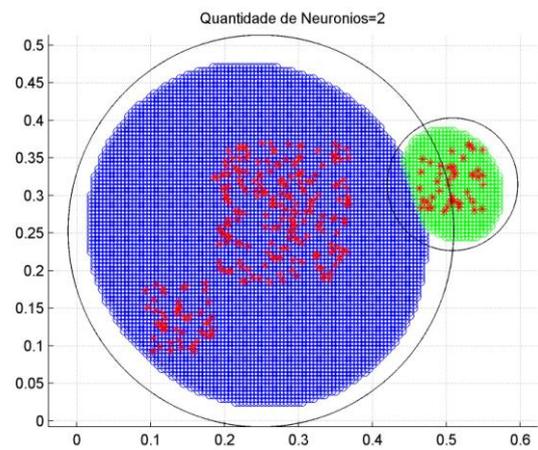


b) 3 Iterações

Figura 7. Simplificação da rede.



a) 4 Iterações.



b) 5 Iterações.

Figura 8. Simplificação da rede.

Finalmente e como é apresentada na Tabela (1) a última rede achada, onde a rotina programada não consegue aumentar o raio dos neurônios sem invadir outras classes, é a quinta rede, a Figura (8.b) apresenta a topologia e o tamanho da classe, já na Figura (9) é observada a diferença entre o tamanho da classe representado pelos neurônios iniciais (Figura (6)) e o tamanho das classes da rede final.

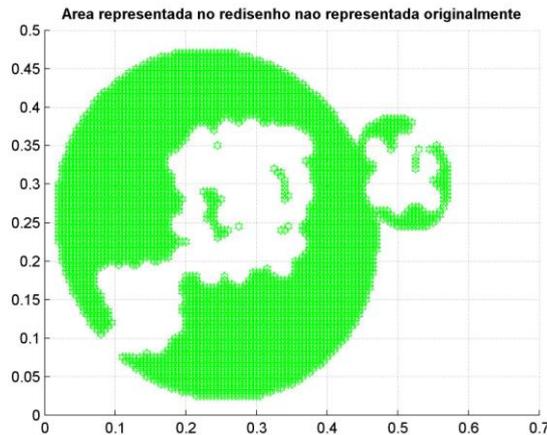


Figura 9. Diferença de cobertura da rede

2.2. Vários números de classes.

Neste exemplo se procura observar o funcionamento do algoritmo utilizando um conjunto de dados que pode ser dividido em diversos tipos, um tipo de aplicação pode ser a identificação de objetos para robótica móvel [6] neste exemplo em particular o número de clusters definidos é de 4.

Numa primeira definição de clusters são necessários mais de 240 neurônios para identificar corretamente todos os elementos e as respectivas classes. A divisão e distribuição dos clusters se pode observar na Figura 10.

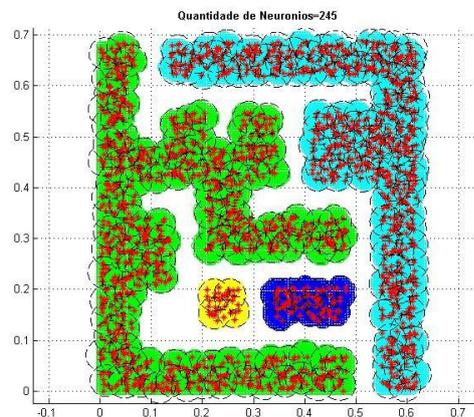


Figura 10. Simplificação da rede, iteração 1

Na segunda iteração, apresentada na Figura 11, é possível observar uma redução de mais de 50% dos neurônios utilizados sem afetar de forma considerável a média dos raios dos neurônios implementados.

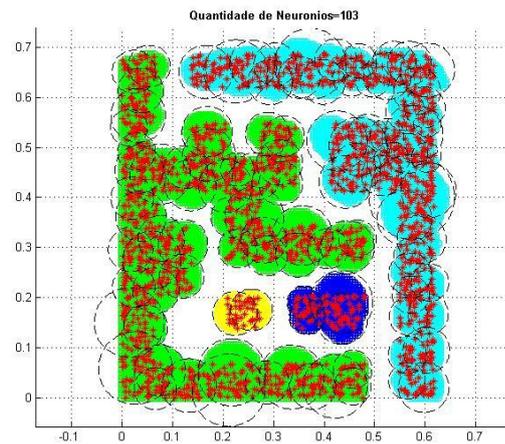


Figura 11. Simplificação da rede, iteração 2.

Finalmente na Figura 14 é possível observar a redução de número de neurônios para um total de 18, o que permite afirmar foi possível reduzir a rede neural a um valor de 7.5% da quantidade de neurônios inicialmente utilizada.

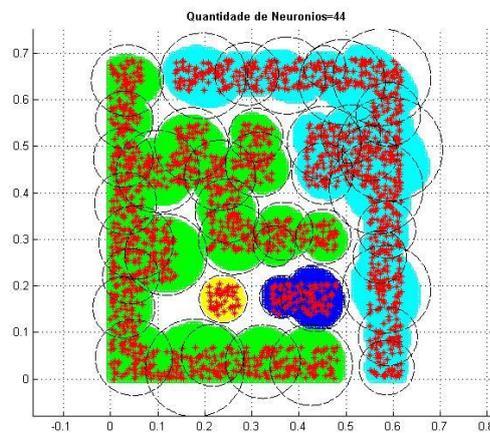


Figura 12. Simplificação da rede, iteração 3.

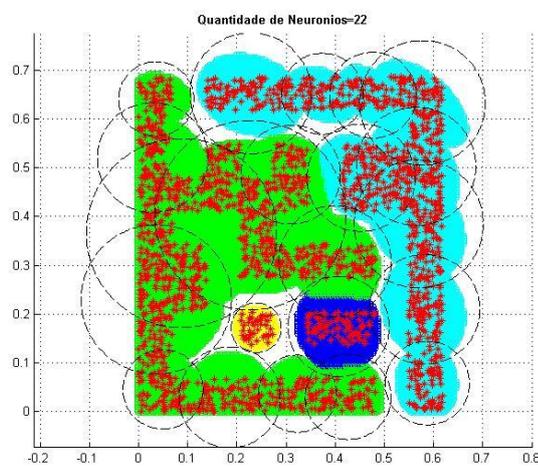


Figura 13. Simplificação da rede, iteração 4.

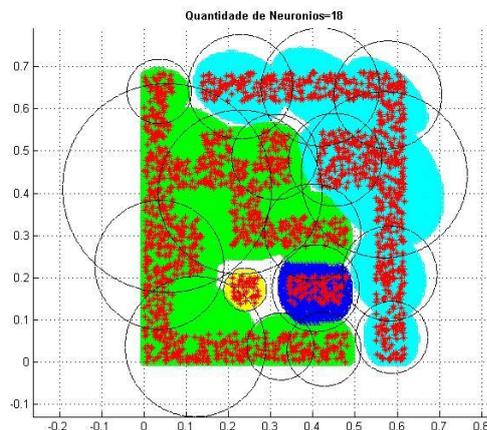


Figura 14. Simplificação da rede, iteração 5.

3. Conclusões

O procedimento de redução ou simplificação de redes neurais de raio variável tipo ART, resultou ser mais sofisticado do previsto inicialmente pelos autores deste documento, a intervenção de diversas variáveis que influencia o sucesso da redução da rede, aumentou a dificuldade do trabalho, mas gerou uma grande satisfação ao observar os resultados finais do algoritmo proposto.

A primeira variável que dentro do procedimento de redução é considerada como "sensível" é a determinação do raio mínimo r_0 , este segue o procedimento apresentado na seção 3.1. Neste procedimento é necessário achar a primeira moda da distribuição do quadrado das distâncias entre os centros dos neurônios, a dificuldade radica em que ao ser um procedimento não supervisionado, pequenas variações no cálculo da moda, podem influir em variações significativas do raio mínimo.

Depois de determinar o valor da moda, é possível determinar o valor de r_0 usando a equação (3) lembrando que este valor é uma sugestão para a inicialização do raio dos neurônios e não corresponde necessariamente, já que depende dos dados de entrada, o valor ideal para a clusterização, porém, a equação (3) mostrou-se válida para a grande maioria dos conjuntos de dados apresentados neste documento.

O segundo passo de importância na hora de simplificação da rede neural é o treinamento da rede, deste dependerá a posição e número de neurônios, portanto, as variáveis envolvidas no treinamento resultam essenciais no desempenho futuro da rede. Os valores usados para estas variáveis foram sempre os mesmos para todas as experiências, demonstrando ser uma escolha conservadora, mas apropriada, para a grande maioria de tipos de dados.

A seguinte etapa do procedimento é a definição dos clusters sendo a etapa mais importante, pois, dela depende a correta limitação dos dados nos seus específicos clusters. O procedimento usado neste trabalho, baseado no conceito de vizinhança, conseguia definir clusters separados com pelo menos um r_0 de distância entre seus neurônios mais próximos, o que permitiu obter resultados bons para classes facilmente separáveis, apresentando dificuldades na definição das classes com os conjuntos de dados menos distanciados.

Sendo definidas as classes é possível iniciar o procedimento de simplificação da rede, onde a única restrição para limitar o crescimento dum neurônio (i) é a não representação de dados de outra classe como a classe do neurônio (i). Esta restrição permitiu manter a representação das classes iniciais na redução da rede.

Para selecionar o que neurônio (i) que inicia o procedimento de ampliação foi usado novamente o critério de vizinhança, este critério apresentou resultados, mas não discriminava entre neurônios internos ou externos numa classe o que permitia a modificação da área de cobertura da nova rede, gerando em alguns dos testes novas áreas de cobertura maiores das apresentadas pela rede original.

Finalmente e como pode ser observado ao longo do trabalho o critério de parada, redução máxima de parada, pode não sempre ser conveniente se o desejo é manter a área de cobertura original da rede, por o que métodos que permitam a simplificação de redes mas mantendo os limites podem ser alvos de novos estudos.

4. Referências Bibliográficas

- [1] Alpaydin, E. & Baraldi, A. Constructive feedforward ART clustering networks part I. *IEEE Transactions on Neural Networks*, vol. 13, no. 3, pp. 645-661, 2002.
- [2] Kong, J. H. L. & Martin, G. P. M. D. A review of a hybrid neural network: Kohonen learning vector quantization and counterpropagation. *Proceedings of the IEEE International Conference on Neural Networks*, vol. 3, pp. 1397-1402, 1995.
- [3] Gestri, G. & Starita, A. Comparison between different neural network architectures for odour discrimination. *Proceedings of the 4th International Conference on Artificial Neural Networks*, pp. 410-414, 1995.
- [4] Kuhlen, T., Frank, T. & Kraiss, K. F. Comparative analysis of fuzzy ART and ART-2a network clustering performance. *IEEE Transactions on Neural Networks*, vol. 9, no. 3, pp. 544-559, 1998.
- [5] Medeiros, A. S. Uma análise da aplicação do modelo de rede neural RePART em comitês de classificadores. Master's thesis, Universidade Federal do Rio Grande do Norte (UFRN), 2008.
- [6] Gavrilov, A.V. Hybrid Neural Network Based on Models of Multi-Layer Perceptron and Adaptive Resonance Theory. *Proceedings of 9th Korean-Russian International Symposium on Science and Technology (KORUS'2005)*, pp. 119-122, 2005.