

# Aplicação de Mapa Auto-Organizável ao Agrupamento de Dados de Refletividade de Radar

Jorge V. R. Bonato, Paulo H. Siqueira,

Programa de Pós-Graduação em Métodos Numéricos, CESEC, UFPR,  
81531-970, Curitiba, PR

E-mail: bonato.ruviaro@gmail.com, paulohs@gmail.com,

Cesar A. A. Beneti

Instituto Tecnológico Simepar

81531-980, Curitiba, PR

E-mail: beneti@simepar.br.

**Resumo:** *A clusterização busca organizar grandes volumes de dados em subgrupos menores, de maneira a facilitar sua análise. Ao aplicá-la em dados meteorológicos de refletividade deseja-se agrupar os dados de maneira a formar grupos representando nuvens. Nesse trabalho usou-se a Rede Neural Artificial SOM de Kohonen.*

**Palavras-chave:** *Rede Neural Artificial, Meteorologia, Agrupamentos*

## 1 Introdução

A clusterização de dados é importante ferramenta utilizada à análise de dados em grande volume. É também utilizada na meteorologia quando, por exemplo, se deseja clusterizar dados de refletividade de radar, para posterior aplicação em técnicas de previsão de desenvolvimento e deslocamento de eventos severos. Em [7] apresenta-se a técnica de *nowcasting* ForTraCC, na qual a clusterização se dá através da identificação de *pixels* contíguos, pertencentes a dado intervalo, que passam então por dilatação e erosão matemáticas. Outras técnicas podem ser vistas em [4] e [6].

Neste trabalho aplicou-se a Rede Neural Artificial, de aprendizado não-supervisionado, SOM (*Self – Organizing Maps*) de Kohonen para clusterizar dados de refletividade vindos de um dos radares do Instituto Tecnológico Simepar, instalado em Teixeira Soares - PR e que atualmente opera com alcances quantitativo e qualitativo, respectivamente de 480 Km e 240 Km.

## 2 Objetivos

O presente trabalho tem por objetivo agrupar dados de radar, em especial os referentes a refletividade, de maneira a gerar núcleos de dados para posterior aplicação à métodos de previsão de desenvolvimento de eventos severos, entre eles o apresentado em [7].

## 3 Dados

Radares meteorológicos possuem 4 componentes principais, sendo eles o transmissor, a antena, o receptor e um display. Seu funcionamento se dá da seguinte maneira: o transmissor emite um sinal eletromagnético, que é direcionado à atmosfera pela antena. Ao atingir uma partícula, parte do sinal é refletido em direção ao receptor do radar e então pode ser visualizado no display.

A partir da detecção de uma partícula, sua posição é dada através da tripla  $(\theta, r, \phi)$ , sendo  $\theta$  o ângulo de azimute (ângulo em relação ao Norte),  $r$  a distância partícula-radar e  $\phi$  o ângulo de elevação da antena (em relação ao solo).

Uma das variáveis passíveis de serem medidas através do uso de radares meteorológicos é chamada de fator refletividade, que está relacionada à quantidade de energia que retorna ao radar após a onda eletromagnética emitida pelo radar atingir um alvo na atmosfera. Esta variável é denotada por  $z$  e definida como:

$$z = \sum_{i=1}^N D_i^6 \tag{1}$$

sendo  $N$  o número de partículas presente na amostra do volume em análise.

Para fins de ilustração segue a figura 1, na qual há presença de 16 partículas com diferentes raios:

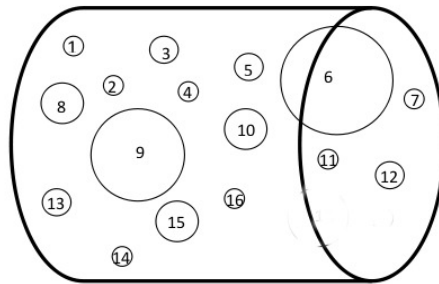


Figura 1: Ilustração de volume com distribuição de gotas

Como pode ser ver em (1) o fator refletividade é proporcional à sexta potência do diâmetro das partículas, podendo variar de pequenos a grandes valores no casos de, respectivamente, névoa e tempestades. De maneira a facilitar seu uso e análise faz-se necessário alterar a escala, de linear para logarítmica, através da seguinte transformação:

$$Z = 10 \log_{10} \left( \frac{z}{1 \text{mm}^6 / \text{m}^3} \right) \tag{2}$$

Ao fim desse processo tem-se que a variável refletividade é media em decibéis de refletividade ( $dBZ$ ).

Segundo [5] os dados volumétricos de radar são geralmente obtidos através da rotação da antena em torno de um eixo vertical, para diferentes ângulos de elevação. Após cada varredura volumétrica tem-se uma matriz de dados em coordenadas polares, que antes de aplicados à rede neural passam por um processo de mudança de coordenadas para o sistema cartesiano, nesse caso com resolução de  $0.5 \text{ Km} \times 0.5 \text{ Km} \times 0.5 \text{ Km}$ , obtendo assim uma matriz de dimensão  $(480, 480, 30)$ . Para cada coluna da matriz de dados cartesianas determinou-se o valor de máxima refletividade, gerando assim uma matriz de dimensão  $(480, 480)$ , que então foi aplicada à rede.

Neste trabalho definiu-se tempestade através dos seguintes passos:

- Identificação de pontos tais que seu valor de refletividade pertenciam ao intervalo entre 30 e 60  $dBZ$ ;
- Busca de regiões contíguas formadas por pontos identificados acima, tais que sua área seja superior a 20  $\text{Km}^2$

O período de dados utilizados para esse estudo compreende os 12 meses, entre Setembro de 2011 e Setembro de 2012 .

## 4 Metodologia

Desenvolvida por [3], a Rede Neural Artificial Mapa de Kohonen possui aprendizado não-supervisionado baseado na competição dos neurônios, cujo principal objetivo é transformar padrões de entrada com dimensão arbitrária em mapas uni ou bi-dimensional. Seu algoritmo genérico é mostrado a seguir:

1. Inicialização: escolha de valores aleatórios para os pesos sinápticos iniciais;
2. Amostragem: aplica-se à rede um vetor de entrada  $x$  pertencente ao conjunto de dados;
3. Combinação por similaridade: determina-se o neurônio  $BMU$  (*Best Matching Unit*), com base em uma função discriminante;
4. Atualização: pesos sinápticos do neurônio vencedor e daqueles pertencentes à sua vizinhança topológica são atualizados;
5. Continuação: repete-se os passos 2,3 e 4 até não haver alterações significativas;

### 4.1 Inicialização

A primeira etapa a ser realizada antes de iniciar o agrupamento dos dados é determinar quais características melhor definem cada grupo, além de escolher uma função discriminante para medir a similaridade entre dados de entrada e os neurônios da rede. Nesse trabalho optou-se por fazer uso do vetor  $(i, j, Z_{ij})$ , sendo  $(i, j)$  a posição da matriz de dados e  $Z_{ij}$  seu respectivo valor de refletividade. A escolha de uma grade com 25 neurônios se deve ao fato de ser pequena a probabilidade de haver um número maior de grupos em uma imagem de radar.

A inicialização dos pesos sinápticos foi feita escolhendo, de maneira aleatória, valores para  $i$  e  $j$  no intervalo de 0 a 479, e então buscou-se o valor  $Z_{ij}$  na matriz de dados. Segundo [1], a normalização dos dados de entrada e dos pesos sinápticos leva a uma melhoria na convergência, assim determinou-se:

$$i^* = \frac{i}{479} \quad j^* = \frac{j}{479} \quad Z_{ij}^* = \frac{Z_{ij}}{\max(Z)} \quad (3)$$

### 4.2 Combinação por Similaridade

Após definido o número de neurônios e características a serem analisadas, e inicialização dos vetores de pesos sinápticos de cada padrão de entrada e neurônio têm, respectivamente, as formas:

$$x = [i^* \quad j^* \quad Z_{ij}^*]^T \quad w_k = [w_k^1 \quad w_k^2 \quad w_k^3]^T \quad (4)$$

Como função discriminante usou-se o quadrado da distância Euclidiana:

$$D(x, w_k) = \sum_{l=1}^3 (x^l - w_k^l)^2 \quad (5)$$

O neurônio vencedor ou  $BMU$  (*Best Matching Unit*) é o neurônio tal que a equação 5 possua valor mínimo.

### 4.3 Atualização

O  $BMU$  define o centro da vizinhança topológica, formada por neurônios que também terão seus pesos sinápticos atualizados. A função vizinhança deve possuir valor máximo no neurônio vencedor e decrescimento monótono com o aumento da distância ao neurônio vencedor. Considerando  $d_{ij}$  a distância entre o  $BMU$  e um neurônio  $j$  pertencente à vizinhança topológica então define-se a função vizinhança  $h_{ij}$ :

$$h_{ij} = e^{\frac{-d_{ij}^2}{2\sigma^2}} \quad (6)$$

sendo  $\sigma$  o raio da vizinhança topológica, que também é atualizado a cada iteração. Seguindo o que é apresentado em [2] fez-se uso do decaimento exponencial como atualizador do raio da vizinhança:

$$\sigma(t) = \sigma_0 e^{\frac{-t}{\tau_1}} \quad (7)$$

sendo  $\sigma_0 = 3$  o raio inicial da vizinhança e  $\tau_1$  o número total de iterações.

Outro parâmetro utilizado na atualização de pesos sinápticos é a taxa de aprendizagem  $\eta$ , que inicia com valor unitário e é atualizada por:

$$\eta(t) = \eta_0 e^{\frac{-t}{\tau_1}} \quad (8)$$

Após a definição dos parâmetros pode-se construir a equação para atualização dos pesos sinápticos  $w_j$ :

$$w_j(t+1) = w_j(t) + \eta(t)h_{ij}(t)(x - w_j(t)) \quad (9)$$

que é aplicada ao neurônio vencedor e a todos os pertencentes à vizinhança topológica. A equação (9) move o vetor de peso sináptico em direção ao vetor de dado de entrada  $x$ , melhorando sua resposta ao apresentar dados semelhantes.

## 5 Resultados

Como definido anteriormente, fez-se uso de uma grade de neurônios quadrada com dimensão 5x5. Os dados foram apresentados de maneira aleatória para garantir a não existência de vício durante o processo de treinamento.

A seguir apresenta-se duas imagens, relativas ao dia 22/10/12.

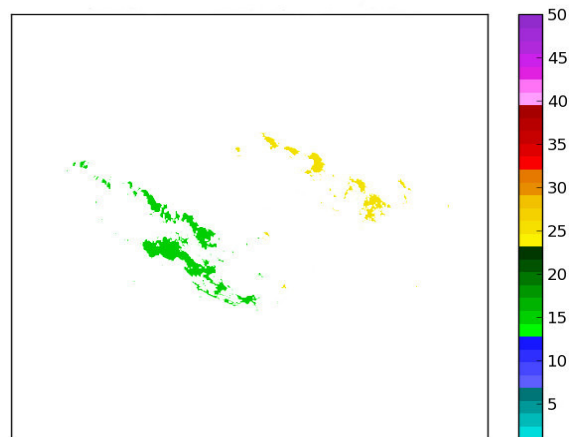


Figura 2: Agrupamento - 22/10/12 às 19h30

Observando a figura 2 verifica-se uma satisfatória separação dos dados em dois grupos principais, que nessa imagem são representados pelas cores verde e amarelo.

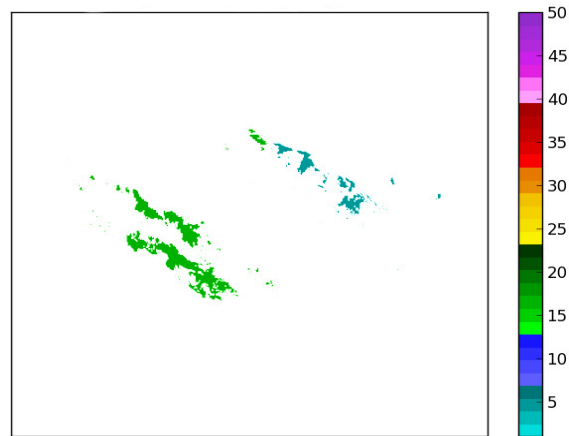


Figura 3: Agrupamento - 22/10/12 às 19h50

Na figura 3 nota-se também boa representação dos dados em grupos distintos, representados em verde e azul.

Tanto na figura 2 quanto na figura 3 as cores de cada grupo representam seu número de identificação, que será utilizado para acompanhamento, determinação de parâmetros e previsão de desenvolvimento e deslocamento futuros.

## 6 Conclusões

Ao final do trabalho verificou-se que a escolha das variáveis, refletividade e respectiva posição na matriz de dados, aliada à normalização dos dados, definição dos parâmetros e escolha das funções de atualização, gerou bons resultados. Em trabalhos futuros pretende-se utilizar outros dados meteorológicos, entre eles as alturas de refletividade máxima e de refletividades significativas (30, 45 e 60 *dBZ*) além de usar o resultado da clusterização afim de buscar melhorias em técnicas de previsão do desenvolvimento de eventos meteorológicos severos, através da escolha de limiares de intensidade e área mais condizentes com a região em estudo.

## Referências

- [1] A. C. P. de L. F. Carvalho and A. P. Braga and T. B. Lauremir, “Fundamentos de Redes Neurais Artificiais”, DCC/IM, Rio de Janeiro, 1998.
- [2] S. Haykin, “Redes Neurais: princípios e prática”, Bookman, Porto Alegre, 2001.
- [3] T. Kohonen, “Self-Organizing Maps”, SPRINGER, New York, 1995.
- [4] Y. Lu and T. Ma and C. Yin and X. Xie and W. Tian and S. Zhong, Implementation of The Fuzzy C-means Clustering Algorithm in Meteorological Data, *International Journal of Database Theory and Application*, 23 (2013) 1-18.
- [5] R. E. Rinehart, “Radar for Meteorologists”, Rinehart Publishing, Nevada, 2004.
- [6] Z. Suraj and J. F. Peters and W. Rzasa, A Comparision of Different Decision Algorithms Used in Volumetric Storm Cells Classification, *Fundamenta Informaticae*, 51 (2002) 201-214.

- [7] D. A. Vila and L. A. T. Machado and H. Laurent and I. Velasco, Forecast and Tracking the Evolution of Cloud Clusters (ForTraCC) Using Satellite Imagery: Methodology and Validation, *Weather and Forecasting*, 23 (2008) 233-245.