

Uso da Rede Neural Artificial Self Organizing Maps (SOM) na clusterização de dados meteorológicos

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

Programa de Pós-Graduação de Métodos Numéricos em Engenharia, PPGMNE, 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: *O agrupamento de dados, a partir de suas características, torna sua análise menos trabalhosa, uma vez que dados pertencentes a um mesmo grupo podem ser tratados de maneira semelhante. Uma de suas aplicações está na meteorologia, quando se deseja agrupar dados de nuvens por exemplo. Nesse trabalho, buscou-se clusterizar dados meteorológicos de refletividade de radar, através do uso da Rede Neural Artificial SOM.*

Palavras-chave: *Rede Neural Artificial, Clusterização, Dados Meteorológicos de Refletividade*

1 Introdução

A clusterização de dados meteorológicos é ferramenta de grande importância na análise de eventos severos. Em Villa (2008) a clusterização das nuvens se dá através da busca de pixels, em imagens de satélite, pertencentes a dado intervalo, e que após identificados, passam por dilatação e erosão matemática. Outras técnicas usadas podem ser vistas em Ester et al (1996) e MacQueen (1967).

No presente trabalho fez-se uso da Rede Neural Artificial SOM (Self-Organizing Maps - Mapas Auto-organizáveis de Kohonen) para clusterizar dados recebidos de um radar meteorológico, do Instituto Tecnológico Simepar, instalado na cidade de Teixeira Soares - Pr que possui alcances quantitativo de 400km e qualitativo de 200km.

Formados por quatro componentes (antena, transmissor, receptor e display), os radares meteorológicos funcionam através do uso de ondas eletromagnéticas, que são emitidas e direcionadas à atmosfera pela antena. Este sinal então viaja a velocidade da luz, até ser refletido em partes por algum objeto presente na atmosfera. Com base no intervalo de tempo, entre emissão e recepção do sinal, determina-se a distância objeto-radar. Também é possível determinar a variável Refletividade, medida em *dBZ* e que varia de acordo com a composição, concentração e estado físico do objeto refletor.

A varredura da atmosfera se dá através da rotação da antena em torno de um eixo vertical. Para cada elevação e ângulo de rotação é emitido um sinal e medido a parcela do sinal refletido. Como resultado da varredura obtem-se dados em coordenadas polares, que para serem aplicados à rede neural fez-se mudança de coordenadas para o sistema cartesiano, com resolução 1km x 1km x 0.5km, gerando então uma matriz de dimensão (200, 200, 10). Após a mudança de coordenadas, determinou-se para cada coluna o maior valor de refletividade, reduzindo assim a dimensão da matriz de dados para (200, 200).

2 Metodologia

Desenvolvida por T Kohonen (1995), a Rede Neural Artificial SOM (Self-Organizing Maps - Mapas Auto-organizáveis de Kohonen) possui aprendizado não supervisionado e apresenta três camadas: a de entrada, a escondida e a de saída. Na camada de entrada introduz-se os dados, na camada escondida está a grade de neurônios, e na de saída apresenta-se o resultado. Por ser do tipo FEED-FORWARD, apresenta somente um caminho para os dados: iniciando na camada de entrada, passando apenas uma vez na camada escondida e então indo à camada de saída.

O aprendizado se baseia em competição e cooperação. Durante o processo de competição os dados são apresentados aos neurônios presentes na camada escondida, e então determina-se o neurônio mais semelhante a cada padrão de entrada. Para cada neurônio vencedor, determina-se uma vizinhança topológica, com neurônios que também terão seus pesos sinápticos atualizados.

Seu algoritmo pode ser descrito pela seguinte sequência:

1. Inicialização dos neurônios e seus pesos sinápticos;
2. Definição de parâmetros;
3. Apresentação dos dados de entrada;
4. Cálculo de similaridade, determinação do neurônio vencedor e definição da vizinhança topológica;
5. Atualização dos pesos sinápticos dos neurônios.

2.1 Definição dos neurônios

A primeira etapa executada durante a clusterização de dados é a definição das características de cada grupo, além da escolha de uma função para medir a semelhança dos dados. Neste trabalho optou-se pelo uso dos dados de refletividade e posição de cada elemento na matriz de dados como padrões de entrada, e do quadrado da distância euclidiana como medida de semelhança. A escolha de uma rede com 25 neurônios se deve ao fato de que há mínima possibilidade de ocorrência de uma quantidade de agrupamentos superior a esse número.

Escolheu-se para cada neurônio, de maneira aleatória, dois valores i e j no intervalo entre 0 e 199, gerando assim um par (i, j) que define uma posição na matriz de dados. Dessa maneira, o peso sináptico do neurônio seria dado pela tripla (Z_{ij}, i, j) , onde Z_{ij} é o valor da posição (i, j) na matriz de dados.

Segundo Carvalho (1998), a normalização dos dados de entrada e pesos sinápticos leva a uma melhoria na convergência da rede SOM. Assim, o peso sináptico de cada neurônio da rede, é dado por: (Z_{ij}^*, i^*, j^*) , onde $i^* = \frac{i}{199}$, $j^* = \frac{j}{199}$ e $Z_{ij}^* = Z_{ij}/\max Z$.

2.2 Parâmetros

Como visto de Kohonen (1995), define-se a taxa de aprendizagem α_0 como tendo valor inicial unitário. Para cada iteração t a taxa de aprendizagem é então dada por:

$$\alpha(t) = \alpha_0 e^{-\frac{t}{T}}$$

sendo $T = 40000$ o número de iterações do processo de treinamento.

Segundo Kohonen (1995), o raio inicial da vizinhança do neurônio deve cobrir praticamente toda a grade de neurônios, e nesse caso definiu-se $r_0 = 3$. Em cada iteração t , tem-se:

$$r(t) = r_0 e^{-\frac{t}{40000}}$$

2.3 Similaridade

A análise de similaridade entre os dados de entrada e os neurônios da camada escondida foi feita usando o quadrado da distância Euclidiana:

$$D(j) = \sum_j (x_j - w_{ij})^2$$

onde i é o índice do neurônio e j as coordenadas do vetor de pesos dos neurônios. O neurônio dito vencedor é aquele cujo valor $D(j)$ é mínimo.

2.4 Vizinhaça

Definido o neurônio vencedor, a próxima etapa é então determinar a sua vizinhaça topológica, formada por todos os neurônios que estão a uma distância menor ou igual ao raio $r(t)$.

2.5 Pesos Sinápticos

A atualização dos pesos sinápticos decresce com o aumento da distância entre os neurônios vizinhos e o neurônio vencedor. Os pesos sinápticos são atualizados segundo a equação:

$$w_{ij}(t+1) = w_{ij}(t) + \alpha(t)\beta(t)(x_i - w_{ij})$$

onde $\alpha(t)$ é a taxa de aprendizagem na iteração t , $r(t)$ é o raio de busca na t -ésima iteração, $\beta(t)$ é definida por:

$$\beta(t) = e^{\frac{-d_k^2}{2r(t)^2}}$$

e d_k é a distância topológica entre o neurônio k e o neurônio vencedor.

A escolha de β , tal qual mostrada acima, se deve ao fato dela ser simétrica em relação ao neurônio vencedor e ter sua amplitude decrescendo monotonicamente com o aumento da distância d_k .

3 Resultados

Três das imagens geradas, para o dia 01/04/11, podem ser vistas a seguir:

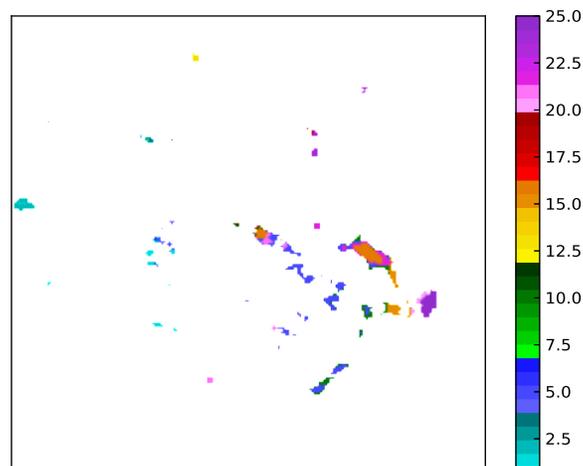


Figura 1: Clusterização - 01/04/11 16h40

Na Figura 1 nota-se uma divisão de clusters em outros menores, causada principalmente pelo deslocamento das nuvens entre o treinamento e a aplicação da rede.

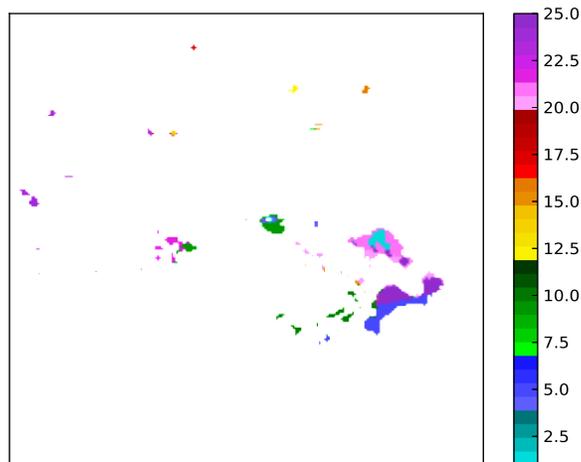


Figura 2: Clusterização - 01/04/11 17h00

Na Figura 2 já há uma melhor definição de cada cluster, com um menor número de divisões em alguns clusters.

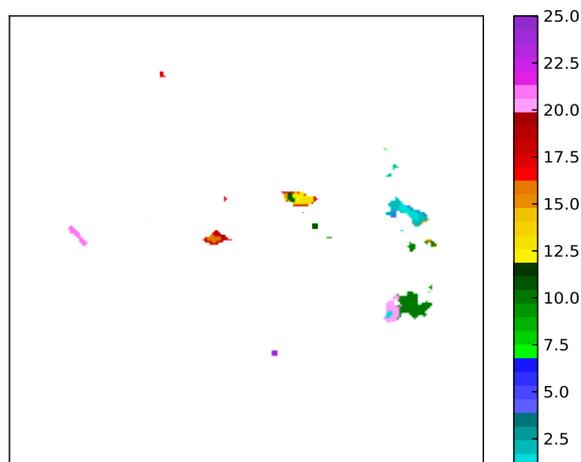


Figura 3: Clusterização - 01/04/11 17h50

Nessa imagem ainda há ocorrência de divisões em cada cluster, mas em alguns dos clusters definidos já nota-se uma melhor classificação, quando comparado às imagens anteriores.

4 Conclusões

Observando as imagens geradas, nota-se divisão de nuvens em mais de um cluster, fato que pode ser explicado pelo deslocamento e decaimento/crescimento (e conseqüente variação nos valores de refletividade e posição) das nuvens entre o momento do treinamento e utilização da rede. Dessa maneira percebe-se a necessidade da inclusão de outras variáveis na rede, entre elas a altura de ocorrência da refletividade máxima e altura do topo da nuvem.

Referências

- [1] A. C. P. de L. F. de Carvalho, "Fundamentos de Redes Neurais Artificiais", DCC/IM, COPPE/Sistemas, NCE-UFRJ, Rio de Janeiro, 1998.
- [2] M. Ester, H-P. Kriegel, J. Sander, X. XU, "A Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noises". In Proceedings of 2nd Int. Con. on Knowledge Discovery and Data Mining (KDD), Portland, 1996, 226-231.
- [3] T. Kohonen, "Self-Organizing Maps", SPRINGER, New York, 1995.
- [4] J. MacQUEEN, "Some Methods for Classification and Analysis of Multivariate Observations". Proceedings of 5th Berkeley Symposium Mathematical Statistics and Probability, University of California Press, 1: 281-297.
- [5] R. E. Rinehart, "Radar For Meteorologists", Rinehart Publishing, Nevada, 2004.
- [6] J. Vesanto, "Clustering of the Self-Organizing Map", *EEE TRANS. ON NEURAL NETWORKS*, 11 (2000) 586-599.
- [7] D. A. Vila, L. A. T. Machado, H. Laurent, I Velasco, "Forecast and Tracking the Evolution of Cloud CLusters (ForTraCC) Using Satellite Infrared Imagery: Methodology and Validation", *Weather and Forecasting*, 23 (2006) 233-245.