

# REDUÇÃO DE DADOS METEOROLÓGICOS APLICADOS A PREVISÃO CLIMÁTICA POR REDES NEURAIAS

*Alex Sandro Aguiar Pessoa<sup>1</sup>, José Demísio Simões da Silva<sup>2</sup> e Hélio Camargo Júnior<sup>1</sup>*

**RESUMO.** Este artigo visa mostrar um método de redução de dados, obtido por meio da mineração de dados, para realização de previsão climática. Embora seja apenas um estudo de caso, o experimento mostra que é possível reduzir as condições de entrada no modelo aplicado neste trabalho, que neste caso é baseado em redes neurais artificiais, sem que os resultados sejam degradados. Este estudo também tem como intuito promover o uso de inteligência artificial na meteorologia como instrumento de auxílio aos métodos físico-matemáticos existentes, sem a pretensão de substituí-los.

**ABSTRACT.** This article aims at to show a method of data reduction, through of the data mining, for make the climatic forecast. Although it is only one study of case, the experiment show that is possible to reduce the initial conditions in the model applied in this work, that in this case is based on artificial neural networks, without the results are degraded. This study also it has as intention to promote the use of artificial intelligence in meteorology as instrument of aid to the methods existing physicist-mathematicians, without the idea of change them.

**Palavras-chave.** Mineração de dados, Redes Neurais e Previsão Climática.

## INTRODUÇÃO

Com o crescente volume de informações e da capacidade computacional, apoiados pelo vertiginoso crescimento tecnológico da computação, a capacidade e disponibilidade de dados não é mais um empecilho para grande parte da comunidade científica. Um problema relevante é o uso de informações de modo eficiente e eficaz.

Na meteorologia o problema não é diferente, pois existem uma enorme quantidade de dados que servem de entrada para os modelos que fazem as previsões, sejam do tempo ou clima, demandado grande quantidade de recursos computacionais e tempo.

Este artigo trata de reduzir a quantidade de informações de entrada de um modelo de previsão climática, chamado de condições iniciais (CI), sem que haja degradação no resultado, ou seja, o resultado do conjunto reduzido de dados das CI tem que apresentar, na pior hipótese, um erro similar ao comparado com as previsões realizadas com o conjunto completo de condições iniciais.

A técnica utilizada para reduzir as CI é conhecida como mineração de dados, que é uma área que contempla a descoberta de padrões escondidos entre os dados, sem que se tenha um

---

<sup>1</sup> Centro de Previsão do Tempo e Estudos Climáticos (CPTEC), Instituto Nacional de Pesquisas Espaciais (INPE), Rodovia Presidente Dutra Km.39, Cachoeira Paulista. CEP:12630-000. Tel.: (12)3186-8541, e-mail: pessoa@cptec.inpe.br, helio@cptec.inpe.br.

<sup>2</sup> Laboratório Associado de Computação e Matemática Aplicada (LAC), Instituto Nacional de Pesquisas Espaciais (INPE), Av. dos Astronautas, 1758, São José dos Campos, CEP: 12227-010. Tel.: (12)3945-6543, e-mail: demisio@lac.inpe.br.

conhecimento prévio destes padrões.

Como modelo para a realização deste experimento optou-se por uso das redes neurais artificiais (RNA), que são estruturas computacionais que visam o aprendizado através do conhecimento, inspirada no funcionamento do cérebro. As redes neurais são amplamente empregadas na inteligência artificial (IA), que é um conjunto de teorias e técnicas empregadas para simular a inteligência humana em sistemas computacionais (Fernandes, 2003).

## **PREVISÃO CLIMÁTICA**

A previsão climática é uma estimativa do comportamento médio da atmosfera com alguns meses de antecedência. Por exemplo, pode-se prever se o próximo verão será mais quente ou mais frio que o normal, ou ainda, mais ou menos chuvoso. Todavia, tal estimativa não pode dizer exatamente qual será a quantidade de chuvas ou quantos graus a temperatura estará mais ou menos elevada. Para efeitos didáticos existe uma distinção entre tempo e clima por parte dos meteorologistas. O principal diferencial entre previsão do tempo e previsão clima é a escala temporal. O clima está associado em geral a previsões de longo prazo, como saber a média de temperaturas do próximo verão. Já a previsão do tempo diz respeito à estimativa para pouco tempo, como horas, dias e semanas.

## **KDD E MINERAÇÃO DE DADOS**

A descoberta de conhecimento em banco de dados (KDD – *Knowledge Discovery in Database*) é o processo de encontrar em um banco de dados padrões que estejam escondidos, sem uma idéia pré-determinada ou hipótese sobre o que são estes padrões (Chen, 2001). Além de prover esta descoberta de informações a KDD também é composta pelas etapas de: definição dos objetivos, limpeza dos dados, transformação dos dados, mineração de dados e interpretação dos resultados

A etapa de maior interesse no processo de descoberta de conhecimento é a mineração de dados, cujo enfoque é a exploração e análise de grandes quantidades de dados para descobrir significativamente modelos e regras. Existem diversas técnicas de mineração de dados, podendo destacar entre tantas: a teoria dos conjuntos aproximativos (*rough sets theory*), teoria dos conjuntos nebulosos (*fuzzy sets theory*), redes neurais artificiais, indução de regras e árvores de decisão.

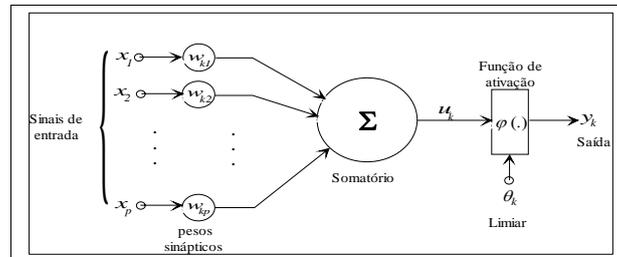
## **REDES NEURAIIS ARTIFICIAIS**

As redes neurais artificiais (RNA) têm sua inspiração no funcionamento do cérebro, tentando assim imitá-lo por técnicas computacionais com o fim de adquirir, armazenar e utilizar conhecimentos (Haykin, 2001). Assim como no modelo biológico a unidade básica de processamento de uma rede neural é o neurônio. A estrutura topológica a qual estes neurônios estão

conectados, em uma RNA é conhecida como arquitetura. Por fim, o processo que adapta a rede a computar uma função desejada é a aprendizagem.

As funções de um neurônio artificial basicamente são:

1. Avaliação dos valores de entrada;
2. Calcular um total para o valor de entradas combinado, ou ponderados pelos pesos  $w_i$ ;
3. Comparar o total com um limiar ( $\theta$ );
4. Determinar a saída.



**Figura 1:** Esquema de um neurônio artificial.

## TEORIA DOS CONJUNTOS APROXIMATIVOS

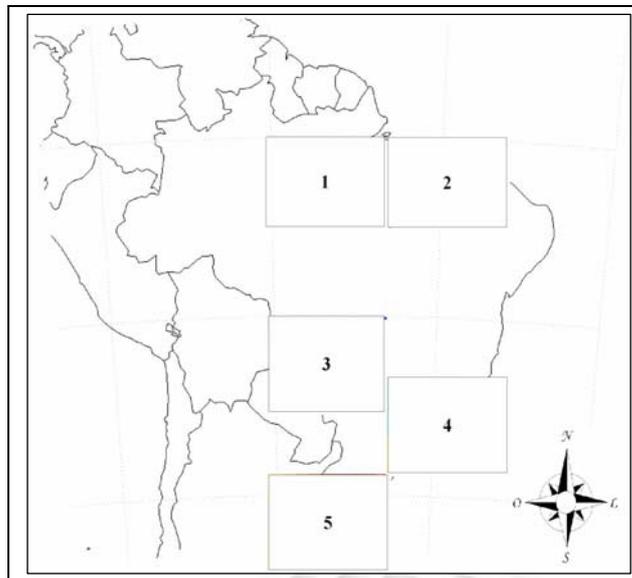
No início da década de 80, surgiu uma teoria, caracterizada pela simplicidade e bom formalismo matemático, o que facilita a manipulação de informações, em especial, incertas, conhecida como **Teoria dos Conjuntos Aproximativos** (TCA), ou do inglês *Rough Set Theory*. A TCA é uma extensão da teoria dos conjuntos, que enfoca o tratamento de incerteza dos dados através de uma relação de *indiscernibilidade* que diz que dois elementos são ditos indiscerníveis, se possui as mesmas propriedades, segundo Leibniz (Scuderi, 2001). A TCA é apoiada sobre dois pilares, um deles já foi mencionado, que é a relação de indiscernibilidade e o outro são as chamadas *reduções*, que é a eliminação de atributos supérfluos para a base de conhecimento analisada. Neste trabalho a TCA foi utilizada para reduzir o número de variáveis de entrada no modelo composto por redes neurais (Pessoa, 2004).

## METODOLOGIA

As análises foram feitas para a previsão de precipitação e temperatura de 1 e 3 estações adiante ( $\Delta$ ). Para a realização das simulações foram coletados dados de Reanálise do NCEP (Kalnay et al., 1996), para o período de janeiro de 1980 a dezembro de 2000 e a área contida entre as latitudes  $[10^\circ \text{ N}, 35^\circ \text{ S}]$  e longitudes  $[80^\circ \text{ W}, 30^\circ \text{ W}]$ , referente à América do Sul, em uma resolução espacial em ambas as dimensões ( $x, y$ ) de  $2.5^\circ$  e resolução temporal ( $t$ ) de 1 mês. As variáveis utilizadas são temperatura do ar, divergência, estação do ano, latitude, longitude, precipitação, umidade específica, pressão da superfície, temperatura, vento zonal - 300, 500 e 850 hPa e vento meridional - 300, 500 e 850 hPa.

Para simplificação e agilidade computacional serão analisadas cinco subáreas da América do

Sul, dispostas a cobrirem uma parte das cinco regiões brasileiras, pois devido a grande extensão territorial do Brasil, há muitos regimes de precipitação e conseqüentemente cada uma com seu clima típico. As regiões são mostradas na Figura 2:



**Figura 2:** Regiões de análise

A análise é constituída de três fases:

1. Calcular as reduções do SI para a região correspondente. Nesta fase são realizadas todas as etapas do processo de KDD. O resultado é um conjunto de reduções onde as variáveis com maior ocorrência são escolhidas como reduções para o treinamento da rede neural.
2. Treinar a rede neural com o SI sem reduções de variáveis, ou seja, com todas as variáveis mostradas anteriormente;
3. Treinar a rede neural com o SI com reduções de variáveis. A partir das reduções provenientes da fase de KDD, a rede é treinada e por fim os resultados são comparados com a rede neural de conjunto de variáveis completa.

## RESULTADOS

A redução de variáveis pode ser vista nas Tabelas 1 e 2, para precipitação e temperatura, respectivamente, ambos para um passo de previsão, ou seja, a partir da entradas é previsto o próximo trimestre. Existe um consenso em relação a todas as regiões de haver uma significativa eliminação de variáveis supostamente supérfluas e as variáveis de espaço-temporal aparecem em todas as reduções, indicando uma correlação entre o resultado e este tipo de variável.

**Tabela 1:** Redução de variáveis (Precipitação e  $\Delta = 1$ )

CO	N	NE	S	SE
estacao	lon	estacao	lon	lon
lon	lat	lat	estacao	lat
lat	estacao	lon	lat	estacao
u500	prec	temp	airt	temp
prec	shum	shum		u300
u850		prec		airt
u300				

**Tabela 2:** Redução de variáveis (Temperatura e  $\Delta = 1$ )

CO	N	NE	S	SE
lat	lon	lon	lon	temp
estacao	estacao	estacao	lat	lon
temp	temp	temp	estacao	estacao
lon	lat	div	airt	lat
airt	u850	prec	temp	
u500	u300			
prec				

As Tabela 3 e 4 apresentam os resultados obtidos para a previsão climática de precipitação, para um e três passos de tempo, onde um passo de tempo é equivalente a um trimestre. São mostrados os erros médios para o conjunto completo (COMP) e reduzido ( $N_a$ ) de variáveis de entrada, assim como o ganho, ganho expressa, em porcentagem, se houve uma redução do erro médio do conjunto treinado com os atributos reduzidos em relação a COMP. A diminuição do erro é maior em  $N_a$ , para as previsões de temperatura com 3 passos de tempos ( $\Delta=3$ ), assim como ocorreu na previsão de precipitação.

**Tabela 3:** Erro médio (Precipitação)

Região	$\Delta=1$		Ganho	$\Delta=3$		Ganho
	COMP	$N_a$	(%)	COMP	$N_a$	(%)
CO	1.72	1.62	6.04	1.42	1.27	10.31
N	1.60	1.42	11.10	1.69	1.36	19.77
NE	1.74	1.78	-2.46	2.09	1.86	11.11
S	1.22	1.30	-6.56	1.44	1.04	27.42
SE	1.61	1.20	25.59	1.83	1.23	32.63

**Tabela 4:** Erro médio (Temperatura)

Região	$\Delta=1$		Ganho	$\Delta=3$		Ganho
	COMP	$N_a$	(%)	COMP	$N_a$	(%)
CO	1.30	1.44	-11.05	1.98	1.35	31.81
N	0.38	0.48	-25.55	0.47	0.42	11.00
NE	0.47	0.50	-7.25	0.55	0.52	6.62
S	1.50	0.93	37.74	1.43	0.96	32.59
SE	1.00	0.62	37.91	0.84	0.73	13.63

## CONSIDERAÇÕES FINAIS

Com a aplicação da redução de variáveis, para  $\Delta = 1$ , em 57% dos casos houve um ganho (diminuição do erro quadrático médio - EQM) do conjunto reduzido ( $N_a$ ) em relação ao conjunto completo de atributos (COMP), para a realização de previsão climática para as variáveis de temperatura e precipitação. Já para  $\Delta = 3$  o ganho na diminuição do EQM ocorre em 100% dos casos. Embora isto possa parecer paradoxal, previsões para alcance maiores sejam melhores, pode ser explicado pelo ciclo sazonal, pois quando se prevê 3 estações adiante, está prevendo quase no final do ciclo, onde as condições climáticas estão voltando a ser como eram no início do ciclo.

Portanto a metodologia utilizada no contexto global, tanto na previsão climática de temperatura como na precipitação mostrou-se eficaz, considerando pontos fortes:

- Possibilidade de implementação para tornar-se operacional e realizar previsões climáticas reais, devido a simplicidade de implementação e baixo ônus financeiro;
- Abertura de novos horizontes, tanto no que diz respeito a trabalhos futuros, quanto na aplicação na previsão climática, no uso conjunto desta metodologia com os modelos numéricos em operação, com o propósito de oferecer novos produtos, como a identificação das regiões mais ou menos chuvosas e mais ou menos quentes, além de oferecer maior respaldo as previsões climáticas de modo a somar os conhecimentos já existentes.

## REFERÊNCIAS BIBLIOGRÁFICAS

Cavalcanti, I. F. A. et al. *Global Climatological Features in a Simulation Using the CPTEC-COLA AGCM*. Journal of Climate, **15**, n 27, p. 2965-2988. 2002

Chen, Z. *Data Mining and Uncertain Reasoning: an Integrated Approach.*, John Wiley & Sons, New York 2001.

Fernandes, A. M. da R. *Inteligência Artificial: Noções Gerais*, Visual Books, Florianópolis, 2003.

Haykin, S. *Redes Neurais: Princípios e Práticas*. Porto Alegre: Bookman, 2001.

Lorenz, E. N. *Deterministic non-periodic flow*. J. Atmos. Sci., v. 20, p. 130-141, 1963.

Kalnay, E; Coauthors, 1996. NCEP/NCAR 40-year: NCEP/NCAR 40-year Reanalysis Project. *Bulletin of the American Meteorological Society*, **77**, 437-471.

Pessoa, A.S.A; *Mineração de dados meteorológicos pela teoria dos conjuntos aproximativos na previsão de clima por redes neurais artificiais*. São Paulo, 2004. 132 p. [Dissertação (Mestrado) – Instituto Nacional de Pesquisas Espaciais].