

REDUÇÃO DA DIMENSIONALIDADE DO VETOR DE ENTRADA DE RNA QUALIFICADA EM PREDIÇÃO DO MÓDULO DE DEFORMAÇÃO LONGITUDINAL SEGUNDO ANÁLISE DE COMPONENTES PRINCIPAIS

Brunelle de Oliveira Santos¹; Koji de Jesus Nagahama²; Paulo Vitor Souza Santos³; Anderson de Souza Matos Gadéa⁴

1. Bolsista PROBIC/UEFS, Graduanda em Engenharia Civil, Universidade Estadual de Feira de Santana, e-mail: brunawek@hotmail.com
2. Orientador, Departamento de Tecnologia, Universidade Estadual de Feira de Santana, e-mail: kjnagahama@gmail.com
3. Bolsista PIBIC/FAPESB, Graduando em Engenharia Civil, Universidade Estadual de Feira de Santana, e-mail: pvss.uefs@gmail.com
4. Co-Orientador, Departamento de Tecnologia, Universidade Estadual de Feira de Santana, e-mail: agadea@gmail.com

PALAVRAS-CHAVE: análise de componentes principais, redução da dimensionalidade, redes neurais artificiais.

INTRODUÇÃO

Apesar de seu elevado potencial de aprendizagem e generalização do conhecimento, a utilização de redes neurais artificiais (RNA) requer a calibragem de seus parâmetros o que inclui a escolha de um projeto estatístico eficiente na etapa de mineração de dados. As contribuições de um projeto estatístico conveniente vão desde o aumento da eficiência durante o processo de treinamento das redes à redução da dimensionalidade do vetor de entrada, quesito fundamental para reduzir a complexidade matemática inerente aos modelos não lineares extraídos das associações sinápticas das redes.

Visando aprimorar os estudos até então desenvolvidos (Dantas, 2011), este trabalho aplicou uma técnica de estatística multivariada, a Análise de Componentes Principais (ACP), para reduzir a dimensionalidade do vetor de entrada da RNA qualificada em predição do Módulo de Deformação Longitudinal (MDL).

METODOLOGIA

A ACP representa uma transformação linear ótima (Anton e Rorres, 2004) que, por meio de compactação de energia, possibilita a identificação de variáveis redundantes ou não influentes. Os dados disponíveis para a qualificação da RNA para predição do MDL foram inicialmente padronizados, uma vez que Míngoti (2007) aponta a conveniência do processo de padronização para evitar que as diferentes escalas ou unidades de medida das variáveis influenciem a análise de sensibilidade proposta pela ACP, dos quais se extraiu a Matriz de

Covariância Amostral posteriormente submetida à análise de vetores próprios e valores próprios da Álgebra Linear.

RESULTADOS E DISCUSSÕES

A ACP é uma técnica estatística poderosa que pode ser utilizada para redução do número de variáveis e para fornecer uma visão estatisticamente privilegiada do conjunto de dados (Varella, 2003). Esta análise fornece as ferramentas adequadas para identificar as variáveis mais importantes no espaço das componentes principais. Para Moita Neto (1998), o entendimento exaustivo do assunto requer o conhecimento de operações com matrizes, mas, o mesmo pode ser apresentado por uma abordagem conceitual usando as noções de Álgebra Linear.

Seja $A_{n \times m}$ a matriz que armazena os n experimentos com m variáveis dos ensaios de laboratório para a obtenção do MDL:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1m} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2m} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nm} \end{bmatrix}$$

A padronização das variáveis proposta por Mingoti (2004) resume-se à operação estatística:

$$Z_{ij} = \frac{a_{ij}}{s(a_{ij})}$$

onde:

$s(\cdot)$ = desvio padrão

Seja $Z_{n \times m}$ a matriz resultante do processo de padronização e $S_{m \times m}$ a Matriz de Covariância Amostral dos dados de Z , os resultados algébricos da equação característica que segue definem a ACP:

$$\det[S - \lambda I] = 0 \quad (1)$$

Dado que λ_i é o auto valor associado ao auto vetor v_i , define-se a i -ésima componente principal através da combinação linear:

$$CP_i = v_{i1}VAR1 + v_{i2}VAR2 + v_{i3}VAR3 + \dots + v_{im}VARm(2)$$

A importância de cada Componente Principal é avaliada a partir da proporção de sua Variância Total Explicada – VTE – podendo-se, através dos coeficientes da combinação linear expressa em (2) (loadings), avaliar a significância de cada variável original. Entende-se por pouco influente, a informação associada à variável com maior projeção no eixo da(s) Componente(s) principal(is) cuja VTE não necessite ser adicionada às anteriores para acúmulo mínimo de 70% (Costa Paiva et al., 2010; Varella, 2003). As variáveis com maior

projeção no eixo das Componentes Principais com zero de VTE são definidas como informação redundante.

O modelo para a predição do MDL proposto por Dantas (2011) foi extraído de uma RNA do tipo *feedforward* treinada com o algoritmo *backpropagation* alimentada por vinte e três variáveis de entrada: Taxa de Absorção do Agregado Reciclado – ABG, Taxa de Absorção de Agregado Miúdo Reciclado – ABM, Percentual de Aditivo – ADT, Percentual de Agregado Reciclado de Argamassa – ARG, Taxa de Compensação do Agregado Graúdo Reciclado – CAG, Taxa de Compensação do Agregado Miúdo Reciclado – CAM, Percentual de Agregado Reciclado de Cerâmica – CER, Quantidade de Cimento – CIM, Percentual de Agregado Reciclado de Concreto – COC, Dimensão Máxima do Agregado Graúdo Natural – DGN, Dimensão Máxima do Agregado Graúdo Reciclado – DGR, Dimensão Máxima do Agregado Miúdo Natural – DMN, Dimensão Máxima do Agregado Miúdo Reciclado – DMR, Fator Água Cimento – FAC, Módulo de Finura do Agregado Graúdo Natural – FGN, Módulo de Finura do Agregado Graúdo Reciclado – FGR, Módulo de Finura do Agregado Miúdo Natural – FMN, Módulo de Finura do Agregado Miúdo Reciclado – FMR, Percentual de Outros Materiais Reciclados – OUT, Percentual de Agregado Graúdo Reciclado – PAG, Percentual de Agregado Miúdo Reciclado – PAM, Teor de Agregados – TAG e Teor de Argamassa – TAR.

Com a aplicação da ACP, a dimensão do vetor de entrada foi reduzido para nove, conforme Figura 1.

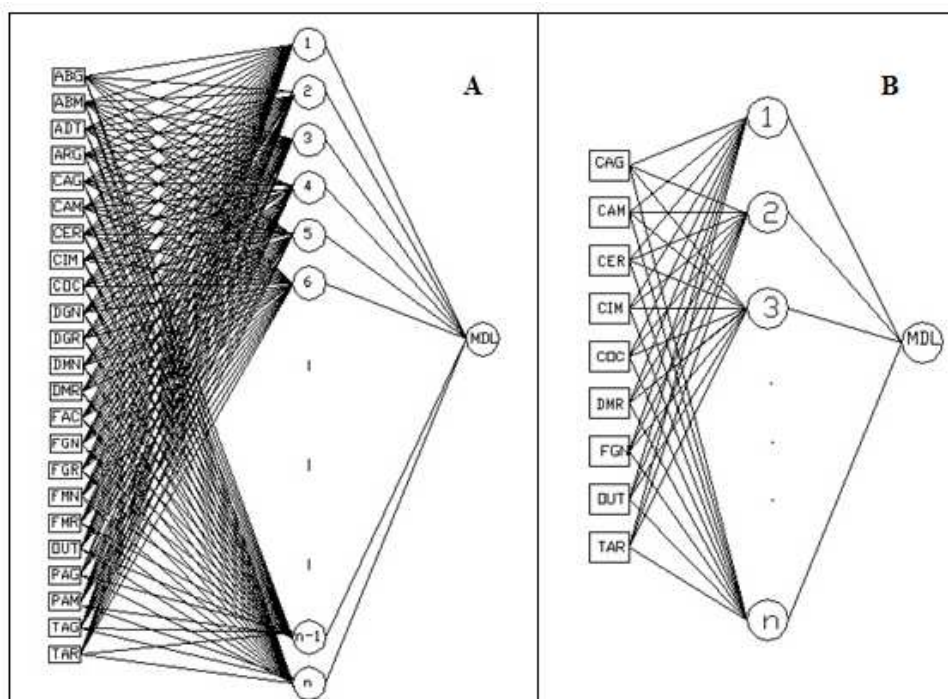


Figura 1. Arquitetura da RNA utilizada para estimativa do MDL. **A** – Dantas (2011). **B** – Redução da dimensionalidade com ACP.

CONSIDERAÇÕES FINAIS

Um modelo com elevada simplicidade algébrica tem a sua precisão comprometida, a determinação de um modelo mais simples e mais preciso possível conduz o aprimoramento da técnica a um problema de otimização onde a análise de sensibilidade das variáveis representa uma etapa fundamental. A ACP mostrou-se uma técnica de mineração eficiente ao aprimoramento dos estudos até então desenvolvidos, dada a dimensão do espaço amostral das variáveis de entrada e sua insensibilidade diante dos dados com características *outliers*. A ACP é uma ferramenta produtiva, prática e eficiente à mineração de dados que alimentam RNA qualificadas para predição de propriedades mecânicas de concretos com RCD constituindo, sobretudo, uma ferramenta fundamental quando se objetiva extrair das RNA modelos de predição com menor complexidade algébrica.

REFERÊNCIAS

- ANTON, H. e RORRES, C. (2004) Álgebra Linear com aplicações, Bookman, Porto Alegre
- COSTA PAIVA, A. L.; TEIXEIRA, R. B.; YAMAKI, M.; MENEZES, G. R. O.; LEITE, C. D. S.; TORRES, R. A. Análise de componentes principais em características de produção de aves de postura. R. Bras. Zootec. vol.39 no.2 Viçosa Feb. 2010
- DANTAS, A. T. A. Utilização de redes neurais artificiais na predição de propriedades físicas e mecânicas de concretos contendo resíduos de construção e demolição. Feira de Santana, 2011. Dissertação (mestrado) – Universidade Estadual de Feira de Santana.
- HAYKIN, Simon S.; ENGEL, Paulo Martins. Redes neurais: princípios e prática. 2.ed. Porto Alegre: Bookman, 2001. 900p. ISBN 857307
- MINGOTI, Sueli Aparecida. Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada. Belo Horizonte: Ed. UFMG, 2007. 295p. ISBN 857041451X (broch.).
- MOITA NETO, J. M.; MOITA, G. C. Uma introdução à análise multivariada de dados. Química Nova, vol. 21, n. 4, 1998. Universidade Federal do Piauí.
- MOITA NETO, J. M. Estatística Multivariada: Uma visão didática-metodológica. Revista de Filosofia – Filosofia da Ciência. Maio de 2004. ISSN 1749-8457
- VARELLA, C. A. A. Análise multivariada aplicada as Ciências Agrárias – Análise de Componentes Principais. Seropédica, 2008 – Universidade Federal Rural do Rio de Janeiro.