

REDES NEURAS ARTIFICIAIS APLICADAS À ANÁLISE DA VARIABILIDADE DE PROCESSOS PRODUTIVOS

Ademyr Gonçalves de Oliveira

Sergio Butkewitsch

Universidade Federal de Uberlândia, Faculdade de Engenharia Mecânica, Laboratório de Projetos Mecânicos, Campus Santa Mônica, Bloco M, 38400-089, Uberlândia, MG, Brasil. E-mail: ademyr@mecanica.ufu.br, sbut@mecanica.ufu.br

Resumo

Este trabalho parte do pressuposto de que processos produtivos são regidos por relações funcionais altamente não-lineares e geralmente indeterminadas entre seus componentes, podendo portanto ser analisados através de técnicas de inteligência artificial, mais especificamente as redes neurais. Inicialmente, para derivar alguns conceitos importantes, os mecanismos de inferência estatística usuais são apresentados. Em seguida, é feito um resumo sobre as características operacionais das redes neurais artificiais, com ênfase nas de maior interesse para a aplicação aqui proposta. Finalmente, é desenvolvido um exemplo ilustrativo baseado na identificação das fontes de variação de um processo produtivo simulado em computador. Nesta aplicação, são usadas, para fins de comparação, tanto as técnicas estatísticas baseadas na análise de regressão quanto as redes neurais artificiais, fornecendo algumas conclusões e sugestões para trabalhos futuros.

Palavras-chave: Redes Neurais Artificiais, Inferência Estatística.

1. INTRODUÇÃO

O estudo científico e sistemático dos processos produtivos (Wheeler, 1993) revela que eles estão, inexoravelmente, sujeitos a causas de variação, que afetam seus elementos constituintes e, conseqüentemente, as características de qualidade dos produtos resultantes. Tais causas de variação podem ser classificadas como na Figura 1 abaixo.

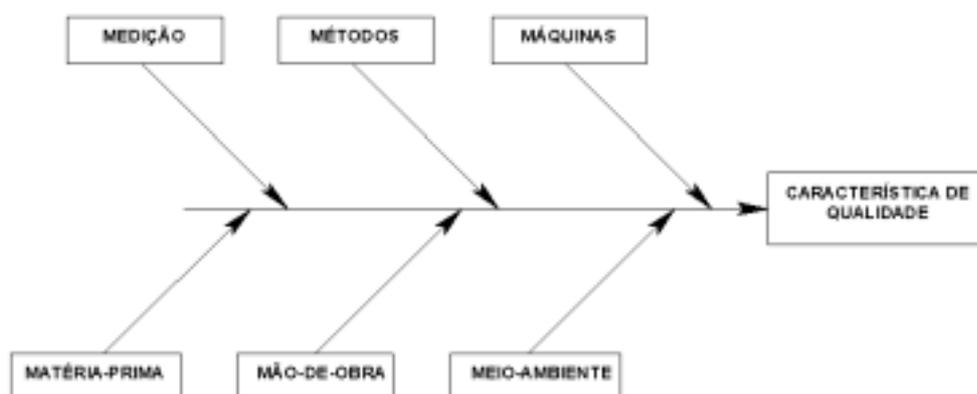


Figura 1. Diagrama de Causa-e-Efeito de Ishikawa

Traduzindo o Diagrama de Causa-e-Efeito de Ishikawa em termos matemáticos, é possível escrever que $CARACTERÍSTICA-DE-QUALIDADE = f(MÁQUINAS, MÉTODO, MEDIÇÃO, MATÉRIA-PRIMA, MÃO-DE-OBRA, MEIO-AMBIENTE)$.

A consequência prática da vigência desta relação funcional é que uma vez que as variáveis independentes (argumentos da função f) variam, a variável dependente, ou seja, qualquer característica de qualidade mensurável do processo produtivo é desviada de sua especificação original, que por sua vez reflete um requisito ao funcionamento adequado do processo produtivo. A melhor maneira de representar este desvio em relação aos requisitos funcionais do processo é através da Função Perda de Taguchi (Oliveira, 1997):



Figura 2. Representação gráfica da função Perda de Taguchi para diferentes especificações

Seja qual for a forma da especificação, o processo é acometido de uma "perda" que cresce quadraticamente à medida em que aumenta a variação.

Assim sendo, torna-se imprescindível que os processos produtivos sejam configurados de forma a resultarem em menor perda, ou seja, que suas características de qualidade mensuráveis apresentem:

- Máxima conformidade em relação a valores de referência especificados;
- Mínima dispersão em torno dos valores de referência estabelecidos;
- Máxima robustez em relação às variações não-controláveis.

O cumprimento destes requisitos (sobretudo os dois primeiros, uma vez que o terceiro envolve procedimentos além do escopo deste trabalho) é tanto mais facilitado quanto maior é o conhecimento disponível acerca da forma da função f , que relaciona a variação nos elementos do processo à variação em suas características de qualidade mensuráveis. Em outras palavras, o interesse reside em responder a perguntas do tipo: "Variando-se o fator M_i do processo produtivo, qual o efeito esperado nos resultados?".

Este tipo de conhecimento, entretanto, é de difícil obtenção, uma vez que as relações de causa e efeito num processo produtivo razoavelmente complexo tendem a ser altamente não-lineares, o que coloca uma série de impedimentos à elaboração de modelos que relacionem as "entradas" ($MÁQUINAS, MÉTODO, MEDIÇÃO, MATÉRIA-PRIMA, MÃO-DE-OBRA, MEIO-AMBIENTE$) às saídas ($CARACTERÍSTICA-DE-QUALIDADE$).

Este trabalho apresenta uma proposta de ferramenta analítica para a elaboração de modelos de variabilidade em processos produtivos. Para tanto, são usados instrumental estatístico de inferência (análise de regressão) e a técnica das Redes Neurais Artificiais, algoritmos matemático-computacionais que, entre outras propriedades, apresentam alta capacidade de reconhecimento de padrões.

2. TEORIAS DE BASE

2.1 Análise de Regressão e Planejamento Experimental

A análise de regressão é uma técnica estatística pela qual um conjunto de dados é usado para estabelecer uma relação funcional entre variáveis dependentes e independentes. Inicialmente, um conjunto de valores de resposta (em termos gerais, "saídas", e no presente contexto, Características de Qualidade) é obtido para uma série de valores de "entrada" correspondentes. Em seguida, a base de dados assim constituída é usada para interpolar uma expressão analítica qualquer (polinomial, exponencial, etc...), ou seja, uma equação, que represente a resposta em função das "entradas".

Para o caso de uma regressão não-linear múltipla (a variável dependente está sujeita a mais de uma variável independente), tal equação é representada por:

$$\{Y\} = [E] \cdot \{B\} + \{\delta\} \quad (1)$$

onde $\{Y\}$ é o vetor das respostas, $[E]$ a matriz das variáveis independentes e $\{B\}$ o vetor dos coeficientes das variáveis independentes, mais um termo livre.

Na verdade, a equação obtida através deste procedimento fornece os estimadores das respostas em função dos dados de entrada. Entre estes estimadores previstos pelo modelo e os valores efetivos das respostas existem diferenças, daí a inclusão de um vetor dos resíduos $\{\delta\}$ na formulação. Para que este processo de estimação seja o mais preciso possível (resíduos mínimos, idêntica, independente e normalmente distribuídos), é necessário definir os coeficientes mais acertados para cada termo da equação interpolada. Via de regra, isto é feito através do Método dos Mínimos Quadrados:

$$\{B\} = \left([E]' \cdot [E] \right)^{-1} \cdot [E]' \{Y\} \quad (2)$$

A equação (2) acima é o modelo para o ajuste generalizado (múltiplo e não-linear) por Mínimos Quadrados. A matriz de variâncias e covariâncias normalizada, dada pelo termo $([E]' \cdot [E])^{-1}$, é uma importante métrica do erro cometido no processo de estimação dos parâmetros do modelo e ressalta a importância dos valores de entrada $[E]$. Todo o processo de estimação é baseado nesta matriz e ela deve, portanto, ser representativa da relação funcional vigente entre as variáveis dependentes e independentes, ao longo de todas as faixas de valores destas variáveis.

Esta representatividade pode ser obtida programando a obtenção dos valores de entrada mediante técnicas de Planejamento Experimental. Experimentos Estatisticamente Planejados são combinações especiais dos fatores cujos efeitos nas respostas se desejam estimar através da análise de regressão.

Diversos pesquisadores vêm propondo uma série de configurações para os arranjos experimentais (Montgomery, 1996). Caso sejam adotadas na coleta de dados para uma análise de regressão, $[E]$ passa a se chamar Matriz Experimental, com a vantagem já apontada no parágrafo anterior. As principais categorias de arranjos experimentais são:

- Experimentos Fatoriais Plenos;
- Experimentos Fatoriais Fracionários;
- Experimentos Centrais Compostos;
- Experimentos Ortogonais;
- Experimentos D-Ótimos e A-Ótimos.

2.2 Redes Neurais Artificiais

Redes neurais são modelos computacionais que podem ser tratados como um dispositivo "caixa preta" que aceita uma entrada e produz uma saída. São potencialmente eficazes nas tarefas de remoção de ruído, classificação, reconhecimento e complementação de padrões.

Uma rede neural compõe-se de um conjunto de unidades de processamento ou neurônios, uma regra de propagação que determina como as informações percorrem os neurônios e uma função de ativação que produzirá um valor de saída(a) do neurônio a partir de um conjunto de valores de entrada(X) e um valor de *offset*(θ).

O neurônio é o elemento responsável pelo processamento numérico da rede, atua de forma isolada dos demais, realiza apenas operações matemáticas predefinidas, só tem acesso às informações locais e produz somente um valor de saída.

As sinapses ou ligações entre neurônios são caminhos lógicos por onde os sinais fluem de um neurônio para outro. Cada sinapse, em particular, tem um peso característico que amplifica ou atenua o sinal transportado como mostra a Figura 3.

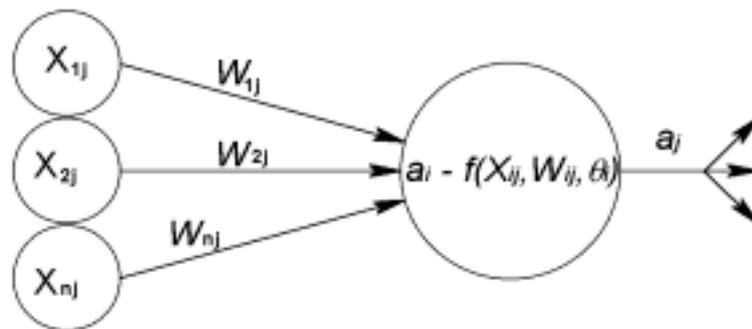


Figura 3. Esquema de processamento de um neurônio

As funções de ativação são expressões matemáticas que convertem um domínio em outro. A maioria das redes neurais empregam funções de ativação simples (Oliveira, 1999): função de ativação linear, degrau, rampa, sigmoideal, gaussiana, etc..

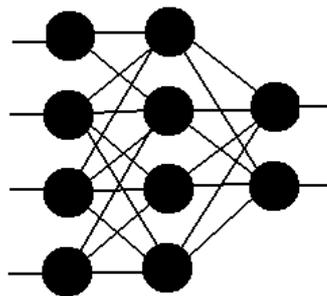


Figura 4. Esquema de sinapses entre neurônios de uma Rede Neural

A topologia das redes neurais refere-se à maneira pela qual a rede neural está organizada: o número de camadas, o tipo de ligação entre os neurônios e como os dados fluem através dela. Na rede neural unidirecional (*feed-forward networks*) os dados entram na rede pela camada de entrada, percorrem as camadas invisíveis, onde são processados e saem pela camada de saída da rede.

Nenhum neurônio recebe sinal de outro da mesma camada, e tampouco de uma camada subsequente. Na rede neural recorrente (*feedback networks*) os neurônios são retroalimentados.

2.3 O Algoritmo *Back-propagation*

Rumelhart et al. (1986) propôs um algoritmo que permite ajustar os pesos das sinapses a partir da camada de entrada para as camadas invisíveis. O erro de cada camada invisível é obtido retrocedendo os erros a partir da camada de saída. Por essa razão este algoritmo é denominado *back-propagation learning rule*. Esse algoritmo pode ser considerado uma generalização da regra delta para funções de ativação não-lineares e rede neural multicamadas (Oliveira, 1999).

O treinamento das redes neurais é obtido ajustando os pesos das ligações entre os neurônios através de um método de otimização. A variação dos pesos é proporcional ao erro residual calculado para cada camada da rede. A constante de proporcionalidade é a taxa de aprendizagem. Uma evolução desse método considera parte da informação do gradiente da iteração anterior, o que permite o otimizador se libertar de mínimos locais pouco significantes. Esta parcela é dita termo de inércia.

Apesar de que o algoritmo de aprendizagem de *Back-propagation* estar baseado numa metodologia sólida, método do gradiente descendente, existem alguns aspectos que não garantem a sua aplicabilidade universal.

O elevado custo computacional de treinamento pode ser provocado por valores inadequados da taxa de aprendizagem e do fator de inércia. A maioria dos métodos avançados de treinamento, baseado no método de *Back-propagation*, têm algum dispositivo de ajuste automático da taxa de aprendizagem. Além disso o número adequado de neurônios é decisivo com relação ao seu treinamento. Muitos neurônios implica em elevado custo computacional de treinamento e rede neural de baixa eficiência. Poucos neurônios implica em instabilidade numérica do processo de treinamento.

A quantidade e a representatividade dos dados apresentados à rede durante o seu treinamento é fator essencial para a obtenção de uma rede neural de qualidade.

3. ESTUDO DE CASO

3.1 Sistema Proposto

Para testar as ferramentas de modelagem de variabilidade apresentadas nas seções precedentes, foi idealizado um processo produtivo com uma resposta dependente de três variáveis de entrada, conforme esquematizado na Figura 5 abaixo:

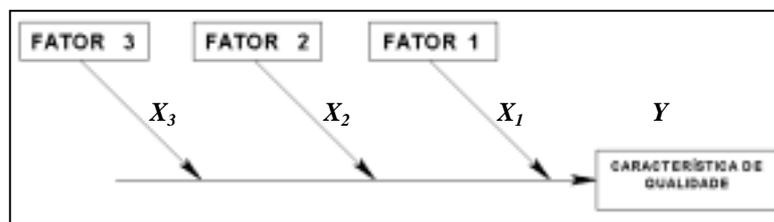


Figura 5. Representação esquemática de um processo de produção para exemplo ilustrativo

Com o intuito de simplificar a verificação da eficácia dos métodos de modelagem, foram supostas três possíveis relações funcionais conhecidas e completamente determinadas entre as entradas e a saída do sistema esquematizado acima:

- Função 1: Relação linear em todas as variáveis independentes mais um termo livre:

$$Y = a_0 + a_1 \cdot X_1 + a_2 \cdot X_2 + a_3 \cdot X_3 \quad (3)$$

- Função 2: Relação quadrática em todas as variáveis independentes mais um termo livre:

$$Y = b_0 + b_1 \cdot X_1^2 + b_2 \cdot X_2^2 + b_3 \cdot X_3^2 \quad (4)$$

- Função 3: Relação com termos lineares, quadráticos e interações de segunda ordem em todas as variáveis independentes, mais um termo livre:

$$Y = c_0 + c_1 \cdot X_1^2 + c_2 \cdot X_2^2 + c_3 \cdot X_3^2 + c_4 \cdot X_1 + c_5 \cdot X_2 + c_6 \cdot X_3 + c_7 \cdot X_1 \cdot X_2 + c_8 \cdot X_1 \cdot X_3 + c_9 \cdot X_2 \cdot X_3 \quad (5)$$

Os valores dos coeficientes das equações acima foram gerados aleatoriamente e retidos para o cálculo dos resultados das funções (vide Tabela 1 abaixo).

Tabela 1. Valores de referência dos coeficientes das funções

Função 1	a0	a1	a2	a3						
	0.95	0.23	0.61	0.49						
Função 2	b0	b1	b2	b3						
	0.89	0.76	0.46	0.02						
Função 3	c0	c1	c2	c3	c4	c5	c6	c7	c8	c9
	0.82	0.44	0.62	0.79	0.92	0.74	0.18	0.41	0.94	0.92

3.2 Dados de Entrada

Os dados de entrada foram dispostos em um experimento do tipo Fatorial Pleno a 3 níveis, para possibilitar regressões não-lineares do segundo grau (três pontos definem uma parábola). As 27 combinações requeridas por este arranjo estão na Tabela 2 abaixo:

Tabela 2. Arranjo experimental empregado no exemplo ilustrativo

COMBINAÇÕES	1	2	3	4	5	6	7	8	9
FATOR 1	-1	-1	-1	-1	-1	-1	-1	-1	-1
FATOR 2	-1	-1	-1	0	0	0	1	1	1
FATOR 3	-1	0	1	-1	0	1	-1	0	1
RESPOSTA	Y ₁	Y ₂	Y ₃	Y ₄	Y ₅	Y ₆	Y ₇	Y ₈	Y ₉
COMBINAÇÕES	10	11	12	13	14	15	16	17	18
FATOR 1	0	0	0	0	0	0	0	0	0
FATOR 2	-1	-1	-1	0	0	0	1	1	1
FATOR 3	-1	0	1	-1	0	1	-1	0	1
RESPOSTA	Y ₁₀	Y ₁₁	Y ₁₂	Y ₁₃	Y ₁₄	Y ₁₅	Y ₁₆	Y ₁₇	Y ₁₈
COMBINAÇÕES	19	20	21	22	23	24	25	26	27
FATOR 1	1	1	1	1	1	1	1	1	1
FATOR 2	-1	-1	-1	0	0	0	1	1	1
FATOR 3	-1	0	1	-1	0	1	-1	0	1
RESPOSTA	Y ₁₉	Y ₂₀	Y ₂₁	Y ₂₂	Y ₂₃	Y ₂₄	Y ₂₅	Y ₂₆	Y ₂₇

onde os códigos -1, 0 e 1 correspondem respectivamente aos valores mínimos, intermediários e máximos das variáveis independentes.

Num primeiro momento, os dados de entrada são utilizados tal qual na Tabela 2. Posteriormente, ruídos randômicos crescentes (1%, 5%, 10%, 50% e 100% da média dos valores de cada fator) são adicionados para avaliar a robustez dos métodos propostos.

Todas as modalidades de dados assim geradas são submetidas a uma família de redes neurais com e neurônios na camada de entrada, o neurônios na camada oculta e sempre três neurônios na camada de saída, cada qual destinado, simultaneamente, a representar uma das funções consideradas neste estudo de caso. Na seção de resultados, cada uma das topologias de rede empregadas estão representadas pelo código $e X o X 3$.

4. APRESENTAÇÃO E ANÁLISE DE RESULTADOS

A métrica adotada para avaliar os resultados da análise de regressão é o percentual da variância explicada pelo modelo, expresso pelo coeficiente de regressão múltipla R^2 :

Tabela 3. Resultados da Variância Explicada (R^2) para a análise de regressão:

Nível de Ruído Aleatório	Função 1	Função 2	Função 3
0% (Variação Determinística)	100.000%	100.000%	100.000%
1% (Variação Aleatória)	99.999%	99.586%	99.999%
5% (Variação Aleatória)	99.968%	99.571%	99.976%
10% (Variação Aleatória)	99.818%	99.613%	99.827%
50% (Variação Aleatória)	97.218%	97.081%	97.787%
100% (Variação Aleatória)	93.012%	92.972%	95.874%

A primeira linha da Tabela 3 apresenta o resultado (óbvio) do caso ideal em que os dados envolvidos são determinísticos. Neste caso, toda a variância é explicada pelo modelo de regressão.

A partir da segunda linha, por outro lado, são listados os casos em que interfere um ruído aditivo aleatório na entrada, cujo efeito é medido na resposta. Percebe-se que a qualidade do modelo de regressão cai com o aumento do ruído, porém de maneira extremamente lenta, numa demonstração de robustez. Entre os modelos, as diferenças de robustez não foram significativas. Estas observações, embora não levem a conclusões inéditas, são extremamente importantes na aplicação do método estatístico aqui exposto.

Os resultados advindos da aplicação da técnica das redes neurais, por sua vez, são avaliados para todos os níveis de ruído introduzidos (de 1% a 100%), computando-se as médias gerais (todas as combinações de variáveis de entrada para todas as três funções consideradas para o estudo de caso) dos erros relativos entre as estimativas das redes neurais e os valores efetivos das respostas:

Tabela 4. Percentuais médios de erro para avaliação do desempenho das redes neurais

Topologia	Ruído= 1%	Ruído= 5%	Ruído= 10%	Ruído= 50%	Ruído= 100%
1 X 4 X 3	-2.7381014e-3	-1.2638002e-2	-3.3938797e-2	-1.4777761e-1	-3.3947241e-1
2 X 4 X 3	-3.1149805e-3	-1.4890376e-2	-4.0613776e-2	-1.5339190e-1	-3.5260005e-1
2 X 6 X 3	-2.9280539e-3	-1.3911295e-2	-3.6051400e-2	-1.5654788e-1	-3.6464360e-1
2 X 8 X 3	-2.2935612e-3	-8.5232753e-3	-2.7642746e-2	-1.4263394e-1	-2.4921967e-1
3 X 3 X 3	-2.4247190e-3	-1.5212011e-2	-3.6319027e-2	-1.5664673e-1	-3.0946620e-1

Os resultados acima espelham, antes de mais nada, o altíssimo grau de robustez das redes neurais para estimação dos resultados advindos da variabilidade dos parâmetros de entrada do

processo. Isto se verifica mesmo quando são consideradas perturbações aleatórias da mesma ordem de grandeza dos fatores principais (ruído de 100%). Evidentemente, a magnitude dos erros aumenta com o ruído.

Além disso, observa-se que não existe nenhuma relação direta significativa entre o tamanho da rede, definido pela quantidade de sinapses resultantes de sua topologia, e a precisão no processo de estimação das respostas. Este resultado é coerente com premissas teóricas que estabelecem que o aumento do número de neurônios da rede não leva necessariamente a incrementos na qualidade dos resultados.

5. CONCLUSÕES E SUGESTÕES PARA TRABALHOS FUTUROS

Através das considerações e resultados expostos neste trabalho é razoável concluir que os tradicionais métodos estatísticos são uma ferramenta robusta e confiável para a análise da variabilidade de processos. Sua operação unida às modernas técnicas matemático-computacionais da inteligência artificial, como as redes neurais, forma um conjunto completo de ferramentas, que fornecem informações em formatos diferentes, mas complementares em conteúdo e utilidade prática.

Enquanto a análise de regressão apresenta como vantagem a disponibilidade de uma expressão analítica (equação) com toda a perspectiva sobre o funcionamento do processo analisado, as redes neurais constituem um preditor de resultados extremamente robusto, com grau de imunidade superior aos métodos estatísticos tradicionais, mesmo sob a intervenção de níveis de ruído altamente significativos. Para a rede empregada na solução do estudo de caso aqui apresentado, cabe salientar que, uma vez superada a fase de treinamento (responsável por grande parte do esforço computacional associado à implementação), a característica de robustez é mantida, mesmo numa topologia idealizada para abstrair, simultaneamente, os resultados de três funções distintas, cada qual associada a um dos neurônios da camada de saída.

Como sugestões para trabalhos futuros, os autores gostariam de apontar:

- Emprego de outros arranjos experimentais, para avaliar a influência da matriz experimental $[E]$, sobretudo no desempenho das redes neurais artificiais;
- Aplicação da metodologia proposta, com diversos arranjos experimentais diferentes, em dados provenientes de um processo produtivo real, sob influência de erro experimental.

6. REFERÊNCIAS

- Montgomery D.C., 1996, "Design and Analysis of Experiments", John Willey & Sons, 4th Edition, New York, NY, USA
- Oliveira, A. G., 1999, "Técnicas de Caracterização de Excitações em Máquinas Rotativas", Tese de doutorado, UFU, Uberlândia.
- Oliveira, C. A., 1997, "Engenharia da Qualidade - Método Taguchi", publicação interna Belgo Mineira, João Monlevade, 239p.
- Rumelhart, D. E.; Hinton, G. E. e Williams, R. J., 1986, "Learning representations by back-propagations errors", Nature 323. pp. 533 - 536.
- Wheeler, D. J., 1993, "Understanding Variation: The key to manage chaos", SPC Press Inc., Knoxville, TN, USA.