



## **Aplicação da Técnica de Mineração de dados para a Tomada de Decisão em Processos Industriais Envolvendo Laminador a Frio**

**Luís Caldas (FEI) – [lcaldas@fei.edu.br](mailto:lcaldas@fei.edu.br)**

**Fabrizio Leonardi (FEI) – [fabrizio@fei.edu.br](mailto:fabrizio@fei.edu.br)**

**Patrícia Prado Belfiore (FEI) – [patricia.belfiore@fei.edu.br](mailto:patricia.belfiore@fei.edu.br)**

***Resumo:** Este artigo propõe o emprego de etapas da técnica de mineração de dados, como ferramenta do setor de manutenção, para a avaliação do desempenho, tomada de decisão e interferência no laminador de chapas a frio. O procedimento tem como entrada uma base de dados reais com as amostras coletadas do processo de laminação e usa ferramentas de análise estatística multivariada para avaliar a dinâmica do sistema de acionamento dos cilindros do laminador. Vários fatores, a exemplo do nível de óleo nos cilindros hidráulicos, podem alterar a dinâmica da planta, fazendo com que o controlador automático perca eficiência na laminação. Entretanto, essas medidas diretas são inviáveis. A proposta deste trabalho é mostrar que a posição dos centros de agrupamentos se presta para identificar eficientemente essas alterações. O método de agrupamento utilizado é o FCM - fuzzy cluster means. Para ilustrar o procedimento proposto, o modelo de um laminador disponível na literatura é usado para representar a planta e os dados, coletados via simulação.*

***Palavras-chaves :** Cluster; Mineração de dados; Tomada de decisão; Qualidade.*

### **1. Introdução**

A exigência do mercado para melhoria da qualidade nos processos produtivos tem aumentado consideravelmente nos últimos anos. Quando uma não conformidade no material produzido é detectada somente no cliente e leva certamente a sérios prejuízos e destrói a imagem da empresa como fornecedora desse produto. É implícita a empresa ter o domínio das tecnologias envolvidas na sua produção tanto em máquinas e/ou equipamentos como do processo produtivo, em síntese, conhecer o que se produz e a forma como se produz.

Muitas vezes a qualidade do produto final é comprometida não pelos requisitos citados, mas pelo mau funcionamento de componentes os quais se encontram fora das exigências técnicas e provocam a variabilidade no processo produtivo.

Este cenário é realidade nos processos produtivos que envolvem os laminadores que são equipamentos de processamento de metais cuja finalidade é reduzir a espessura do material de entrada a cada passe até a espessura desejada. Essa redução é conseguida por meio de cilindros de trabalho acionados por atuadores hidráulicos montados de maneira a controlar a abertura de passagem do material em processo conhecido como *gap*.

O índice de desempenho está vinculado a precisão na abertura e fechamento dos cilindros que operam, normalmente, compensando a variação da espessura do material de entrada através da variação da força de laminação.



Vários fatores internos e externos podem comprometer a qualidade da resposta do atuador e conseqüentemente alteração da força de laminação, resultando em uma não-conformidade pois apresenta problema na espessura final. Um dos fatores que provoca esse impacto na força de laminação é a variação do nível de óleo no cilindro hidráulico em virtude de vazamentos internos (SIEMENS, 1976), (SIEMENS, 1979), (SIEMENS, 1998). Essa variação do nível do óleo no atuador reflete-se como uma diminuição do curso de trabalho no atuador e, conseqüentemente, na força de laminação. O resultado é a produção de materiais não conformes com problemas nas espessuras finais pois estão fora da tolerância.

Este artigo propõe o uso de ferramentas de análise estatística multivariada para avaliar a dinâmica do sistema de acionamento dos cilindros de trabalho do laminador e, portanto, servindo à tomada de decisão sobre a interferência no equipamento pela equipe de manutenção.

Os dados de entrada compõem uma base de dados reais com as amostras coletadas do processo de laminação. Mostra-se que a posição dos centros de agrupamentos ou *clusters* se presta para identificar eficientemente essas alterações. O método de agrupamento utilizado é o *fuzzy cluster means*. Para ilustrar o procedimento proposto, o modelo de um laminador disponível na literatura é usado para representar a planta e os dados, coletados via simulação.

Este artigo está organizado da seguinte forma. A seção 2 apresenta uma breve introdução sobre mineração de dados e a ferramenta de agrupamento. Na seção 3 apresenta-se a metodologia proposta baseada nos dados levantados e na aplicação da ferramenta de agrupamento. A seção 4 a amostragem dos dados, a seção 5 apresenta e discutem os resultados obtidos e a seção 6 as conclusões do trabalho.

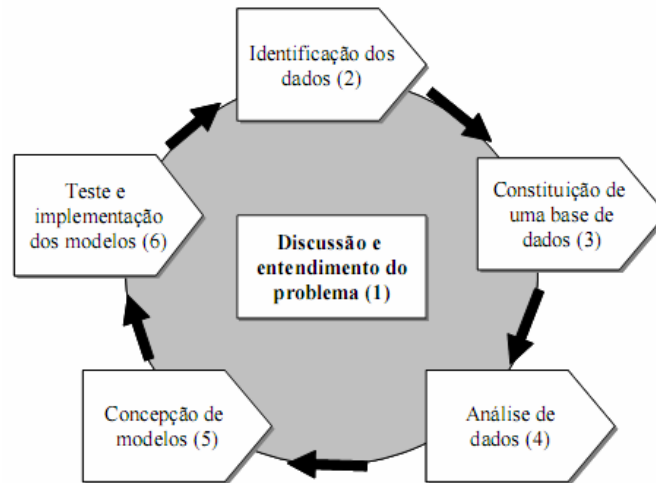
## **2. Mineração de Dados**

Para Bispo (1998), a raiz da mineração de dados está na década de sessenta e tem como base a estatística tradicional. A estatística convencional é baseada em pequenas amostras e utiliza dados comportados e estáticos. A mineração de dados originou-se de técnicas estatísticas clássicas como *regressão*. Com o passar dos anos, um conjunto muito maior de técnicas, como lógica nebulosa (*fuzzy logic*), redes neurais, árvores de decisão e outras técnicas de inteligência artificial foram agregadas à estatística tradicional. A mineração de dados tem as seguintes características: grande volume de dados, dados não comportados, dinâmicos e inconsistentes.

A mineração de dados pode ser definida como a aplicação de técnicas estatísticas e de inteligência artificial em grandes quantidades de dados, com o objetivo de descobrir relações, tendências e padrões relevantes entre os dados (Fayyad, 1996).

Almeida et al. (2005) definem o processo de mineração de dados em seis etapas, conforme mostra a figura 1.

Figura 1: O processo de mineração de dados



Fonte: Almeida et al. (2005)

Segundo os autores, o processo de mineração de dados inicia a *primeira etapa* com a definição do problema, incluindo objetivo e ações para tomada de decisão. A *segunda etapa* de identificação de dados deve determinar as variáveis, a quantidade de dados, o local de extração e o custo dos dados a serem coletados.

Os dados devem ser armazenados de forma adequada, através de banco de dados *data warehouse* (DW) que consiste na *terceira etapa*. A análise estatística de dados (*quarta etapa*) consiste na quantificação da informação, inferência e estimação, de forma a reduzir a incerteza dos dados. Nesta etapa inicia-se o processo de exploração de conhecimento. A *quinta e sexta* etapas constituem as funções e finalidades da mineração de dados. São elas: descrição e visualização, classificação, previsão e agrupamento.

Segundo Passari (2003), a previsão lida com comportamentos futuros. Já a descrição e a visualização têm como objetivo aumentar o grau de compreensão sobre um fenômeno complexo, através de técnicas de estatística descritiva e ferramentas de visualização gráfica. A classificação consiste em identificar as características de um objeto e colocá-los em uma série de classes pré-definidas. Finalmente, o agrupamento ou segmentação consiste em agrupar grupos homogêneos ou *clusters* de uma população heterogênea. Esta técnica também é chamada *análise de clusters*.

A análise de *cluster* é uma das técnicas de mineração de dados. Além da análise de *clusters*, têm-se as técnicas de árvores de regressão, redes neurais, lógica *fuzzy*, entre outras. Neste trabalho utilizaremos a técnica *fuzzy cluster* que utiliza conceitos da análise de *clusters* e da lógica *fuzzy*.

A lógica *fuzzy* é utilizada para capturar informações vagas, geralmente de forma natural, e convertê-las em um formato numérico, de forma a facilitar a sua análise. A técnica trabalha com a teoria dos conjuntos nebulosos (*fuzzy set*) que, em muitos casos, vem alcançando resultados superiores quando comparada à estatística convencional (BISPO, 1998).

A análise de *clusters* ou agrupamento classifica objetos ou indivíduos em relação a algum critério de seleção pré-determinado. Os agrupamentos resultantes de objetos devem



então exibir elevada homogeneidade interna dentro dos agrupamentos e elevada heterogeneidade externa entre agrupamentos. Desta forma, se a classificação for bem sucedida, os objetos dentro dos agrupamentos estarão próximos entre si, e os objetos de diferentes agrupamentos estarão distantes entre si (HAIR et al., 1998).

Como afirmam JOHNSON e WICHERN (1992, p. 573), o método corresponde também a uma importante técnica exploratória que busca identificar uma estrutura de agrupamentos com o intuito de avaliar a dimensionalidade dos dados, identificar *outliers* e fornecer interessantes hipóteses acerca de associações.

Dentre os métodos de análise de *cluster* têm-se os métodos não-hierárquicos e os hierárquicos. No método hierárquico, cada elemento é isolado em um *cluster*. Em seguida há agregações de *clusters* até que se obtenha um único *cluster* com todos os elementos. No método não-hierárquico é preciso definir, a priori, o número de *clusters*. Em geral, utiliza-se uma sub-amostra e o método hierárquico para se obter uma estimativa deste número. Dentre os métodos não-hierárquicos mais utilizados baseados em um ponto central tem-se o método *k-means* (média dos atributos dos objetos).

Segundo Álvares (2003), o método *k-means* exige a definição prévia do número de *cluster* e do posicionamento do centro de cada *cluster* no espaço de atributos. O centro do *cluster* é o chamado centróide, e corresponde ao ponto médio mais central do *cluster*. Esse algoritmo é sensível ao ruído, mas em termos de desempenho é relativamente eficiente para grandes bases de dados.

Álvares (2003) define os passos do algoritmo *k-means*:

- Passo Um : Selecionar dos objetos que serão centros iniciais dos *k cluster*;
- Passo Dois : Associar cada objeto a um *cluster*;
- Passo Três : Recalcular os centros dos *clusters* em função dos atributos de todos os objetos pertencentes ao *cluster*.
- Passo Quatro : Retornar ao passo dois até que os centros dos *clusters* se estabilizem.

Segundo o autor, a cada interação, os objetos são agrupados em função do centro do *cluster* mais próximo e, conseqüentemente, os centros dos *clusters* são reavaliados (passo três), o que resulta no deslocamento dos centros médios. O algoritmo é interrompido quando as médias não mais são deslocadas, ou há uma insignificante realocação de objetos entre os *clusters*.

O objetivo deste artigo consiste na implementação de técnicas de análise estatísticas multivariada baseadas no algoritmo *fuzzy cluster means* para o apoio e a tomada de decisão associada ao controle de processos industriais MISO múltiplas variáveis de entrada e uma variável de saída (processo de laminação, por exemplo). Por meio dessa técnica, propõe-se ;

a) Identificar as variações no sistema de acionamento de cilindros de trabalho em virtude de fatores de mau funcionamento;

b) Quantificar o desempenho do laminador.



O *fuzzy clusters means* (FCM) é um algoritmo de agrupamento de dados no qual cada dado pertence a um agrupamento ou *cluster* com um nível especificado pelo grau de pertinência. Bezdek propôs este algoritmo em 1973. O algoritmo FCM particiona uma coleção de  $n$  vetores  $\mathbf{x}_i$  ( $i = 1, 2, \dots, n$ ) em  $c$  grupos *fuzzy*, e encontra o centro de cada *cluster* em cada grupo, minimizando a função objetivo pela medida da dissimilaridade.

O FCM emprega o particionamento *fuzzy* onde um ponto pode pertencer a vários *clusters* com grau de pertinências diferentes. Para representar o particionamento *fuzzy*, utiliza-se uma matriz  $U$  de pertinência, cujos elementos estão na faixa  $[0, 1]$ . É razoável assumir que pontos na região média entre *clusters*, tenham um grau de pertinência em ambos *clusters*. A pertinência do total de pontos de todos os *clusters*, contudo, deve ser igual a 1 para manter as propriedades da matriz  $U$ .

O algoritmo *fuzzy clusters means* (JANG, 1997), determina os centros dos *clusters*  $c_i$  e a matriz  $U$  de pertinência usando os seguintes passos :

- Passo Um : Inicializar a matriz  $U$  de pertinência com valores aleatórios entre 0 e 1 segundo a restrição;
- Passo Dois : Calcular o *fuzzy-centro*  $c$  dos *clusters*  $c_i$  ( $i = 1, 2, \dots, c$ );
- Passo Três : Determinar a função objetivo : Parar o processamento se o resultado for abaixo de um certo nível ou abaixo de um certo valor de tolerância;
- Passo Quatro : Determinar a nova matriz  $U$ ;
- Passo Cinco : Voltar ao passo dois se os centros dos *clusters* não alteraram de posições.

### 3. Metodologia Proposta

A metodologia utilizada para avaliação do desempenho do laminador extrai padrões de uma base de dados. As formas de aquisições dos conhecimentos de uma base de dados podem ser realizadas de várias formas. Como o laminador é um sistema dinâmico, deve-se incorporar a dependência temporal da resposta do sistema. Isto pode ser feito por meio de regressores aplicados aos estados do modelo.

A base de dados será composta de uma excitação fixa de entrada e da resposta do atuador hidráulico. A base de dados é formada por uma série de *cem* pontos coletados para cada variável durante o processo de laminação e a cada passe, os dados são lidos por meio de dispositivos de medida da espessura final. Os sensores devidamente alocados no equipamento fazem parte do sistema de controle automático de espessura, conhecido como *AGC*. Estes sensores além de enviar os dados ao controle, estes ainda são transmitidos continuamente para um sistema de aquisição de dados que faz a armazenagem.

Estes dados de entrada e saída coletados no processo são normalizados a fim de serem aplicados à ferramenta de análise estatística multivariada, pela técnica de agrupamento. A seguir a ferramenta de agrupamento é aplicada e o resultado obtido são as posições dos

centros dos *clusters* que representam a partição do espaço. O trabalho final para o técnico de manutenção se resume em avaliar se houve mudança da dinâmica da planta pelo deslocamento dos centros.

Para verificar a eficácia da ferramenta e avaliar os resultados foram coletados dois conjuntos de dados do laminador sendo : o primeiro conjunto de dados para referência, os dados com o laminador operando dentro de condições normais e o segundo conjunto de dados com o laminador submetido a fatores que influenciam na resposta dinâmica do sistema.

O laminador utilizado para avaliação é um laminador a frio reversível do tipo quádruplo, sendo dois cilindros de trabalho e dois cilindros de apoio e atuador hidráulico para o controle de abertura e do fechamento do *gap* entre cilindros. A dinâmica do laminador é representada por um sistema dinâmico de segunda ordem e a base de dados coletada proveio de simulação numérica do modelo. O sinal *wxe* representa o sinal de entrada  $u(k)$  aplicado ao modelo do laminador.

Foi mantida, para análise e avaliação do desempenho, a mesma excitação de entrada em todos os ensaios e foram variados parâmetros do modelo dinâmico do atuador hidráulico do laminador.

Os gráficos a seguir, a figura 2a apresenta a resposta do sistema de acionamento dos cilindros de laminação  $x(k)$  e  $x(k+1)$  e a figura 2b apresenta um scatter com  $x(k)$ ,  $x(k+1)$  e a entrada  $u(k)$ .

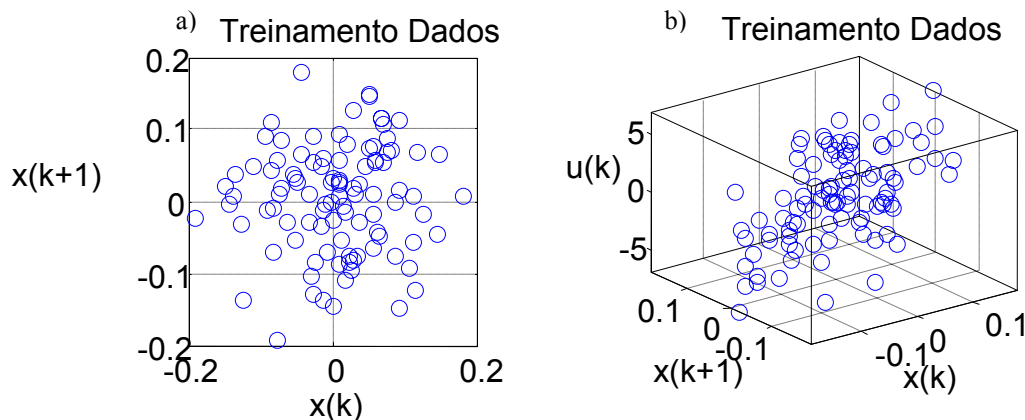


Figura 2 : a) Distribuição dos dados coletados no tempo b) *Scatter* com  $u(k)$ ,  $x(k)$  e  $x(k+1)$ .

#### 4. Amostragem

Foram coletados dois tipos de amostras : a primeira com o laminador operando em condições normais ou seja a sua produção atendendo aos níveis de tolerâncias exigidas conforme padrão de qualidade e a segunda com o laminador operando com as não conformidades de alguns componentes. Para cada uma das coletas realizadas o número de amostras utilizadas para análise e tratamento foram iguais a *cem* sendo que cada amostra continha duas variáveis a saber  $u(k)$ ,  $x(k)$ . A variável  $x(k+1)$  é a resposta do sistema dinâmico que se encontra no estado atual  $x(k)$  recebe uma excitação de entrada  $u(k)$ , em suma  $x(k+1)$  é o estado futuro de  $x(k)$ . O aplicativo para análise dos clusters é um *toolbox* que roda no

ambiente *Matlab*, o qual utiliza o algoritmo FCM – (Mathworks, 1996). Fixando-se em dois *clusters* e adotando-se a distância euclidiana com métrica, foram feitas as simulações para determinação dos centros de cada *cluster*. As duas coletas de dados foram utilizadas para análise dos *clusters* e a seguir os resultados foram comparados entre si para a avaliação da correlação entre as posições dos centros e a dinâmica da planta. A figura 3 mostra os dados de uma das coletas dos dados para a análise dos *clusters* pela FCM.

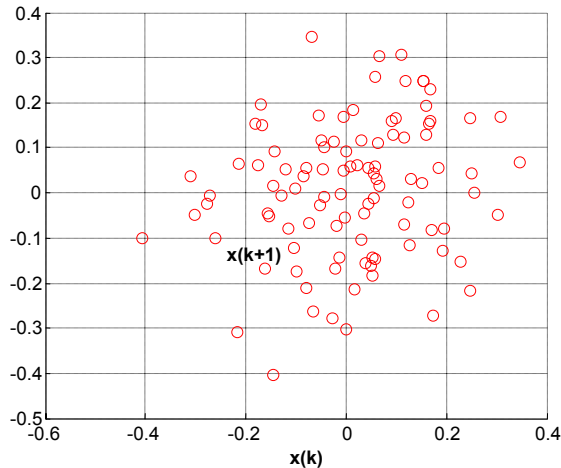


Figura 3 : Amostra de dados coletados para análise dos *clusters* pela técnica FCM.

## 5. Análise dos Resultados

O parâmetro mais significativo na análise foi a frequência natural  $\omega_n$  do modelo. O gráfico dos pontos coletados e os centros para uma operação em condições normais do atuador hidráulico está na figura 4 e o gráfico para uma operação do atuador hidráulico fora das condições normais (alteração de  $\omega_n$ ) está na figura 5.

Pode-se notar pelas figuras 4 e 5 que os centros foram alterados de posições sugerindo uma mudança no comportamento dinâmico da planta em função da alteração da frequência natural  $\omega_n$  num nível suficiente para a tomada de decisão de parada do equipamento para um técnico e corrigir o problema.

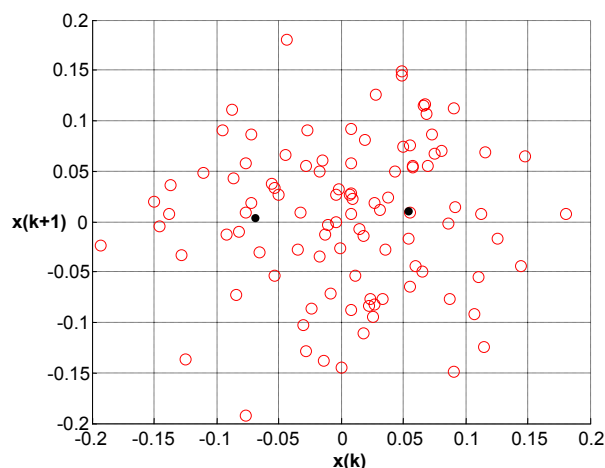


Figura 4 : Distribuição dos pontos do sistema dinâmico em conformidade.

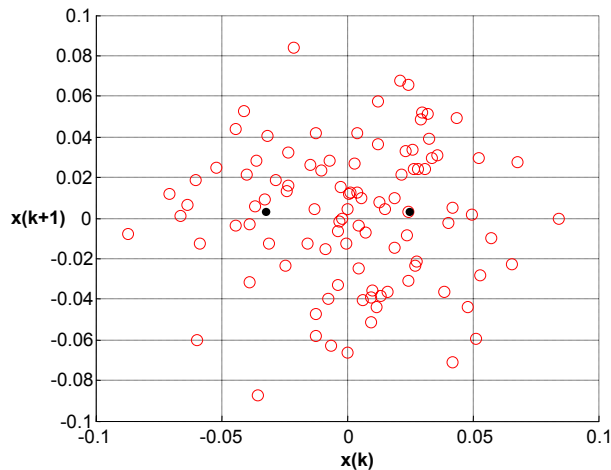


Figura 5 : Distribuição dos pontos do sistema dinâmica fora de conformidade.

## 6. Considerações Finais

O artigo propôs uma solução para um problema bastante conhecido no campo da laminação, para o apoio à tomada de decisão, determinar o instante da intervenção no equipamento para não impactar o processo produtivo mais que o necessário. Algumas soluções focam primordialmente na questão do nível de óleo, que é um dos fatores que podem alterar a dinâmica da planta. É um tanto complicado identificar todos os fatores que são quase ocultos e influenciam significativamente na resposta dinâmica do sistema de atuação.

Muitas tentativas de soluções foram experimentadas com sucesso relativo e outras falharam quando aplicadas, algumas tecnicamente e outras em virtude da operação, pois o ambiente é agressivo e pode ficar contaminado com partículas metálicas oriundas da parede ou da camisa dos cilindros hidráulicos.

A solução proposta neste trabalho está baseada na leitura dos dados em processo e em seguida aplica uma técnica de análise de *cluster*, permitindo inferir alterações na dinâmica da planta de forma simples e, aparentemente, sensível. Pelo gráfico gerado, é visual o nível de deslocamento e posição dos centros dos *clusters*, sendo fácil para o técnico tomar a decisão do instante onde ele deve interferir sobre o equipamento a fim de garantir a qualidade produzida através do restabelecimento das condições mínimas para a operação.

A utilização de uma ferramenta de análise estatística multivariada que é uma etapa da mineração de dados para a tomada de decisão abre uma perspectiva para novos trabalhos no campo industrial e se torna uma ferramenta de apoio para o controle estatístico do processo.

## Referências

Almeida, F. C.; Siqueira, J. O.; Onusic, L. M. Data Mining no Contexto de Customer Relationship Management. *Caderno de Pesquisas em Administração*, v.12, n.2, p.85-97, 2005.

Álvares, L. O. C. *Algoritmos e Ferramentas de Descoberta de Conhecimento em Bancos de Dados Geográficos*. Dissertação (Mestrado). Universidade Federal do Rio Grande do Sul, Porto Alegre, 2003.





Bezdek, J. C. *Fuzzy Mathematics in Pattern Classification*. Tese (Doutorado), Applied Math. Center, Cornell University, Ithaca, 1973.

Bispo, C. A. F. *Uma Análise da Nova Geração de Sistemas de Apoio à Decisão*. Dissertação (Mestrado). Universidade de São Paulo, São Carlos, 1998.

Fayyad, U.; Piatetsky-Shapiro, G.; Smith, P. From Data Mining to Knowledge Discovery: an Overview. In: *Advance in Knowledge Discovery & Data Mining*, p.1-34, 1996.

Hair, J. F.; Anderson, R. E.; Tatham, R. L.; Black, W. C. *Multivariate Data Analysis*. 5 ed. New Jersey: Prentice Hall, 1998.

Jang, J. S. R.; Sun, C. T.; Mizutani, E. *Neuro Fuzzy and Soft-Computing*. Prentice Hall, 1997.

Johnson, R.; Wichern, D. *Applied Multivariate Statistical Analysis*. 3 ed. New Jersey: Prentice Hall, 1992.

Mathworks Inc. *MATLAB Control System Toolbox – User's Guide*, 1996.

Passari, A. F. L. *Exploração de Dados Atomizados para Previsão de Vendas no Varejo utilizando Redes Neurais*. Dissertação (Mestrado). Universidade de São Paulo, 2003.

Siemens AG Industrial and Building Group, Cold Rolling Mills, Processing Lines. *Electrical Equipment and Automation for single stand mills – Solutions for Metals, Mining and More*, 1976.

Siemens AG Industrial and Building Group, Cold Rolling Mills, Processing Lines. *Electrical Equipment and Automation for Cold Strip Tandem Mills – Solutions for Metals Mining and More*, 1979.

Siemens AG Industrial and Building Group, Cold Rolling Mills, Processing Lines. *New Thickness Control Mode Based on Mass Flow Principle Increases Cold Rolling Accuracy – Ideas for Steel*, 1998.