

Aplicação de redes neurais artificiais e regressão por vetores de suporte na modelagem da fração de produtos não conformes de um processo produtivo

Simone Massulini Acosta (UTFPR) – smacosta@utfpr.edu.br
Anderson Levati Amoroso (UTFPR) – andersonl@utfpr.edu.br

Resumo

Para manterem-se competitivas as empresas buscam a qualidade de seus produtos, para que estes atendam às especificações com a menor variabilidade possível. Em um processo industrial muitas vezes não se consegue controlar todas as causas ou fatores de variação que produzem determinado efeito sobre as características da qualidade dos produtos. Como resultado pode-se ter a produção de produtos não conformes às especificações, que podem ser mensurados através da fração de produtos não conformes. A modelagem da fração de produtos não conformes pode ser realizada utilizando-se diferentes técnicas e modelos de regressão. As redes neurais artificiais e a regressão por vetores de suporte são técnicas de aprendizado de máquina muito utilizadas para a modelagem de processos. O objetivo deste artigo é modelar a fração de produtos não conformes às especificações de uma indústria curtidora de couro utilizando as técnicas de redes neurais artificiais e de regressão por vetores de suporte e comparar os resultados obtidos com outros modelos de regressão desenvolvidos para o mesmo processo produtivo. Os modelos de regressão desenvolvidos apresentaram melhor desempenho no ajuste do modelo aos dados do processo quando comparados ao modelo de regressão linear, modelos de regressão lineares generalizados e modelo de regressão Beta.

Palavras-Chaves: Modelagem de dados, aprendizado de máquina, redes neurais artificiais, regressão por vetores de suporte.

1. Introdução

Os processos produtivos estão sujeitos a variabilidade, que está relacionada com as diferenças existentes entre as unidades produzidas. Em um processo industrial, muitas vezes não se consegue controlar todas as causas ou fatores de variação que produzem determinado efeito sobre as características da qualidade dos produtos, pois certas causas são inerentes ao processo. As causas de variação que podem ser controladas podem interferir em um processo podendo gerar produtos com características da qualidade não conformes às especificações

preestabelecidas, que podem ser mensuradas através da fração de produtos não conformes (MONTGOMERY, 2004).

A fração não conforme é definida como a razão entre o número de unidades não conformes da amostra e o tamanho da amostra e podem compreender a razão entre dois números discretos, denominada de percentual, ou a razão entre dois números contínuos, denominada de proporção. As características da qualidade do tipo fração podem ser aproximadas pela distribuição Binomial e são observações expressas no intervalo $[0,1]$ (MONTGOMERY, 2004).

A análise de regressão consiste na modelagem e investigação da relação entre as características da qualidade e as variáveis de controle do processo. Segundo Sant'Anna (2009), os modelos de regressão linear não são adequados para modelar as características da qualidade do tipo fração de produtos não conformes, pois este modelos pressupõem que a característica da qualidade possui distribuição Normal e possibilitam a previsão de valores fora do limite do intervalo $[0,1]$.

Para a modelagem da característica da qualidade do tipo fração podem ser utilizados os modelos pertencentes à classe dos Modelos Lineares Generalizados (MLGs), tais como os modelos de distribuição Binomial e os modelos de quase-verossimilhança (MQVs). Nos MQVs as variáveis são consideradas independentes sem ser necessário especificar qualquer distribuição de probabilidade para a variável resposta (CORDEIRO; DEMÉTRIO, 2008). Outra alternativa é o Modelo de Regressão Beta (MRB), que se baseia na suposição de que a fração segue a distribuição Beta, que assume valores restritos ao intervalo $[0,1]$ (FERRARI; CRIBARI-NETO, 2004).

O aprendizado de máquina trata do desenvolvimento de técnicas computacionais sobre o aprendizado e a construção de sistemas capazes de aprender e melhorar seu desempenho baseado em experiências acumuladas através da solução de problemas anteriores (MITCHELL, 1997). As técnicas de aprendizado de máquina utilizadas para a modelagem do processo neste artigo são as redes neurais artificiais e as máquinas de vetores de suporte.

As Redes Neurais Artificiais (RNAs) são modelos de processamento de dados que são inspirados no que se conhece das redes de neurônios biológicos. Segundo Haykin (2009), a principal propriedade de uma RNA é a sua habilidade de aprender a partir de seu ambiente e de melhorar o seu desempenho através do aprendizado. As RNAs utilizam o princípio de minimização do risco empírico (*Empirical Risk Minimization*, ERM) baseado no princípio da

indução para inferir uma função que minimize o erro sobre os dados de treinamento conhecidos e, desta forma, espera-se que esse procedimento leve também a um menor erro sobre os dados de teste (VAPNIK, 1998).

As Máquinas de Vetores de Suporte (*Support Vector Machine*, SVM) são fundamentadas na teoria de aprendizado estatístico, desenvolvida visando a proposição de técnicas de aprendizado de máquina que buscam maximizar a capacidade de generalização (VAPNIK, 1998). A aplicação de SVMs foi originalmente desenvolvida para problemas de classificação de padrões, sendo posteriormente estendida para o tratamento de problemas de regressão e denominada Regressão por Vetores de Suporte (*Support Vector Regression*, SVR) (SCHÖLKOPF; SMOLA, 2004).

As SVMs utilizam o princípio de minimização do risco estrutural (*Structural Risk Minimization*, SRM) que envolve a minimização de um limite superior sobre o erro de generalização. Desta forma, modelos de aprendizado de máquina baseados no princípio SRM tendem a apresentar uma maior habilidade para generalizar bem frente a dados de teste não observados (ALVES, 2012). Vapnik (1999) demonstrou que o princípio SRM é superior ao princípio ERM, utilizado em redes neurais artificiais.

O objetivo deste artigo é modelar a fração de produtos não conformes às especificações de uma empresa curtidora de couro utilizando redes neurais artificiais e regressão por vetores de suporte e comparar os resultados obtidos com outros modelos de regressão desenvolvidos para o mesmo processo produtivo.

2. Referencial teórico

Nesta sessão são apresentados os referenciais teóricos de redes neurais artificiais e regressão por vetores de suporte.

2.1. Redes neurais artificiais

Segundo Braga, Carvalho e Ludermir (2012), as RNAs são sistemas paralelos distribuídos compostos por unidades de processamento simples (neurônios artificiais), dispostas em uma ou mais camadas e interligadas por um grande número de conexões, que calculam determinadas funções matemáticas (normalmente não lineares). Na maioria dos modelos as conexões estão associadas a pesos, que armazenam o conhecimento adquirido pelo modelo e

servem para ponderar a entrada recebida por cada neurônio da rede.

Existem várias funções de ativação que podem ser utilizadas para o projeto de RNAs. As funções de base radial (por exemplo, gaussiana) e sigmoide (por exemplo, logística e tangente hiperbólica) são totalmente diferenciáveis e facilitam o processo de aprendizado da rede (HAYKIN, 2009).

A arquitetura de uma RNA define a forma como seus neurônios estão arranjados, uns em relação aos outros, através do direcionamento das conexões sinápticas destes neurônios. A topologia de uma RNA pode ser definida como sendo as diferentes formas de composições estruturais que esta poderá assumir.

Para resolver problemas não linearmente separáveis utilizando RNAs, a alternativa mais utilizada são as redes *feedforward* de camadas múltiplas. As camadas escondidas utilizam funções de ativação não lineares. Entre os principais tipos de redes com esta arquitetura está o Perceptron Multicamadas (*Multilayer Perceptron*, MLP) (FACELI et al., 2011).

Segundo Cybenko (1989), o teorema de Kolmogorov garante que uma rede *feedforward* com uma camada escondida, com um número suficiente de neurônios, pode aproximar qualquer função contínua com qualquer grau de precisão. Baseado no teorema de Kolmogorov, o teorema de Hecht-Nielsen apresenta que qualquer função contínua pode ser aproximada por uma rede *feedforward* com m nós de entrada, $2m+1$ neurônios escondidos e k neurônios de saída (FACELI et al., 2011).

Segundo Haykin (2009), a propriedade de importância primordial para uma RNA é a sua habilidade de aprender a partir de seu ambiente e de melhorar o seu desempenho através da aprendizagem. Uma RNA aprende através de um processo iterativo de ajustes dos pesos sinápticos e níveis de bias, tornando-se mais instruída sobre seu ambiente após cada iteração do processo de aprendizagem. Um conjunto preestabelecido de regras para a solução de um problema de aprendizagem é denominado de algoritmo de aprendizagem, tendo-se como objetivo final a generalização de soluções a serem produzidas pelas saídas.

Conforme Braga, Carvalho e Ludermir (2012) para as redes MLP, que se caracterizam por utilizarem funções de ativação sigmoidais nas camadas escondidas e sigmoidais ou lineares na saída, é utilizado o algoritmo de retropropagação de erros ou *error backpropagation*. Este algoritmo utiliza a descida de encosta e estima o erro das camadas escondidas por meio de uma estimativa do efeito que estas causam no erro da camada de saída. Existem várias modificações do algoritmo *backpropagation* que visam acelerar seu tempo de treinamento e

melhorar seu desempenho, sendo uma destas modificações a Levenberg-Marquardt.

2.2. Regressão por vetores de suporte

O algoritmo de regressão por vetores de suporte (SVR) utiliza uma função de perda ε -insensível que ignora erros que estão além de uma certa distância dos valores considerados válidos, ou seja, erros são permitidos somente se forem menores do que ε . A função de perda ε -insensível também garante a esparsidade da variável dual. Assim, na SVR tem-se dois conceitos importantes, que são o ε -tubo e a função de perda ε -insensível (ALVES, 2012).

Seja \mathbf{X} um conjunto de treinamento com n objetos $x_i \in X$ com seus respectivos rótulos (respostas) $y_i \in Y$. Considerando a função linear da equação (1), na SVR busca-se uma função com pequeno w , o que pode ser conseguido pela minimização da norma $\|w\|$ (FACELI *et al.*, 2011).

$$f(x) = \mathbf{w} \cdot \mathbf{x} + b \quad (1)$$

onde w é o vetor de pesos, x é o vetor de entrada e b o termo de limiar (bias).

O problema de minimização pode ser relaxado com a introdução de variáveis de folga (ξ_i), permitindo assim lidar com ruídos e *outliers* nos objetos, permitindo que alguns exemplos fiquem fora da região entre $-\varepsilon$ e $+\varepsilon$. A função ε -insensível é apresentada na equação (2) e a Figura 1 apresenta esta situação graficamente, sendo a área sombreada correspondente ao ε -tubo (SCHÖLKOPF; SMOLA, 2004).

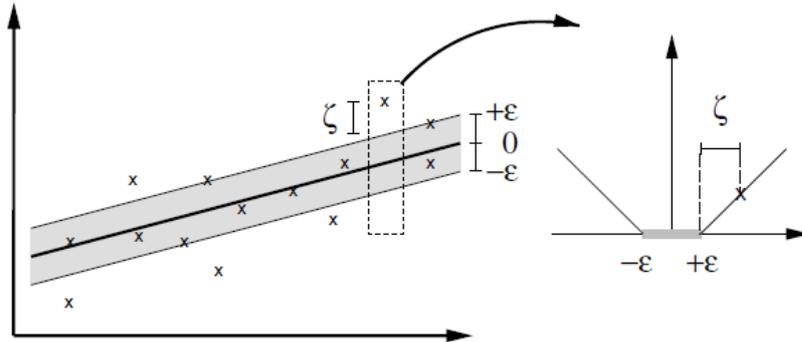
$$|\xi|_{\varepsilon} = \begin{cases} 0 & \text{se } |\xi| \leq \varepsilon \\ |\xi| - \varepsilon & \text{caso contrário} \end{cases} \quad (2)$$

A inequação (3) apresenta a introdução das variáveis de folga (ξ_i) e as restrições são apresentadas em (4). A variável de folga ξ_i está associada com os dados localizados abaixo da margem inferior e ξ_i^* com os localizados acima da margem superior (SCHÖLKOPF; SMOLA, 2004).

$$\underset{w, b, \xi, \xi^*}{\text{Minimizar}} \frac{1}{2} \|w\|^2 + C \left(\sum_{i=1}^n (\xi_i + \xi_i^*) \right) \quad (3)$$

$$\text{com as restrições } \begin{cases} y_i - w \cdot x_i - b \leq \varepsilon_i + \xi \\ w \cdot x_i + b - y_i \leq \varepsilon_i + \xi^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \quad (4)$$

Figura 1 - Função de perda ε -insensível



Fonte: Schölkopf e Smola (2004)

Na inequação (3), a constante C é a constante de regularização e determina a quantidade de erros de treinamento que serão permitidos e o custo para obtenção da região delgada do ε -tubo (SCHÖLKOPF; SMOLA, 2004). Os pontos fora do ε -tubo da Figura 1 são chamados de vetores de suporte (SVs), porque estabelecem os fundamentos para a função de regressão estimada. Isso significa que todos os outros pontos não são incluídos no modelo e podem ser removidos após a construção do modelo SVR (KECMAN, 2001; ALVES, 2012).

Quando um modelo não linear é necessário para modelar adequadamente os dados é necessário estender o ε -SVR linear para a regressão não linear. A utilização de uma função kernel possibilita que os dados de entrada originais sejam mapeados no espaço de características de elevada dimensão, onde uma regressão linear pode ser utilizada.

Utilizando o método do multiplicador de Lagrange pode ser obtido o problema dual apresentado na inequação (5), descrito em termos de produtos internos entre os objetos, sendo as restrições apresentadas em (6). A função objetivo é apresentada na equação (7). As variáveis de Lagrange associadas a todos os exemplos que se encontram dentro da margem entre $-\varepsilon$ e $+\varepsilon$ são nulas (SCHÖLKOPF; SMOLA, 2004).

$$\text{Maximizar}_{\alpha, \alpha^*} -\frac{1}{2} \sum_{i,j=1}^n (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) k(x_i, x_j) - \varepsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) + \sum_{i=1}^n y_i (\alpha_i - \alpha_i^*) \quad (5)$$

$$\text{com as restrições} \begin{cases} \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0 \\ \alpha_i, \alpha_i^* \in [0, C] \end{cases} \quad (6)$$

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) k(x_i, x_j) + b \quad (7)$$

onde os termos α_i e α_i^* representam as variáveis de Lagrange, $k(x_i, x_j)$ é a função kernel e b o termo de limiar (*bias*).

As funções kernel mais utilizadas são a linear, polinomial, função de base radial (*Radial Basis Function*, RBF) e sigmoideal. Neste artigo foi utilizado o kernel RBF, apresentado na equação (8), onde $\gamma = 1/2\nu^2$ e $\nu > 0$ é o parâmetro que define a largura do kernel.

$$k(x_i, x_j) = \exp\left(-\frac{1}{2\nu^2} \|x_i - x_j\|^2\right) = \exp\left(-\gamma \|x_i - x_j\|^2\right) \quad (8)$$

A capacidade de generalização do modelo SVR depende, durante a etapa de treinamento, da correta seleção dos parâmetros C e ε e do tipo de função kernel e seus parâmetros. Das diversas propostas de seleção de parâmetros existentes na literatura, a busca em grade (*grid search*) e a validação cruzada (*cross validation*) são as mais utilizadas (FACELI et al., 2011).

2.3. Procedimento para o desenvolvimento dos modelos

O procedimento utilizado neste artigo para o desenvolvimento dos modelos RNA e SVR segue os seguintes passos: obtenção dos dados, análise e preparação dos dados, seleção das variáveis, divisão dos dados, escolha da arquitetura para a RNA ou escolha da função kernel para a SVR, seleção dos melhores parâmetros, simulações de treinamento e de teste e análise do desempenho. A Figura 2 apresenta um fluxograma simplificado do procedimento para a obtenção dos modelos RNA e SVR.

Para a análise do desempenho e avaliação da capacidade de generalização dos modelos desenvolvidos podem ser utilizadas estratégias de minimização do erro de predição, também denominado de erro residual. O erro residual (e_i) é a diferença entre o valor real (y_i) e o valor estimado pelo modelo (\hat{y}_i), conforme apresentado na equação (9). A equação (10) apresenta o erro médio absoluto (*Mean Absolute Error*, MAE), a equação (11) apresenta o erro médio

quadrático (*Mean Squared Error*, MSE) e a equação (12) apresenta a raiz do erro médio quadrático (*Root Mean Squared Error*, RMSE), onde n representa o número de amostras.

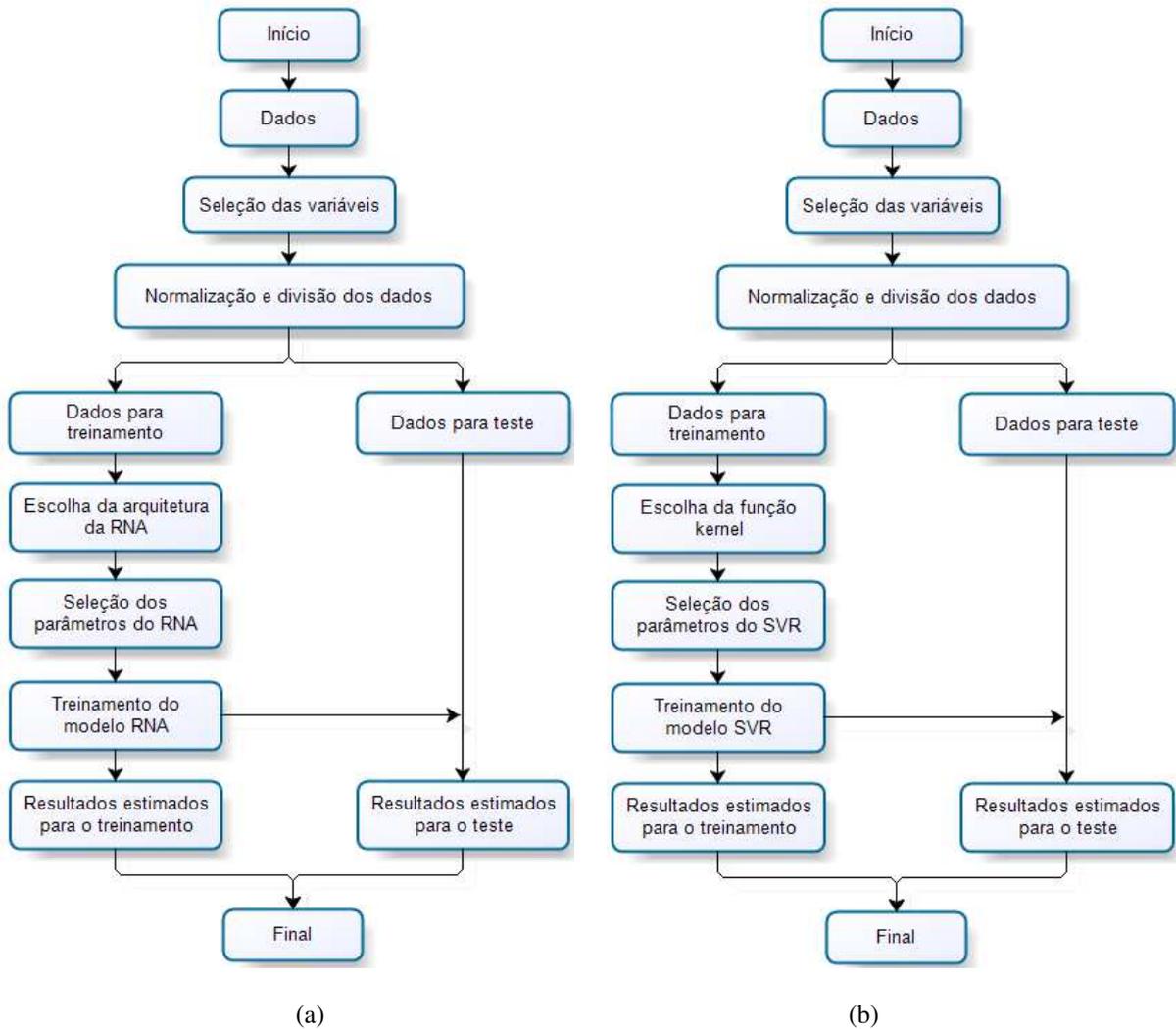
$$e_t = y_i - y_t \quad (9)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - y_t| = \frac{1}{n} \sum_{i=1}^n |e_t| \quad (10)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - y_t)^2 = \frac{1}{n} \sum_{i=1}^n (e_t)^2 \quad (11)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y_t)^2} = \sqrt{\frac{1}{n} \sum_{i=1}^n (e_t)^2} \quad (12)$$

Figura 2 - Fluxograma para o desenvolvimento do: (a) modelo RNA e (b) modelo SVR



Fonte: Autoria própria (2019)

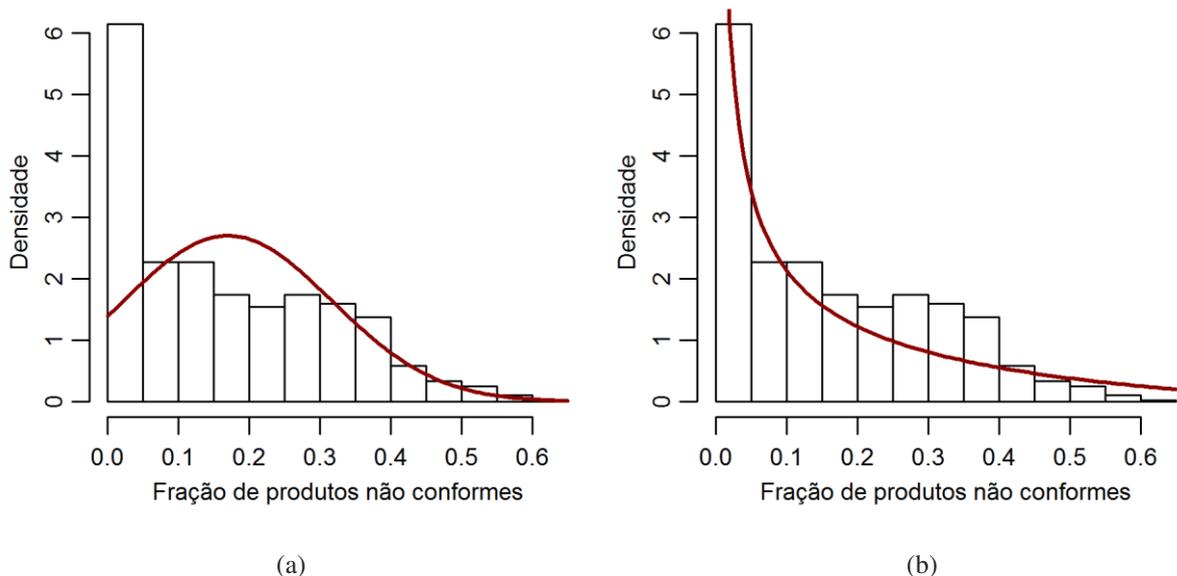
3. Estudo aplicado

Neste artigo foi desenvolvida a modelagem da fração de produtos não conformes às especificações de uma empresa curtidora de couro, produtora de couro acabado e fornecedora para as indústrias de calçados e artefatos em couro.

A etapa *wet blue* do processo produtivo avaliado consiste em: o classificador recebe um lote de diferentes tamanhos contendo as matérias-primas e verifica se as características de qualidade satisfazem às especificações, por métodos cognitivos. As matérias-primas que não satisfazem às especificações são classificadas como produtos não conformes e a fração de produtos não conformes às especificações, por lote, é considerada a variável dependente (característica da qualidade). Os dados coletados contemplaram uma amostra de 713 lotes.

A fração de produtos não conformes do processo segue a distribuição Beta, conforme apresentado na Figura 3. A Figura 3 (a) apresenta o histograma da fração de produtos não conformes com a distribuição Normal sobreposta e a Figura 3 (b), com a distribuição Beta sobreposta.

Figura 3 - Histogramas da fração de produtos não conformes com distribuição sobreposta: (a) Normal, (b) Beta



Fonte: Autoria própria (2019)

Os fatores controláveis definidos como variáveis independentes para a modelagem da fração de produtos não conformes foram: a seleção da matéria prima conforme qualidade e preço, com cinco níveis diferentes; a procedência da matéria-prima adquirida pela empresa, com cinco níveis; o classificador que inspeciona as matérias-primas, com três níveis, e o estado de rebaixamento da matéria-prima, com dois níveis.

Para o desenvolvimento dos modelos de regressão as variáveis independentes qualitativas seleção, procedência, classificador e rebaixamento foram substituídas pelas variáveis *dummy*. As novas variáveis independentes foram definidas como: seleção tipo 2 (x_1), seleção tipo 3 (x_2), seleção tipo 4 (x_3), seleção tipo 5 (x_4), procedência 2 (x_5), procedência 3 (x_6), procedência 4 (x_7), procedência 5 (x_8), classificador 2 (x_9), classificador 3 (x_{10}) e rebaixamento (x_{11}).

Para a definição das variáveis independentes relevantes para o desenvolvimento do modelo de regressão foi realizado o teste de correlação de Pearson. Após o teste foram selecionadas as variáveis seleção (tipo 2, tipo 3, tipo 4 e tipo 5), classificador (2 e 3) e rebaixamento como estatisticamente significativas para explicar a variável dependente fração de produtos não conformes.

Para o desenvolvimento dos modelos baseados em RNAs e SVR foi utilizado o programa R[®]. Após a análise e preparação dos dados o conjunto total de dados foi dividido aleatoriamente em dois subconjuntos: 70% (499 observações) para o grupo de treinamento e 30% (214 observações) para o grupo de teste.

A arquitetura selecionada para a RNA foi a MLP com sete variáveis de entrada, uma camada oculta e uma variável de saída. Após diversas simulações, os parâmetros selecionados para as RNAs foram: algoritmo de treinamento Levenberg-Marquardt, função de ativação logística para a camada oculta e linear para a camada de saída e taxa de aprendizado de 0,01.

Para a obtenção do número de neurônios da camada oculta foi utilizado o teorema de Hecht-Nielsen e foram analisados MSE, RMSE e MAE para redes com diferentes números de neurônios na camada oculta. A Tabela 1 apresenta os resultados de algumas simulações com os dados de treinamento para a obtenção do número de neurônios da camada oculta da RNA. Em função dos resultados obtidos para os erros foi selecionada a RNA com 19 neurônios na camada oculta.

Para o desenvolvimento do modelo SVR foi utilizada a biblioteca para máquinas de vetores de suporte denominada LIBSVM. O método utilizado para a seleção dos parâmetros do

modelo SVR foi a busca em grade em conjunto com a validação cruzada (10-*fold*) nos dados de treinamento. Para a seleção dos parâmetros do modelo SVR com o kernel RBF foi utilizado o espaço de busca: $C \in [1;50]$, $\varepsilon \in [0,001;1]$ e $\gamma \in [0,001;1]$. Os melhores valores obtidos para os parâmetros do modelo SVR foram: $C = 10$, $\varepsilon = 0,03$ e $\gamma = 0,2$.

Tabela 1 - Resultados das simulações para obtenção do número de neurônios da camada oculta da RNA

Neurônios	14	15	16	17	18	19	20	21
MSE	0,01407	0,01285	0,01288	0,01281	0,01852	0,01249	0,01269	0,01271
RMSE	0,1186	0,1134	0,11349	0,1132	0,1361	0,11173	0,11265	0,11275
MAE	0,09502	0,08939	0,08934	0,08763	0,1138	0,08796	0,08727	0,08878

Fonte: Autoria própria (2019)

Após a seleção da arquitetura e dos parâmetros da RNA e da seleção dos melhores parâmetros para o modelo SVR foi realizada a fase de treinamento e, após, os modelos obtidos foram utilizados para estimar os valores das características da qualidade utilizando os dados do grupo de teste. A Tabela 2 apresenta os valores dos erros calculados para os dados do grupo de treinamento e do grupo de teste para os modelos RNA e SVR, conforme equações (9) a (12). Os valores dos erros obtidos pelos modelos para os grupos de treinamento e de teste não diferem significativamente, indicando não haver sobreajuste dos modelos.

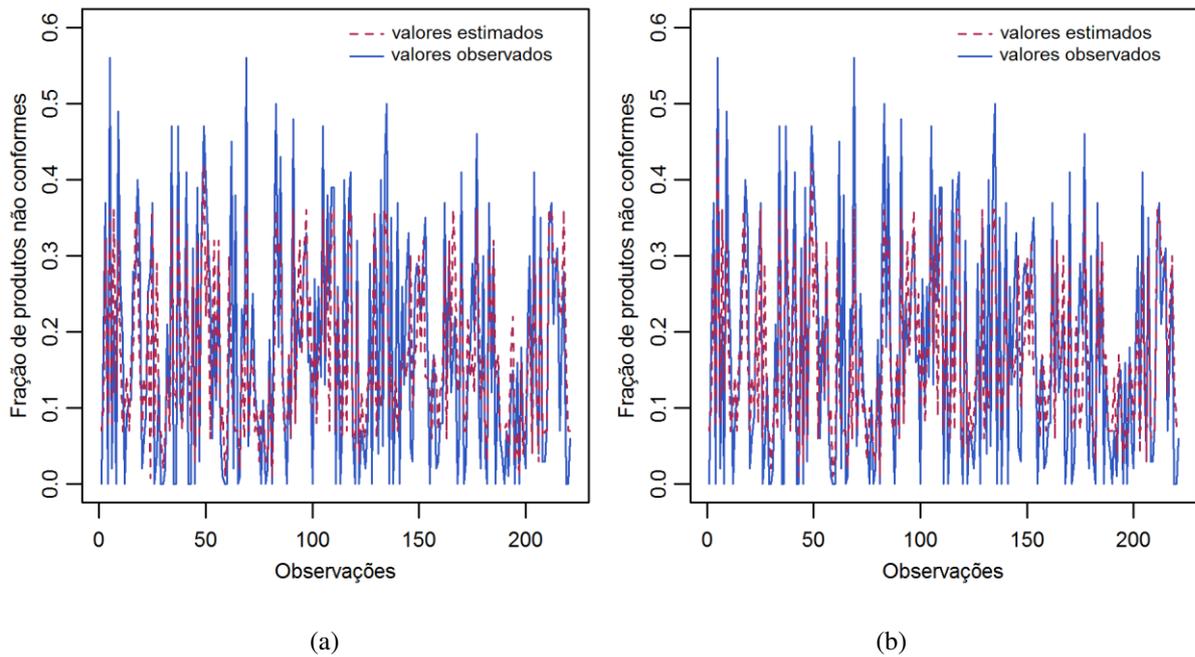
Tabela 2 - Valores dos erros calculados para os grupos de treinamento e de teste

Erro	RNA		SVR	
	Treinamento	Teste	Treinamento	Teste
MSE	0,01249	0,01391	0,01305	0,01355
RMSE	0,1117	0,1179	0,1142	0,1164
MAE	0,08796	0,09431	0,08462	0,08951

Fonte: Autoria própria (2019)

A Figura 4 apresenta os valores observados do processo e os valores estimados pelos modelos RNA e SVR para os dados do grupo de teste, onde verifica-se o bom desempenho preditivo dos modelos, pois os valores estimados se aproximam dos valores experimentais do processo. Analisando a Figura 4 e a Tabela 2 pode-se considerar que os modelos RNA e SVR representam adequadamente a fração de produtos não conforme do processo.

Figura 4 - Valores observados e valores estimados para o grupo de teste pelo modelo: (a) RNA e (b) SVR



Fonte: Autoria própria (2019)

Na sequência foi realizada a comparação dos modelos RNA e SVR com os modelos de regressão linear, linear generalizado de distribuição Binomial, de Quase-verossimilhança e Beta. A equação (13) apresenta a forma de regressão para o modelo linear, a equação (14) apresenta o modelo de regressão linear generalizado (MLG) de distribuição Binomial com função de ligação *logit*, a equação (15) apresenta a forma de regressão para o modelo de Quase-verossimilhança (MQV) com função de ligação *logit* e variância do tipo $V(\mu) = \mu(1 - \mu)$, e a equação (16) apresenta o modelo de regressão Beta com função de ligação *logit* e método de estimação por máxima verossimilhança.

$$\text{Linear} \Rightarrow NC = 0,2288 - 0,0178 * x_1 - 0,0563 * x_2 - 0,1326 * x_3 - 0,2006 * x_4 + 0,0300 * x_9 + 0,0946 * x_{10} - 0,1010 * x_{11} \quad (13)$$

$$\text{MLG Binomial} \Rightarrow NC = -1,2550 - 0,1041 * x_1 - 0,3767 * x_2 - 0,9376 * x_3 - 2,3871 * x_4 + 0,2625 * x_9 + 0,6222 * x_{10} - 0,7599 * x_{11} \quad (14)$$

$$\text{MQV} \Rightarrow NC = -1,3567 - 0,0051 * x_1 - 0,2566 * x_2 - 0,8531 * x_3 - 2,1960 * x_4 + 0,2766 * x_9 + 0,6500 * x_{10} - 0,7508 * x_{11} \quad (15)$$

$$\text{Beta} \Rightarrow NC = -1,2799 - 0,1888 * x_1 - 0,4243 * x_2 - 1,4856 * x_3 - 1,7678 * x_4 + 0,0325 * x_9 + 0,6201 * x_{10} - 0,6480 * x_{11} \quad (16)$$

Os dados do grupo de teste foram utilizados com os modelos linear, linear generalizado de distribuição Binomial, de Quase-verossimilhança e Beta. A Tabela 3 apresenta os valores dos erros calculados para todos os modelos, conforme equações (9) a (12).

Tabela 3 - Valores dos erros calculados para o grupo de teste com os modelos

Modelo	MSE	RMSE	MAE
RNA	0,01391	0,1179	0,09431
SVR	0,01355	0,1164	0,08951
Linear	0,01619	0,1272	0,10453
MLG Binomial	0,01580	0,1257	0,10017
Quase-verossimilhança	0,01578	0,1256	0,10208
Beta	0,01570	0,1253	0,10203

Fonte: Autorial própria (2019)

Analisando os resultados da Tabela 3 verifica-se que: (i) o modelo SVR apresenta os menores valores de erros entre os modelos; (ii) o modelo SVR apresenta desempenho um pouco melhor do que o modelo RNA; (iii) as duas técnicas de aprendizado de máquina apresentam desempenho melhor do que as quatro técnicas estatísticas; (iv) o modelo Beta apresenta desempenho um pouco melhor do que os modelos linear, linear generalizado de distribuição Binomial e de Quase-verossimilhança; (v) os modelos SVR e RNA apresentam desempenho melhor do que o modelo Beta, sem a necessidade de assumir uma distribuição de probabilidade específica para a implementação dos modelos.

4. Conclusões

Os efeitos das variáveis independentes sobre a característica da qualidade em processos industriais podem ser analisados através da modelagem dos dados destes processos usando modelos de regressão. Para esta modelagem podem ser utilizadas técnicas estatísticas, como os modelos lineares, modelos lineares generalizados e o modelo de regressão Beta, e técnicas de aprendizado de máquina, como as redes neurais artificiais e a regressão por vetores de suporte.

Neste artigo foram utilizadas as redes neurais artificiais e a regressão por vetores de suporte para a modelagem da fração de produtos não conformes às especificações de uma indústria curtidora de couro em que a característica da qualidade é mensurada no intervalo $[0,1]$. Os modelos RNA e SVR obtidos foram comparados com quatro modelos de regressão desenvolvidos com técnicas estatísticas para o mesmo processo produtivo.

Analisando os valores dos erros obtidos para os modelos verifica-se que o modelo SVR apresentou desempenho próximo ao do modelo RNA e melhor desempenho no ajuste do modelo aos dados do que os modelos de regressão linear, modelos de regressão lineares generalizados e modelo de regressão Beta.

Através dos resultados verifica-se que os modelos obtidos utilizando as duas técnicas de aprendizado de máquina (RNA e SVR) obtiveram um desempenho melhor do que os modelos utilizando as quatro técnicas estatísticas (modelo linear, modelos lineares generalizados e modelo Beta). Estes resultados estão de acordo com trabalhos disponíveis na literatura onde os autores concluem que as redes neurais artificiais e a regressão por vetores de suporte apresentam desempenho similar ou superior quando comparadas às técnicas estatísticas convencionais. Apesar destas técnicas de aprendizado de máquina apresentarem bons resultados no modelamento de processos, estas ainda são mais complexas do que as técnicas estatísticas convencionais.

Referências

ALVES, Julio Cesar Laurentino. **Máquinas de vetores de suporte aplicada a dados de espectroscopia NIR de combustíveis e lubrificantes para o desenvolvimento de modelos de regressão e classificação**. Campinas: Unicamp, 2012. 272 p. Tese (Doutorado) – Instituto de Química, Universidade Estadual de Campinas, Campinas, 2012.

BRAGA, Antônio de P.; CARVALHO, André P. de L. F.; LUDERMIR, Teresa B. **Redes neurais artificiais: teoria e aplicações**. 2. ed. Rio de Janeiro: LTC, 2012.

CORDEIRO, Gauss M.; DEMÉTRIO, Clarice G. B. **Modelos lineares generalizados**. São Paulo: Editora da USP, 2008.

CYBENKO, G. Approximations by superpositions of sigmoidal functions. **Mathematics of Control, Signals, and Systems**, v. 2, n. 4, p. 303-314, 1989.

FACELI, Katti; LORENA, Ana C.; GAMA, João; CARVALHO, André C. P. L. F. **Inteligência artificial: uma abordagem de aprendizado de máquina**. Rio de Janeiro: LTC, 2011.

FERRARI, S.; CRIBARI-NETO, F. Beta regression for modelling rates and proportions. **Journal of Applied Statistics**, v. 31, n. 7, p. 799-815, 2004.

HAYKIN, Simon. **Neural networks and learning machines**. 3rd ed. New York: Prentice Hall, 2009.

KECMAN, Vojislav. **Learning and soft computing: support vector machines, neural networks and fuzzy logic models**. London: MIT Press, 2001.

MITCHELL, Tom M. **Machine learning**. New York: McGraw-Hill, 1997.

MONTGOMERY, Douglas C. **Introdução ao controle estatístico da qualidade**. Rio de Janeiro: LTC, 2004.

SANT'ANNA, Ângelo. M. O. **Ferramentas para modelagem e monitoramento de características de qualidade do tipo fração**. 2009. 150 f. Tese (Doutorado em Engenharia de Produção) – Escola de Engenharia, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2009.

SCHÖLKOPF, Bernhard; SMOLA, Alex J. A tutorial on support vector regression. **Statistics and Computing**, Netherlands, vol. 14, n° 3, p. 199-222, august, 2004.

VAPNIK, Vladimir N. **Statistical learning theory**. New York: John Wiley & Sons, 1998.

VAPNIK, Vladimir N. An overview of statistical learning theory. **IEEE Transactions On Neural Networks**, vol. 10, n° 5, september, 1999.