

Aplicação da regressão por vetores de relevância 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: Em um processo produtivo pode ser definido um conjunto de causas que produzem determinado efeito sobre as características de qualidade dos produtos. Estas causas podem resultar em produtos que são não conformes, ou seja, produtos que não atendem uma ou mais das especificações preestabelecidas e 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. A regressão por vetores de relevância (RVR) é uma técnica de aprendizado de máquina utilizada nos últimos anos para a modelagem de processos. Este artigo apresenta a aplicação da regressão por vetores de relevância na modelagem da fração de produtos não conformes às especificações de uma indústria curtidora de couro. O modelo RVR obtido foi comparado com modelos de regressão obtidos com redes neurais artificiais, regressão por vetores de suporte, modelos lineares generalizados, regressão Beta, regressão Ridge e regressão LASSO, desenvolvidos para o mesmo processo produtivo. Através dos resultados verifica-se que o modelo RVR apresenta melhor desempenho no ajuste do modelo aos dados do processo do que os demais modelos analisados e o modelo RVR representa adequadamente a fração de produtos não conformes às especificações do processo produtivo.

Palavras chave: Modelagem de processos, fração de produtos não conformes, aprendizado de máquina, regressão por vetores de relevância.

Relevance vector regression applied to modeling the fraction nonconforming in a production process

Abstract: In a production process, one can enumerate causes that act on the quality characteristics of products. These causes can result in nonconforming items, i.e., products are those that fail to meet one or more of its pre-established specifications. Nonconforming items can be measured using the fraction nonconforming. Modeling the fraction of nonconforming items can be performed using different techniques and regression models. Relevance vector regression (RVR) is a machine learning technique used for process modeling in recent years. This paper proposes an application of relevance vector regression for modeling the fraction nonconforming items from a tanning process. We compared the performance of RVR model with that obtained from the artificial neural network model, support vector regression model, generalized linear models, Beta regression model, Ridge regression model, and LASSO regression model, developed from the same process. The comparative analysis showed that the RVR model has better performance in fitting the model to the process data than other models used, and the RVR model is an adequate tool for modeling the process.

Key-words: Process modeling, fraction nonconforming, machine learning, relevance vector regression.

1. Introdução

O conceito de qualidade tem sido definido de diferentes maneiras na literatura e a mais usual consiste em que a qualidade é inversamente proporcional à variabilidade e que a melhoria da qualidade é a redução da variabilidade nos processos e produtos. Em um processo industrial

certas causas são inerentes ao processo e, desta forma, muitas vezes não se consegue controlar todas as causas de variação que produzem determinado efeito sobre as características da qualidade. As causas de variação que podem ser controladas podem resultar na produção de produtos 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 do tipo fração não conforme podem ser aproximadas pela distribuição Binomial e são observações expressas no intervalo $[0,1]$ (SANT'ANNA, 2009).

A modelagem da fração não conforme é uma relevante estratégia para investigar a relação entre as causas de variação que interferem no processo e as características da qualidade dos produtos. Conforme Montgomery e Runger (2003), um modelo de regressão que apresenta um bom ajuste usualmente permite gerar boas estimativas dos efeitos dos fatores, pois é possível prever a fração não conforme em função do ajuste das variáveis do processo.

Modelos de regressão baseados em técnicas de aprendizado de máquina têm sido propostos na literatura. O aprendizado de máquina diz respeito ao 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).

Tipping (2000) introduziu as Máquinas de Vetores de Relevância (*Relevance Vector Machine, RVM*), que são modelos probabilísticos baseados em métodos kernel com forma funcional similar as Máquinas de Vetores de Suporte (*Support Vector Machine, SVM*) (VAPNIK, 1998). A RVM pode ser utilizada para problemas de classificação e de regressão, sendo esta última denominada de Regressão por Vetores de Relevância (*Relevance Vector Regression, RVR*).

O objetivo deste artigo é modelar a fração de produtos não conformes às especificações de uma empresa curtidora de couro utilizando regressão por vetores de relevância. O modelo obtido foi comparado com os modelos de regressão obtidos com outras duas técnicas de aprendizado de máquina (redes neurais artificiais e regressão por vetores de suporte) e os obtidos com cinco técnicas estatísticas (modelos lineares generalizados, regressão Beta, regressão Ridge e regressão LASSO), desenvolvidos para o mesmo processo produtivo.

2. Referencial teórico

Nesta sessão apresenta-se uma breve descrição dos referenciais teóricos das técnicas utilizadas para o desenvolvimento dos modelos de regressão deste artigo: regressão por vetores de relevância, redes neurais artificiais, regressão por vetores de suporte, modelos lineares generalizados, regressão Beta, regressão Ridge e regressão LASSO.

2.1 Regressão por vetores de relevância

A RVR utiliza um conjunto de pares de dados de entrada e saída (alvo) $\{\mathbf{x}_r, t_r\}_{r=1}^n$, que segue a formulação probabilística padrão e se assume que as saídas (alvos) são observações de um modelo com ruído aditivo, conforme equação (1) (TIPPING, 2001)

$$t_r = \sum_{r=1}^n \omega_r k(x_i, x_j) + b + \epsilon_r \quad (1)$$

onde ω_r são os pesos do modelo, $k(x_i, x_j)$ é uma função kernel, b é o termo de limiar (*bias*) e ϵ_r são amostras independentes de ruído consideradas como ruído gaussiano. Então, tem-se $p(t_r | \mathbf{x}) = \mathcal{N}(t_r | y(\mathbf{x}_r), \sigma_r^2)$, onde a notação especifica uma distribuição Gaussiana sobre t_r com média $y(\mathbf{x}_r)$ e variância σ_r^2 .

A RVR adota uma estrutura probabilística e utiliza a probabilidade *a priori* sobre os pesos do modelo regidos por um conjunto de hiperparâmetros, um associado a cada peso, cujos valores mais prováveis são estimados de forma iterativa a partir dos dados de treinamento.

A RVR é o tratamento Bayesiano da equação (1) (TIPPING, 2001). A introdução de um hiperparâmetro α_r separado para cada um dos parâmetros de peso ω_r é a característica principal da RVR e que a diferencia da Regressão por Vetores de Suporte (*Support Vector Regression*, SVR), que possui um único hiperparâmetro compartilhado. O hiperparâmetros α_r representa a precisão do parâmetro correspondente ω_r (BISHOP, 2006).

Com a distribuição *a priori*, utiliza-se a regra de Bayes para determinar a distribuição *a posteriori* sobre os pesos, equação (2) (TIPPING, 2001)

$$p(\mathbf{w} | \mathbf{t}, \boldsymbol{\alpha}, \sigma_r^2) = (2\pi)^{-(n+1)/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{w} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{w} - \boldsymbol{\mu})\right\} \quad (2)$$

onde a matriz de covariância ($\boldsymbol{\Sigma}$) e o vetor média ($\boldsymbol{\mu}$) *a posteriori* são estimados, respectivamente, pelas equações (3) e (4)

$$\boldsymbol{\Sigma} = (\sigma_r^{-2} \boldsymbol{\Phi}^\top \boldsymbol{\Phi} + \mathbf{A})^{-1} \quad (3)$$

$$\boldsymbol{\mu} = \sigma_r^{-2} \boldsymbol{\Sigma} \boldsymbol{\Phi}^\top \mathbf{t} \quad (4)$$

onde $\mathbf{A} = \text{diag}(\alpha_{t_0}, \alpha_{t_1}, \dots, \alpha_{t_n})$. A verossimilhança marginalizada dos hiperparâmetros é obtida pela integração dos pesos, conforme equação (5)

$$p(\mathbf{t} | \boldsymbol{\alpha}, \sigma_r^2) = (2\pi)^{-n/2} |\sigma_r^2 \mathbf{I} + \boldsymbol{\Phi} \mathbf{A}^{-1} \boldsymbol{\Phi}^\top|^{-1/2} \exp\left\{-\frac{1}{2} \mathbf{t}^\top (\sigma_r^2 \mathbf{I} + \boldsymbol{\Phi} \mathbf{A}^{-1} \boldsymbol{\Phi}^\top)^{-1} \mathbf{t}\right\} \quad (5)$$

O σ_r^2 é estimado como um hiperparâmetro do modelo. Para $\boldsymbol{\alpha}$, a partir da equação (5), obtêm-se a equação (6) (TIPPING, 2001)

$$\alpha_r^{\text{nov}} = \frac{\delta_r}{\mu_r^2} \quad (6)$$

onde μ_r é o r -ésimo componente do vetor média do peso *a posteriori* da equação (4) e a quantidade δ_r é definida por $\delta_r = 1 - \alpha_r \sum_{rr}$, onde \sum_{rr} é o r -ésimo elemento da diagonal da matriz de covariância do peso *a posteriori* da equação (3), calculados com os valores atuais de $\boldsymbol{\alpha}$ e σ_r^2 . Cada $\delta_r \in [0, 1]$ pode ser interpretado como uma medida de quanto "bem determinado" está sendo cada parâmetro ω_r pelos dados.

Para a variância do ruído σ_r^2 , a diferenciação da equação (5) leva à reestimativa apresentada pela equação (7), onde n se refere ao número de observações (TIPPING, 2001)

$$(\sigma_r^2)^{novo} = \frac{\|\mathbf{t} - \Phi\boldsymbol{\mu}\|^2}{n - \sum_r \delta_r} \quad (7)$$

O algoritmo de aprendizado prossegue pela aplicação repetida das equações (6) e (7), concomitante à atualização da estatística *a posteriori* $\boldsymbol{\Sigma}$ e $\boldsymbol{\mu}$ das equações (3) e (4), até que alguns critérios de convergência adequados tenham sido satisfeitos, como um limiar pré-determinado ou um certo número de iterações tenha sido atingido.

Como resultado do processo de otimização, a proporção dos hiperparâmetros α_r é conduzida para valores elevados e, portanto, os parâmetros de peso ω_r correspondentes a estes hiperparâmetros possuem distribuições *a posteriori* com média e variância nulas (BISHOP, 2006). Assim, os parâmetros infinitos e nulos resultante da maximização são removidos do modelo e as funções de base $\phi(x_r)$ correspondentes podem ser podadas e a esparsividade é obtida. No caso dos modelos com a forma da equação (1), as entradas x_n correspondentes aos pesos remanescentes não nulos são denominados vetores de relevância (RVs) e são análogos aos vetores de suporte (SVs) de uma SVR.

Uma característica da RVR é que, embora com capacidade de generalização comparável a uma SVR equivalente, o número de RVs é, na maioria dos casos, significativamente menor que o número de SVs usados por uma SVR para resolver o mesmo problema. Na RVR os parâmetros que controlam a complexidade e a variância do ruído (α_r 's e σ_r^2) são automaticamente estimados pelo procedimento de aprendizado, enquanto na SVR é necessário ajustar os parâmetros constante de regularização (C) e função de perda insensível (ε) (TIPPING, 2000). Um kernel $k(x_i, x_j)$ é uma função que recebe dois pontos no espaço de entrada original e calcula o produto escalar desses objetos no espaço de características de elevada dimensão. Dessa forma, o kernel pode transformar um conjunto de dados não linearmente separáveis em um conjunto de dados linearmente separáveis (CRISTIANINI & SHAW-TAYLOR, 2000).

As funções kernel mais utilizadas são a linear, polinomial, função de base radial (*Radial Basis Function*, RBF) e sigmoideal. Neste artigo foi utilizada a função kernel RBF, apresentada 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 RVR depende, durante a etapa de treinamento, da correta seleção do parâmetro da função kernel. 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).

A busca em grade se refere ao processo de busca exaustiva sobre um subconjunto do espaço de trabalho, sendo que a busca é realizada em um espaço formado pelos parâmetros de interesse e seu objetivo é encontrar pontos nos quais a acurácia seja a maior possível. A validação cruzada é uma técnica utilizada para avaliar a capacidade de generalização de um algoritmo quando exposto a um novo conjunto de dados. Um tipo comum de validação cruzada é a *k-fold cross-validation*, uma técnica de reamostragem na qual os dados são divididos em k conjuntos de tamanhos aproximadamente iguais (FACELI et al., 2011).

2.2 Redes neurais artificiais

Segundo Haykin (2009), a principal propriedade de uma Rede Neural Artificial (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).

Entre os principais tipos de RNAs com a arquitetura *feedforward* de camadas múltiplas está o Perceptron Multicamadas (*Multilayer Perceptron*, MLP). Para as redes MLP pode ser utilizado o algoritmo *error backpropagation* e existem várias modificações deste algoritmo que visam melhorar seu desempenho, sendo uma destas a Levenberg-Marquardt.

Algumas das dificuldades na utilização de RNAs são a definição de sua arquitetura (funções de ativação e topologia da rede) e a escolha dos melhores parâmetros para a arquitetura da rede, pois o desempenho de uma RNA depende muito dos parâmetros. Para mais informações sobre RNAs sugere-se consultar Vapnik (1998), Haykin (2009) e Faceli et al. (2011).

2.3 Regressão por vetores de suporte

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). 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. Vapnik (1998) demonstrou que o princípio SRM é superior ao princípio ERM, utilizado em RNAs.

A aplicação de SVMs foi originalmente desenvolvida para problemas de classificação, sendo posteriormente estendida para o tratamento de problemas de regressão e denominada Regressão por Vetores de Suporte (*Support Vector Regression*, SVR) (SCHÖLKOPF & SMOLA, 2004). Na SVR tem-se dois conceitos importantes, que são o ε -tubo e a função de perda ε -insensível, que ignora erros que estão além de uma certa distância dos valores considerados válidos. As amostras fora do ε -tubo correspondem aos vetores de suporte (SVs).

Quando um modelo não linear é necessário utiliza-se uma função kernel, sendo que neste artigo foi utilizada a função kernel RBF apresentada na equação (8). A capacidade de generalização do modelo SVR depende da seleção, durante a etapa de treinamento, dos parâmetros constante de regularização (C), da função de perda insensível (ε) e do parâmetro (γ) da função kernel RBF. Para mais detalhes sobre SVMs recomenda-se consultar Vapnik (1998), Kecman (2001) e Schölkopf e Smola (2004).

2.4 Modelos lineares generalizados

Segundo Sant'Anna (2009) os modelos de regressão linear não são adequados para modelar as características da qualidade do tipo fração não conforme, pois estes 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 a distribuição de probabilidade para a característica da qualidade (CORDEIRO & DEMÉTRIO, 2008). Para mais detalhes sobre modelos lineares generalizados recomenda-se consultar Hamada e Nelder (1997), Cordeiro e Demétrio (2008) e Sant'Anna (2009).

2.5 Modelo de regressão Beta

De acordo com Sant'Anna (2009), para modelar as características da qualidade do tipo fração não conforme pode ser utilizado o Modelo de Regressão Beta (MRB). Ferrari e Cribari-Neto (2004) propuseram o MRB, cuja estrutura do modelo de regressão permite modelar as relações lineares e não-lineares entre as variáveis independentes e a variável dependente. Na construção destes modelos as variáveis independentes podem ser de natureza quantitativa ou qualitativa. O MRB se baseia na suposição de que a fração segue a distribuição Beta, que assume valores restritos ao intervalo $[0,1]$. Para mais detalhes sobre o modelo de regressão Beta recomenda-se consultar Ferrari e Cribari-Neto (2004) e Sant'Anna (2009).

2.6 Regressão Ridge

A regressão Ridge (HOERL & KENNARD, 1970) objetiva tratar o problema de multicolinearidade das variáveis explicativas adicionando uma penalização ao estimador dos coeficientes de regressão encontrados pelo método de quadrados mínimos. O estimador Ridge possui um parâmetro λ cujo valor pode ser escolhido através da validação cruzada. Para mais detalhes sobre a regressão Ridge recomenda-se consultar Hoerl e Kennard (1970), Marquardt e Snee (1975) e Casagrande (2016).

2.7 Regressão LASSO

A regressão LASSO (*Least Absolute Shrinkage and Selection Operator*) é um método utilizado para reduzir os efeitos dos atributos que não contribuem para a identificação da variável resposta (TIBSHIRANI, 1996). O estimador LASSO possui um parâmetro λ , sendo que cada valor de λ leva a um conjunto diferente de coeficientes estimados.

As principais diferenças entre a regressão Ridge e a LASSO estão na penalização utilizada para estimar os coeficientes de regressão e no fato da regressão LASSO fazer com que as estimativas de alguns coeficientes sejam iguais a zero resultando em um menor número de covariáveis (CASAGRANDE, 2016). Para mais detalhes sobre regressão LASSO recomenda-se consultar Tibshirani (1996) e Friedman et al. (2010).

3. Procedimento para o desenvolvimento do modelo RVR e análise de desempenho

O procedimento utilizado neste artigo para o desenvolvimento do modelo RVR segue os seguintes passos: obtenção dos dados, análise e preparação dos dados, seleção das variáveis, divisão dos dados, seleção do melhor parâmetro para a função kernel RBF, simulações de treinamento e de teste e análise do desempenho.

Para a análise de desempenho e avaliação da capacidade de generalização do modelo desenvolvido podem ser utilizadas estratégias de minimização do erro de predição, também denominado de erro residual. O erro residual (e_t) é a diferença entre o valor real (y_i) e o valor estimado pelo modelo (y_t), 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.

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

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

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

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

4. 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. A Figura 1 (a) apresenta o histograma da fração de produtos não conformes com a distribuição Normal sobreposta e a Figura 1 (b), com a distribuição Beta sobreposta.

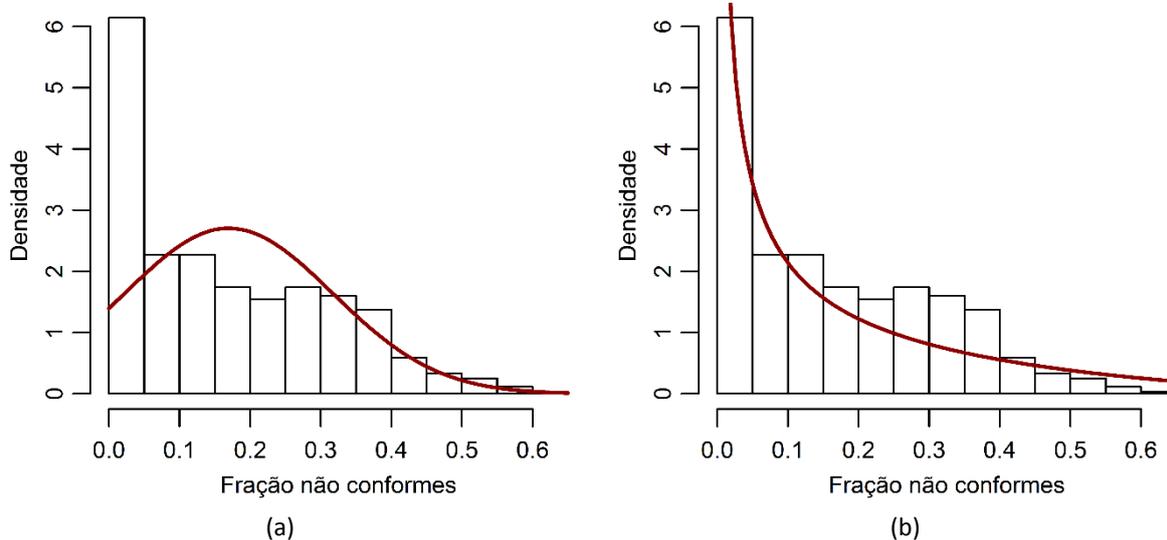


Figura 1 – Histogramas da fração de produtos não conformes com distribuição sobreposta: (a) Normal, (b) Beta

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), conforme Tabela 1.

Fatores Controláveis	Número de Níveis	Níveis Codificados				
Seleção	5	1	2	3	4	5
Procedência	5	1	2	3	4	5
Classificador	3	1	2	3	-	-
Rebaixamento	2	1	2	-	-	-

Tabela 1 – Caracterização dos níveis dos fatores controláveis

Para a construção 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 a construção dos modelos de regressão foi realizado o teste de correlação de Pearson, sendo selecionadas as variáveis seleção (tipo 2, tipo 3, tipo 4 e tipo 5), classificador (2 e 3) e rebaixamento como significativas estatisticamente para explicar a variável dependente fração de produtos não conformes.

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. Os cálculos e as simulações necessárias foram desenvolvidas com o programa R®.

Para a seleção do parâmetro do kernel RBF do modelo RVR o espaço de busca foi definido em $\gamma \in [0,001;1]$. O método de seleção utilizado foi a busca em grade em conjunto com a validação cruzada (*10-fold*). O melhor valor obtido para o parâmetro foi $\gamma = 0,2127$.

Após a seleção do melhor parâmetro para o kernel RBF foi realizada a fase de treinamento e, após, o modelo RVR obtido foi utilizado para estimar os valores da característica da qualidade utilizando os dados do grupo de teste. A Tabela 2 apresenta os valores dos erros obtidos para os dados do grupo de treinamento e de teste para o modelo RVR, utilizando as equações (9) a (12). Os valores dos erros obtidos para os grupos de treinamento e de teste não diferem significativamente, indicando não haver sobreajuste do modelo RVR.

Erro	Grupo de Treinamento	Grupo de Teste
MSE	0,01257	0,01348
RMSE	0,11190	0,11610
MAE	0,08650	0,09060

Tabela 2 – Valores dos erros obtidos para os grupos de treinamento e de teste para o modelo RVR

A Figura 2 apresenta os valores observados do processo e os valores estimados pelo modelo RVR para o grupo de teste, onde verifica-se o bom desempenho preditivo do modelo pois os valores estimados se aproximam adequadamente dos valores experimentais do processo.

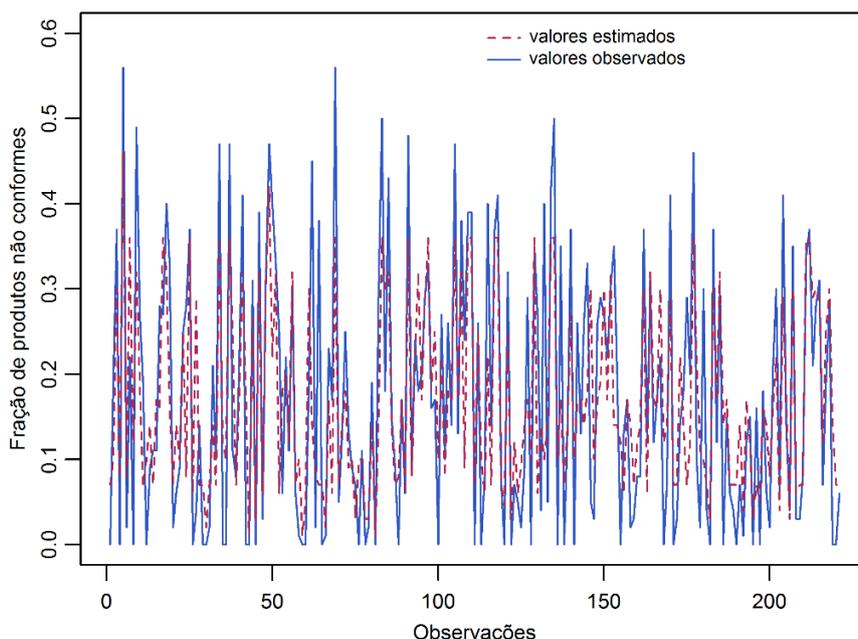


Figura 2 – Valores observados e valores estimados pelo modelo RVR para o grupo de teste

Na sequência, para o mesmo processo produtivo, foram desenvolvidos os modelos de regressão baseados em redes neurais artificiais, regressão por vetores de suporte, modelos lineares generalizados, regressão Beta, regressão Ridge e regressão LASSO.

A arquitetura selecionada para a RNA foi a MLP com uma camada oculta. Os parâmetros selecionados para a RNA foram: algoritmo de treinamento Levenberg-Marquardt, função de ativação logística para a camada oculta e linear para a camada de saída, taxa de aprendizado de 0,01 e 19 neurônios na camada oculta.

O método de seleção dos parâmetros do modelo SVR foi a busca em grade em conjunto com a validação cruzada (10-fold) nos dados de treinamento. O espaço de busca definido para os parâmetros foi: $C \in [1;50]$, $\varepsilon \in [0,001;1]$ e $\gamma \in [0,001;1]$. Os melhores valores obtidos para os parâmetros foram: $C = 10$, $\varepsilon = 0,03$ e $\gamma = 0,2$ (ACOSTA et al., 2016).

A forma de regressão para o modelo linear generalizado (MLG) Binomial com função de ligação *logit* é apresentada na equação (13) e, na equação (14), para o modelo de Quase-verossimilhança (MQV) com função de ligação *logit* e variância do tipo $V(\mu) = \mu(1 - \mu)$

$$\text{MLG Binomial} \Rightarrow NC = -1,2550 - 0,1041 * x_1 - 0,3767 * x_2 - 0,9376 * x_3 - 2,3871 * x_4 + 0,2625 * x_9 + 0,6222 * x_{10} - 0,7599 * x_{11} \quad (13)$$

$$\text{MQV} \Rightarrow NC = -1,3567 - 0,0051 * x_1 - 0,2566 * x_2 - 0,8531 * x_3 - 2,1960 * x_4 + 0,2766 * x_9 + 0,6500 * x_{10} - 0,7508 * x_{11} \quad (14)$$

A equação (15) apresenta o modelo de regressão Beta com função de ligação *logit* e método de estimação por máxima verossimilhança

$$\text{Beta} \Rightarrow NC = -1,2799 - 0,1888 * x_1 - 0,4243 * x_2 - 1,4856 * x_3 - 1,7678 * x_4 + 0,0325 * x_9 + 0,6201 * x_{10} - 0,6480 * x_{11} \quad (15)$$

A otimização do parâmetro (λ) na regressão Ridge foi realizada usando validação cruzada para computar o modelo com menor erro, sendo obtido $\lambda = 0,006844$. Na regressão LASSO o valor de λ também foi selecionado através da validação cruzada, sendo obtido $\lambda = 0,0001618$.

Após a seleção dos parâmetros e dos coeficientes dos modelos 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.

Na sequência, o modelo RVR foi comparado com o modelo SVR, modelo RNA, modelo Beta, modelo de Quase-verossimilhança (MQV), modelo linear generalizado Binomial (MLG Binomial), modelo LASSO e modelo Ridge. A Tabela 3 apresenta os valores dos erros obtidos com os modelos para os dados do grupo de teste, calculados com as equações (9) até (12).

Modelo	MSE	RMSE	MAE
RVR	0,01348	0,1161	0,09060
SVR	0,01355	0,1164	0,08951
RNA	0,01391	0,1179	0,09431
Beta	0,01570	0,1253	0,10203
MQV	0,01578	0,1256	0,10208
MLG Binomial	0,01580	0,1257	0,10017
LASSO	0,01619	0,1272	0,10460
Ridge	0,01621	0,1273	0,10568

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

Considerando os valores dos erros apresentados na Tabela 3 pode-se considerar que todos os modelos representam adequadamente a fração de produtos não conforme do processo. Analisando os resultados da Tabela 3 verifica-se que:

- O modelo RVR apresenta os menores valores de erros entre os modelos analisados;
- O modelo RVR apresenta desempenho melhor do que os modelos obtidos com as outras duas técnicas de aprendizado de máquina (SVR e RNA);
- Os modelos obtidos com as três técnicas de aprendizado de máquina (RVR, SVR e RNA) apresentam desempenho melhor do que os modelos obtidos com as cinco técnicas estatísticas (Beta, MQV, MLG Binomial, LASSO e Ridge);
- O modelo Beta apresenta desempenho melhor do que os modelos obtidos com as outras quatro técnicas estatísticas (MQV, MLG Binomial, LASSO e Ridge);
- Os modelos RVR, 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.

Das três técnicas de aprendizado de máquina analisadas, o modelo RVR apresentou melhor desempenho. Deve-se levar em consideração, também, que o projeto e o treinamento de RNAs geralmente resultam em uma tarefa complexa na qual vários parâmetros devem ser ajustados; no modelo SVR devem ser selecionados os parâmetros constante de regularização (C), função de perda insensível (ε) e o parâmetro (γ) da função kernel RBF; enquanto no modelo RVR é necessário selecionar apenas o parâmetro (γ) da função kernel RBF.

5. Conclusões

Os efeitos das variáveis independentes (fatores controláveis) sobre a variável dependente (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 e técnicas de aprendizado de máquina.

Neste artigo foi utilizada a regressão por vetores de relevância 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 $[0,1]$. O modelo RVR foi comparado com dois modelos de regressão baseados em técnicas de aprendizado de máquina e com cinco modelos de regressão desenvolvidos com técnicas estatísticas.

Analisando os valores dos erros verifica-se que o modelo RVR apresentou desempenho melhor no ajuste do modelo aos dados do processo do que os modelos RNA, SVR, modelo linear generalizado Binomial, modelo de Quase-verossimilhança, Beta, Ridge e LASSO.

Através dos resultados pode-se considerar que os modelos obtidos utilizando as técnicas de aprendizado de máquina (RVR, SVR e RNA) obtiveram desempenho melhor do que os modelos utilizando as técnicas estatísticas (modelo linear generalizado Binomial, modelo de Quase-verossimilhança, Beta, Ridge e LASSO) e que todos os modelos representam adequadamente a fração de produtos não conformes do processo produtivo.

Estes resultados estão de acordo com trabalhos disponíveis na literatura onde os autores concluem que as técnicas de aprendizado de máquina (tais como RVR, RNA e SVR) apresentam desempenho similar ou superior quando comparadas às técnicas estatísticas. 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

ACOSTA, S. M.; AMOROSO, A. L.; SANT'ANNA, Â. M. O. Modelagem de um processo produtivo utilizando regressão por vetores suporte. In: ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO, 2016, João Pessoa. **Anais...** Rio de Janeiro: ABEPRO, 2016.

BISHOP, C. M. **Pattern recognition and machine learning**. New York: Springer, 2006.

CASAGRANDE, M. H. **Comparação de métodos de estimação para problemas com colinearidade e/ou alta dimensionalidade ($p > n$)**. São Carlos, 65p., 2016. Dissertação (Mestrado) – Universidade Federal de São Carlos.

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

CRISTIANINI, N.; SHAWE-TAYLOR, J. **An introduction to support vector machines and other kernel-based learning methods**. United Kingdom: Cambridge University Press, 2000.

FACELI, K.; LORENA, A. C.; GAMA, J.; CARVALHO, A. 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.

FRIEDMAN, J.; HASTIE, T.; TIBSHIRANI, R. Regularization paths for generalized linear models via coordinate descent. **Journal of Statistical Software**, v. 33, n. 1, p. 1-22, 2010.

HAMADA, M.; NELDER, J.A. Generalized linear models for quality-improvement experiments. **Journal of Quality Technology**, v. 29, p. 292-304, 1997.

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

HOERL, A. E.; K., R. W. Ridge regression: biased estimation for nonorthogonal problems. **Technometrics**, v. 12, n. 1, p. 55-67, 1970.

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

MARQUARDT, D. W.; SNEE, R. D. Ridge regression in practice. **The American Statistician**, v. 29, n. 1, p. 3-20, 1975.

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

MONTGOMERY, D. C.; RUNGER, G. C. **Estatística aplicada e probabilidade para engenheiros**. Rio de Janeiro: LTC, 2003.

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

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

SCHÖLKOPF, B.; SMOLA, A. J. A tutorial on support vector regression. **Statistics and Computing**, v. 14, n. 3, p. 199-222, 2004.

TIBSHIRANI, R. Regression shrinkage and selection via the lasso. **Journal of the Royal Statistical Society. Series B (Methodological)**, p. 267-288, 1996.

TIPPING, M. E. The relevance vector machine. In: SOLLA, S. A.; LEEN, T. K.; MÜLLER, K. R. (Org.), **Advances in neural information processing systems**, London: MIT Press, v. 12, p. 652-658, 2000.

TIPPING, M. E. Sparse Bayesian learning and the relevance vector machine. **Journal of Machine Learning Research**, v. 1, p. 211-244, 2001.

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