

DESENVOLVIMENTO DE SISTEMA DE DETECÇÃO DE FALHAS BASEADO EM APRENDIZADO ESTATÍSTICO DE MÁQUINAS DE VETORES DE SUPORTE

M. H. GRANZOTTO¹ e L. C. OLIVEIRA-LOPES²

¹Universidade Federal dos Vales do Jequitinhonha e Mucuri, Instituto de Ciências e Tecnologia, Engenharia Química

² Universidade Federal de Uberlândia, Faculdade de Engenharia Química
E-mail para contato: ¹mhgranzotto@gmail.com, ²lcol@ufu.br

RESUMO – As máquinas de vetores de suporte são uma poderosa ferramenta de reconhecimento de padrões. Elas representam um conjunto de metodologias do aprendizado supervisionado estatístico que é usado para classificação e análise de regressão de dados. Os vetores de suporte utilizam um hiperplano com máxima margem para separar as diferentes classes de dados. Este trabalho apresenta um sistema de detecção que utiliza as máquinas de vetores de suporte para detectar falhas em processos e realizar o diagnóstico das mesmas. A metodologia baseia-se em informações históricas quantitativas e qualitativas com máquinas de vetores de suporte treinadas e sintonizadas pelo método da evolução diferencial. O sistema de detecção de falhas foi avaliado em processos químicos para verificar a adequabilidade da proposta para a detecção e classificação de falhas com dados históricos e qualitativos. Métricas de desempenho do sistema apontam para a eficiência do sistema de detecção nos cenários investigados.

1. INTRODUÇÃO

A utilização de sistemas de detecção e diagnóstico de falhas é significativa e vêm crescendo de maneira constante, principalmente para manutenção de controladores e isolamento de perturbações em processos. O grande crescimento na utilização de sistemas de controle automático está atualmente ligado ao monitoramento de falhas em tempo real. Através da grande quantidade de dados que podem ser coletados em diferentes processos químicos industriais, é possível analisar esses dados de forma a estabelecer quando uma falha ocorre ou não em um sistema. Um método de aprendizado de máquina baseada na teoria de aprendizado estatístico, a máquina de vetores de suporte (*Support Vector Machine* - SVM), vêm se mostrando muito promissor para a detecção e diagnóstico de falhas. A técnica do SVM já foi utilizada em várias aplicações como reconhecimento facial (Lu *et al.*, 2002), previsão de séries temporais (Ahn *et al.*, 2011), detecção de falhas (Gryllias e Antoniadis, 2012; Park *et al.*, 2011) e modelagem de sistemas dinâmicos não lineares (Wu, 2011).

Esse trabalho apresenta os resultados da detecção e diagnóstico de falhas em um sistema de reação para a produção de ciclopentenol em um CSTR (*Continuous Stirred Tank Reactor*), utilizando a metodologia de aprendizado estatístico de máquinas de vetores de suporte, ou SVM.

2. SISTEMA DE DETECÇÃO DE FALHAS

O sistema de detecção de falhas proposto nesse trabalho consiste de duas etapas: (a) treinamento de máquinas de detecção e classificação com dados históricos de plantas de interesse; (b) utilização das máquinas de detecção *online* para monitoramento e classificação de eventos de falhas. A detecção de falhas se dá com a separação de medidas de processo na máquina de vetores de suporte (SVM, *support vector machine*), de forma a especificar a qual categoria de dados (normais ou de respectiva falha) ele está intimamente ligado. A Figura 1 apresenta a estrutura do sistema de detecção proposto que utiliza as metodologias SVM-C (máquina de vetores de suporte para classificação) e SVM-R (máquina de vetores de suporte para regressão) para detectar e diagnosticar as falhas em um processo controlado. As falhas a serem avaliadas pertencem a três categorias: (i) Falha tipo 1 - perturbação não medida (que nesse caso é tratada com uma falha); (ii) Falha tipo 2 - falha em sensor e (iii) Falha tipo 3 - falha em atuador.

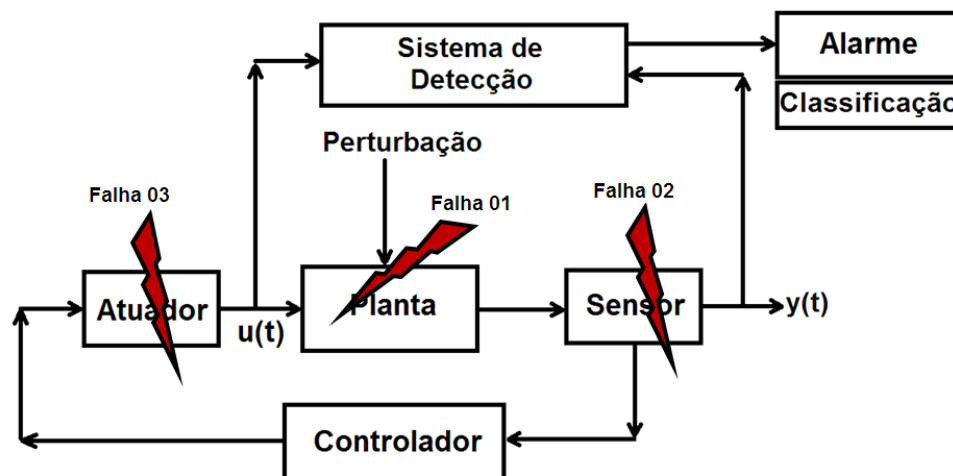


Figura 1 – Fluxograma de detecção de falhas.

As máquinas de vetores de suporte (SVM) são modelos de aprendizado supervisionado com algoritmos de aprendizagem associados que analisam os dados e reconhecem padrões, utilizados para classificações e análise de regressões. O SVM básico toma um conjunto de dados de entrada, e faz a predição, para cada entrada conhecida, de qual das duas possíveis classes constitui a saída, fazendo deste um classificador linear binário não probabilístico.

2.1. Máquinas de vetores de suporte para classificação (SVM-C)

O SVM para análise de classificação (SVM-C) utiliza os dados obtidos através da operação normal do sistema e as operações com os padrões de falhas na mesma matriz de aprendizado, designando cada padrão de falha, assim como o padrão normal, a categorias diferentes. Nesse problema o objetivo é separar as duas classes através de uma função que é induzida a partir dos dados disponíveis.

Hiperplano de separação ótimo: Considere o problema de separar o conjunto de vetores de treinamento pertencentes a duas classes distintas da Equação 1.

$$D = \{(x^1, y^1), \dots, (x^i, y^i)\}, \quad x \in \mathbb{R}^n, y \in \{-1, 1\}, \quad (1)$$

em que D apresenta os dados de treinamento do SVM, y^i indica a classe a que o dado x^i pertence, com um hiperplano (Equação 2),

$$\langle w, x \rangle + b = 0 \quad (2)$$

em que $\langle w, x \rangle$ apresenta o produto escalar e w o vetor normal ao hiperplano.

O conjunto de vetores é dito otimamente separado por um hiperplano se for separado sem erros e a distância entre o vetor mais próximo e o hiperplano for máxima. Existe alguma redundância na Equação 2, e sem perda de generalidade, é apropriado considerar um hiperplano canônico (Vapnik, 1995), em que os parâmetros w e b são restringidos pela Equação 3.

$$\min_i |\langle w, x^i \rangle + b| = 1 \quad (3)$$

Esta restrição incisiva na parametrização é preferível à alternativa no sentido de simplificar a formulação do problema. Em palavras, ele afirma que: *a norma do vetor de peso deve ser igual ao inverso da distância, do ponto mais próximo do conjunto de dados para o hiperplano.*

Então, o classificador rígido é, de acordo com a Equação 4,

$$f(x) = \text{sgn}(\langle w^*, x \rangle + b) \quad (4)$$

Alternativamente, um classificador suave pode ser utilizado, que interpola linearmente a margem, de acordo com a Equação 5,

$$f(x) = h(\langle w^*, x \rangle + b) \text{ em que } h(z) = \begin{cases} -1: z < -1 \\ z : -1 \leq z \leq 1 \\ +1: z > 1 \end{cases} \quad (5)$$

Este pode ser mais apropriado que o classificador rígido da Equação 4, porque produz um valor real entre -1 e 1 quando o classificador é consultado dentro da margem, onde não reside nenhum dado de treinamento.

A utilização de funções *kernel* possibilita a operação das máquinas de vetores de suporte em espaços de características implícitos de alta dimensão, sem necessitar o cálculo das coordenadas desses dados no espaço em questão, porém simplificando através do cálculo do produto interno entre as imagens de todos os pares de dados no espaço de característica. Essa operação é frequentemente menos custosa computacionalmente que o cálculo explícito das coordenadas. Esse método de

aproximação é conhecida como *kernel trick*. As funções *kernel* foram introduzidas para classificar dados sequenciais, gráficos, textos, imagens, assim como vetores.

O SVM é um dos algoritmos capazes de operar com *kernels*, assim como os processos gaussianos, a análise de componentes principais (*Principal Component Analysis – PCA*), entre outros algoritmos. Qualquer modelo linear pode ser transformado em um modelo não linear através da aplicação do *kernel trick* ao modelo: trocando suas características (preditores) por uma função de *kernel*.

Os *kernels* possuem alguns parâmetros que devem ser ajustados para melhor adaptar o hiperplano de separação aos dados de treinamento. De acordo com o melhor ajuste, existiram valores para o parâmetro C , chamado de parâmetro de custo, que determina um ponto de equilíbrio razoável entre a maximização da margem e a minimização do erro de classificação, assim como um melhor *kernel* que representou melhor os dados de treinamento. O *kernel* RBF (ou Gaussiano) possui o parâmetro custo C e o parâmetro γ (gamma), sendo que o parâmetro γ é responsável por ajustar os dados de treinamento ao modelo.

2.2. Máquinas de vetores de suporte para regressão (SVM-R)

O SVM para análise de regressão (SVM-R) cria um modelo para os dados de condições operacionais do sistema, e faz a predição dos resultados para todas as entradas aplicadas no modelo, resultando em uma diferença entre o valor real e o valor predito para as variáveis de saída. Os SVMs podem também ser aplicados para problemas de regressão através da introdução de uma função alternativa de perda (Smola, 1996). A função de perda convencional deve ser modificada para incluir a medição de distância.

De forma similar ao problema de classificação, um modelo não linear é usualmente necessário para adequar os dados a um modelo de interesse à indústria. Da mesma maneira que a aproximação SVM-C não linear, um mapeamento não linear pode ser utilizado para mapear os dados em um espaço característico de maior dimensão em que a regressão linear é realizada. A abordagem de *kernel* é novamente empregada para resolver os problemas da dimensionalidade. A solução SVM-R não linear, utilizando uma função de perda ϵ -insensitiva, é dada pela Equação 6,

$$\max_{\alpha, \alpha^*} W(\alpha, \alpha^*) = \max_{\alpha, \alpha^*} \sum_{i=1}^l \alpha_i^* (y_i - \epsilon) - \alpha_i (y_i + \epsilon) - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j) K(x_i, x_j) \quad (6)$$

em que α_i são os multiplicadores de Lagrange, e cujas restrições são dadas pela Equação 7,

$$0 \leq \alpha_i, \alpha_i^* \leq C, \quad i = 1, \dots, l$$

$$\sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \quad (7)$$

Resolvendo a Equação 6 com restrições da Equação 7, determina-se os multiplicadores de Lagrange, α_i , α_i^* , e a função de regressão é dada pela Equação 8,

$$f(x) = \sum_{SVs} (\bar{\alpha}_i - \bar{\alpha}_i^*) K(\mathbf{x}_i, \mathbf{x}) + \bar{b} \quad (8)$$

$$\langle \bar{w}, x \rangle = \sum_{i=1}^l (\bar{\alpha}_i - \bar{\alpha}_i^*) K(x_i, x_j)$$

$$\bar{b} = -\frac{1}{2} \sum_{i=1}^l (\bar{\alpha}_i - \bar{\alpha}_i^*) (K(x_i, x_r) + K(x_i, x_s)) \quad (9)$$

Tal como acontece com o SVM-C, a restrição de igualdade pode ser descartada se o *kernel* contém um termo de polarização, b sendo acomodado dentro da função *kernel*, e a função de regressão é apresentada na Equação 10,

$$f(x) = \sum_{i=1}^l (\bar{\alpha}_i - \bar{\alpha}_i^*) K(x_i, x) \quad (10)$$

Os critérios de otimização para as outras funções de perda são obtidas da mesma forma através da substituição do produto escalar com uma função *kernel*. A função de perda ε -insensitiva é atraente porque, ao contrário das funções de custo quadrática e de Huber, onde todos os pontos de dados serão vetores de suporte, a solução SV pode ser esparsa. A função de perda quadrática produz uma solução que é equivalente a regressão no cume, ou uma regularização de ordem zero, onde o parâmetro de regularização é igual a $\lambda = 1/2C$.

3. ESTUDO DE CASO

Um reator de produção de ciclopentenol que segue a dinâmica da reação de *van der Vusse* é investigado para três diferentes padrões de falhas. Considera-se o mecanismo de reação conhecido como reação de *van der Vusse*, descrito em Klatt e Engell (1998). A reação principal é a transformação do reagente ciclopentadieno (A) no produto ciclopentenol (B). A reação paralela ocorre produzindo um produto secundário, o componente dicitlopentadieno (C). Além disso, ciclopentadieno reage novamente formando um produto indesejado, o componente ciclopentanediol (D). Todas essas reações podem ser descritas através do seguinte esquema reacional (Equação 11).



A corrente de entrada do reator contém somente o reagente A na concentração C_{A0} . Considerando que a densidade do líquido é constante e uma distribuição de um tempo de residência ideal dentro do reator, as equações dinâmicas de *van der Vusse* do reator (Figura 2) são apresentados em Reis e Oliveira-Lopes (2007). As taxas k_1 , k_2 e k_3 dependem exponencialmente da temperatura T do reator de acordo com a Lei de Arrhenius.

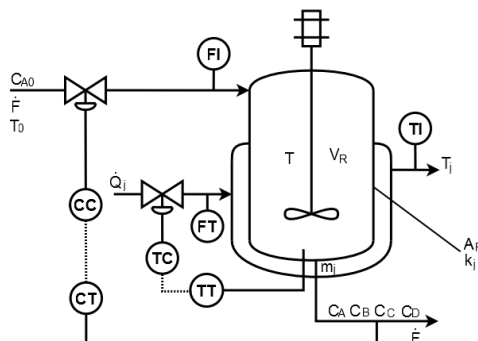


Figura 2 – Reator de *van der Vusse* (Reis e Oliveira-Lopes, 2007).

Considera-se que a temperatura do reator, a concentração do ciclopentenol no reator, a temperatura da jaqueta de resfriamento, a vazão de calor removido e de reagente são obtidos através de instrumentos de medições. Para propósito de simulação, foi adicionado um ruído Gaussiano de média zero e variância 1×10^{-5} para concentração e 1×10^{-3} para as outras medidas. A vazão de entrada do reagente A, \dot{F} , e a quantidade de calor removido pelo refrigerante, \dot{Q}_j , são as variáveis manipuladas e sujeitas as restrições apresentadas na Equação 12.

$$\begin{aligned} 50 \frac{L}{h} \leq \dot{F} = u_1 \leq 350 \frac{L}{h} \\ -8500 \frac{kJ}{h} \leq \dot{Q}_j = u_2 \leq 0 \frac{kJ}{h} \end{aligned} \quad (12)$$

Os valores dos parâmetros do modelo podem ser encontrados no artigo de Klatt e Engell (1998). Foi mostrado que para uma taxa constante de calor removido e uma variação de 50 – 1500 l/h de alimentação de reagente do reator, esse processo existe em seis regiões com diferentes graus de não linearidade (Ogunnaike e Ray, 1994).

3.1. Resultados da Detecção de Falhas

Nesse trabalho, foi escolhido restringir a vazão entre 50 e 350 l/h para os limites físicos de forma que o volume do reator simulado não seja excedido. Para controlar o processo foram designados dois controladores PID (Proporcional-Integral-Derivativo), um para controlar a concentração de produto na saída do reator e um para a temperatura do reator. O valor de *setpoint* para a concentração de A igual a 0,69883 mol/l e 407,031 K para a temperatura do reator. Esses valores são relativos a operação em estado estacionário, gerada através da escolha de uma taxa de fluxo de alimentação do reagente de 112 l/h (u_1) e uma taxa de calor removido -2856,91 kJ/h (u_2). O processo e a subsequente detecção e diagnóstico de falhas no processo de produção de A foram realizados através de simulação computacional com o *software* matemático livre SciLab® 5.4.1. De forma ilustrativa, considera-se a existência de três conhecidas falhas no processo (Reis e Oliveira-Lopes, 2007).

Dados do sistema foram coletados durante a operação normal e durante operações anormais ocorridas durante as falhas. Esses dados foram organizados e armazenados em matrizes, de acordo com a Equação 1, para uso posterior. Os dados então passaram por pré-tratamento: variáveis com nenhuma informação relevante para o processo de monitoramento foram removidas e os dados foram escalonados para evitar que variáveis particulares dominassem o método de monitoramento de processos, e foram removidos *outliers* que poderiam significativamente influenciar a estimativa dos parâmetros estatísticos e outros parâmetros relacionados a alguma medição em particular.

Os perfis da operação normal para as variáveis de entrada e de saída estão apresentados na Figura 3(a). A falha 01 considera que ocorreu uma mudança degrau na concentração do reagente A na alimentação do reator, de 5,1 mol/l para 4,1 mol/l, aplicada no instante de 8h. A Figura 3(b) apresenta os perfis das variáveis de entrada e de saída, quando a falha 01 é aplicada. A falha 02 considera que o sensor de temperatura do reator está danificado a partir de um determinado momento da simulação e a temperatura é conhecida incorretamente 10% maior que a última medida correta produzida pelo sensor mais um ruído randômico gerado de uma distribuição Gaussiana com média 0 e variância 1×10^{-3} . A Figura 3(c) apresenta os comportamentos das variáveis de saída e de entrada quando a falha 02 é aplicada no instante de oito horas. A falha 03 apresenta um emperramento da válvula de alimentação do reagente A, na posição que produz vazão 30% menor que o estado estacionário. A Figura 3(d) apresenta os comportamentos das variáveis de entrada e saída quando aplicada a falha 03 no instante de 8h.

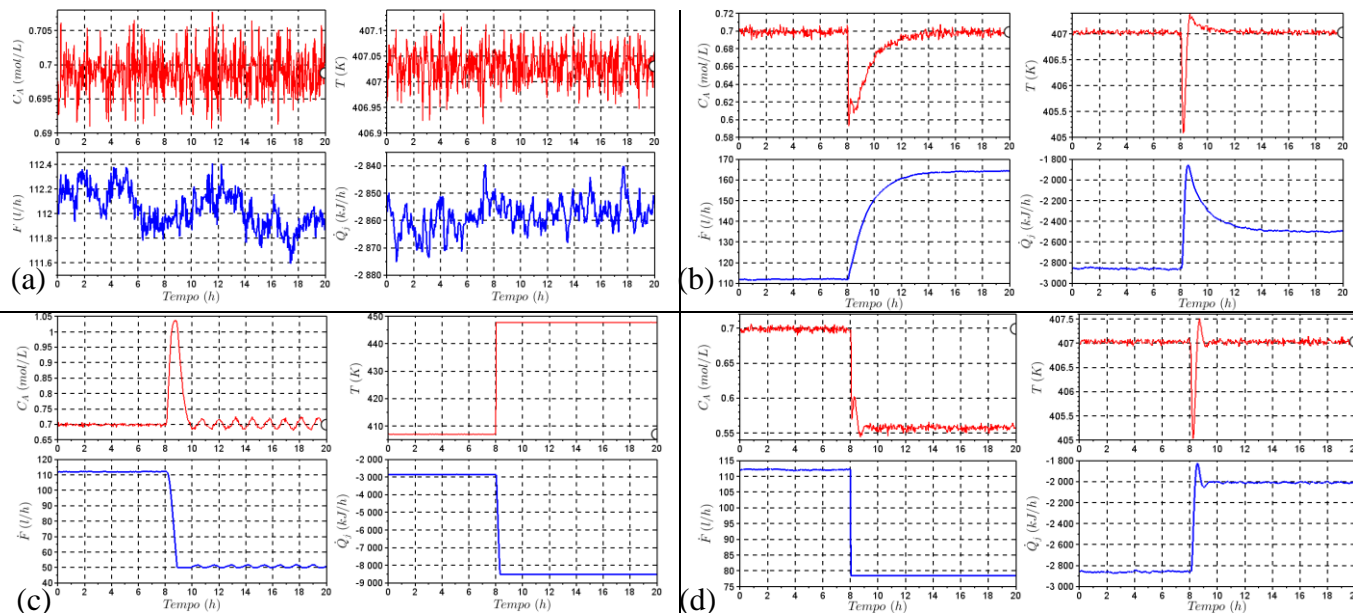


Figura 3 – Comportamento das variáveis de entrada (\dot{F} e \dot{Q}_j) e variáveis de saída (C_A e T) para falhas aplicadas no instante de 8 h. (a) Condição Normal, (b) Falha 01, (c) Falha 02 e (d) Falha 03.

A Figura 4 apresenta a variável de saída C_A , real e predita através do SVM-R, para a condição normal e as falhas 01, 02 e 03.

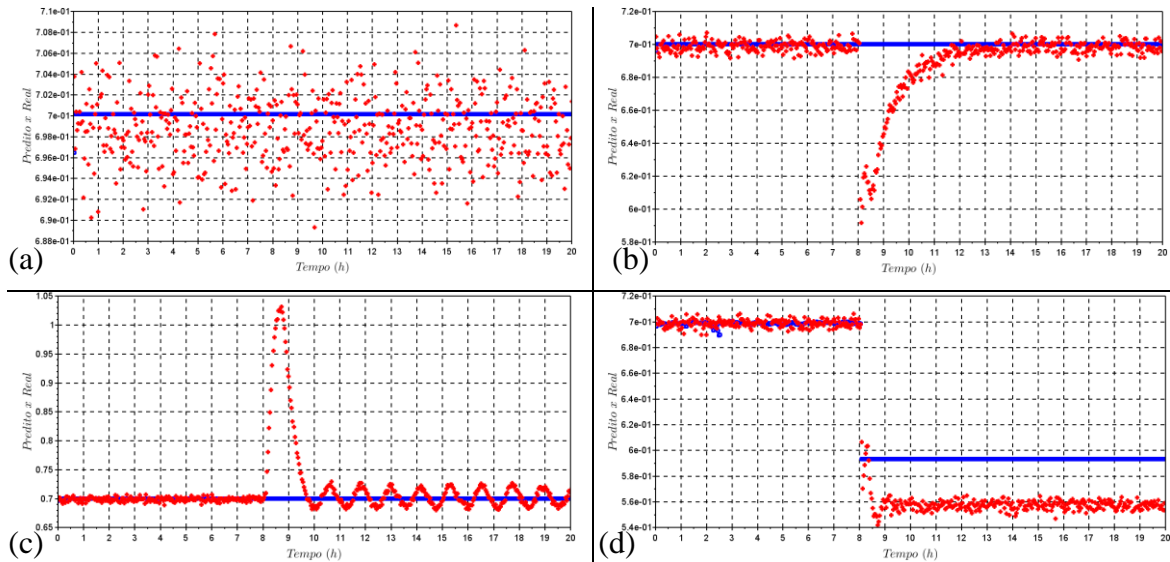


Figura 4 – Comportamento da variável de saída real versus predita (C_A) para falhas aplicadas no instante de 8 h. (a) Condição Normal, (b) Falha 01, (c) Falha 02 e (d) Falha 03.

A Figura 5 apresenta a variável de saída T , real e predita através do SVM-R, para a condição normal e as falhas 01, 02 e 03.

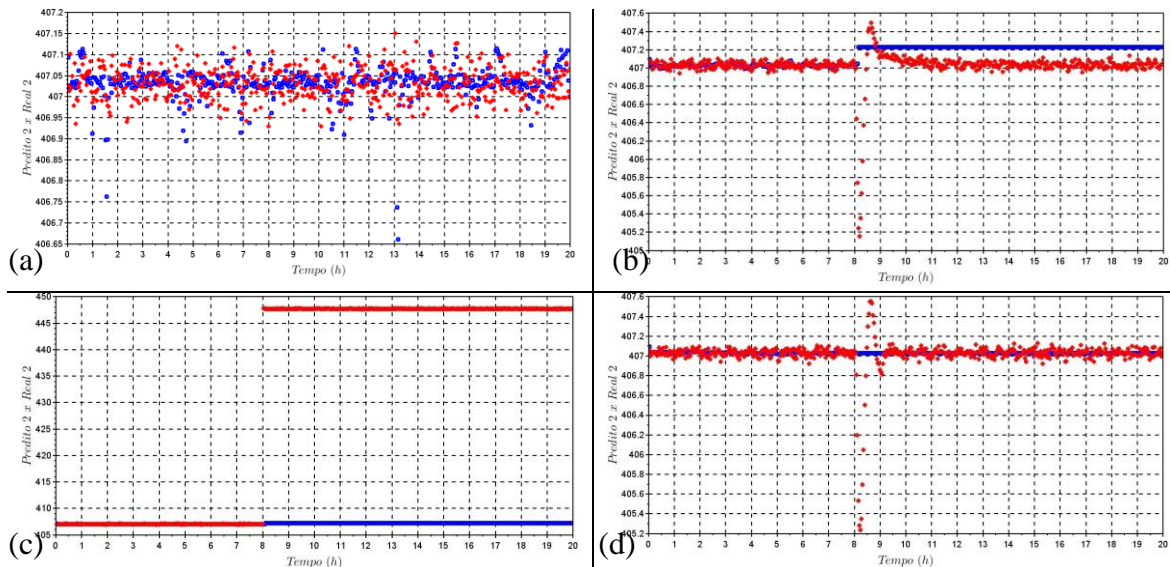


Figura 5 – Comportamento da variável de saída real versus predita (T) para falhas aplicadas no instante de 8 h. (a) Condição Normal, (b) Falha 01, (c) Falha 02 e (d) Falha 03.

Os resultados para as condições operacionais investigadas para o reator CSTR estão relacionados na Tabela 1. Nessa tabela, os resultados contêm métricas para o diagnóstico e detecção de falhas para o SVM-C e o SVM-R. Os algoritmos SVM-C e SVM-R foram aplicados através do pacote LIBSVM (Chang e Lin, 2011) para o SciLab[®]. Os parâmetros para o SVM-C são $C=137,19$ e

o *Kernel* dado pela função de base radial $K(u,v) = \exp(-\gamma|u-v|^2)$ em que $\gamma = 191085$. Esses parâmetros foram encontrados através de um escaneamento, direcionado a encontrar o modelo para o SVM-C que melhor se ajustasse aos dados de treinamento. Para o SVM-R, os parâmetros foram estabelecidos como $C = 100$ e o *Kernel* dado pela função de base radial em que $\gamma = 0,5$ e $\varepsilon = 0,05$.

Foram medidos para os métodos o instante em que a falha foi reconhecida (TF) e o instante em que a falha foi diagnosticada corretamente pela primeira vez (TFR). Para medir a qualidade dos métodos propostos para detectar a falha, foi calculado o atraso da detecção (TAD) em minutos, que foi a quantidade de tempo que se passou desde o instante em que a falha foi aplicada no sistema. Nesse trabalho, foram utilizados os índices apresentados nas Equações 13 e 14 para medir a eficiência dos métodos de detecção e diagnóstico de falhas. FD apresenta a falha diagnosticada pelo sistema de detecção.

Tabela 1 - Resultados das simulações. 20 horas de operação, tempo amostral de 2,4 minutos e falha ativada após 8 horas de operação.

Operação	Método	TFR (h)	TF (h)	D _{F/F}	D _{N/N}	FD	TAD (min)
Falha 01	SVM-C	8,28	8,24	98,33	100,00	1	14,4
	SVM-R	8,08	8,04	99,67	100,00	-	4,8
Falha 02	SVM-C	8,08	8,04	99,67	100,00	2	4,8
	SVM-R	8,04	8,00	100,0	100,00	-	2,4
Falha 03	SVM-C	8,08	8,04	99,67	100,00	3	4,8
	SVM-R	8,08	8,04	99,67	100,00	-	4,8

$$D_{F/F} = \frac{\# \text{ instantes detectados com falhas}}{\# \text{ instantes com falhas}} \times 100 \quad (13)$$

$$D_{N/N} = \frac{\# \text{ instantes detectados sem falha}}{\# \text{ instantes em falhas}} \times 100 \quad (14)$$

Os resultados para a detecção e diagnóstico de falhas indicam que as máquinas de vetores de suporte com análises de classificação e regressão utilizadas possuem uma maior confiabilidade e uma detecção rápida e eficiente, com a utilização de poucas informações do processo. Os métodos de SVM utilizados para classificação de falhas parecem entregar resultados condizentes e robustos para os cenários investigados.

4. CONCLUSÃO

O método de máquinas de vetores de suporte mostrou que pode, de forma eficiente, reconhecer padrões sob as condições desenvolvidas. Ambos os métodos levaram mais de um instante amostral para identificar corretamente a falha em questão. Isto é devido ao próprio algoritmo, que requer uma sucessão de mais de um instante amostral para detectar a condição diferente da normal. As

metodologias baseadas em SVM provaram ser confiáveis e eficientes, detectando falhas em tempo hábil, de acordo com cada uma das operações anormais. O SVM-R provou ser mais eficiente do que o SVM-C para a detecção de falhas nos casos tratados, provavelmente devido a sintonia e ao fato da metodologia utilizar somente os dados da operação normal para a detecção de falhas. A sintonia dos métodos, entretanto, depende de um procedimento de otimização dos parâmetros da máquina de vetores de suporte. Por outro lado a metodologia SVM-C não apenas é capaz de detectar, mas também de classificar, ou seja, diagnosticar a falha encontrada dentre os padrões treinados. O acoplamento das máquinas reúne no sistema proposto os lados positivos de ambas metodologias.

5. AGRADECIMENTOS

Os autores agradecem à UFU, à CAPES e à FAPEMIG (Processo PCE-00089-14) pelo apoio concedido.

6. REFERÊNCIAS

- AHN, J.J.; OH, K.J.; KIM, T.Y.; KIM, D.H. Usefulness of support vector machine to develop an early warning system for financial crisis, *Expert Systems with Applications*, v. 38, 4ª Ed., 2966-2973, 2011.
- GRYLLIAS, K.C.; ANTONIADIS, I.A. A Support Vector Machine approach based on physical model training for rolling element bearing fault detection in industrial environments, *Engineering Applications of Artificial Intelligence*, v. 25, 2ª Ed., 326-344, 2012
- KLATT, K.U., ENGELL, S. Gain-scheduling trajectory control of a continuous stirred tank reactor. *Computers & Chemical Engineering*, 491–502, 1998.
- LU, J.; PLATANIOTIS, K.N.; VENETSANOPOULOS, A.N. Face recognition using feature optimization and mu-support vector learning, neural networks for signal processing XI, em *Proceedings of the IEEE signal, processing society workshop*. 373-382, 2001.
- OGUNNAIKE, B.A., RAY, W.H. *Process Dynamics, Modeling, and Control*. Oxford University Press, New York, 1994.
- PARK, J.; KWON, I.-H.; KIM, S.-S. BAEK, J.-G. Spline regression based feature extraction for semiconductor process fault detection using support vector machine, *Expert Systems with Applications*, v. 38, 5ª Ed., 5711-5718, 2011.
- REIS, L.L.G.; OLIVEIRA-LOPES, L.C. Controle PID Tolerante a Falhas de um CSTR. em *V Seminário Nacional de Controle e Automação Industrial, Elétrica e de Telecomunicações*, Salvador. V SNCA, 1-7, 2007.
- SMOLA, A. J. Regression estimation with support vector learning machines. *Master's thesis, Technische Universität München*, 1996.
- WU, Q. Car assembly line fault diagnosis model based on triangular fuzzy Gaussian wavelet kernel support vector classifier machine and genetic algorithm, *Expert Systems with Applications*, v. 38, 12ª Ed., 14812-14818, 2011.