

Desenvolvimento de Inferências utilizando o Método de Otimização Colônia de Formigas

L. RANZAN, C. RANZAN, L. F. TRIERWEILER e J. O. TRIERWEILER

¹ Universidade Federal do Rio Grande do Sul, Departamento de Engenharia Química

E-mail para contato: lucas.ranzan, cassiano, luciane, jorge@enq.ufrgs.br

RESUMO – O controle de processos industriais depende diretamente da capacidade em se medir ou prever as variáveis de interesse. Em sistemas com elevado número de variáveis medidas, existe a dúvida sobre quais destas são mais significativas para a construção de inferidores de estado. Esta seleção é usualmente baseada em técnicas de Busca Exaustiva (BE) que são impraticáveis para sistemas complexos. Este trabalho sugere uma nova metodologia para seleção de variáveis com o uso de algoritmos de otimização de Colônia de Formigas e modelos multivariáveis (VS-ACO), minimizando o custo computacional. Para análise da metodologia, diversos modelos foram desenvolvidos e comparados através do coeficiente de correlação e RMSE. Como estudo de caso são utilizados dados reais de uma coluna fracionadora de GLP. Os resultados obtidos mostram que VS-ACO obtém a mesma seleção de variáveis em um tempo cerca de 40 vezes menor que BE, comprovando a aplicabilidade da técnica em sistemas reais complexos.

1. INTRODUÇÃO

Inferidores de estado são baseados em modelos matemáticos preditivos construídos a partir de medidas de um sistema, que tem por objetivo gerar informações rápidas e freqüentes de outra propriedade que em geral tem uma difícil medição direta. As inferências tem por sua vez ampla aplicação na indústria de processos, atuando como ferramentas para monitoramento, detecção de falhas, back-up de outros sensores e no controle de processos (Kano *et al.*, 2000).

Na literatura existem diversos estudos sobre o uso de inferidores de estado como ferramentas para o controle de processos. A automação dos processos químicos, além de garantir a estabilidade e a segurança da unidade industrial, também colabora diretamente para a qualidade e garantia de especificação do produto final (Fortuna, Graziani e Xibilia, 2005).

A construção de um inferidor pode ser dividida em partes, entre elas: a avaliação inicial dos dados, seleção de períodos em estado estacionário, tratamento dos dados (normalização para média zero e desvio padrão unitário) e por fim a escolha do tipo de modelo a ser ajustado, além de quantificar seu desempenho.

A etapa de seleção de variáveis é de fundamental importância para a construção de qualquer

modelo, e está diretamente associada ao sucesso da modelagem. Incluir um número excessivo de variáveis pode levar a problemas como ruído, aumentando a variabilidade da resposta. (Warne *et al.*, 2004)

Diversas técnicas simples podem ser usadas para uma avaliação preliminar das variáveis disponíveis. Análises de gráficos entre variáveis dependentes versus independentes permite uma busca visual por possíveis relações existentes. Porém, esta análise simples ignora o efeito das outras variáveis candidatas no conjunto. Assim, em um caso multivariável, a verdadeira correlação pode ser ocultada pelo efeito de outras variáveis.

Surge então a necessidade de uma metodologia mais complexa capaz de definir com precisão quais das variáveis disponíveis geram os modelos com melhores resultados, sobretudo em sistemas de elevada ordem, onde diversas variáveis estão disponíveis para inferir variáveis de interesse

A forma mais simples de seleção de variáveis consiste na avaliação de todas as combinações possíveis das variáveis auxiliares para obter assim o melhor modelo para descrever o sistema. Esta metodologia, também conhecida como busca exaustiva (BE), exige uma grande carga computacional, uma vez que o número de modelos a serem analisados é exponencialmente proporcional ao número de variáveis auxiliares na proporção 2^q (sendo q o número de variáveis auxiliares testadas). Esta limitação computacional fez com que esta metodologia fosse condicionada a ser aplicada apenas em sistemas com baixa ordem. A evolução tecnológica das últimas décadas tornou o gasto computacional uma variável de menor impacto na seleção de variáveis, e, assim, a metodologia de busca exaustiva será nesse trabalho avaliada como a melhor alternativa para sistemas com poucas variáveis auxiliares e o modelo encontrado servirá como base para comparação entre outras metodologias. (Warne *et al.*, 2004)

Existem na literatura diversos métodos de otimização que buscam a seleção de variáveis de forma a não ser necessário o uso de todas as combinações de modelos possíveis. Pode-se citar aqui trabalhos como (Facchin, 2005) que compara o uso de técnicas de algoritmos genéticos com BE sem a obtenção de bons resultados.

Neste trabalho será apresentada uma nova metodologia para seleção de variáveis auxiliares baseada em técnicas de otimização Colônia de Formigas (VS-ACO). Esta técnica, desenvolvida para resolver problemas de otimização, é baseada no comportamento real de formigas, que por meio de comunicação indireta através de trilhas de feromônios, conseguem encontrar o menor caminho entre seu ninho e a fonte de alimento (Tabakhi, Moradi e Akhlaghian, 2014).

Os modelos desenvolvidos durante este estudo serão comparados através de cálculos de coeficiente de correlação (R^2) e da raiz do erro médio quadrático (root mean square error - RMSE), de acordo com as equações 1 e 2. Nas equações apresentadas, o subíndice p refere ao vetor de variáveis preditas, resultantes da avaliação do modelo, o subíndice m ao vetor de variáveis medidas, N o número de medidas e y é o vetor das variáveis de interesse (Kaneko, Arakawa e Funatsu, 2011).

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_{p,i} - y_{m,i})^2}{\sum_{i=1}^N (y_{m,i} - \bar{y}_m)^2} \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_{p,i} - y_{m,i})^2}{N}} \quad (2)$$

O escopo final deste trabalho é então a comparação de metodologias de seleção de variáveis por BE contra VS-ACO, para determinar se esta última converge para os mesmos modelos de ótimo, e confirmar sua aplicabilidade em sistemas com elevado número de variáveis disponíveis.

Se comprovado o uso da nova metodologia apresentada para o sistema proposto com baixo número de variáveis auxiliares, é possível a expansão da técnica para sistemas de maior grau de complexidade, onde o alto número de variáveis auxiliares impossibilita o uso da técnica de BE, como, por exemplo, sistemas baseados em dados de espectroscopia.

2. Estudo de Caso

Os dados utilizados neste trabalho são provenientes de uma unidade real de processamento de gás natural (UPGN). Um estudo detalhado do funcionamento da planta, coleta e tratamento de dados pode ser visto no trabalho de (Fleck, 2012).

A unidade consiste, resumidamente, em duas colunas de destilação, onde na primeira ocorre a separação do metano e do etano como produtos de topo e na segunda é obtido como produto de topo propano e butanos (GLP) e no fundo a gasolina natural.

O estudo de caso deste trabalho é focado na coluna fracionadora de GLP, que possui três principais requisitos para especificação de produto. Primeiramente, o GLP deve condensar facilmente sobre pressão para transporte e armazenamento, e também deve evaporar com facilidade nas condições ambientes para queima. O primeiro é garantido pela PVR (Pressão de Vapor Reid), que indica a pressão em que deve ser armazenado o produto e é medida indireta da presença de hidrocarbonetos leves. O segundo requisito é avaliado pelo teste de intemperismo, que qualifica a dificuldade em vaporizar o GLP na pressão atmosférica e é uma medida indireta da presença de hidrocarbonetos pesados. O terceiro requisito é a qualidade e teor de etano, que deve ser limitado para evitar o deslocamento da chama durante seu uso no fogão.

3. Metodologia

O desenvolvimento deste estudo foi segmentado em três partes: a colheita e tratamento inicial dos dados, a seleção de variáveis auxiliares por BE e VS-ACO e a comparação entre os resultados.

Todas as implementações e cálculos neste trabalho foram realizados com o uso de MATLAB software (Ver. 7.8.0.347, The Mathworks, Inc., Natick, USA).

Os dados foram divididos em dois conjuntos, um de validação, e outro de calibração, utilizando a técnica y-rank. Durante a comparação de resultados, uma distinção entre dados de calibração e validação serão apresentados.

3.1 Coleta e Tratamento de Dados

Para a coleta de dados, primeiramente foram analisados os fluxogramas de processo e todas as variáveis que apresentaram certo grau de correlação com a variável de interesse, a qual se deseja modelar, foram selecionadas. Com esta pré-seleção de variáveis, foram adquiridos os dados reais da planta, pelo maior tempo de amostragem disponível.

Os dados brutos passam então por um pré-tratamento para garantir sua qualidade e repetibilidade. Períodos de parada de planta ou falha em sensores são identificados e retirados. *Outliers* (observações não consistentes) foram descartados. Após, foram identificados estados estacionários, a partir dos quais os modelos estacionários foram obtidos.

Por fim, um conjunto com 124 pontos (amostragens) de vinte de uma propriedades diferentes (variáveis independentes) é obtida. A partir desta matriz de dados experimentais é realizado o estudo para obtenção de inferidores de estado, capazes de converter dados de processo em variáveis de interesse, através de suas utilizações em modelos do tipo MISO (*Multiple Input Single Output*).

3.2 Seleção de Variáveis por Busca Exaustiva - BE

A partir da literatura, é possível encontrar diversos trabalhos que focam seu estudo na busca pela dimensão ideal de modelos para problemas de inferências em colunas de destilação. Trabalhos como (Kano *et al.*, 2000) e (Fleck, 2012) apresentam bons resultados para o uso de modelos com seis variáveis independentes, baseados em técnicas de redução de dimensionalidade. Sendo assim, a comparação entre metodologias neste trabalho vai se focar na busca por modelos que tenham como base seis variáveis independentes.

Com posse da tabela de dados selecionados, o algoritmo matemático de BE foi aplicado, analisando todos os possíveis modelos de dimensão seis, e selecionando o melhor modelo encontrado com base em valores de R^2 e RMSE.

Com base em um universo de variáveis independentes compostos por 20 variáveis medidas, combinando-as seis a seis, 38760 possíveis combinações, ou modelos, são possíveis, sem repetições de variáveis dentro dos modelos.

3.3 Seleção de Variáveis por VS-ACO

A implementação de VS-ACO neste estudo é uma versão discreta, que baseia a seleção de variáveis auxiliares em um fator randômico associado a uma função de densidade de feromônios

(variável qualitativa auxiliar ao método de otimização). A cada nova iteração, o vetor de feromônio é atualizado para as variáveis auxiliares testadas, em função do resultado apresentado pela função do erro residual entre as variáveis de interesse medidas e previstas pelo modelo. Uma visão mais profunda do algoritmo aplicado pode ser encontrada no trabalho de (Ranzan *et al.*, 2014).

O algoritmo pode ser dividido em quatro fases principais: inicialização das variáveis, inicialização da solução, rotina de busca e apresentação dos resultados.

Na fase inicial, são inicializadas as variáveis necessárias para dar início à resolução do problema de otimização. São carregadas as matrizes de dados de processo, o vetor da variável observada, a escolha do número de ciclos, do tamanho do exército de formigas, do tamanho e tipo do modelo, do valor base inicial para trilha de feromônios, e a taxa de evaporação de feromônios a cada ciclo.

Na segunda fase é feita a inicialização do vetor de soluções, resolvendo a função objetivo com uma seleção aleatória do conjunto de variáveis auxiliares disponíveis. Esta etapa inicializa o vetor de soluções que será utilizado como forma de comparação futura.

A terceira fase se refere ao núcleo do processo de otimização. Nesta fase, o programa determina a melhor combinação de variáveis auxiliares para prever a variável de interesse, através da estrutura de modelo selecionada. Nesta busca são propostas diferentes possíveis soluções que são comparadas com o melhor resultado obtido até o momento. Durante cada ciclo do algoritmo, o exército de formigas varre o universo de possibilidades que forneça o menor valor possível para função objetivo (somatório do erro quadrático entre o valor medido e o previsto). Cada elemento do exército de formigas escolhe um grupo de variáveis e o submete ao teste da função objetivo. Caso o resultado seja menor ao previamente armazenado, este então o substitui e toma seu lugar no vetor de soluções.

A seleção de cada candidato é baseado em dois fatores, um aleatório, que garante que a busca não fique retida em mínimos locais, e um baseado no vetor da trilha de feromônios (que é atualizada a cada ciclo, onde cada candidato recebe um incremento inversamente proporcional ao erro quadrático entre os valores observados e previstos apresentado pelo modelo com ele gerado).

A última etapa apresenta a melhor solução obtida durante o processo de otimização. Este resultado corresponde aos melhores conjuntos de variáveis auxiliares, combinadas na estrutura de modelo proposta, capazes de prever os valores da variável de interesse, apresentando os menores erros entre a variável medida e a variável prevista.

Assim como na BE, a metodologia VS-ACO buscou os melhores modelos de dimensão seis, baseados em comparativos de R^2 e RMSE. Por uma questão comparativa, modelos de dimensão entre um e vinte serão também avaliados, mostrando a capacidade da metodologia no quesito redução de esforço computacional.

4. Resultados

As Tabelas 1 e 2 apresentam respectivamente os resultados de R^2 de calibração e predição,

assim como os valores de RMSE de calibração (RMSEC) e predição (RMSEP), para os modelos obtidos com as metodologias BE e VS-ACO.

A Tabela 1 apresenta o melhor resultado obtido através da busca exaustiva por modelos compostos por seis variáveis de entrada. Desta forma, todas as possíveis combinações das 20 variáveis disponíveis são feitas, seis à seis, e o melhor resultado obtido para predição da variável de interesse é apresentado nesta tabela.

Tabela 1 - Melhor resultado para predição da variável de interesse utilizando Busca Exaustiva

Dimensão do Modelo	RMSEC	R ² Cal.	RMSEP	R ² pred.	Tempo Computacional (s)
6	0,4955416	0,732700	0,548317	0,732576	75,39253519

De forma equivalente à Tabela 1, a Tabela 2 apresenta os melhores resultados obtidos da combinação das variáveis de entrada disponíveis para este sistema, mas avaliando diferentes tamanhos de modelos, desde modelos MISO compostos por uma única variável de entrada, até modelos que utilizam as 20 variáveis de entrada disponíveis.

Os resultados apresentados na Tabela 2 mostram que mesmo com o incremento significativo do tamanho de modelo buscado pela metodologia VS-ACO, o incremento no tempo computacional não é diretamente proporcional.

Tabela 2 - Calibração e Teste de modelos MISO usando VS-ACO.

Dimensão do Modelo	RMSEC	R ² Cal.	RMSEP	R ² pred.	Tempo Computacional (s)
1	0,759516	0,372069	0,781305	0,484037	0,6516018066
3	0,587783	0,623927	0,702231	0,562533	1,1831075338
4	0,553481	0,666541	0,650712	0,624034	1,4174773586
6	0,495542	0,732701	0,548317	0,732577	1,8888247198
7	0,479196	0,750044	0,615423	0,671224	2,0867103742
9	0,465267	0,764364	0,561083	0,722000	2,8048372816
10	0,457584	0,772082	0,610409	0,674617	3,0317884236
12	0,450867	0,778724	0,527724	0,752454	3,5437587060
13	0,446869	0,782631	0,491042	0,785527	3,6259067512
15	0,440192	0,789078	0,510359	0,767952	4,1211368316
16	0,439275	0,789956	0,504889	0,773091	4,4068557156
18	0,438494	0,790702	0,515701	0,763560	4,6525674530
20	0,438328	0,790860	0,512761	0,766130	5,1241727863

A variação do tamanho de modelo também permite uma comparação mais precisa com relação à qualidade de predição dos modelos. Valores de RMSEC e R^2 de calibração aumentam com o incremento do tamanho de modelo, entretanto, a avaliação destes mesmos parâmetros com relação à qualidade da predição não segue o mesmo padrão. O aumento do número de variáveis de entrada provoca uma maior correlação do modelo com o conjunto de amostras de calibração, fazendo com que o tamanho do modelo apresente um tamanho ótimo, a partir do qual a predição perde eficiência. No estudo de caso abordado, a Tabela 2 indica que este ótimo seria para modelos com 13 variáveis de entrada, diferentemente do resultado indicado pela análise PCA, em (Fleck, 2012).

A Tabela 3 apresenta um comparativo entre os tempos de calibração e ajuste de modelos com dimensão seis, para as duas metodologias de avaliadas. Além disto, esta tabela também apresenta as nomenclaturas de processo das variáveis selecionadas por cada estratégia.

Tabela 3 - Modelos ótimos selecionados para a predição da variável de interesse.

Tipo de Método	Tempo (s)	Variáveis Selecionadas
BE	75,39253519	FIC_01 , TI_06 , PI_01 , PDI_01 , TI_11 , PIC_02
VS-ACO	1,8888247198	FIC_01 , TI_06 , PI_01 , PDI_01 , TI_11 , PIC_02

Através da análise dos resultados presentes na Tabela 3, fica visível que ambas as metodologias selecionaram o mesmo conjunto de variáveis, utilizando o mesmo conjunto de calibração e teste. Porém , é significativo a diminuição do tempo computacional apresentado pela estratégia VS-ACO, com um tempo computacional cerca de 40 vezes menor.

Apesar da diminuição do tempo computacional não ser um fator significativo para o sistema em questão, esta variável para a ter importância significativa no momento em que inferidores de estado necessitam ser ajustados em sistemas com elevado número de variáveis medidas.

5. Conclusões

O controle industrial de processos depende diretamente da capacidade de se medir ou prever variáveis de interesse. Durante o desenvolvimento de inferidores de estado, a etapa de seleção de variáveis é de extrema importância, impactando diretamente na qualidade da variável predita.

As pesquisas apresentadas neste trabalho mostram que a metodologia de busca e ajuste de modelos baseada uso de otimização Colônia de Formigas é capaz de obter a mesma seleção de variáveis que metodologias de Busca Exaustiva. Assim sendo, surge a possibilidade da extrapolação da técnica para sistemas de maior ordem, onde o tempo computacional inviabiliza o uso de BE.

O requerimento de tempo computacional pela nova metodologia apresentada foi de aproximadamente 40 vezes menor que o tempo da metodologia convencional, um valor expressivo que para sistemas de altas dimensões pode representar uma economia significativa em recursos, além de possibilitar o acompanhamento de variáveis de interesse de forma mais rápida e precisa.

6. REFERÊNCIAS

FACCHIN, S. **Técnicas de Análise Multivariável Aplicadas ao Desenvolvimento de Analisadores Virtuais**. PPGEQ: UFRGS 2005.

FLECK, T. D. **Nova Metodologia para Desenvolvimento de Inferências Baseadas em Dados**. Porto Alegre: PPGEQ, UFRGS 2012.

FORTUNA, L.; GRAZIANI, S.; XIBILIA, M. G. Soft sensors for product quality monitoring in debutanizer distillation columns. **Control Engineering Practice**, v. 13, n. 4, p. 499-508, 4// 2005. ISSN 0967-0661. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S0967066104000899> >.

KANEKO, H.; ARAKAWA, M.; FUNATSU, K. Novel soft sensor method for detecting completion of transition in industrial polymer processes. **Computers & Chemical Engineering**, v. 35, n. 6, p. 1135-1142, 6/9/ 2011. ISSN 0098-1354. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S0098135410003078> >.

KANO, M. et al. Inferential control system of distillation compositions using dynamic partial least squares regression. **Journal of Process Control**, v. 10, n. 2-3, p. 157-166, 4// 2000. ISSN 0959-1524. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S095915249900027X> >.

RANZAN, C. et al. Wheat flour characterization using NIR and spectral filter based on Ant Colony Optimization. **Chemometrics and Intelligent Laboratory Systems**, v. 132, n. 0, p. 133-140, 3/15/ 2014. ISSN 0169-7439. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S0169743914000203> >.

TABAKHI, S.; MORADI, P.; AKHLAGHIAN, F. An unsupervised feature selection algorithm based on ant colony optimization. **Engineering Applications of Artificial Intelligence**, v. 32, n. 0, p. 112-123, 6// 2014. ISSN 0952-1976. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S0952197614000621> >.

WARNE, K. et al. Statistical and computational intelligence techniques for inferential model development: a comparative evaluation and a novel proposition for fusion. **Engineering Applications of Artificial Intelligence**, v. 17, n. 8, p. 871-885, 12// 2004. ISSN 0952-1976. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S0952197604000971> >.