

A identificação estrutural como proposta de classificação de variáveis para a reconciliação de dados industriais

A. M. O. Júnior¹; E. L. Lima²; J. C. C. S. Pinto²

¹Programa de Pós-graduação em Engenharia Química, Universidade Federal de Sergipe, 49100-000, São Cristóvão-Se, Brasil

²Programa de Engenharia Química, Universidade Federal do Rio de Janeiro, 21941-972, Rio de Janeiro-RJ, Brasil
antonio_martins@pq.cnpq.br

(Recebido em 05 de abril de 2011; aceito em 23 de novembro de 2011)

A reconciliação dos dados é fortemente afetada pela formulação do problema, pelo desempenho da otimização e interpretação estatística dos resultados, devendo ser precedida por uma cuidadosa classificação das variáveis. Devido à complexidade dos atuais processos integrados e a grande quantidade de dados disponíveis em plantas automatizadas de alto desempenho, o uso de algoritmos de classificação vem crescendo dia a dia. Eles são aplicados no projeto de novas unidades, *revamps* e sistemas de monitoramento para reduzir a dimensão do problema de reconciliação de dados. É descrito um novo algoritmo que poderá ajudar o engenheiro a encontrar eficientes estratégias para o problema de classificação, aliado à formulação matemática. Algumas propriedades estruturais são discutidas e ilustradas, bem como é descrito o algoritmo de identificação estrutural. Existem incentivos econômicos para a classificação robusta de variáveis, pois um procedimento deficiente irá requerer uma instrumentação adicional.

Palavras-chave: Reconciliação de dados, Classificação de variáveis, Processos industriais.

Data reconciliation is strongly affected by formulation problem, statistical results interpretation and optimization performance. This must be valued by a carefully variable classification. Due to of the complexity of integrated process and the large volume of available data in highly automated plants; classification algorithms are increasing used nowadays. They are applied to the *revamps*, design and monitoring systems and to reduce the dimension of the data reconciliation problem. A new algorithm is described which can help the engineer find efficient strategies for the classification problem allied with the mathematical formulation. Some structural properties are discussed and illustrated. The new structural identification algorithm is described. There is a large economical incentive for the variable robust classification, because a defective procedure will request an additional instrumentation.

Keywords: Data reconciliation, Variable classification, Industrial Process.

1. INTRODUÇÃO

Para uma rede de processos operando no estado estacionário, as variáveis são restringidas pelas equações de conservação. Para um dado processo com conjunto de instrumentos distribuídos ao longo da planta, caso se realize uma mudança de variável sem ser detectada (observada) pelos instrumentos, a variável é dita não observável. Usando esta definição, toda variável seria observável se todas as variáveis fossem medidas. Como isso quase nunca acontece por razões de custo, conveniência ou viabilidade técnica, o que se tem na realidade é um conjunto de medições incompletas que gostaríamos de conhecer se cada variável é ou não observável. Esse problema é chamado de classificação de observabilidade.

A primeira etapa de classificação de variáveis envolve a determinação de quais variáveis são observáveis ou não observáveis e quais são redundantes ou indeterminadas. Vários autores têm publicado algoritmos para classificação de variáveis (Crowe, 1986, Stanley e Mah, 1981, 1990). Aquelas variáveis que são indeterminadas não são disponíveis para melhoria dos dados. Os erros grosseiros são identificados e removidos (Oliveira Jr, 2006; Mah, 1990; Benqlilou, 2004)

Com o desenvolvimento recente de monitoramento em linha e armazenamento de dados, os conjuntos de informações disponíveis nas plantas industriais têm crescido significativamente. No entanto, somente alguns desses dados são tratados. Dados os custos e dificuldades envolvendo os processos de medições, as medidas de cada variável dos processos químicos não

estão disponíveis e são irrealis. Quatro categorias de variáveis podem ser definidas como resultado da identificação estrutural dos processos e estudos sobre localização de sensores conforme Figura 1.

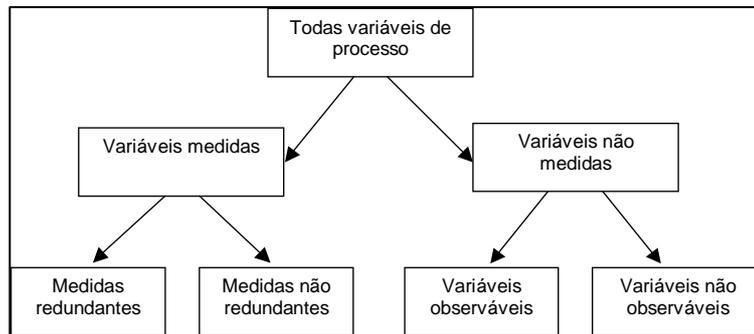


Figura 1 – Classificação de variáveis

Devido ao conjunto incompleto de instrumentos em uma unidade industrial, as variáveis não medidas são divididas em determináveis ou estimáveis e indetermináveis ou inestimáveis. As variáveis não medidas são determináveis ou estimáveis, se os valores podem ser calculados usando as medidas existentes. Já as variáveis medidas são classificadas em redundantes e não redundantes. Uma medida é redundante se a mesma permanece determinável quando a sua medida é excluída.

O objetivo do trabalho proposto é definir um novo algoritmo de classificação de todas as variáveis de um processo estudado para futuro procedimento de reconciliação de dados

2. CLASSIFICAÇÃO DE VARIÁVEIS

Crowe (1989), Stanley e Mah (1981), desenvolveram métodos baseados, para processos estacionários, com as condições necessárias e suficientes para observabilidade e redundância. Sanchez *et al* (1999) apresentaram a decomposição Q-R, através de fatorizações ortogonais, para analisar, decompor e resolver o problema de reconciliação de dados lineares e bilineares. Foi mostrado que a decomposição forneceu informações adicionais na identificação estrutural, bem como singularidades na topologia do processo. Um fluxograma simples de técnica de melhoria de medições de dados pode ser mostrado na Figura 2 subdividido em três etapas (Edgard *et al*, 2001).

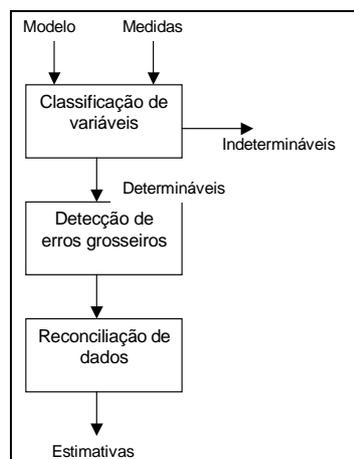


Figura 2 – Etapas para melhoria da qualidade dos dados

A primeira etapa de classificação de variáveis envolve a determinação de quais variáveis são observáveis ou não observáveis e quais são redundantes ou indeterminadas. Vários autores têm publicado algoritmos para classificação de variáveis.

Observabilidade e redundância são atributos desejáveis para as variáveis de processos. A sua importância foi discutida em trabalhos anteriores, Stanley e Mah (1981) e Kretsovalis e Mah (1987). As análises de redundância e observabilidade permitem determinar a viabilidade de se realizar um balanço de massa ou energia através da identificação de quais variáveis não observáveis. A etapa de classificação deve ser a primeira etapa da reconciliação de dados. Através dessa análise pode-se determinar uma estratégia de amostragem para garantir que todo o conjunto de variáveis não medidas seja observável, ou então remover as variáveis não observáveis que não são de interesse para evitar problemas de indeterminação. Utilizaremos a definição dada por Stanley e Mah (1981):

Redundante: Uma medida é denominada redundante se, e somente se, a observabilidade do sistema não for alterada pela sua exclusão.

Observável: uma variável não medida é chamada observável se, e somente se, pode ser calculada através de outras medições.

Equação redundante: Uma equação de balanço é chamada redundante, se e somente se, variáveis medidas estiverem presentes na equação.

A violação das equações de conservação ocorre também porque existem variáveis medidas redundantes. As variáveis não redundantes, em geral, não influem no processo de ajuste de uma função objetivo por mínimos quadrados utilizado em reconciliação de dados, bem como as variáveis não medidas concorrem apenas para aumentar o tamanho do sistema sob análise. Geralmente, a redundância trata de conflitos entre os dados e os balanços, sendo necessário reconciliar os dados. Tal procedimento também fornece uma checagem da confiabilidade das medidas de processo. Se o balanço for não redundante, não existirá possibilidade de conflito entre as equações de conservação desde que as medidas ausentes possam ser escolhidas de forma a satisfazer o balanço (PAPADOKONSTANTAKIS *et. al.*, 2005).

O algoritmo proposto difere dos existentes na literatura aberta pelo fato de não utilizar nem a decomposição Q-R, nem o enfoque grafológico. A proposta baseia-se no enfoque matricial, com sistema booleano, e identificação de variáveis de iteração como utilizado nos algoritmos de *tearing* (teia) e *partitioning* (partição) em projetos de indústrias químicas.

3. ESTRATÉGIA UTILIZADA

O problema de otimização com restrição, para ajuste das medidas, é resolvido em intervalo delimitado. Nesse contexto, a classificação de variáveis é utilizada para reduzir o conjunto de restrições, eliminando as variáveis não medidas e não redundantes. A redução do sistema de restrições permite uma resolução matemática rápida e fácil do problema.

Considera-se o processo contendo K unidades dados por $k=1, \dots, K$, e J correntes $j=1, \dots, J$ com C componentes $c=1, \dots, C$. A topologia do processo é representada pela matriz de incidência A, com as linhas correspondendo às unidades e as colunas representando as correntes. Então,

$$\begin{aligned} a_{jk} &= 1 && \text{se a corrente } j \text{ entra no nó } k \\ a_{jk} &= -1 && \text{se a corrente } j \text{ sai do nó } k \\ a_{jk} &= 0 && \text{em qualquer outro caso} \end{aligned} \quad (1)$$

As equações de balanço de massa e energia multicomponente podem ser representadas por:

$$\sum_{j=1}^J a_{j,k} \mu_j = 0, \quad k = 1, 2, \dots, K. \quad (2)$$

Nestas equações, o μ_j representa o parâmetro da corrente j (vazões individuais, vazão total, entalpia). a_{jk} é o elemento da matriz de incidência. Processos com reações químicas podem ser

facilmente inseridas nesse balanço de equação através da inserção de correntes artificiais internas ou externas. A Equação 2, também, pode ser escrita em notação matricial como segue:

$$A\mu = 0 \quad (3)$$

O vetor μ possui $n+m$ dimensões que pode ser particionado em dois vetores: o vetor de dimensão n de variáveis medidas, e o vetor de dimensão m de variáveis não medidas. As medidas de x são representadas por \tilde{x} . A diferença entre o valor medido e o valor verdadeiro é denominado de erro, representado por δ , isto é:

$$\delta = x - \tilde{x} \quad (4)$$

O valor estimado de x é representado por \hat{x} e a diferença entre o valor estimado e o valor medido será denominado de correção do parâmetro e representado por $\hat{\delta}$, ou seja:

$$\hat{\delta} = \hat{x} - \tilde{x} \quad (5)$$

Na Equação 2 foi considerado um sistema linear, porém as situações reais são normalmente diferentes. Por exemplo, considere um instrumento que sirva para medida de vazão e determinação de composição das correntes. As vazões do componente desejadas são obtidas pelo produto das duas leituras. Por conseguinte, constatamos a existência de termos bilineares $x_j x_k$, pois os termos também são encontrados nas equações de balanço. Termos similares aparecem nas equações de balanço de energia. Assim, em geral, as equações de balanço individual por componente em torno da unidade k , toma a seguinte forma:

$$\sum_{j=1}^J a_{jk} M_j x_{ji} = 0, \quad i = 1, \dots, I \quad e \quad k = 1, \dots, K \quad (6)$$

com

$$\sum_{i=1}^I x_{ji} = 1 \quad (7)$$

M_j é a vazão da corrente j e x_{ji} é a fração mássica ou molar do i ésimo componente da j ésima corrente j . Analogamente para o balanço de energia, nós temos:

$$\sum_{j=1}^J a_{jk} M_j H_j = 0, \quad \text{para } k = 1, \dots, K \quad (8)$$

onde H_j é a entalpia específica da corrente j .

Baseado nas formulações acima os seguintes problemas são definidos e serão analisados nesse trabalho:

- Classificação das variáveis medidas em redundante e não redundante
- Classificação das variáveis não medidas em observáveis e não observáveis
- Definição da seqüência e conjunto de equações que serão usadas para retificação dos valores medidos

As unidades de processo industriais são compostas de diversas unidades contendo vários componentes e várias correntes. Normalmente, os conjuntos de equações de balanço e energia

constituem um conjunto de equações não lineares (Prata, 2009). A estrutura topológica das variáveis de processo e das equações de balanço é o ponto fundamental na etapa de identificação estrutural e classificação de variáveis e correção de erros, porventura, existentes. A topologia das equações de balanço será representada por uma matriz de ocorrência onde:

- as linhas da matriz de ocorrência correspondem as equações e as colunas correspondem às variáveis de processo tanto medidas quanto não medidas
- os elementos da matriz S_{ij} irá assumir o valor booleano 1 ou 0, isto é:

$$S_{ij} = \begin{cases} 1, & \text{se a variável } j \text{ aparece na equação } i \\ 0, & \text{qualquer outra possibilidade} \end{cases} \quad (9)$$

De forma a classificar as variáveis do processo, deve-se primeiro estabelecer qual informação da equação deve ser fornecida, ou seja, estabelecer a seqüência de cálculo das equações utilizando a informações disponíveis.

4. ALGORITMO DE IDENTIFICAÇÃO ESTRUTURAL E RESULTADOS

As etapas do algoritmo utilizado são:

- a) Construção de uma matriz de ocorrência, S , para o sistema de equações a ser resolvido preenchendo as posições i,j , com 1 se a variável j aparece na equação i , ou zero em qualquer outro caso. Preparar para construir uma matriz de ocorrência reordenada que quando finalizada irá indicar a estratégia ótima de solução.
- b) Encontrar uma linha na matriz de ocorrência que tenha apenas uma entrada. Essa entrada representa um subconjunto removível. Remove-se o subconjunto deletando a linha e a coluna onde ele ocorre. Indica-se que essa coluna (variável) é para ser resolvida com a respectiva linha (equação) para o usuário.
- c) Repetir a etapa b até que não haja mais linhas com apenas uma ocorrência, eliminando sempre qualquer linha que não tenha ocorrência.
- d) Encontra-se a coluna com maior número de ocorrências. Substituem-se os valores onde foi encontrado ocorrência por 2. Define-se essa variável como variável de corte (iteração)
- e) Repete-se a etapa b até que não haja mais linhas com apenas uma ocorrência de valor 1.

Segue um exemplo de uma polimerização via radical livre, com as etapas usuais de iniciação, propagação e terminação. O subsistema possui 20 variáveis e 10 equações. Associa-se a isso, o fato de nove variáveis serem mensuradas. São elas: 1, 2, 3, 4, 6, 8, 10, 12, 14, 19. Com a informação de variáveis medidas, a matriz de ocorrência é dada por:

$$S = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

Imediatamente, pode-se observar que as variáveis 5, 9 e 13 podem ser calculadas através das equações 1, 4 e 7 respectivamente. Posteriormente, escolhe-se a variável 16 como variável de iteração que ocasionará o cálculo das variáveis 20, 7, 11 e 15 pelas equações 2, 3, 6 e 9. Em seguida, após a escolha de outra variável de corte, a variável 19 pôde ser calculada através da equação 10.

Após o final da primeira iteração, observa-se que: o grau de liberdade do processo é 2; que as equações 5 e 8 não foram utilizadas e que as variáveis 16, 17 e 18 necessitam ser informadas para o conhecimento completo do sistema.

Devido ao grau de redundância do processo, permite-se prosseguir com as iterações com a variável 16 (antiga variável de iteração) podendo ser calculada através da equação 5.

Entretanto, mesmo reduzindo-se o grau de redundância do processo para 1, as variáveis 17 e 18 continuam inobserváveis, pois não existe nenhuma variável que possa ser calculada através da equação 8. Assim, a matriz final é dada por:

$$S = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 2 & 0 & 0 \end{bmatrix}$$

Após a obtenção da classificação de todas as variáveis tem-se uma grande quantidade de informações sobre o processo em estudo. A questão, agora, passa a ser como usar as variáveis classificadas e todas as informações disponíveis do processo após a classificação das variáveis. Em um processo industrial temos os mais diversos tipos de problemas a resolver, e os objetivos variarão de um processo para outro. Algumas possibilidades encontradas podem ser:

- I. Em um processo existem parâmetros que por alguma razão são necessários. Nesse caso, queremos checar a precisão das medidas e corrigi-las quando for necessário. Então, através das variáveis redundantes selecionamos quais variáveis devem ser corrigidas;
- II. Minimização da quantidade de instrumentação existentes no processo, de tal forma que todas as variáveis não medidas sejam determinadas. Por outro lado, poderíamos selecionar o número mínimo de medidas de modo a tornar as variáveis não medidas em variáveis determinadas;
- III. Deseja-se que nem todos os parâmetros sejam determináveis, mas somente aqueles imprescindíveis ao bom andamento do processo, conseqüentemente devem ser verificados quais variáveis devem obrigatoriamente ser medidas.

5. CONCLUSÃO

O conhecimento do status de cada variável através do procedimento e classificação de variáveis permitirá: inferir com maior segurança sobre a qualidade dos dados disponíveis; identificar como e onde se podem efetuar modificações visando melhorias em geral; remover antecipadamente do sistema variáveis que não são de interesse; obter problemas de dimensão reduzida os quais requerem memória e tempo computacional menores.

Tanto a observabilidade quanto a redundância são altamente dependentes da escolha do conjunto de equações que serão utilizadas para formulação do problema de reconciliação. O algoritmo proposto difere dos demais utilizados na literatura aberta, pois após as exclusões iniciais das ocorrências individuais das linhas e colunas procura-se subconjuntos otimizados através da variável de corte. Isto possibilita superar dificuldades combinatoriais associadas a grandes sistemas onde todas as combinações possíveis de linhas e colunas elegíveis provocariam um consumo de tempo proibitivo.

-
1. CROWE, C.M., Reconciliation of Process Flow rates by Matrix Projection. Part II: The Nonlinear Case, *AIChE Journal*, v. 32, p. 616-623 (1986).
 2. CROWE, C.M., Observability and Redundancy of Process Data for Steady State Reconciliation, *Chemical Engineering Science*, v. 44, p. 2909-2917 (1989).
 3. EDGAR, T.F., HIMMELBLAU, D.M., LADSON, N., Optimization of Chemical Processes. 2 Ed. New York. McGraw-Hill, 2001.
 4. KRETISOVALIS, A., MAH, R. S., Effect of Redundancy on Estimation Accuracy in Process Data reconciliation, *Chem. Eng. Science*, v.42, p. 2115-2121 (1987).
 5. MAH, R. S. H., Chemical Process Structures and Information Flows, 1 Ed. Stoneham. Butterworth, 1990.
 6. OLIVEIRA JR, A . M., *Estimação de parâmetros em modelos de processo usando dados industriais e técnicas de reconciliação de dados*. Tese de Doutorado do Programa de Engenharia Química da Universidade Federal do Rio de Janeiro, 2006, 197 p.
 7. PAPADOKONSTANTAKIS, S.; MACHEFER, S., SCHNITZLEIN, K., LYGEROS, A. I., Variable Selection and data pre-processing in NN modeling of complex chemical process, *Computers and Chemical Engineering*, vol. 29, p. 1647-1659 (2005).
 8. SANCHEZ, M.; ROMAGNOLI, J.A., JIANG, Q., BAGAJEWICZ, M., Simultaneous Estimation Of Biases And Leaks In Process Plants, *Computers and Chemical Engineering*, vol. 23, p. 841-857 (1999).
 9. STANLEY, G.M., MAH, R. S. H., Observability and Redundancy classification in Process Networks”, *Chemical Engineering Science*, v. 12, p. 1941-1954 (1981).
 10. BENQLILOU, C., GRAELLS, M., PUIGJANER, L., Decision-Making Strategy and Tool for Sensor Network Design and Retrofit, *Industrial and Engineering Chemistry Research*, v. 43, p. 1711-1722 (2004)
 11. PRATA, D.M., SCHWAAB, M., LIMA, E. L., PINTO, J.C., Nonlinear dynamic data reconciliation and parameter estimation through particle swarm optimization: Application for an industrial polypropylene reactor, *Chemical Engineering Science*, v. 64, p. 3953-3967 (2009).