

O USO DA METODOLOGIA DE SUPERFÍCIE DE RESPOSTA PARA ADEQUAR OS PARÂMETROS DE ENTRADA DE UM ALGORITMO GENÉTICO

Pedro Alberto Barbeta

Departamento de Informática e de Estatística – CTC – UFSC
Caixa Postal 476 - Florianópolis - SC, 88.040-900

Cristiano Lehrer

Curso de Pós-Graduação em Ciência da Computação – UFSC
Caixa Postal 476 - Florianópolis - SC, 88.040-900

Abstract: This paper describes the response surface methodology (RSM), quite common in industrial processes optimization. It shows how the RSM being applied to optimize the choice of input parameters for a genetic algorithm. The application is very specific, but the same processes may be applied to more general computer systems.

Key Words: experimental design, response surface methodology, genetic algorithm

1 – Introdução

Desde os primórdios da investigação científica, realizavam-se experimentos para validar teorias e para levantar novas hipóteses sobre um mundo desconhecido. Mas foi a partir da década de 20, com os trabalhos de R. A. Fisher, que passou-se a ter uma metodologia confiável para planejar experimentos e analisar os seus dados. Grande parte dos trabalhos de Fisher eram voltados à agricultura, onde buscavam encontrar uma combinação adequada dos elementos de um fertilizante considerando o cultivo, o clima e o tipo de solo.

Na área industrial, especialmente nas duas últimas décadas, as técnicas de planejamento de experimentos passaram a fazer parte de programas de melhoria da qualidade. Esta prática deve-se muito aos trabalhos Box et al. (1978) e Taguchi (1987) que popularizaram os chamados projetos experimentais do tipo fatorial fracionado, os quais permitem avaliar grande número de fatores com poucos ensaios. Montgomery (1997) oferece uma apresentação bastante didática destes projetos.

A aplicação da metodologia de planejamento de experimentos na avaliação de sistemas computacionais é razoavelmente recente. Contudo, Jain (1991) já elegia o planejamento de experimentos como uma das principais ferramentas na avaliação de desempenho de sistemas computacionais. Muitos trabalhos, também, apóiam-se em sistemas computacionais para simular dados de algum processo, como descrito em Welch (1992), Mayer e Benjamin (1992) e Shang e Tadikamalla (1998), os quais projetam a simulação e realizam a análise dos dados sob a metodologia de planejamento de experimentos.

Barbeta (1998) e Barbeta et al. (2000) também apresentaram uma aplicação pouco comum de planejamento de experimentos ao aperfeiçoar um procedimento estatístico. A

metodologia foi aplicada sob dois enfoques. Num primeiro momento, identificou-se os fatores mais relevantes do processo (fase de caracterização) e, depois, buscou-se a combinação de níveis dos fatores mais relevantes para maximizar o desempenho do processo (fase de otimização).

2 – Otimização de processos através da metodologia de superfície de resposta

A metodologia de superfície de resposta (RSM) pode ser entendida como uma combinação de técnicas de planejamento de experimentos, análise de regressão e métodos de otimização. Tem larga aplicação nas pesquisas industriais, particularmente em situações onde um grande número de variáveis de um sistema influencia alguma característica fundamental deste sistema.

Box et al. (1978) apresentam a RSM como uma metodologia que consiste num grupo de técnicas usadas em estudos empíricos. Estas técnicas relacionam uma ou mais respostas, tais como produção, tonalidade e viscosidade com variáveis de entrada, tais como tempo, temperatura, pressão e concentração. A RSM tem sido usada para responder questões do tipo:

- Como uma resposta particular é afetada por um conjunto de variáveis de entrada sobre uma região de interesse?
- Quais conjuntos de entradas geram um produto satisfazendo simultaneamente várias especificações desejadas?
- Quais os valores de entradas que produzem o valor máximo para uma resposta específica e como se comporta a resposta em torno deste máximo?

Dado um processo ou sistema, com várias variáveis (ou fatores) de entrada, $x = (x_1, x_2, \dots, x_k)'$ e uma variável de saída (ou resposta) y , a RSM normalmente consiste em:

- planejar um experimento que permita estimar uma equação de regressão (superfície de resposta) do tipo

$$E(y/x) = f(x, \beta), \quad x \in D \subset \mathfrak{R}^k$$

- onde:
- $E(y/x)$ representa o valor esperado de y , condicionado a pontos $x \in D$;
 - f é uma função (usualmente um polinômio de grau não superior a 2);
 - x é o vetor de variáveis de entrada;
 - β é um vetor de parâmetros a ser estimado a partir dos dados do experimento e
- $D \subset \mathfrak{R}^k$ é a região especificada para a investigação.

- com os dados do experimento, estimar o vetor de parâmetros β , avaliando a significância estatística de seus componentes e
- estudar o comportamento da função f (por exemplo, se é desejado o maior valor possível para a resposta y , um dos interesses é obter $x_0 \in D$ que maximiza f , caso ele exista).

3 – Modelo para a variância

Considere um processo com valor ideal τ , mas, devido a efeitos aleatórios, o resultado obtido é um valor y . Por exemplo, uma empacotadora de café pode estar programada para encher pacotes com $\tau = 500g$, mas ao medir o peso de um pacote de café

o valor observado deve ser $y \neq 500g$. A função perda quadrática é uma forma bastante usada para medir a perda da qualidade devida a variação em torno do valor ideal e é dada por

$$L(y) = k(y - \tau)^2, \quad \text{para } y \in [\text{LIC}, \text{LSC}] \quad (1)$$

onde k é uma constante de proporcionalidade fazendo com que a função L assumia valores em termos monetários e $[\text{LIC}, \text{LSC}]$ é o intervalo de especificação da característica y (ver figura 1).

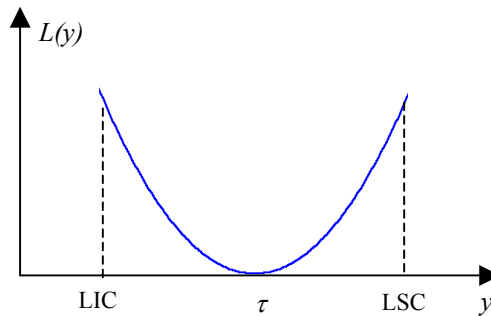


Figura 1 - Ilustração da função perda quadrática.

É natural o objetivo de minimizar o valor esperado da função perda, que é dado por

$$\mu_L = E\{L(y)\} = k[(\mu_y - \tau)^2 + \sigma_y^2] \quad (2)$$

onde μ_y e σ_y^2 são o valor esperado e a variância de y , respectivamente.

Considerando que tanto a média quanto a variância sofrem influência de um vetor $x' = (x_1, x_2, \dots, x_k)$ de fatores do processo, busca-se encontrar o vetor x^0 que minimiza a perda esperada ou, equivalentemente, que ajusta $\mu_y = f(x)$ ao alvo τ e minimiza $\sigma_y^2 = g(x)$. Um passo importante deste processo de otimização é a construção de modelos para μ_y e σ_y^2 em função de $x' = (x_1, x_2, \dots, x_k)$.

Para se construir um modelo para a média, $\mu_y = f(x)$, geralmente se aplica a análise de regressão tradicional, onde existe o método dos mínimos quadrados que é considerado pelos engenheiros quase como um padrão para efetuar a estimação dos parâmetros do modelo. O método pode ser aplicado diretamente sobre as observações de y ou sobre as médias, \bar{y}_i ($i = 1, 2, \dots, n$), calculadas em cada ponto experimental. Por outro lado, não se tem um procedimento de uso comum para estimar os parâmetros de um modelo para a variância, $\sigma_y^2 = g(x)$.

Quando o número de replicações for grande (dez ou mais, segundo Bartlett e Kendall, 1946), um procedimento eficiente para construir o modelo para a variância do processo é calcular a variância amostral em cada ponto experimental ensaiado com replicações: S_i^2 ($i = 1, 2, \dots, n$). Após uma transformação logarítmica, $\log(S_i^2)$, usam-se estas estatísticas como se fossem observações numa análise de regressão tradicional. Outras abordagens são discutidas em Carroll e Ruppert (1988), Barbetta (1998) e Barbetta et al. (2000).

O modelo da variância pode ser usado isoladamente como uma forma objetiva para se conhecer os fatores que mais influenciam a variância e, com isto, buscar robustez em

relação aos ruídos. Mas, também, o modelo da variância pode ser usado em conjunto com o modelo da média. Neste caso, busca-se uma solução de compromisso entre os objetivos de se levar a média ao alvo e de se minimizar a variância. Vining e Myers (1990) e Ribeiro e Elsayed (1995), por exemplo, apresentam técnicas interessantes para a obtenção de soluções de compromisso.

4 – O uso da metodologia de superfície de resposta numa aplicação de algoritmos genéticos

Um problema típico em Engenharia de Produção é o chamado *problema do caixeiro viajante*, que pode ser assim descrito: dadas n cidades, as distâncias entre elas e a cidade de origem e chegada, deve-se determinar a rota de menor distância total. Porém, quando n é grande, o número de possíveis rotas (soluções) torna-se extremamente grande e a avaliação de todas as rotas pode ser uma tarefa demorada até para os equipamentos computacionais atuais.

Para se obter uma boa solução e em pequeno tempo pode-se lançar mão de um algoritmo genético. Considera-se todas as possíveis combinações do conjunto $\{1, 2, \dots, n\}$ como a seqüência de cidades a percorrer. É comum considerar o mesmo valor na primeira e última posição, que corresponde a cidade de origem e de destino final do caixeiro viajante. Na linguagem de algoritmos genéticos, uma combinação de $\{1, 2, \dots, n\}$ é chamada de *cromossomo*. E o algoritmo pode ser descrito assim:

- a) Inicia-se com um conjunto (população) de k cromossomos viáveis (k permutações de $\{1, 2, \dots, n\}$);
- b) Avalia-se o desempenho de cada cromossomo (total das distâncias entre cidades a percorrer);
- c) Selecionam-se pares de cromossomos para participarem da nova geração, onde os cromossomos com melhor desempenho têm maior probabilidade de seleção;
- d) Aplicam-se, com certa probabilidade, operações de *crossover* entre pares de cromossomos selecionados. Esta operação consiste em dividir um par de cromossomo em um ponto aleatoriamente escolhido e emendar cada um deles com a parte do outro, formando um novo par de cromossomos;
- e) Aplicam-se, com certa probabilidade, a operação de *mutação* sobre algum *gene* do cromossomo gerado. Isto é, pode haver a alteração aleatória de um gene (cidade) do novo cromossomo;
- f) Os novos cromossomos gerados constituem uma nova população e volta-se para o passo (b), a menos que um certo critério de parada seja satisfeito.

No presente trabalho, o número n de cidades foi fixado em 26, correspondendo às capitais brasileiras com exceção de Macapá. O número de iterações do algoritmo foi 10.000 e em cada iteração considerou-se 200 rotas possíveis. Já as probabilidades de *crossover* e mutação foram avaliadas por um projeto fatorial 3×3 , ou seja, usou-se 3 diferentes probabilidades de *crossover* (fator A) e de mutação (fator B) cruzando-as, conforme ilustra a figura 2.

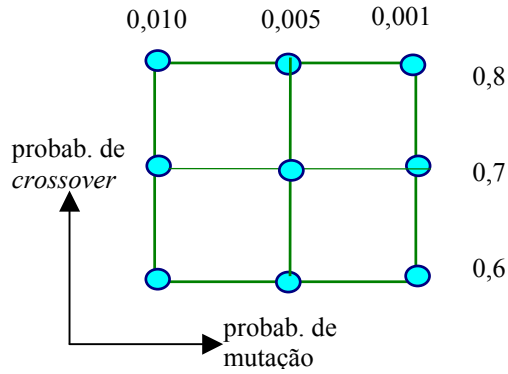


Figura 2 – Níveis dos fatores ensaiados.

Em cada ponto experimental foram feitas 100 realizações do algoritmo e, para cada uma delas, anotou-se a solução (a menor rota a percorrer) e a correspondente distância total y . Considerou-se duas estatísticas baseadas nas 100 realizações de cada ponto experimental: a média aritmética (\bar{y}_i , $i = 1, 2, \dots, 9$) e o logaritmo natural da variância amostral ($\log(S_i^2)$, $i = 1, 2, \dots, 9$). O processo ainda foi repetido quatro vezes, tendo, na verdade, quatro valores de \bar{y} e de $\log(S^2)$ em cada ponto experimental, conforme mostra a tabela 1.

ensaio	probab. de crossover	probab. de mutação	resposta \bar{y}	resposta $\log(S^2)$	ensaio	probab. de crossover	probab. de mutação	resposta \bar{y}	resposta $\log(S^2)$
1	0,6	0,010	22829	15,20	19	0,6	0,010	23120	14,87
2	0,6	0,005	24122	15,23	20	0,6	0,005	24141	14,99
3	0,6	0,001	25888	15,39	21	0,6	0,001	25797	15,34
4	0,7	0,010	22960	14,63	22	0,7	0,010	23036	14,65
5	0,7	0,005	24126	15,17	23	0,7	0,005	25972	15,48
6	0,7	0,001	25997	15,48	24	0,7	0,001	25532	15,34
7	0,8	0,010	22869	14,69	25	0,8	0,010	22593	14,52
8	0,8	0,005	24308	15,45	26	0,8	0,005	24029	15,14
9	0,8	0,001	25368	15,54	27	0,8	0,001	25900	15,51
10	0,6	0,010	25444	15,33	28	0,6	0,010	25599	15,71
11	0,6	0,005	24135	15,10	29	0,6	0,005	23908	15,21
12	0,6	0,001	25759	15,67	30	0,6	0,001	25747	15,39
13	0,7	0,010	25666	15,56	31	0,7	0,010	22581	14,56
14	0,7	0,005	24159	15,22	32	0,7	0,005	24147	15,11
15	0,7	0,001	25985	15,66	33	0,7	0,001	26129	15,33
16	0,8	0,010	23025	14,59	34	0,8	0,010	23170	14,81
17	0,8	0,005	25697	15,68	35	0,8	0,005	24038	14,97
18	0,8	0,001	25838	15,43	36	0,8	0,001	25800	15,67

Tabela 1 – Pontos experimentais ensaiados e estatísticas dos resultados alcançados.

Busca-se conhecer a combinação das duas probabilidades de entrada (mutação e *crossover*) que levam a valores pequenos de \bar{y} e de $\log(S^2)$. Para tanto, ajustou-se, por mínimos quadrados, funções quadráticas para as duas respostas consideradas, como segue:

$$\bar{y} = 17651 + 22529 X_1 + 42801 X_2 - 14399 X_1^2 + 20946486 X_2^2 - 745422 X_1 X_2 \quad (3)$$

$$\log(S^2) = 16,4 - 4,3 X_1 + 226,4 X_2 - 4,2 X_1^2 + 261,9 X_2^2 - 415,2 X_1 X_2 \quad (4)$$

onde X_1 e X_2 representam as probabilidades de *crossover* e de mutação, respectivamente. A qualidade do ajuste, medido pelo coeficiente de determinação R^2 , foi de 0,63 para a equação (3) e 0,54 para a equação (4). A figura 3 mostram as curvas de níveis das equações (3) e (4), que permitem encontrar as melhores probabilidades de *crossover* e de mutação para um problema do tipo em estudo.

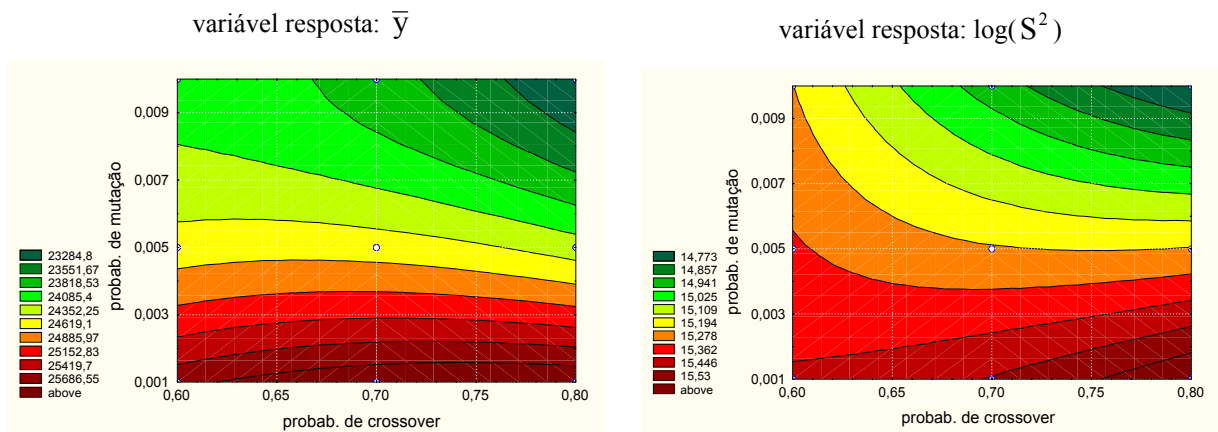


Figura 3 – Curvas de níveis das superfícies de resposta das equações (3) e (4)

Como no presente caso é desejável que tanto \bar{y} e quanto $\log(S^2)$ sejam os menores possíveis, verifica-se que os melhores resultados são alcançados com probabilidade de *crossover* em torno de 0,8 e probabilidade de mutação em torno de 0,01.

5 – Considerações finais

O principal objetivo deste trabalho é mostrar como a metodologia planejamento de experimentos e de superfície de resposta podem ser usada para adequar parâmetros de entrada em sistemas computacionais. A aplicação do algoritmo genético no problema do caixeiro viajante foi feita com vários dados de entrada fixados, mas pretende-se elaborar experimentos mais complexos, considerando, por exemplo, o número de cidades e as distâncias também como fatores variáveis no experimento. Com isso, pretende-se obter equações que, a partir dos dados de entrada de um particular problema, possa-se ter estimativas dos valores ideais dos parâmetros controláveis.

Referências Bibliográficas

- BARBETTA, P. A. - *Construção de modelos para médias e variâncias na otimização experimental de produtos e processos*. Tese (Engenharia de Produção), Florianópolis, 1998.
- BARBETTA, P. A., RIBEIRO, J. L. D., SAMOHYL, R. W. – *Variance regression models in Experiments with few replications*. Quality And Reliability Engineering International, Inglaterra, v.16, 2000, p. 397 – 404.
- BARTLETT, M. S., KENDALL, D. J. - *The statistical analysis of variance-heterogeneity and the logarithmic transformation*. Journal of the Royal Statistical Society, Série B, v. 8, 1946, p. 128-138.
- BOX, G. E. P., HUNTER, W. G., HUNTER, J. S. - *Statistics for experimenters*. USA: John Wiley & Sons, 1978.
- CARROLL, R. J., RUPPERT, D. - *Transformation and weighting in regression*. USA: Chapman and Hall, 1988.

- JAIN, R. – *The art of computer systems performance analysis: techniques for experimental design, measurement, simulation, and modeling*. USA: John Wiley & Sons, 1991.
- LEHRER, C. – *Operador de seleção para algoritmos genéticos baseado no jogo hawk-dove*. Dissertação (Ciência da Computação), Florianópolis, 2000.
- MAYER, R. J., BENJAMIN, P. C. – *Using the Taguchi paradigm for manufacturing system design using simulation experiments*. Computers and Industrial Engineering, v. 22, n. 2, 1992, p. 195 – 209.
- MONTGOMERY, D. C. - *Design and analysis of experiments*, 4 ed., USA: John Wiley & Sons, 1997.
- MYERS, R. H., CARTER, W. H. - *Response surface techniques for dual response systems*. Technometrics, v. 15, n. 2, 1973, p. 301 - 317.
- MYERS, R. H., MONTGOMERY, D. C. - *Response surface methodology: process and product optimization using designed experiments*. USA: John Wiley & Sons, 1995.
- RIBEIRO, J. L. D., ELSAYED, E. A. - *A case study on process optimization using the gradient loss function*. International Journal of Production Research, v. 33, n. 12, 1995, p. 3233-3248.
- SHANG, J. S., TADIKAMALLA, P. R. – *Multicriteria design and control of a cellular manufacturing system through and optimization*. International Journal of Production Research, v. 36, n. 6, 1998, p. 1515 – 1528.
- TAGUCHI, G. - *System of experimental design: engineering methods to optimize quality and minimize costs*. 2 ed. USA: UNIPUB, 1987.
- VINING, G. G., MYERS, R. H. - *Combining Taguchi and response surface philosophies: a dual response approach*. Journal of Quality Technology, v. 22, n. 1, 1990, p. 38 - 45.
- WELCH, W. J., BUCK R. J., SACKS, J., WYNN, H. P., MITCHELL, T. J., MORRIS, M. D. - *Screening, predicting, and computer experiments*. Technometrics, v. 34, n. 1, 1992, p. 15 - 25.