APLICAÇÃO DE UM VNS NO PROCESSO DE SELEÇÃO DE VARIÁVEIS DE UM MODELO DE REGRESSÃO LINEAR MÚLTIPLA

Felipe Teles
Universidade Estadual Paulista "Júlio de Mesquita Filho"
fernandesfelipe13@hotmail.com

Daniela R. Cantane[1]
Universidade Estadual Paulista "Júlio de Mesquita Filho"
daniela.cantane@unesp.br

Helenice Florentino[2]
Universidade Estadual Paulista "Júlio de Mesquita Filho"
helenice.silva@unesp.br

Vieira, J. M.[1]
Universidade Estadual Paulista "Júlio de Mesquita Filho"
janielly.matos@unesp.br

RESUMO

Em diversas situações-problema, interessa-se em explicar o comportamento de uma variável observada y em relação a um conjunto de características mensuradas desta variável, denotadas por X_i, para i=1,...,p características. Este procedimento, que é conhecido como Regressão, permite estabelecer um modelo linear y=beta_0+beta_1X_1+beta_2X_2+...+beta_pX_p que possa informar qual o efeito de cada regressora X_i no comportamento de y, sendo β_i as contribuições de cada regressora e beta_0 o intercepto do modelo. O diagnóstico do modelo pode indicar a existência de uma ou mais regressoras que possuem baixa ou nenhuma relevância para a explicação do comportamento de y. Deste modo, é necessário selecionar m das p características para que o modelo de regressão ajustado seja simples e informativo.
\par Para a simplificação do modelo, deve-se combinar estratégias de seleção com critérios de qualidade. Dentre os critérios mais utilizados, citam-se os critério de informação de Akaike, critério de informação Bayesiano e o C_p de Mallows, os quais podem ser combinados com estratégias de seleção backward, forward, stepwise, entre outras. De forma geral, a seleção de variáveis com a minimização da falta de ajuste do modelo pode ser classificada como um problema de otimização combinatória.
Considerando que o número de variáveis regressoras X_i para a explicação de uma resposta y pode ser muito grande, o que custará caro computacionalmente para o processo de seleção, este trabalho propõe a implementação de uma heurística de busca em vizinhança variável para determinar quais variáveis do conjunto de regressoras devem ser escolhidas para que o modelo seja parcimonioso e com mínima falta de ajuste. Os resultados computacionais da aplicação da heurística, bem como sua comparação com a resolução exata por Branch & Bound com apoio do software R, serão apresentados no evento.

PALAVRAS CHAVE: Modelos lineares com múltiplas regressoras, seleção de variáveis, Metaheurísticas

Tópico: EST – Estatística e Modelos Probabilísticos