



DESENVOLVIMENTO E MELHORAMENTO DE ESTRATÉGIAS DE SELEÇÃO DE VARIÁVEIS

Weverton D Lucas Moura Marinho¹
Adriano de Araújo Gomes²

Agência Financiadora: CNPq

1. INTRODUÇÃO

Ferramentas quimiométricas são muito úteis no que diz respeito a otimização de condições experimental e compilação de dados químicos para tomada de decisões (ANTONIJE, 2016). Aplicações da quimiometria vem sendo reportada nas diversas áreas da química e Ciências correlatas. Contudo, para uma maior difusão destas ferramentas é preciso o desenvolvimento de software amigáveis e de fácil operação. Técnicas instrumentais, tais como espectroscópicas e imagens digitais, combinação com modelos multivariados tem permitido o desenvolvimento novas metodologias analíticas que portam uma série de vantagens como: rapidez, baixo custo, robustez, determinações simultâneas e menores consumo de amostras e reagentes. Entretanto, a qualidade final de um modelo quimiométrico está diretamente ligada com as variáveis de entrada (PESSOA, 2015). As técnicas instrumentais modernas são conhecidas por gerar uma grande quantidade de informação por amostra em um pequeno intervalo de tempo (GOMES, 2015), parte da informação gerada não é informativa ou é redundante, o que faz necessário o uso de estratégias de seleção de variáveis como etapa previa na construção dos modelos.

Neste contexto o método de otimização por Colônia de Formigas, que simula via software o comportamento das formigas em busca de alimentos foi proposto por Dorigo *et al* (SHAMSIPUR; ZARE-SHAHABADI; HEMMATEENEJAD; AKHOND, 2007), vem demonstrando potencial como ferramenta de seleção variável, na modelagem de químicas que envolvem dados de calibração multivariada através dos mínimos quadrados parciais (PLS) (ALLEGRINI; OLIVIERI, 2011; TABAKHI; MORADI, 2015). Contando não há registros do seu uso em problemas de classificação.

Portanto, neste trabalho é proposto o desenvolvimento de um método de seleção de variáveis baseado em Colônia de Formigas (AOC) acoplado a modelos de Análise Discriminante Linear (LDA). O método proposto, denominado (AOC), foi implementado em ambiente MatLab em forma de linhas de comandos e implemento em forma de GUI (graphical user interfaces) para disponibilização para comunidade científica na forma de um software de fácil manuseio. O método desenvolvido neste trabalho, foi avaliado em estudos de caso envolvendo a detecção de adulteração de amostras de açaí empregando dados de espectroscopia no infravermelho.

2. MATERIAIS E MÉTODOS

2.1 Método Proposto

O método proposto neste trabalho, o algoritmo de colônia de formigas combinado com análise discriminante linear (AOC-LDA) é composto de cinco etapas: codificação binária (1 para variáveis incluídas no modelo e 0 para o caso contrário), inicialização em modo randomico, avaliação com base na função de custo (G_{cost}) que representa o risco médio de um dados modelo LDA baseado un subconjunto de variáveis cometer uma classificação incorreta e por fim é conduzido a atualização da trilha de feromônios empregando taxa de evaporação padrão de 65% e depósito proporcional a qualidade do ajuste. A cada ciclo uma nova colônia é gerada baseado no método da roleta e em uma proporção de formigas cegas, previamente definido pelo usuário.

2.2 Estudo De Caso

Foram adquiridas 60 amostras de açaí in natura, em feiras locais da Cidade de Marabá-PA. Para todos os casos, foi solicitado, ao produtor, que o açaí fosse processado no ato da compra, esta ação visa assegurar que o açaí adquirido não seja alvo de adulteração, o que afetaria os resultados deste estudo. Amostras adulteradas foram geradas em laboratório considerando a farinha trigo e a fécula de mandioca como

¹Graduando do Curso de Licenciatura em Química (FAQUIM/ICE/Unifesspa). wevertondlucas@unifesspa.edu.br.

²Professor da Universidade Federal do Sul e Sudeste do Pará. E mail. adrianogmes@gmail.com.

adulterantes. Foram preparadas misturas binárias açai/fécula (AF) de mandioca e açai/trigo (AT) na proporção mássica de 90/10%. Esta proporção de açai /adulterante não provoca alteração de cor ou sabor detectáveis de forma trivial. Os espectros de infravermelho (IR) em modo de refletância total atenuada (ATR) foram registrados em um equipamento AGILENT CARY 630 FT em triplicata na faixa de 650 a 4000 cm^{-1} , com resolução de 4 cm^{-1} . O método AOC foi configurado para operar com 100 formigas, por 200 ciclos empregando a taxa de formigas cegas de 40% na atualização das colônias. A taxa de evaporação de feromônio foi mantida fixa em 65%.

3. RESULTADOS E DISCURSÕES

3.1 Método Proposto

Na Figura 1 é exibido a interface gráfica principal do **LINEAR DISCRIMINANT ANALYSIS-VARIABLE SELECTION TOOLBOX** desenvolvida em ambiente MatLab® (2010b) para construção de modelos LDA com ou sem associação a métodos de seleção de variáveis.

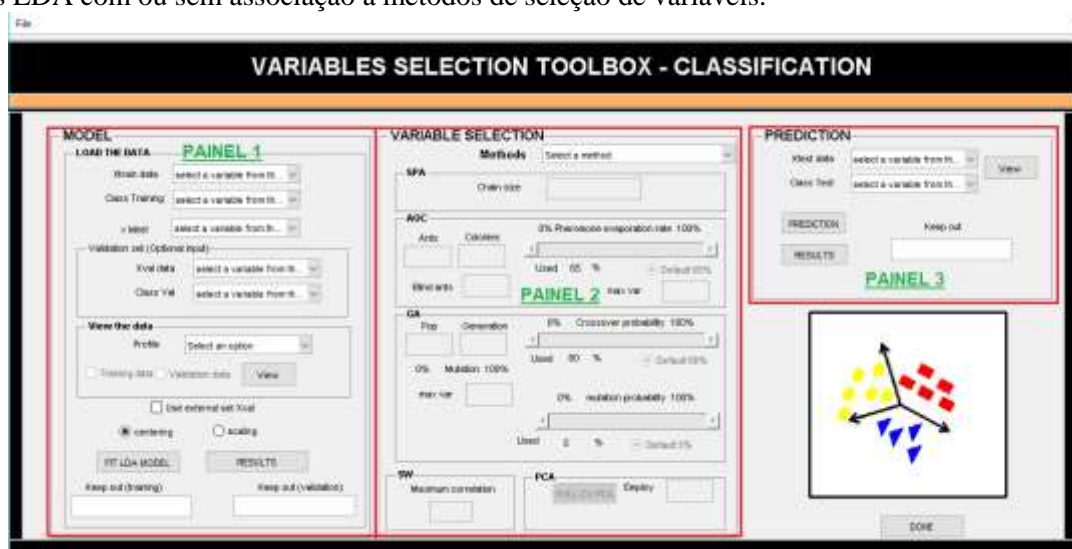


Figura 1: Interface principal do toolbox.

No painel 1 da interface, o usuário poderá carregar os dados, visualizar e excluir amostras, executar o ajuste do modelo LDA e visualizar os resultados do ajuste. No painel 2 pode ser selecionado e configurado o método de seleção de variáveis que se deseja associar ao modelo LDA ou indicar que não se deseja fazer a seleção. Ao selecionar o método (AOC), será habilitado o painel destinado ao respectivo método, em que o usuário poderá configurar os parâmetros: (i) quantidade de formigas que deseja usar, (ii) ciclos de geração, (iii) taxa de evaporação de feromônio, (iv) quantidade máxima de variáveis que uma formiga pode portar e (v) taxa de atualização de formigas cegas na colônia. A interface permite ao usuário ter acesso as principais métricas de desempenho do modelo de classificação, como por exemplo, matriz de confusão, sensibilidade, seletividade, taxa correta de classificação e a classe na qual a amostra foi alocada. Não menos importante, o pacote LDA-VS gera diversas saídas gráficas, tais como escores, pesos e variáveis selecionadas.

3.2 Estudo De Caso

Na Figura 2A são apresentados os espectros brutos das amostras de açai, é possível notar que ocorre uma forte sobreposição dos mesmos, não sendo possível distinguir amostras adulteradas de não adulteradas baseado apenas na inspeção visual dos espectros. É possível observar ainda, que em virtude do teor elevado de água nas amostras de açai, a maior parte da informação contida nos dados de ATR-FTIR são relativos banda de OH. A inspeção dos espectros de açai, comparados ao perfil espectral exibido pela água deionizada (Figura 1B), mostra que, embora em baixa magnitude de sinal, informações do açai e adulterantes estão contidos nos espectros registrados. Subsequente a inspeção dos espectros, foi conduzida a partição do conjunto de amostras em treinamento (30 amostras por classe), validação (10 amostras por classe) e teste (20 amostras por classe) empregando o método Kernnard-Stone (KS), para fins de construção de modelos de reconhecimento de padrões baseados em análise discriminante linear com seleção de variáveis via algoritmo de colônia de formigas.

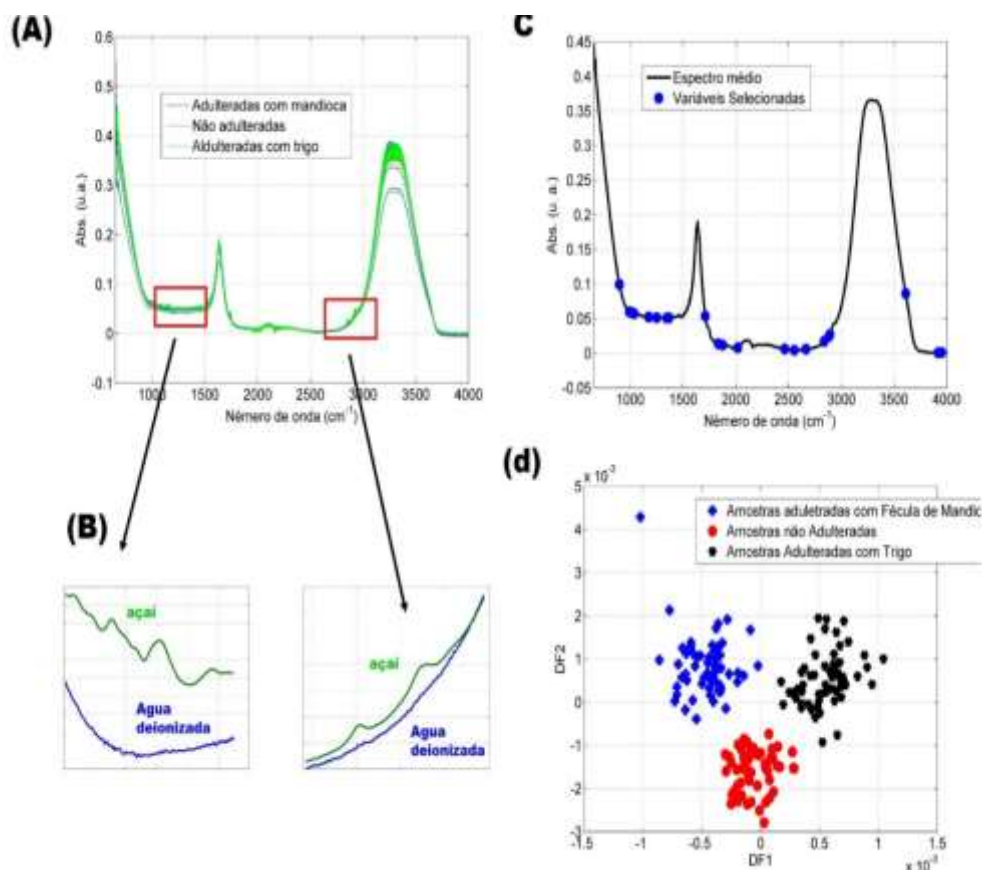


Figura 2: Em (A) espectros das amostras de açaí, (B) perfil do espectral da água e do açaí, (C) variáveis selecionadas e (d) gráfico de escores.

As variáveis selecionadas pelo AOC-LDA, na Figura 1C, estão majoritariamente nas regiões espectrais destacadas na Figura 1B. Os modelos LDA resultante classificou corretamente todas amostras como pode ser visualizado no gráfico de dispersão (DF1 versus DF2) dos escores de Fisher.

4. CONSIDERAÇÕES FINAIS

O método colônia de formigas combinado com análise discriminante linear é um material útil para seleção de variáveis em estudos de classificação. A interface desenvolvida nesse trabalho mostrou-se de fácil manuseio e amigável para geração de modelos LDA baseado em seleção de variáveis. Os espectros ART-FTIR portam a informação química necessária para discriminar amostras de açaí adulteradas de amostras não adulteradas, em adição o método de seleção de variáveis AOC acessou corretamente as variáveis mais informativas levando a modelos LDA com 100% de classificação correta.

REFERÊNCIAS

- ALLEGRI, F.; OLIVIERI, A. C. A new and efficient variable selection algorithm based on ant colony optimization. Applications to near infrared spectroscopy/partial least-squares analysis. *Analytica Chimica Acta*, v. 699, n. 1, p. 18–25, maio 2011.
- ANTONIJE, E. O. **Chemometric Approach to the Experiment Optimization and Data Evaluation in Analytical Chemistry**. [s.l.] Faculty of Technology and Metallurgy, University of Belgrade Karnegijeva, 2016.
- GOMES, A. DE A. **Algoritmo das projeções sucessivas para seleção de variáveis em calibração de segunda ordem**. [s.l.] Universidade Federal da Paraíba, 2015.
- PESSOA, Carolina de Marco. **Aperfeiçoamento do Algoritmo de Colônia de Formigas para o Desenvolvimento de Modelo Quimiométricos**. Universidade Federal do Rio Grande do Sul, 2015.
- SHAMSIPUR, MOJTABA; ZARE-SHAHABADI, VALI; HEMMATEENEJAD, BAHRAM; AKHOND, M. Identification of genes involved in radiation-induced G 1 arrest y. *Journal of Chemometrics*, p. 398–405, jan. 2007.

TABAKHI, S.; MORADI, P. Relevance-redundancy feature selection based on ant colony optimization. **Pattern Recognition**, v. 48, n. 9, p. 2798–2811, abr. 2015.