

## METODO PARA SELEÇÃO DE ATRIBUTOS EM PERFIS SENSORIAIS COM BASE EM FERRAMENTAS DE MINERAÇÃO DE DADOS

**Karina Rossini**

Programa de Pós Graduação em Engenharia de Produção – UFRGS  
Av. Osvaldo Aranha, 99 – 5° andar, 90035-190 - Porto Alegre– RS  
e-mail: rossini@producao.ufrgs.br

**Flávio S. Fogliatto**

Programa de Pós Graduação em Engenharia de Produção – UFRGS  
Av. Osvaldo Aranha, 99 – 5° andar, 90035-190 - Porto Alegre– RS  
e-mail: fflogliatto@producao.ufrgs.br

**Michel J. Anzanello**

Programa de Pós Graduação em Engenharia de Produção – UFRGS  
Av. Osvaldo Aranha, 99 – 5° andar, 90035-190 - Porto Alegre– RS  
e-mail: [anzanello@producao.ufrgs.br](mailto:anzanello@producao.ufrgs.br)

### RESUMO

A seleção dos atributos a serem avaliados em uma análise sensorial é uma importante etapa no planejamento de painéis sensoriais. Na seleção, deseja-se obter um conjunto de atributos de reduzido tamanho e elevado poder de informação, evitando assim a imposição de fadiga aos membros do painel, minimizando custos e tempo. Este artigo apresenta um método para seleção de atributos e avaliadores em painéis sensoriais baseados em avaliações descritivas das amostras, tais como a Análise Descritiva Quantitativa (ADQ) e o protocolo Spectrum. O método proposto integra a Análise de Componentes Principais, o algoritmo de classificação dos  $k$  vizinhos mais próximos (KVP) e a análise de Pareto Ótimo. Objetiva-se (i) identificar o conjunto de atributos que melhor discrimina as amostras analisadas no painel, e (ii) indicar o grupo de avaliadores que fornece avaliações consistentes. O método é testado em um estudo de caso onde cubos de carne bovina com molho, utilizados como ração de combate pelo exército norte-americano, são caracterizados em painéis sensoriais utilizando o protocolo Spectrum.

**PALAVRAS CHAVE.** Seleção de variáveis. Painel sensorial. Ferramentas de mineração de dados. Outras aplicações de PO.

### ABSTRACT

Selection of attributes to be evaluated in a sensory analysis study is an important issue when planning data collection in sensory panels. In attribute selection it is desirable to reduce the list of those to be presented to panelists to avoid fatigue, minimize costs and save time. In this paper we present a data mining-based method for attribute and panelist selection in descriptive sensory panels, such as those used in the Quantitative Descriptive Analysis and the Spectrum protocol. The proposed method is implemented using Principal Component Analysis and the  $k$ -Nearest Neighbor classification technique, in conjunction with Pareto Optimal analysis. Objectives are (i) to identify the set of attributes that best discriminate samples analyzed in the panel, and (ii) to indicate the group of panelists that provide consistent evaluations. The method is illustrated through a case study where beef cubes in stew, used as combat ration by the American Army, are characterized in sensory panels using the Spectrum protocol.

**PALAVRAS CHAVE.** Variable selection. Sensory panel. Data mining tools. Other OR applications.

## 1. Introdução

Métodos de análise descritiva (AD) visam fornecer perfis sensoriais de produtos. Em essência, os protocolos de AD para a avaliação de produtos baseiam-se em um grande número de atributos, com base nos quais os produtos são avaliados, sendo os resultados expressos em uma escala numérica contínua.

Embora amplamente utilizados, os métodos AD apresentam limitações: (i) o número de atributos a serem avaliados geralmente é elevado, tornando a coleta de dados cansativa, demorada e de alto custo (CARBONELL et al., 2007); (ii) nem sempre o perfil obtido para os produtos inclui atributos com poder de discriminação, embora em algumas aplicações este seja o objetivo da coleta de dados (GRANITTO et al. 2007); e (iii) métodos AD não oferecem uma forma estruturada de avaliação do desempenho dos avaliadores.

A seleção de atributos é um importante tema de pesquisa na área de avaliação sensorial. Autores como Dijksterhuis et al. (2002), Westad et al. (2003) e Sahmer e Qannari (2008) propuseram o uso de diferentes métodos de análise multivariada para selecionar um subconjunto de atributos relevantes e/ou não redundantes partindo de uma amostra composta por múltiplos atributos. Em oposição a esses autores e próximo da proposta do presente artigo, Granitto et al. (2007) introduziu o uso de ferramentas de mineração de dados (mais especificamente, a técnica de Floresta Randômica – Random Forest) para selecionar os atributos que melhor discriminam os produtos. Em todos os casos, métodos AD foram utilizados para caracterizar os produtos.

Este artigo integra métodos de análise multivariada a ferramentas de mineração de dados para selecionar atributos relevantes no perfil sensorial de produtos. Um atributo relevante em uma investigação sensorial é aquele cuja avaliação apresenta diferenças sistemáticas e significativas entre os produtos investigados, permitindo a identificação de formulações com características distintas. O método proposto inicialmente classifica os avaliadores utilizando, de forma inédita, o Alfa de Cronbach como índice de consistência; o índice é calculado utilizando uma abordagem “omite os dados de um avaliador por vez” (*leave-one-out-at-a-time*). Na sequência, a análise de componentes principais (ACP) é aplicada sobre os dados sensoriais oriundos de todos os avaliadores, gerando-se índices de importância dos atributos com base nos pesos do ACP. Inicia-se então um processo iterativo de classificação e eliminação: classificam-se os produtos através do algoritmo dos  $k$  vizinhos mais próximos (KVP) e calcula-se a acurácia da classificação; na sequência, o atributo com o menor índice de importância é removido do conjunto de dados, os produtos são classificados novamente e um novo valor de acurácia é calculado. O processo é repetido até que reste apenas um atributo. Ao final do processo, obtém-se um gráfico relacionando acurácia de classificação e número de atributos retidos.

Na sequência, eliminam-se as avaliações sensoriais propostas pelo avaliador com o menor índice de consistência, e repete-se o processo iterativo de seleção de atributos descrito acima. Os resultados gerados para cada conjunto de avaliadores são armazenados em um único gráfico de acurácia, e o melhor subconjunto de atributos e avaliadores é determinado através de análise de Pareto Ótimo (PO).

Existem três contribuições relevantes no método proposto. A primeira é a utilização do alfa de Cronbach para avaliar o desempenho de avaliadores em painel sensorial. Na utilização proposta neste artigo, o alfa de Cronbach mede a similaridade entre os perfis de diferentes avaliadores, apontando aqueles que avaliam de forma inconsistente relativamente aos demais membros do grupo. Embora amplamente utilizado para avaliação de consistência de questionários (FORNELL E LARCKER, 1981; BROWN, 2006), a utilização do Alfa de Cronbach em dados sensoriais é inédita. A segunda contribuição é a integração de ACP com o algoritmo de classificação KVP para obtenção de método eficiente de seleção de atributos sensoriais. O KVP é selecionado como técnica de classificação devido ao seu desempenho satisfatório em aplicações práticas, simplicidade conceitual e disponibilidade em pacotes computacionais; ver Chaovalitwongse et al. (2007) e Anzanello et al. (2009). A terceira contribuição consiste na utilização da análise de PO para identificar um número limitado de soluções distintas que maximizam a acurácia da classificação e minimizam o número de atributos retidos, bem como o número de avaliadores. A análise de PO tem sido empregada em

uma grande variedade de aplicações, como na avaliação do ciclo de vida de produtos químicos (AZAPAGIC, 1999), seqüenciamento de operações (TABOADA E COIT, 2008), e otimização de confiabilidade na transmissão de energia (TABOADA E COIT, 2007). A aplicação de PO em sistemáticas de seleção de atributos, no entanto, é bastante limitada.

O método de seleção de atributos proposto foi aplicado em um estudo de caso envolvendo um produto alimentício utilizado pelo exército americano (FOGLIATTO et al., 1999). O produto é constituído de cubos de carne com molho embalados em *pouches* termoestáveis, preparado em uma planta piloto na Rutgers University, EUA. Oito variantes do produto (doravante designadas por produtos) foram avaliadas por um painel sensorial com 9 avaliadores e 26 atributos sensoriais. O método proposto reduziu o número de atributos e avaliadores necessários para caracterizar as amostras para 7 e 6, respectivamente.

Este artigo está organizado em quatro seções, além desta introdução. A Seção 2 apresenta uma breve revisão da literatura sobre abordagens propostas para seleção de atributos no campo sensorial, e faz uma breve descrição das ferramentas utilizadas no método. A Seção 3 apresenta a seqüência de etapas para a aplicação do método. A Seção 4 traz os resultados da aplicação do método no estudo de caso, seguido de conclusões na Seção 5.

## 2. Fundamentação Teórica

O referencial teórico deste artigo aborda dois assuntos: estudos para seleção de atributos em análise sensorial e ferramentas utilizadas na implementação do método proposto. Com exceção de Granitto et al. (2007), onde ferramentas de mineração de dados são aplicadas, todas as abordagens são baseadas em combinações de técnicas de análise multivariada e Procrustes.

Dijksterhuis et al. (2002) aplicou um método baseado em ACP proposto originalmente por Krzanowski (1987) na seleção de atributos em dados sensoriais. No artigo, o autor investiga como subconjuntos de atributos preservam as informações do conjunto de dados originais usando a perda de Procrustes como um índice de desempenho. As proposições são testadas em dois conjuntos de dados reais.

Sahmer et al. (2008) estenderam a pesquisa apresentada por Dijksterhuis et al. (2002) testando seis estratégias multivariadas para selecionar os melhores atributos em conjuntos de dados sensoriais descritivos: duas estratégias baseadas em ACP, duas no critério de Procrustes e duas no agrupamento de atributos. O objetivo foi preservar a estrutura de dados multivariados, reduzindo o número total de atributos apresentados ao painel. Os autores concluíram que as estratégias baseadas no agrupamento levaram a resultados superiores quando aplicadas a dados reais e simulados.

Por sua vez, Westad et al. (2003) sugeriram um método baseado em estimativas de incerteza da validação cruzada, a qual é aplicada sobre a ACP de dados sensoriais descritivos. O método sugerido enfatiza o cálculo da significância do atributo nos componentes. Para tanto, a análise de Procrustes é realizada para girar as cargas e os escores na ACP. As vantagens e desvantagens desta ação são ilustradas através da análise de dois conjuntos de dados.

Valendo-se da ferramenta de mineração de dados conhecida por Floresta Randômica (*Random Forest* - FR), Granitto et al. (2007) propuseram um método de seleção de atributos para discriminação de amostras de queijo. A função discriminante gerada pela ferramenta FR atribui pesos de importância aos atributos; estes pesos podem ser utilizados para selecionar um número reduzido de atributos com elevado poder discriminatório. Salienta-se que os autores foram os primeiros a propor o uso de ferramentas de mineração de dados no âmbito da caracterização sensorial.

No que segue, uma breve revisão sobre a ferramenta *k* vizinhos mais próximos (KVP) e a análise de Pareto Ótimo (PO) é apresentada.

A ferramenta KVP tem sido utilizada em uma ampla variedade de contextos, incluindo análise de padrões de textos (WEISS et al., 1999), detecção de atividade cerebral anormal (CHAOVALITWONGSE et al., 2007) e classificação de lotes de produção (ANZANELLO et al., 2009).

Considere observações em um conjunto de dados  $J$ -dimensional, correspondentes a  $J$  atributos, e duas classes de produtos (A ou B). O objetivo é classificar uma nova observação (produto) em A ou B com base apenas em atributos. Considere os  $k$  vizinhos mais próximos da nova observação, onde a distância entre as observações é medida pela distância Euclidiana. Para cada observação contida em  $k$ , indique uma categoria, A ou B. A nova observação é inserida na classe A se a maioria dos seus  $k$  vizinhos mais próximos pertencer à A. O número  $k$  de vizinhos é escolhido utilizando procedimentos de validação cruzada em bancos de dados onde a classe de cada observação é conhecida *a priori*. Detalhes sobre técnica de classificação KVP podem ser obtidos em Wu et al. (2008).

A técnica KVP possui a vantagem de ser conceitualmente mais simples e intuitiva do que outras ferramentas de classificação, além de amplamente disponível em pacotes computacionais. Paralelamente, a KVP requer apenas um parâmetro ( $k$ ) e a acurácia da classificação não é significativamente sensível a esta escolha dentro de um intervalo razoável.

A análise de Pareto Ótimo (PO) identifica um conjunto de soluções distintas em aplicações com múltiplas funções objetivo. Estas funções normalmente não apresentam uma solução única, mas um conjunto de soluções adequadas. As aplicações de seleção de variáveis, onde a classificação das medidas de desempenho é maximizada e o número de variáveis retidas é minimizado, são exemplos de cenários com várias soluções possíveis. Por conta de sua vasta aplicabilidade, a análise de PO tem sido amplamente integrada a algoritmos e abordagens para fins de otimização, como relatado em Deb et al. (2002a,b).

As soluções identificadas pela análise de PO são chamadas não-dominadas, o que significa que elas não podem ser superadas por outras soluções vizinhas com relação aos objetivos estabelecidos. Essa condição permite a redução significativa do número de soluções potenciais, fazendo com que a análise posterior concentre-se em um conjunto reduzido de soluções efetivamente superiores. Essas soluções geralmente são ilustradas em um limite chamado Fronteira de Pareto. A identificação da melhor solução pode depender de informações subjetivas, tornando-se complexa com o aumento do número de funções objetivo; maiores detalhes sobre a análise de PO estão disponíveis em Horn et al. (1994), Zitzler e Thiele (1999) e Taboada et al. (2008).

### 3. Método

O método para selecionar atributos com vistas à classificação das amostras em formulações é operacionalizado em seis etapas: 1. Avalie a coerência entre avaliadores usando o um índice adequado. 2. Aplique uma técnica multivariada no conjunto de atributos sensoriais; o método proposto utiliza Análise de Componentes Principais (ACP), mas outras técnicas podem ser consideradas. 3. Calcule um vetor de índices de importância dos atributos com base nos pesos da ACP. 4. Classifique o conjunto de dados sensoriais utilizando KVP e calcule a acurácia da classificação. Na sequência, elimine o atributo com o menor índice de importância, classifique o conjunto de dados novamente, e recalcule a acurácia. Repita o processo iterativo até restar um único atributo. 5. Construa um gráfico relacionando acurácia e número de atributos retidos. 6. Elimine o avaliador menos consistente e realize uma nova seleção de atributos, repetindo as etapas 2 a 6. Estas etapas são detalhadas nas subseções a seguir.

Considere  $p$  ( $p = 1, \dots, P$ ) avaliadores,  $j$  ( $j = 1, \dots, J$ ) atributos e  $i$  ( $i = 1, \dots, I$ ) produtos analisados em um painel sensorial. Cada produto é associado a uma formulação, sendo essa correspondência previamente conhecida. Os julgamentos dos avaliadores do painel para cada produto podem ser repetidos  $d$  ( $d = 1, \dots, D$ ) vezes. O arranjo dos dados sensoriais pode ser representado através uma matriz  $\mathbf{X}$  composta por  $(P \times I \times D)$  linhas e  $J$  colunas, sendo cada elemento referido como  $x_{pid,j}$ .

#### Passo 1: Medição de consistência dos avaliadores

Avalie a consistência dos avaliadores usando um índice adequado. O método proposto utiliza o alfa de Cronbach ( $\alpha$ ), considerando os atributos de cada avaliador como variáveis na

análise. Esse índice tem sido amplamente utilizado para avaliar quão bem um conjunto de itens (avaliadores neste estudo) mede uma estrutura latente unidimensional (ALLEN E YEN, 2002). Como o valor de  $\alpha$  é proporcional às correlações entre os avaliadores, o índice fornece uma medida de consistência dos avaliadores do painel ao explicar um atributo ou perfil de atributos. Quanto mais próximo de 1, mais consistente é o conjunto de avaliadores.

Para calcular o alfa de Cronbach dos avaliadores, a matriz de dados  $\mathbf{X}$  é reorganizada para apresentar  $(I \times J \times D)$  linhas e  $P$  colunas, ou seja, haverá uma coluna por avaliador contendo as suas avaliações. O alfa de Cronbach é calculado a partir da seguinte expressão:

$$\alpha = \frac{P\bar{c}}{\bar{v} + (P-1)\bar{c}} \quad (1)$$

onde  $P$  denota o número total de avaliadores,  $\bar{c}$  é a covariância média entre avaliadores e  $\bar{v}$  é a variância média de todos os avaliadores. A equação (1) fornece um valor de alfa para o grupo de avaliadores. Para obter o índice de consistência do  $p$ -ésimo avaliador ( $\alpha_p$ ), a  $p$ -ésima coluna de  $\mathbf{X}$ , contendo os julgamentos daquele avaliador, deve ser omitida e a equação (1) aplicada à matriz de dados resultante. Um alto valor de  $\alpha_p$  indica que os julgamentos dos demais membros do painel apresentam maior consenso quando os julgamentos do avaliador  $p$  são omitidos do conjunto de dados. Os avaliadores são então ordenados de acordo com os valores de alfa obtidos no procedimento acima, sendo o avaliador com maior  $\alpha_p$  o primeiro a ser retirado da análise. A remoção do avaliador é seguida por um processo de seleção de atributos, detalhado nas etapas de 2 a 6.

## Passo 2: Aplicação da técnica multivariada no conjunto de dados de atributos sensoriais

Aplique uma técnica multivariada na matriz  $\mathbf{X}$  (Figura 1), considerando os atributos como variáveis na análise. O método proposto utiliza a ACP. Os parâmetros de interesse da ACP são os pesos dos componentes retidos ( $w_{jr}$ ) e o percentual de variância explicada por cada componente  $r$  ( $r = 1, \dots, R$ ) retido. O valor de  $R$  é definido com base na quantidade de variância explicada pelos componentes retidos. Maiores detalhes sobre a ACP podem ser obtidos em Montgomery et al. (2001), entre outros.

## Etapa 3: Geração do índice de importância dos atributos

Gere um índice de importância dos atributos com vistas à remoção de atributos irrelevantes e de ruído, como em Anzanello et al. (2009). O índice do atributo  $j$  é denotado por  $z_j$  ( $j = 1, \dots, J$ ). Quanto maior o valor de  $z_j$ , mais importante é o atributo correspondente na classificação dos produtos em classes de formulação.

O índice  $z_j$  é gerado com base nos pesos  $w_{jr}$  da ACP, como na equação (2). Tendo em vista que os pesos representam a importância dos atributos na concepção do componente principal, pesos elevados denotam atributos detentores de alta variabilidade e que, por consequência, permitem uma melhor discriminação dos produtos em classes de formulação (DUDA et al., 2001).

$$z_j = \sum_{r=1}^R |w_{jr}|, j = 1, \dots, J \quad (2)$$

## Passo 4: Classificação do conjunto de dados usando KVP e eliminação de atributos irrelevantes e ruidosos

Categorize as  $(P \times I \times D)$  observações da matriz  $\mathbf{X}$  em  $I$  classes de formulação usando KVP, e calcule a acurácia da classificação. A acurácia é definida como a razão entre o número de classificações corretas e o número total de classificações realizadas. O parâmetro  $k$  demandado pelo algoritmo KVP é selecionado pela validação cruzada no conjunto de dados

sensoriais, como em Chaovaitwongse et al. (2007). Observe que esta classificação utiliza todos os  $J$  atributos.

Na sequência, elimine o atributo com o menor valor de  $z_j$ , classifique novamente as  $(P \times I \times D)$  observações descritas pelos  $J - 1$  atributos remanescentes, e calcule a acurácia da classificação. Este procedimento de classificação e eliminação é repetido até que haja apenas um atributo remanescente.

### Passo 5: Construção do gráfico de acurácia

Construa um gráfico relacionando a acurácia de classificação ao número de atributos retidos. O ponto de máximo de acurácia indica o melhor subconjunto de atributos a serem retidos em procedimentos de classificação. No caso de haver subconjuntos alternativos com valores idênticos de acurácia, escolhe-se o subconjunto com menor número de atributos retidos.

### Passo 6: Remoção do avaliador menos consistente e realização de uma nova seleção de atributos

Remova o avaliador com o maior  $\alpha_p$  e repita as etapas 2 a 6 para os dados sensoriais compostos pelos  $J$  atributos originais e avaliadores remanescentes. Adicione o novo perfil de acurácia ao gráfico resultante da Etapa 5. Perceba que a eliminação de cada avaliador conduz a uma nova seleção de atributos baseada nas avaliações dos membros remanescentes no painel sensorial, a qual irá gerar um conjunto de perfis de acurácia. Repita este procedimento iterativo até atingir o limite inferior no número de avaliadores remanescentes, previamente definido.

A solução final é obtida pela identificação da acurácia máxima global (pico) no conjunto de perfis de acurácia, como exemplificado na Figura 1. O pico identifica tanto o melhor grupo de avaliadores quanto o melhor subconjunto de atributos a ser considerado em procedimentos de classificação. A análise de Pareto Ótimo (PO) também pode ser usada para identificar um número limitado de soluções que maximizam a acurácia da classificação e minimizam o número de atributos retidos. Tal análise pode ser particularmente útil quando o gráfico apresenta múltiplos picos.

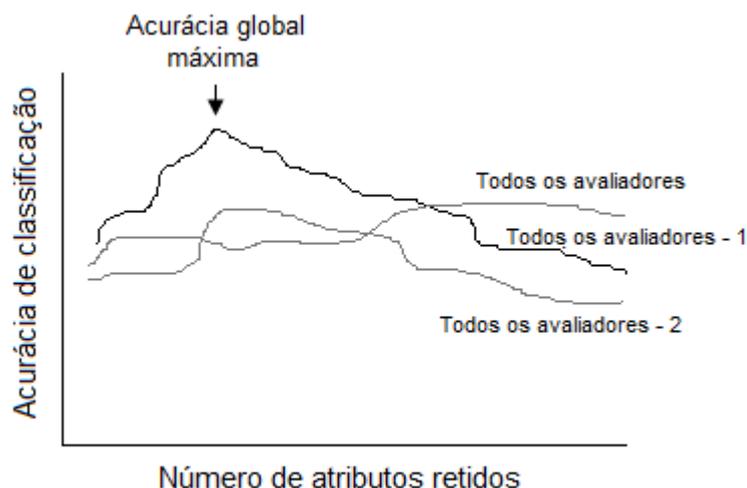


Figura 1: Perfis de precisão para eliminação de atributos e avaliadores

## 4. Estudo de Caso

Nesta seção, aplica-se o método descrito para seleção de atributos e avaliadores a um conjunto de dados sensoriais. Os 26 atributos listados na Tabela 1 são avaliados por nove

avaliadores treinados em um painel sensorial. Os atributos são relacionados à aparência, sabor e textura. Os produtos analisados são diferentes formulações de cubos de carne ao molho, acondicionados em *pouches* termoestáveis.

As avaliações são realizadas segundo o protocolo Spectrum, uma técnica de ADQ. No Spectrum, os produtos são avaliados individualmente por cada avaliador com respeito a um conjunto de atributos sensoriais [ver Meilgaard et al. (1999) para uma descrição completa do método]. As amostras foram preparadas pelo *Combat Ration Advanced Manufacturing Technology Development Program* no *Food Manufacturing Facility* em Piscataway, Nova Jersey (EUA), em 1994. Oito formulações, baseadas em especificações militares, foram testadas.

Nas análises a seguir atributos marcados com (\*) não foram considerados devido à falta de observações por alguns avaliadores. Cada avaliador repetiu as avaliações dos produtos 4 vezes para 24 atributos, gerando um total de 768 observações para cada avaliador. A dimensão da matriz **X** contendo os julgamentos de todos os avaliadores é (288 × 24). Na análise cada avaliador *p* é referido como *Pp*.

<b>Atributos de Aparência</b>	<b>Atributos de Sabor</b>	<b>Atributos de Textura</b>
(1) Proporção de molho na carne	(4) Aroma de carne cozida	(15) Viscosidade do molho
(2) Espessura visual do molho	(5) Aroma de caldo de carne	(16) Elasticidade da carne
(*) Cor do molho	(6) Aroma de carne crua	(17) Coesividade inicial da carne
(*) Cor dos cubos da carne	(7) Aroma de proteína vegetal hidrolisada	(18) Densidade da carne
(3) Uniformidade de tamanho e forma da carne	(8) Carne com sangue coagulado	(19) Firmeza da carne
	(9) Espessura	(20) Maciez da carne
	(10) Aroma de carne queimada	(21) Fibrosidade da carne
	(11) Aroma de gordura	(22) Estratificação da carne
	(12) Sal	(23) Umidade da carne
	(13) Sensação metálica	(24) Película oleosa
	(14) Sensação de calor	

Tabela 1: Atributos sensoriais avaliados no experimento

Os resultados da aplicação do alfa de Cronbach sobre os avaliadores são apresentados na Tabela 2. Cada valor de alfa é gerado após excluírem-se as avaliações de um julgador e os dados da matriz **X** serem reorganizados. O avaliador 9 (P9) é o menos consistente e o primeiro a ser removido do processo de seleção, seguido pelo P7 e assim similarmente para os demais. Em geral, os avaliadores exibiram valores de alfa elevados, indicando um bom consenso do grupo nas avaliações realizadas.

Identificação do Avaliador	Alfa de Cronbach
P9	0,9842
P7	0,9420
P6	0,9415
P2	0,9413
P3	0,9410
P1	0,9409
P5	0,9406
P8	0,9403
P4	0,9401

Tabela 2: Valores de alfa de Cronbach para os avaliadores

Em seguida, aplicou-se uma ACP no conjunto de dados. Foram retidos 2 componentes principais após a remoção de cada avaliador. Este número de componentes explicou pelo menos 63% da variância dos atributos analisados. Neste estudo, o valor  $R$  foi definido pelo pesquisador, tomando como base a variância total explicada pelos componentes retidos.

A seleção de atributos é realizada após eliminação de cada avaliador, levando a um conjunto de perfis de acurácia apresentados na Figura 2. A eliminação dos avaliadores é realizada até atingir um limite inferior de 5 avaliadores no painel sensorial, valor esse definido pelo pesquisador. Os Pontos de Fronteira a partir da análise de PO são identificados no gráfico de acurácia, onde FP1 é o Ponto de Fronteira localizado no extremo esquerdo. Estes pontos também são detalhados na Tabela 3.

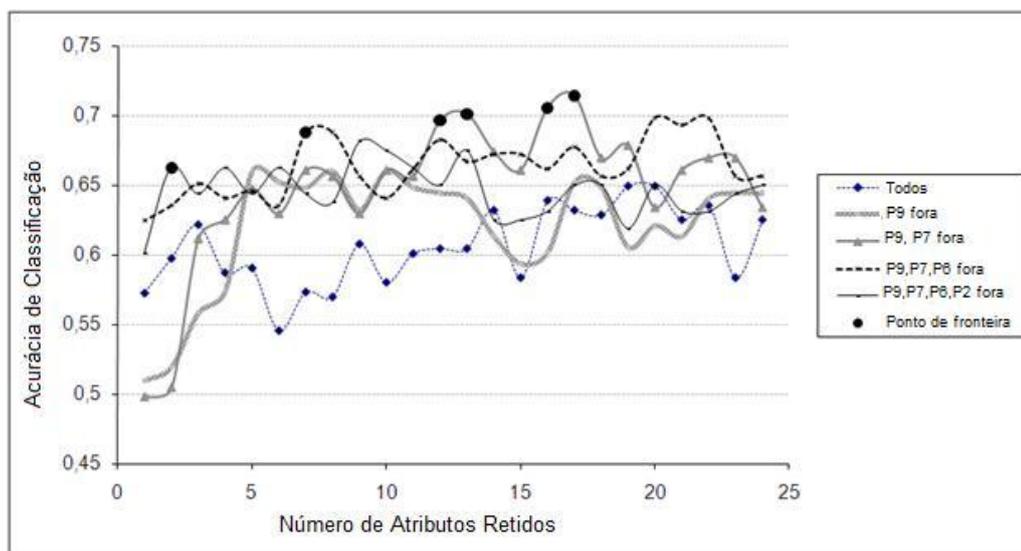


Figura 2: Perfis de acurácia para eliminação de atributos e avaliadores

A acurácia máxima de 71,4% é obtida quando os avaliadores P9 e P7 são removidos do conjunto de dados original, e 17 dos 24 atributos originais são retidos. A utilização dos atributos selecionados leva a classificações 12,6% mais precisas quando comparando à aplicação de KVP nos 24 atributos originais. Os atributos retidos, em ordem decrescente de relevância, são 16, 19, 18, 22, 4, 20, 21, 17, 11, 1, 2, 15, 3, 12, 6, 13 e 5. Observe que a acurácia de classificação com 24 atributos e 9 avaliadores é menor do que o obtido quando 17 atributos e 7 avaliadores são considerados. Isso é explicado pela remoção de atributos ruidosos e avaliadores inconsistentes do conjunto de dados.

Soluções alternativas são representadas pelos Pontos de Fronteira na Tabela 4. Verifique a existência de uma solução promissora: FP2 mostra uma situação em que 6 avaliadores avaliando apenas 7 atributos atingem uma acurácia de classificação de 68,8%. A redução de 70% do grupo de atributos original na avaliação sensorial, sem comprometer significativamente a acurácia de classificação, é altamente desejável uma vez que (i) a fadiga de avaliadores é minimizada, e (ii) há redução no custo da avaliação e preparação do painel sensorial.

## 5. Conclusão

Reduzir o número de atributos a serem analisados em experimentos para determinação de perfis sensoriais de produtos tem sido objeto de recente investigação. O objetivo é identificar um subconjunto de atributos que permita a discriminação das amostras testadas pelo painel. Esse processo visa minimizar o tempo e a fadiga imposta aos avaliadores, levando a procedimentos de coleta de dados menos dispendiosos.

O método proposto para seleção de atributos e avaliadores com propósito de classificação de amostras em formulações distintas consiste em seis passos: (i) avalie a consistência dos avaliadores usando um índice adequado; (ii) aplique ACP no conjunto de dados de atributos

sensoriais; (iii) compute um vetor de índices de importância dos atributos com base em pesos da ACP; (iv) classifique o conjunto de dados sensoriais usando a técnica KVP e calcule a acurácia da classificação; na sequência, elimine o atributo com o menor índice de importância, classifique novamente o conjunto de dados, e recalcule a acurácia; (v) construa um gráfico de acurácia; (vi) elimine o avaliador menos consistente e realize uma nova seleção de atributos, repetindo os passos (ii) a (vi).

O método proposto foi aplicado em um conjunto de dados de análise descritiva composto por julgamentos de 9 avaliadores sobre 24 atributos e 8 produtos. A máxima acurácia de 71,4% é obtida quando dois avaliadores são removidos do conjunto de dados original, e 17 dos 24 atributos originais são mantidos. Uma solução alternativa retendo menor número de avaliadores é identificada através de análise de Pareto Ótimo: 6 avaliadores avaliando apenas 7 atributos atingem uma acurácia de classificação de 68,8%.

Ponto de Fronteira (FP)	Avaliadores Retidos	Atributos Retidos ID	Acurácia de Classificação
1	P1, P3, P4, P5, P8	18, 20	0,663
2	P1, P2, P3, P4, P5, P8	16, 18, 20, 19, 22, 4, 17	0,688
3	P1, P2, P3, P4, P5, P6, P8	16, 19, 18, 22, 4, 20, 21, 17, 11, 1, 2, 15	0,696
4	P1, P2, P3, P4, P5, P6, P8	16, 19, 18, 22, 4, 20, 21, 17, 11, 1, 2, 15, 3	0,701
5	P1, P2, P3, P4, P5, P6, P8	16, 19, 18, 22, 4, 20, 21, 17, 11, 1, 2, 15, 3, 12, 6, 13	0,705
6	P1, P2, P3, P4, P5, P6, P8	16, 19, 18, 22, 4, 20, 21, 17, 11, 1, 2, 15, 3, 12, 6, 13, 5	0,714

Tabela 4. Informações sobre pontos de fronteira do Pareto Ótimo

Pesquisas futuras incluem o desenvolvimento de abordagens alternativas para selecionar o melhor subconjunto de atributos para classificação de produtos. Pretende-se gerar e testar outros índices de consistência para classificar os atributos e avaliadores, bem como aplicar ferramentas alternativas de mineração de dados, como *Support Vector Machine* e *Probabilistic Neural Networks*. Métodos destinados a selecionar atributos com o propósito de previsão também estão entre os objetivos dos próximos estudos.

## Referências

- Allen, M.; Yen, W.** *Introduction to Measurement Theory*. Long Grove, IL: Waveland Press, 2002.
- Anzanello, M.J.; Albin, S.L.; Chaovalitwongse, W.** Selecting the best variables for classifying production batches into two quality classes. *Chemometrics and Intelligent Laboratory Systems*, 97(2), 111-117, 2009.
- Azapagic, A.** Life cycle assessment and its application to process selection, design and optimization. *Chemical Engineering Journal*, 73(1), 1-21, 1999.
- Brown, T.A.** *Confirmatory Factor Analysis for Applied Research*. New York: Guilford, 2006.
- Carbonell, L.; Izquierdo, L.; Carbonell, I.** Sensory analysis of Spanish mandarin juices: selection of attributes and panel performance. *Food Quality and Preference*, 18, 329-341, 2007.
- Chaovalitwongse, W.; Fan, Y.; Sachdeo, C.** On the time series *k*-nearest neighbor classification of abnormal brain activity. *IEEE Transactions on System and Man Cybernetics A*, 37(6), 1005-1016, 2007.

- Deb, K.; Pratap, A.; Agarwal, S.; Meyarivan, T.** A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2), 182-197, 2002.
- Deb, K.; Thiele, L.; Laumanns, M.; Zitzler, E.** Scalable multi-objective optimization test problems. *Proceedings of the 2002 Congress on Evolutionary Computation*, 1, 825-830, 2002.
- Dijksterhuis, G.; Frost, M.B.; Byrne, D.V.** Selection of a subset of variables: minimization of Procrustes loss between a subset and the full set. *Food Quality and Preference*, 13, 89-97, 2002.
- Duda, R.; Hart, P.; Stork, D.** *Pattern Classification*, second ed., New York: Wiley-Interscience, 2001.
- Fogliatto, F.S.; Albin, S.L.; Tepper B.J.** A hierarchical approach to optimizing descriptive analysis multiresponse experiments. *Journal of Sensory Studies*, 14(4), 443-465, 1999.
- Fornell, C.; Larcker, D.F.** Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18 (1), 39-50, 1981.
- Granitto, P.M.; Gasperi, F.; Biasioli, F.; Trainotti, E.; Furlanello, C.** Modern data mining tools in descriptive sensory analysis: a case study with a Random forest approach. *Food Quality and Preference*, 18, 681-689, 2007.
- Horn, J.; Nafpliotis, N.; Goldberg, D.** A niched pareto genetic algorithm for multiobjective optimization. In: *Proceedings of the First IEEE Conference on Evolutionary Computation, IEEE World Congress on Computational Intelligence*, 1, 82-87, 1994.
- Krzanowski, W.J.** Selection of variables to preserve multivariate data structure, using principal components. *Applied Statistics*, 36(1), 22-33, 1987.
- Meilgaard, M.; Civille, G.V.; Carr, B.T.** *Sensory evaluation techniques* (3<sup>rd</sup> ed.), CRC Press: Boca Raton, 1999.
- Montgomery, D.; Peck, E.; Vining, G.** *Introduction to Linear Regression Analysis*, Willey: New York, 2001.
- Sahmer, K.; Qannari, E.M.** Procedures for the selection of a subset of attributes in sensory profiling. *Food Quality and Preference*, 19, 141-145, 2008.
- Taboada, H.; Coit, D.** Data clustering of solutions for multiple objective system reliability optimization problems. *Quality Technology & Quantitative Management Journal*, 4, 35-54, 2007.
- Taboada, H.; Coit, D.** Multi-objective scheduling problems: Determination of pruned Pareto sets. *IIE Transactions*, 40, 552-564, 2008.
- Weiss, S.; Apte, C.; Dameray, D.; Johnson, D.; Ples, F.; Goetz, T.; Hampp, T.** Maximizing text-mining performance. *IEEE Intelligent Systems*, 14(4), 63-69, 1999.
- Westad, F.; Hersleth, M.; Lea, P.; Martens, H.** Variable selection in PCA in sensory descriptive and consumer data. *Food Quality and Preference*, 14, 463-472, 2003.
- Wu, X.; Kumar, V.; Quinlan, J.; Ghosh, J.; Yang, Q.; Motoda H.; McLachlan, G.; Ng, A.; Liu, B.; Yu, P.; Zhou, Z.; Steinbach, M.; Hand, D.; Steinberg, D.** Top 10 algorithms in data mining. *Knowledge and Information Systems*, 14(1), 1-37, 2008.
- Zitzler, E.; Thiele, L.** Multiobjective evolutionary algorithms: a comparative case study and the strength pareto approach. *IEEE Transactions on Evolutionary Computation*, 3(4), 257-271, 1999.