

Classificação de Leveduras para o Controle Microbiano em Processos de Produção de Etanol

Lia N. B. Quinta, Jaime H. F. S. Queiroz, Kleber P. Souza, Hemerson Pistori e Marney P. Cereda
INOVISAO/Biotecnologia Aplicada à Agroindústria
Universidade Católica Dom Bosco
Av. Tamandaré, 6000, Jardim Seminário
79117-900 Campo Grande - MS
(lianaraq, jaime.cpp, kleber.padovani)@gmail.com, (pistori, cereda)@ucdb.br

Resumo

Uma das mais importantes etapas na produção de etanol é a fermentação, que acontece por ação de leveduras (fermento) sobre o caldo de cana diluído com água. A produtividade de uma usina de álcool pode ser significativamente incrementada a partir de um controle regular de leveduras e bactérias durante o processo. Esse controle é geralmente realizado a partir da identificação e contagem, em imagens microscópicas, de leveduras inviáveis e viáveis. Este artigo apresenta experimentos com algumas técnicas de visão computacional com o intuito de automatizar esse processo de identificação e contagem de leveduras. Nos experimentos com o algoritmo de casamento de modelos (template matching) 75% das leveduras (viáveis e inviáveis) foram reconhecidas. Já com as leveduras reconhecidas, experimentos usando a extração usando atributos de cor foram aplicados com o intuito de reconhecer leveduras viáveis e inviáveis, os resultados apontaram uma porcentagem de acerto de até 98%.

1. Introdução

Com uma preocupação recorrente em relação ao meio ambiente e com o fim das fontes de energia não renováveis, como o petróleo, a procura por fontes alternativas de energia tornou o etanol um produto bastante requisitado. O Brasil encontrou nesse setor uma importante fonte de desenvolvimento e crescimento financeiro sem se descuidar das questões ambientais [6].

Para a produção do etanol, a cana-de-açúcar passa pelo processo de moagem, quando o caldo é extraído. Após outras operações, ocorre a fermentação desse caldo com adição de fermento biológico de leveduras da espécie *Saccharomyces cerevisiae*. A eficácia da fermentação depende

de vários fatores. Uma das principais características que reduz o rendimento da fermentação é a grande concentração de açúcar no caldo. Por isso, esse caldo é geralmente diluído em água, produzindo o mosto.

No processo de colheita até a extração do caldo, a cana-de-açúcar é exposta à contaminação por bactérias. Devido à essa contaminação é necessário realizar o controle dos microorganismos presentes no mosto, a partir da análise visual de amostras coletadas periodicamente. Para auxiliar na contagem das leveduras, a amostra é misturada com água e tratada com corante metileno azul, que colore em azul as leveduras inviáveis. Uma levedura é considerada inviável quando não desempenha mais seu papel na fermentação alcoólica. Posteriormente, a amostra é transferida para uma placa em forma retangular e analisada através do microscópio. Nessa análise, realiza-se a contagem visual de leveduras e bactérias [1]. Através de cálculos estatísticos, a quantidade de leveduras e bactérias por litro é estimada. Ao constatar que o número de bactérias está alto, o especialista pode realizar o tratamento desses microorganismos, a fim de controlá-los, já que eles representam os contaminantes do mosto. Esse tratamento pode ser realizado através da inserção de antibióticos no mosto, por exemplo.

O projeto BioViC tem como objetivo oferecer às usinas de álcool um conjunto de ferramentas para facilitar e otimizar a realização do controle microbiano e inclui um módulo de visão computacional responsável pela contagem automática de microorganismos. Neste artigo, são relatados resultados com algumas técnicas de visão computacional com o objetivo de realizar a contagem de leveduras. Foram aplicadas técnicas de extração de atributos de cor e reconhecimento de padrões através de casamento de modelos. No casamento de modelos a medida euclidiana foi usada.

Os experimentos com casamento de modelos ainda são exploratórios e buscaram identificar alguns parâmetros iniciais para otimização da técnica para aplicação neste problema. Nos experimentos com atributos de cor houve a ex-

tração de atributos *pixel a pixel* e os resultados foram usados como treinamento para algoritmos de aprendizagem de máquina, que foram testados em um conjunto de imagens independente.

Nas próximas duas seções são apresentadas as técnicas de reconhecimento de padrões e extração de atributos usadas nesse trabalho. A quarta seção apresenta os experimentos realizados, resultados e análise. Por fim, a última seção apresenta as considerações finais.

2. Reconhecimento de padrões

O reconhecimento de padrões refere-se à etapa de identificação de objetos citados anteriormente. Por exemplo, neste trabalho, estima-se realizar o reconhecimento de duas classes: leveduras viáveis e inviáveis. Por isso, é necessário aplicar algoritmos que realizam a tarefa de reconhecer as classes desses objetos [4]. Dentre as técnicas de reconhecimento de padrões, encontra-se o casamento de modelos, que foi utilizado neste trabalho. De forma geral, no casamento de modelos é realizada uma busca na imagem por ocorrências de um determinado objeto representado por um modelo qualquer. Diversos tipos de modelos e medidas de similaridade para comparação entre trechos da imagem e estes modelos podem ser utilizados. Na sua forma mais simples, testada neste trabalho, o modelo é formado por um conjunto fixo de exemplos de imagens do objeto a ser buscado, em diferentes escalas e rotações. Uma janela de varredura, nos tamanhos dos exemplos de objetos que formam o modelo, percorre toda a imagem e para cada posição é calculado um valor de similaridade entre a janela e os exemplos. Um limiar de similaridade, definido empiricamente, é utilizado para determinar em que posições da imagem os objetos de interesse ocorrem. A medida de similaridade utilizada neste trabalho foi a distância euclidiana [2].

3. Extração de atributos de cor

A extração de atributos visa obter informações relacionadas com objetos presentes em uma imagem que facilitem a caracterização daqueles que pertencem a uma mesma classe. Por exemplo, os valores obtidos pela extração de atributos para caracterizar a classe das leveduras viáveis deverá ser diferente dos obtidos para caracterizar a classe das leveduras inviáveis. Desse modo, será possível discriminar mais facilmente as duas classes. Os modelos de cores utilizados para a fase de extração de atributos neste trabalho foram baseados em tons de cinza, RGB e HSB [5][8][9].

4. Experimentos, resultados e análise

Para a realização dos experimentos foi aplicada a técnica de casamento de padrões. Com os padrões encontrados, técnicas de extração de atributos de cor foram aplicadas para discriminar as classes viáveis e inviáveis.

4.1. Casamento de modelos

Como analisado anteriormente a técnica de casamento de modelos pode realizar o casamento das leveduras (*templates*) capturados nas imagens do projeto. Para medir a semelhança entre as classes foi aplicada a distância euclidiana.

Os modelos originais foram capturados de maneira manual, desse modo a dimensão desses modelos não estavam uniformes. Ao aplicar as medidas de similaridade, houve a necessidade de estabelecer um limiar usado como referência para o resultado obtido das medidas de similaridade. Para definir esse limiar as leveduras da mesma classe foram comparadas entre si. Para essa comparação a dimensão das leveduras deveriam ser iguais. Por isso, os modelos foram redimensionados. Como não era conhecida a dimensão que seria mais interessante para o problema, foram estabelecidas três dimensões. Esses valores foram obtidos com a ordenação dos modelos de forma crescente. Com as dimensões ordenadas, foram capturadas a menor, maior e média dimensão. Com esses valores os grupos foram criados e redimensionados.

Com o maior e menor valor das distâncias obtidas com a comparação das amostras, o limiar foi estimado capturando uma porcentagem da diferença entre os resultados máximos e mínimos. Para esses experimentos foram capturadas 35 amostras de leveduras.

Quando um modelo é comparado a uma imagem, este realiza uma varredura nas imagens variando um *pixel* no eixo *X* e um *pixel* no eixo *Y*. Os valores do deslocamento dos eixos foram estimados experimentalmente. Ao analisar uma imagem, a cada posição dessa imagem, um modelo com as dimensões do modelo original é extraído. O modelo extraído possui as mesmas dimensões do modelo original. Com as duas imagens é calculada a medida de similaridade entre eles. Os experimentos foram realizados com as imagens em tons de cinza e com modelos de cores RGB.

O limiar foi usado como referência para os experimentos com uma imagem real, ou seja, sem marcações. Se o valor obtido da comparação entre os dois modelos atuais apresentassem resultado dentro do limiar proposto, eram considerados como pertencentes à classe em questão. O mesmo raciocínio foi seguido para os experimentos usando modelos com componentes RGB. A distinção diz respeito ao cálculo de três diferenças e três limiares para cada modelo, já que foram analisados três componentes, sendo eles o R, G

e B. Por exemplo, o componente R(*red*) do modelo de cor RGB do modelo original foi comparado com o componente R do novo modelo criado da imagem a ser analisada. Experimentalmente a porcentagem escolhida para definir o limiar de leveduras foi de 60%.

4.2. Extração de atributos de cor

Como mencionado anteriormente, nas amostras de leveduras é adicionado o corante azul de metileno. Com a inserção desse corante as leveduras são diferenciadas por cor. Analisando essa informação, experimentos foram realizados com algoritmos de extração de atributos simples, baseado em cores e tons de cinza e aprendizagem de máquina. Para os experimentos, foram usadas nove imagens. Dessas nove imagens, foram obtidas amostras das classes viável e inviável. Dessas amostras, foram extraídos os tons de cinza, os componentes HSB e RGB de cada *pixel*. Com as informações extraídas, foi criado um arquivo arff (*Attribute-Relation File Format*) que armazena as informações obtidas e as relaciona com classes as quais elas pertencem.

O arquivo arff é um tipo de extensão que é reconhecido pelo software Weka (*Waikato Environment for Knowledge Analysis*). O Weka possui um conjunto de algoritmos de aprendizagem de máquina, sendo que alguns desses foram utilizados nesse trabalho. Como para este problema foi possível a obtenção de informação sobre a classificação dos objetos de interesse na imagem, foram utilizados algoritmos de aprendizagem supervisionada.

Para os experimentos, foram utilizados os algoritmos K2 [3] e o C4.5 [7], com os parâmetros *default*, ou seja, com as configurações originais dos algoritmos, sem variações de parâmetros. O primeiro induz modelos baseados em Redes Bayesianas e o segundo em Árvores de Decisão. Foram utilizadas quatro imagens para o treinamento e cinco para os testes. Foram extraídas amostras das classes viável e inviável. Com essas amostras foram extraídas informações em relação à tonalidade dos *pixels* utilizando os modelos propostos neste trabalho. No treinamento, os algoritmos de aprendizagem foram treinados com as informações contidas no arff. Finalizado o treinamento, foram extraídas características das imagens de teste. As características extraídas seguiram a mesma metodologia da extração das imagens de treinamento. Foram extraídas informações *pixel* a *pixel* das imagens de teste e essas informações foram inseridas nos algoritmos de aprendizagem e classificadas de acordo com a classe pertencente.

Nos experimentos, foram utilizadas 50498 amostras de inviáveis e 75737 amostras de viáveis. Nas imagens usadas para o treinamento, foi observado que as leveduras inviáveis conseguiram ser segmentadas em maior quantidade quando comparadas às leveduras viáveis. Também foi constatado

que os métodos de segmentação usando os componentes RGB e HSB apresentaram dados semelhantes em relação à matriz de confusão. Como mencionado anteriormente, as amostras foram classificadas pelos dois algoritmos mencionados. Na etapa de classificação, as amostras tinham que ser classificadas nas classes corretas, porém foram detectados alguns erros na classificação dessas amostras. Por exemplo, na Tabela 1 a segunda coluna refere-se à classe das leveduras inviáveis e a terceira à classe das leveduras viáveis. Na segunda coluna e na segunda linha estão as classificações corretas para a classe inviáveis e na terceira coluna e segunda linha estão o número de leveduras classificadas incorretamente, ou seja, classificadas como viáveis. O mesmo raciocínio é realizado para as outras Tabelas. As Tabelas 1, 2 e 3 apresentam os resultados obtidos com o algoritmo C4.5 e K2.

Nos experimentos utilizando informações sobre o tom de cinza dos *pixels*, foi detectado uma maior confusão em relação à classificação dos atributos, como mostra a Tabela 3. Como apenas uma informação sobre o *pixel* das imagens é analisada, esse tipo de confusão torna-se ainda maior, visto que nesse modelo o número de variações de tonalidades é inferior quando comparados aos outros modelos propostos neste trabalho.

Método de Aprendizagem	a	b	classe
C4.5	50314	184	a=inviável
C4.5	2026	73711	b=viável
K2	34952	15546	a=inviável
K2	5138	70599	b=viável

Tabela 1. Matriz de confusão resultante da segmentação por componente RGB com o algoritmo C4.5 e K2

Método de Aprendizagem	a	b	classe
C4.5	50310	188	a=inviável
C4.5	1927	73810	b=viável
K2	40873	9625	a=inviável
K2	2287	73450	b=viável

Tabela 2. Matriz de confusão resultante da segmentação por componente HSB com o algoritmo C4.5 e K2

A porcentagem de acerto foi fornecida pelo Weka assim como a matriz de confusão. Essa porcentagem refere-se ao

Método de Aprendizagem	a	b	classe
C4.5	32708	17790	a=inviable
C4.5	4550	71187	b=viable
K2	33508	16990	a=inviable
K2	5545	70192	b=viable

Tabela 3. Matriz de confusão resultante da segmentação por tons de cinza com o algoritmo C4.5

Método de segmentação	C4.5	K2
TDC	82,30%	82,14%
RGB	98,25%	83,61%
HSB	98,32%	90,56%

Tabela 4. Porcentagem de acerto dos métodos utilizados.

número de amostras classificadas corretamente pelos dois algoritmos usados nesse experimento para os modelos de cor HSB, RGB e tons de cinza (TDC). A Tabela 4 apresenta a porcentagem de acerto obtida através dos algoritmos de aprendizagem de máquina resultante dos experimentos de cada componente HSB, RGB e tons de cinza. Com esses resultados, foi observado que o algoritmo C4.5 apresentou uma melhor porcentagem de acerto.

Ao realizar a comparação dos resultados dos métodos RGB e HSB, observa-se que, nos diferentes tipos de dados apresentados, os componentes HSB apresentam melhor resultado. Esses experimentos foram realizados utilizando validação cruzada com dez dobras. Nos experimentos, estavam contidos dados obtidos das amostras das classes presentes nas imagens de treinamento.

O resultado do casamento de padrões com a extração de atributos de cor apresenta a classificação de algumas leveduras e a presença de alguns falsos/positivos e falsos/negativos, as Figuras 1 e 2 (b) ilustram as imagens após a classificação. A Figura 1 (b) mostra uma imagem com o reconhecimento de leveduras viáveis. Os círculos vermelhos são leveduras que não foram classificadas e os quadrados pretos são das células marcadas corretamente. A Figura 2 (b) mostra o reconhecimento de leveduras inviáveis, seguindo a mesma legenda usada para as leveduras viáveis.

5. Conclusão

O algoritmo de casamento de modelos apresentou uma taxa de acerto de 75%. Com os modelos de cores e os algoritmos de aprendizagem de máquina utilizados nos exper-

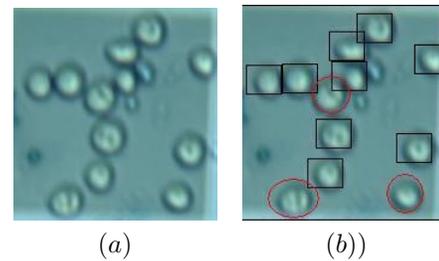


Figura 1. (a) Imagem original, (b) Classificação de leveduras viáveis.

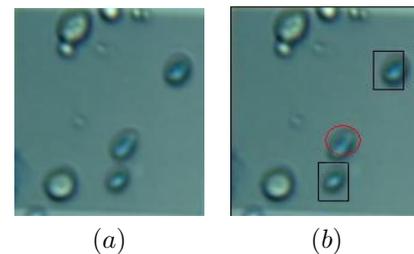


Figura 2. (a) Imagem original, (b) Classificação de leveduras inviáveis.

imentos, obteve-se melhor resultado com o modelo de cor HSB e algoritmo de aprendizagem de máquina C4.5, com a porcentagem de acerto de 98,32%.

Já com os modelos encontrados no primeiro experimento foram aplicadas as técnicas de extração de atributos de cor. A classificação baseada em cores apresentou resultado superior a 80% de acerto. Com esses experimentos observou-se que a união das técnicas de casamento de modelos e extração de atributos de cor foram consideradas satisfatórias na distinção de leveduras.

Como trabalhos futuros pretende-se expandir o número de técnicas que trabalham com o algoritmo de casamento de modelos, já que uma simples técnica aplicada trouxe uma boa distinção de classes, outra meta futura será a classificação de bactérias.

Referências

- [1] S. R. C. Antonini. Métodos de análises e monitoramento microbiológico em laboratório de destilaria. Apostila para o curso de treinamento ministrado nas unidades de Iguatemi-PR, no período de 19 a 21 de fevereiro de 2004, e de Ivaté-PR, no período de 16 a 18 de fevereiro de 2004, pertencentes à Usina de Açúcar Santa Terezinha Ltda., 2004.

- [2] R. O. Borges, R. A. A. Silva, and S. S. Castro. Utilização da classificação por distância euclidiana no mapeamento dos focos de arenização no setor sul da alta bacia do rio araguaia. In *XIII Simpósio Brasileiro de Sensoriamento Remoto*, 2007.
- [3] G. F. Cooper and T. Dietterich. A bayesian method for the induction of probabilistic networks from data. In *Machine Learning*, pages 309–347, 1992.
- [4] L. L. e S. K. Mitra. Color image segmentation: A state-of-the-art survey, 2001.
- [5] R. C. Gonzalez and R. E. Woods. *Processamento de Imagens Digitais*. primeira edition, 2000.
- [6] M. A. Guadagnini and M. S. S. Almeida. Fontes alternativas de energia uma visão geral. Pós Graduação Executiva em Meio Ambiente. 09 agosto de 2005 / 01 agosto de 2006, 14^a Turma, COPPE/UFRJ., 2006.
- [7] J. R. Quinlan. Improved use of continuous attributes in C4.5. *Journal of Artificial Intelligence Research*, 4:77–90, 1996.
- [8] R. C. B. Rodrigues and R. H. R. Viana. Máquinas de vetores de suporte aplicadas à classificação de defeitos em couro bovino. Qualificação de Bacharel em Engenharia de Computação na Universidade Católica Dom Bosco Campo Grande - MS - Brasil, 2007.
- [9] J. A. Silva and W. N. Gonçalves. Modelos ocultos de markov aplicados na identificação de comportamento de animais. Qualificação de Bacharel em Engenharia de Computação na Universidade Católica Dom Bosco Campo Grande - MS - Brasil, 2007.