

14634 - Identificação computadorizada de tipos polínicos através de “Bag of Words”

Identification of pollen grains computed by Bag of Words

GONÇALVES, Ariadne Barbosa¹; RODRIGUES, Carolini Nascimento Martins¹;
CEREDA, Marney Pascoli¹; PISTORI, Hemerson¹

1 Universidade Católica Dom Bosco - UCDB, Av. Tamandaré n° 6000, CEP 79117-900, Jardim Seminário, Campo Grande - MS. ariadne.gon@gmail.com, carolvidaboa@gmail.com, cereda@ucdb.br, pistori@ucdb.br

Resumo: A apicultura é uma atividade do agronegócio vinculada com a preservação do meio ambiente, pois as abelhas contribuem para conservação de espécies nativas e polinização de culturas. Em uso alimentar do mel a identificação da origem é importante para estabelecer a ligação do produto de extrativismo das abelhas com a região de origem que pode ser realizado através do pólen carregado pelas abelhas e esta presente no mel. Também propicia estabelecer e acompanhar plantas apícolas de determinada região, usando como marcadores os polens encontrados no mel. Para isso é necessário estabelecer a frequência e identificação polínica que é um processo demorado e depende de um especialista. A automação é possível usando visão computacional. Para isso foi avaliada a técnica de *Bag of Words* com a extração de atributos a partir de um vetor de atributos que são extraídos da imagem ou conjunto de imagens. O índice de acerto na classificação dos grãos de pólen foi de 70% através do uso do algoritmo classificador SMO.

Palavras-chave: Apicultura; mel; pólen; autenticação; visão computacional

Abstract: Beekeeping is an activity of agribusiness linked to preservation of the environment, because the bees contribute to conservation of native species and crop pollination. Use as food honey origin identification is important to connect the product extraction bees with the region of origin that can be done through pollen carried by bees and is present in honey. It allowed establish and monitor honey plants of a particular region, using as markers the pollens found in honey. It is necessary to establish the frequency and pollen identification which is a time consuming process and depends on an expert. Automation is possible using computer vision. For this technique was evaluated with Bag of Words feature extraction from a vector of attributes that are extracted from the image or set of images. The rate of correct classification of pollen grains was 70% through the use of SMO classifier algorithm.

Keywords: Beekeeping; honey; bee pollen; authentication; computer vision

Introdução

Apicultura é a atividade de criação de abelhas do gênero *Apis* em colmeias artificiais visando à produção de mel. Além de boa opção para o agronegócio em seu produto principal, o mel, a apicultura é atividade ambientalmente sustentável, contribuindo para a manutenção das espécies nativas e aumento da produção agrícola (LOPES et al., 2001).

Como alimento, o mel é produzido pelas abelhas apícolas a partir do néctar das flores, das secreções de partes vivas das plantas ou ainda de excreções de insetos sugadores das mesmas. As abelhas recolhem estes produtos e os transformam,

combinam com substâncias específicas próprias, armazenam e deixam maturar nos favos da colmeia (BRASIL, 2000). Mesmo não sendo intencionais, ao procurarem estas secreções que geralmente estão próximas das flores, as abelhas podem carregar os grãos de polens aderidos às suas cerdas para o mel, fato este que pode ser usado na caracterização da sua origem botânica.

A análise polínica é importante para diagnosticar as espécies de plantas nas quais as abelhas buscam alimento, permitindo monitorar as plantas poliníferas através do pólen, que apresenta formato, tamanho e textura específicos a diferentes espécies. Atualmente a contagem dessas espécies vegetais é feita visualmente, o que é muito trabalhoso e por isso sujeito a erros devido ao desgaste humano.

A rastreabilidade de um produto quando ligada ao conceito de qualidade passa a ser uma ferramenta fundamental para verificação da qualidade nas diversas etapas de produção e no produto final (CAMARGO, 2012). Por isso a rastreabilidade evolui mundialmente para ser uma exigência em produtos e derivados de origem animal. Além de atender as exigências de legislações internacionais e nacionais, permite agregar valor aos processos e produtos. Especialmente para o Brasil, que possui grande parte da sua flora conservada, a produção de mel e outros produtos apícolas é uma possibilidade de agregar valor ao produto com a imagem de preservação ambiental como forma de destacar o produto.

A automatização no processo de identificação e contagem de grãos de pólen tornam a análise mais rápida e menos trabalhosa, pois permite contar com praticidade um número maior de grãos e é mais precisa na identificação polínica (LANGFORD et al., 1990).

A variabilidade de formato e textura dos grãos de pólen facilita a individualização dos mesmos e a ligação com plantas de origem. O campo da visão computacional é dedicado, entre outras coisas, à criação de algoritmos para extrair informações de uma determinada imagem. O uso da técnica de *Bag of Words* (BOW), que faz extração de atributos de uma imagem e é utilizada na área de visão computacional, permite gerar um histograma de cada imagem que contém as características da imagem, que é associada a sua respectiva classe com auxílio de um algoritmo de aprendizagem supervisionada. Ao analisar uma nova imagem os pontos de interesse das imagens são associados a cada palavra visual, de acordo com a menor distância entre o ponto analisado e a média de cada grupo.

O objetivo deste trabalho foi avaliar o uso da técnica de *Bag of Words* acoplada a um algoritmo de aprendizagem supervisionada com a utilização de imagens microscópicas do produto para automatização da identificação dos padrões de grãos de pólen no mel. Para isso foram investigados 4 algoritmos de aprendizagem automática para se identificar qual deles oferece melhores resultados quando utilizados em conjunto com o BOW na solução do problema.

Metodologia

A pesquisa foi desenvolvida nos Laboratórios do Centro de Tecnologia e Análise do Agronegócio (CeTeAgro) e no ambiente do Laboratório do Grupo de Pesquisa INOVISÃO da Universidade Católica Dom Bosco-MS. Méis de 11 municípios do Estado de Mato Grosso do Sul foram processados pelo método de acetólise como

proposto por Jones e Bryant-Junior (2004), para a extração dos grãos de pólen do produto, pois esta técnica melhora a individualização da estrutura do grão de pólen. Posteriormente foi feita a captura de imagens dos grãos de pólen em microscópio digital LCD micro Bresser com lentes de aumento de 40X. De cada tipo polínico foram capturadas 30 imagens que foram utilizadas para aprendizagem do programa. Para o experimento foram utilizados os nove tipos (Figura 1) que compõem o banco de imagem POLEN9E com 270 imagens, sendo estes tipos polínicos uns dos mais frequentes encontrados nas amostras de méis do Estado.

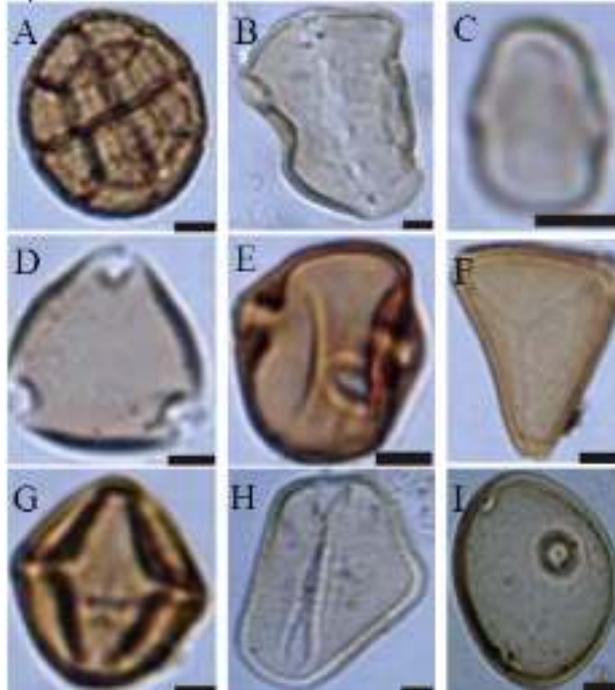


Figura 1. Imagens dos nove tipos polínicos presentes no banco de imagens e das plantas que os originaram. A- *Anadenanthera colubrina*; B- *Arecaceae*; C- *Cecropia pachystachya*; D- *Myrcia* sp.; E- *Protium* sp.; F- *Serjania* sp.; G- *Schinus* sp.; H- *Syagrus oleracea*; I- *Poaceae*. (Escala = 10 µm).

Com as imagens de aprendizagem o BOW extraiu os pontos de interesse de cada imagem de grãos de pólen, o qual foi relevante para a distinção entre as imagens de outros tipos polínicos.

A partir dos pontos de interesse detectados pelo BOW foi gerado um dicionário de 2048 descritores das características presentes em cada pólen e então realizados experimentos com algoritmo de aprendizagem de máquina e o número de atributos a serem extraídos do conjunto de imagens. Para tanto foram utilizados os seguintes algoritmos: SMO, IBK, J.48 e AdaBoost. Para saber qual dos algoritmos apresenta melhor desempenho na classificação dos tipos polínicos foram realizadas comparações entre eles através das seguintes métricas: área sob a curva ROC (Area Under Roc), tempo de uso da CPU para classificação (User CPU Time Testing), porcentagem de acerto (Percent Correct) e a média ponderada da medida F (Weighted Average F Measure).

Através do algoritmo que obteve o melhor desempenho na classificação dos tipos polínicos foi produzida uma matriz de confusão que permitiu uma análise mais detalhada dos tipos de erros cometidos pelo classificador.

Resultados e discussões

As métricas porcentagem de acerto, curva ROC e média ponderada da medida F indicaram melhor desempenho do classificador SMO, que obteve 71% de acerto na classificação dos tipos polínicos (Figura 2). A métrica CPU apresentou melhor resultado do classificador IBK, no entanto os resultados desta métrica foram próximos ao do classificador AdaBoost. A métrica CPU faz uma análise do tempo que leva para o algoritmo testar o conjunto de dados gerado a partir das imagens, como dois algoritmos obtiveram resultados parecidos é preciso analisar outros aspectos para verificar um melhor desempenho.

A classificação correta das imagens de cada tipo polínico realizada pelos classificadores é visualizada na Tabela 1, onde é percebido que o algoritmo SMO proporcionou o melhor desempenho em relação aos demais classificadores. O segundo melhor foi o J.48 com 45% de acerto, mesmo assim é um percentual não satisfatório visto que teve mais erros que acertos.

Tabela 1: Porcentagem de amostras classificadas corretamente utilizando os classificadores presentes neste trabalho.

| Espécies | SMO | J.48 | IBK | ADABOOST |
|--------------------------------|-------|-------|-------|----------|
| <i>Anadenanthera colubrina</i> | 91.7 | 66.7 | 04.2 | 0 |
| Arecaceae | 86.2 | 31.0 | 0.0 | 58.6 |
| <i>Cecropia pachystachya</i> | 100.0 | 90.3 | 100.0 | 96.8 |
| <i>Myrcia</i> sp. | 63.00 | 33.3 | 0.0 | 0 |
| <i>Protium</i> sp. | 57.7 | 26.9 | 03.8 | 0 |
| <i>Schinus</i> sp. | 31.8 | 18.2 | 0 | 0 |
| <i>Serjania</i> sp. | 72.0 | 44.0 | 08.0 | 0 |
| <i>Syagrus oleraceae</i> | 50.0 | 37.5 | 0 | 0 |
| Poaceae | 66,7 | 42.90 | 0 | 0 |

```

a b c d e f g h i <- classificada como
22 0 0 0 0 1 1 0 0 | a = Anadenanthera colubrina
0 25 0 1 1 0 0 0 2 | b = Arecaceae
0 0 31 0 0 0 0 0 0 | c = Cecropia pachystachya
0 0 8 17 0 2 0 0 0 | d = Myrcia sp.
0 0 4 0 0 15 6 1 0 | f = Protium sp.
0 0 5 4 0 6 7 0 0 | g = Schinus sp.
0 0 0 1 1 4 1 18 0 | h = Serjania sp.
0 5 0 0 3 0 0 0 8 | i = Syagrus oleraceae
0 2 0 2 14 1 0 0 2 | e = Poaceae

```

Figura 2: Matriz com o resultado da classificação realizada pelo algoritmo SMO.

A identificação polínica é essencial para a caracterização do território apícola regional sendo o ponto de partida essencial para a sua valorização junto ao consumidor final, e também onde é possível demonstrar de forma sustentável as bases de um programa de certificação de denominação de origem e qualidade, bem como desenvolver uma expressiva imagem de marca do mel, através da certificação polínica em conjunto com a caracterização da flora e da vegetação do território apícola (TERRA-SCENICA, 2012). Com um programa de computador que torna mais confiável, rápida, acessível e prática a classificação dos tipos polínicos, o

apicultor pode ter a certificação polínica do mel que valoriza o produto concomitantemente com a denominação de origem do mel.

Conclusões

Foi possível confirmar com mais do que 70% de acerto a classificação de nove tipos polínicos, portanto a utilização da técnica de *Bag of Words* reforça a viabilidade de automatização da determinação de grãos de pólen. Novas técnicas deverão ser avaliadas visando aumentar ainda mais esta taxa de acerto.

Agradecimentos

Este trabalho recebeu apoio financeiro da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ).

Referências bibliográficas

- BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Defesa Animal. Legislação. **Sistema de Consulta a Legislação. Módulo de Legislação Agropecuária**. Instrução Normativa n. 11, de 20 de outubro de 2000. Regulamento técnico de identidade e qualidade do mel. Disponível em: <<http://sistemasweb.agricultura.gov.br/sislegis/action/detalhaAto.do?method=consultarLegislacaoFederal>>. Acesso em: 29 jun. 2012.
- CAMARGO, R.C.R. **Rastreabilidade na cadeia produtiva apícola**. Disponível em: <<http://www.cpamn.embrapa.br/apicultura/rastreabilidade.php>>. Acesso em: 19 ago. 2012.
- JONES, G. D.; BRYANT, V. M. Jr. The use of ETOH for the dilution of honey. **Grana**, v. 43, p. 174–182, 2004.
- LANGFORD, M., TAYLOR, G.E., FLENLEY, J.R. Computerized identification of pollen grains by texture analysis. **Review of Palaeobotany and Palynology**, n. 64, p. 197-203, 1990.
- LOPES, M.T.R. et al. **Apicultura. Empresa Brasileira de Pesquisa Agropecuária Meio-Norte**. Teresina, PI, 2001, 2 p.
- TERRA-SCENICA. **Pólen, Mel e Território**. Disponível em: <<http://www.terra-scenica.pt/investigacao/ecologia/polenmel.htm>>. Acesso em: 19 ago. 2012.