

Floresta de Caminhos Ótimos na Classificação de Pólen

Lia Nara Balta Quinta¹, Willian Paraguassu Amorim², Marcelo Henriques de Carvalho³,
Marney Pascoli Cereda¹ e Hemerson Pistori¹

1 - Universidade Católica Dom Bosco - UCDB

INOVISAO e Departamento de Biotecnologia - Campo Grande, Brasil

2 - Universidade Federal da Grande Dourados - UFGD

INOVISAO e FACET - Faculdade de Ciências Exatas e Tecnologia - Dourados, Brasil

3 - Universidade Federal de Mato Grosso do Sul - UFMS

FACOM - Faculdade de Computação - Campo Grande, Brasil

Abstract

In the production of honey some pollen grains are deposited on the bees. On returning to the hive, the grains of the species fall into the product and can be viewed through a microscope. Currently for identifying the product, it is necessary to count the pollen grains present in the honey. This work aims to use techniques of computer vision and artificial intelligence in microscopic images, to classify pollen grains present in the honey. To perform the classification of pollen grains was applied a technique called Optimal-Path Forest (OPF). We extracted attributes of color (RGB and HSB), shape (form factor, roundness, aspect ratio and density) and texture (co-occurrence matrix and interaction maps). The OPF was compared with traditional classifiers C4.5, SVM and KNN. According to Friedman test can be affirmed that there is no difference between the classifiers in relation to the result of classification and that the SVM has more training time than the other classifiers.

1. Introdução

Antigamente poucos produtores possuíam capacitação adequada para implantar e produzir mel. Porém, com a maior procura do produto, normas de padronização foram criadas com o objetivo de estabelecer uma identificação e qualidade do produto produzido [9]. A Instrução Normativa N° 11, de 20 de outubro de 2000 é um exemplo da padronização supracitada [4].

Para produzir o mel, as abelhas fêmeas, responsáveis pela fabricação do produto, capturam néctar das flores e, no momento da captura do néctar, alguns grãos de pólen ficam depositados sobre o inseto. Ao retornar à colmeia, al-

guns grãos de pólen estão sobre ele e, por esse motivo, ao analisar, através de um microscópio, uma amostra de mel os grãos das espécies contidas no produto podem ser visualizados. De acordo com o tipo de néctar capturado, as características do produto sofrem alterações, como por exemplo no sabor e coloração.

No momento da captura do néctar, as abelhas realizam um voo com um raio de 2 quilômetros de distância da colmeia. Por isso, o pólen das árvores presentes ao redor da colmeia possui grande influência no mel produzido naquela colmeia.

Há diversos tipos de mel, como os florais e melato. Nesta pesquisa somente os méis florais serão analisados por apresentarem grãos de pólen. Os méis florais são classificados de acordo com a quantidade de espécies vegetais (grãos de pólen) presentes no produto [3][1]. Uma das características usadas para determinar a origem floral do mel é o grão de pólen.

Atualmente, a contagem das espécies vegetais presentes no mel é realizada visualmente. Essa contagem é uma tarefa minuciosa e muito cansativa, por esse motivo, erros na contagem e classificação de espécies podem ocorrer devido ao desgaste humano. Nesta proposta, a tarefa de classificação será realizada via software acarretando maior agilidade sem perda de eficácia na contagem.

Neste trabalho, para realizar a classificação dos grãos de pólen foi aplicada uma técnica denominada Floresta de Caminhos Ótimos (OPF) [7]. Inicialmente foi necessário extrair informações das imagens de grãos de pólen, para isso técnicas de extração de atributos foram aplicadas. Com os dados resultantes da extração de atributos, o classificador foi utilizado para distinguir as classes (espécies de grãos de pólen) analisadas. O OPF foi comparado com os classificadores C4.5, SVM e KNN. Esses classificadores fo-

ram selecionados por serem classificadores tradicionais na área de visão computacional e por estarem presentes em vários trabalhos em que são comparados com o OPF [7][2]. Para avaliar o resultado dos classificadores e analisar se há ou não diferença na aplicação das técnicas supracitadas na classificação de grãos de pólen serão avaliados os resultados que o classificador apresenta, através dos testes de Friedman e *T-Students* aplicados no resultado da taxa de acerto, *recall*, *precision* e *f-score*.

Experimentos iniciais foram realizados com as imagens presentes no banco de imagens do projeto. Três grupos de experimentos foram realizados, o primeiro utilizava atributos de forma e cor com 3 classes (espécies de grãos de pólen). O segundo experimento adicionava atributos de textura aos atributos já extraídos no experimento 1. Além de adicionar estes atributos, houve um aumento no número de classes presente no experimento. No experimento 3, o valor dos parâmetros obtidos no experimento 2 foi utilizado e aplicado aos 4 conjuntos de imagens que serão descritos posteriormente. Experimentos analisando a taxa de acerto, *recall*, *precision* e *f-score* foram realizados e, através do teste de Friedman pode-se afirmar que em relação à estas métricas, não há variação entre os classificadores, ou seja, de acordo com os resultados das métricas pode-se afirmar que eles são iguais.

2. Floresta de Caminhos Ótimos - OPF

O problema de reconhecimento de padrões pode ser modelado para o uso de floresta de caminhos ótimos em um grafo, definido em seu espaço de atributos. Os nós são as amostras, sendo representadas pelos seus respectivos vetores de atributos, e os arcos definidos por uma relação de adjacência entre as amostras [6][5]. Tanto os nós quanto os arcos, podem ser ponderados, tendo a possibilidade de aplicação de diversas funções de custo. Com isso, é possível o particionamento do grafo em árvores de caminhos ótimos, sendo enraizadas pelos seus respectivos protótipos (sementes) na fase de treinamento, e a classificação de uma amostra qualquer sendo o mesmo rótulo do protótipo mais fortemente conexo. A seguir apresentaremos a técnica OPF, seguindo a heurística de análise dos dados sobre um grafo completo e seleção dos protótipos como sendo as amostras mais próximas de classes distintas.

2.1. Definição

O cálculo de caminhos ótimos para as demais amostras são computadas pelo algoritmo de Transformada Imagem Floresta (IFT)[2]. A técnica IFT é essencialmente o algoritmo de Dijkstra, modificado para receber várias fontes e aplicação de outras funções de custo. Seu comportamento inicialmente identifica os mínimos da função de custo como

os nós de origem e se propaga em ordem não decrescente a fim de melhores caminhos, particionando o grafo em uma floresta de caminhos ótimos onde as raízes são os protótipos selecionados.

Atualmente Papa et al[7], apresentou um novo modelo de classificador supervisionado. A técnica usa como modelo de relação de adjacência um grafo de conexão completa, uma função de custo f_{max} , dada pela Equação 1, em que o peso dos arcos $w(s, t)$, é a distância $d(s, t)$ entre os vetores de características $\vec{v}(s)$ e $\vec{v}(t)$, π_s um caminho qualquer, $\pi_s \cdot \langle s, t \rangle$ a concatenação do caminho π_s com término em s e o arco (s, t) , e escolhe os protótipos a partir da árvore resultante gerada pela técnica Árvore de Custo Mínimo (MST).

$$f_{max}(\pi_s \cdot \langle s, t \rangle) = \max\{f_{max}(\pi_s), d(s, t)\}, \quad (1)$$

2.2. Classificador OPF baseado em grafo completo

Dado os seguintes conjuntos de dados representados por grafos completos, Z_1 dados do treinamento, Z_2 dados de avaliação e Z_3 , conjunto de dados que serão utilizados para a classificação dos dados. Seja $Z = Z_1 \cup Z_2 \cup Z_3$, $\lambda(s)$ um função que atribui o rótulo correto $i \in \{1, 2, \dots, c\}$ da classe i para qualquer amostra $s \in Z$, S um conjunto de protótipos de todas as classes, sendo $S \subset Z_1$, e v um algoritmo que extrai n atributos como: (textura, cor, formas geométricas) de qualquer amostra $s \in Z$ retornando um vetor $\vec{v}(s)$. A distância $d(s, t) \geq 0$, entre duas amostras, s e t , é somente um valor entre os vetores de características $\vec{v}(s)$ e $\vec{v}(t)$. Pode-se usar qualquer função distância, mas para a abordagem de Papa et al[7] sobre grafo completo, utiliza a distância Euclidiana $\|\vec{v}(t) - \vec{v}(s)\|$.

O problema de classificação nada mais é do que rotular a classe correta $\lambda(s)$, para qualquer amostra $s \in Z_3$. A fase de treinamento consiste em encontrar um subgrupo de protótipos $S^* \subset Z_1$, e a classificação de qualquer amostra $s \in Z_3$, é feita através da avaliação de forma incremental atribuindo a ele o rótulo do protótipo mais fortemente conexo.

2.3. Fase de Treinamento

Dado um grafo completo Z_1 o treinamento consiste em encontrar protótipos S^* , através da Árvore de Custo Mínimo (MST) e Floresta de Caminhos Ótimos com raízes em S^* , de modo que os elementos sejam considerados os mais representativos. Várias heurísticas podem ser adotadas, entretanto o modo de seleção pode afetar o desempenho do classificador. Um processo de seleção sugerido por Papa et al[7], para seleção de protótipos sobre grafos completos, é o uso do grafo gerado sobre um *MST* (Árvore Geradora de Custo Mínimo). Computando uma *MST* no grafo completo Z_1 , obtemos um grafo conexo acíclico, cujos nós são todas as amostras em Z_1 , e os arcos são não direcionados

e ponderados. A Figura 1(a), ilustra o resultado deste processo.

A soma dos pesos de seus arcos é mínima, se comparada a outras árvores geradoras no grafo completo, com isso obtemos uma árvore geradora ótima. Dessa forma, os protótipos a serem selecionados, são os elementos mais próximos de classes diferentes, ilustrado na Figura 1(b). Removendo-se os arcos entre as classes diferentes, essas amostras adjacentes serão os protótipos em S^* . Um ponto importante, é que uma classe pode conter vários protótipos (árvore de caminhos ótimos), mas devemos sempre garantir que toda classe, deva ter no mínimo um protótipo.

Assumindo que conhecemos o conjunto S^* , o Algoritmo 1, de Floresta de Caminhos Ótimos (OPF), propaga os rótulos dos protótipos para todas as amostras de suas respectivas árvores de caminhos ótimos, formando um mapa de rótulos $L(s) \in \{1, 2, \dots, c\}$. As linhas de 1 – 4, inicializam as variáveis e inserem protótipos em Q. O laço principal inicializado na linha 5, calcula os caminhos ótimos de S^* para toda amostra s , a partir de uma ordem crescente de valores. A cada iteração, um caminho π_s de valor ótimo $V(s)$ é obtido em P , quando removemos seu último nó s de Q (linha 6). Empates podem ocorrer, sendo resolvidos pela política FIFO em Q . Nestes casos, quando dois caminhos ótimos encontram uma mesma amostra s , esta amostra é associada ao primeiro caminho que a encontrou. Na linha 7, $V(t) > V(s)$ é falso, quando s não poder modificar os atributos de t e que $V(t) \neq +\infty$ na linhas 10 e 11, apenas quando $t \in Q$. As demais linhas verificam se o caminho $\pi_s.(s, t)$ seja melhor que o caminho atual $\pi_t \in P$. Se for, atualizam Q , $V(t)$, $L(t)$ e $P(t)$.

2.4. Fase de Classificação

A classificação de uma nova amostra t , considera todas as conexões entre t e as amostras s de treinamento, como se t fizesse parte do grafo original. A partir de todos os possíveis caminhos entre o conjunto S^* até t , o objetivo na classificação está em encontrar o caminho ótimo $P^*(t)$, de S^* até t , com a classe $\lambda(R(t))$ de seu protótipo $R(t) \in S^*$, mais fortemente conexo. O caminho pode ser encontrado incrementalmente, avaliando o valor de custo ótimo $V(t)$, pela Equação 2 aplicado $\forall s \in Z_1$.

$$V(t) = \min\{\max\{V(s), d(s, t)\}\}. \quad (2)$$

Seja $s^* \in Z_1$, que satisfaz a Equação 2, (predecessor $P(t) = s^*$). Dado que $L(s^*) = \lambda(R(t))$, a classificação é realizada associando o rótulo de $L(s^*)$ como sendo a classe de t .

Algoritmo 1: Algoritmo OPF

Entrada: Conjunto Z_1 , protótipos $S^* \subset Z_1$ e par de valores (v, d) para extração de características e cálculos de distâncias.

Saída: Floresta de caminhos ótimos P , mapa de valores ótimos V , e mapa de rótulos L .

Dados: Fila Q de prioridades e variável tmp

```

1 para cada  $s \in Z / S^*$  faça
2    $V(s) \leftarrow +\infty$ ,
3 para cada  $s \in S^*$  faça
4    $V(s) \leftarrow 0$ ,  $P(s) \leftarrow nil$ ,  $L(s) \leftarrow \lambda(s)$ , e
   insira  $s$  em  $Q$ .
5 enquanto  $Q$  não estiver vazia faça
6   Remova de  $Q$  uma amostra  $s$  tal que  $V(s)$  seja
   mínima.
7   para cada  $t \in Z$  tal que  $t \neq s$  e  $V(t) > V(s)$ 
   faça
8     Calcule  $tmp \leftarrow \max\{V(s), d(s, t)\}$ .
9     se  $tmp < V(t)$  então
10      se  $V(t) \neq +\infty$  então
11        Remova  $t$  de  $Q$ .
12         $P(t) \leftarrow s$ 
13         $L(t) \leftarrow L(s)$ 
14         $V(t) \leftarrow tmp$ 
15        Insira  $t$  em  $Q$ .
16 retorna  $P, R, V$ 

```

3. Materiais e Métodos

Para que a imagem seja obtida, é necessário capturar uma amostra de mel. Essa amostra é capturada diretamente do pote de mel e transferida para uma placa. Esse processo é chamado de plaqueamento e é através dele que a amostra de mel pode ser analisada microscopicamente. Com a amostra inserida no microscópio, o especialista realiza capturas de imagens de diferentes posições. Assim que ele consegue capturar imagens de 200 grãos a obtenção de imagens é interrompida. A quantidade de amostras capturadas foi definida previamente pelo especialista. A Figura 2 ilustra um exemplo de imagem de mel visualizada por ele. É através de imagens deste tipo que a contagem de grãos de pólen é realizada.

Para a captura de imagens dos experimentos dois microscópios foram usados, um obtinha a iluminação sob a imagem e outro apresentava a iluminação sobre a imagem. Para facilitar a compreensão, o primeiro microscópio será chamado de microscópio A e o segundo de microscópio B. O microscópio A possui a seguinte configuração: microscópio tradicional Marca Bel Equipamentos analíticos Ltda, acoplado com o dispositivo de captura Marca Proscope HR com lente de 400X. Enquanto o microscópio B é

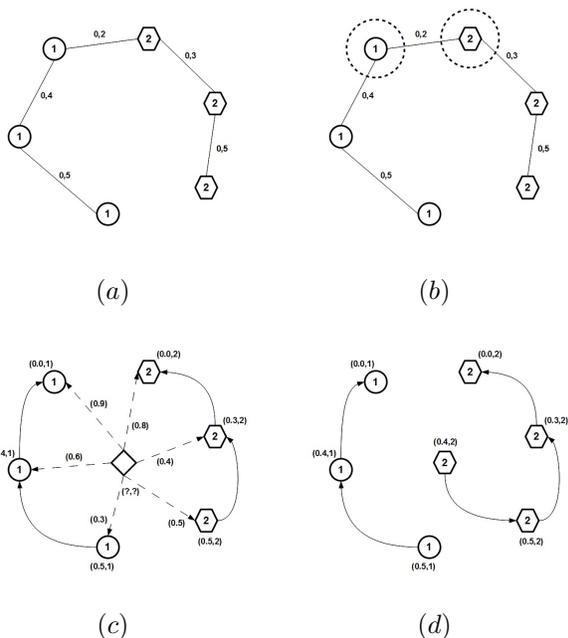


Figura 1. (a) Grafo MST referente ao grafo completo. (b) Protótipos selecionados a partir da heurística de elementos adjacentes de classes distintas na MST. (c) Amostra de teste e suas conexões em todos os elementos do grafo de treinamento. (d) Caminho ótimo do protótipo mais fortemente conexo.



Figura 2. Imagem obtida de amostras de mel contendo grãos de pólen de diferentes espécies.

um Monotela LCD marca LCD Micro Bresse com aumento de 40 a 1600 X.

Como os dois microscópios presentes no projeto possuem configurações distintas, as imagens resultantes de cada microscópio possui características específicas. Por

exemplo, as imagens capturadas pelo microscópio A possui resolução de 640 X 480 *pixels* enquanto as imagens capturadas pelo microscópio B possuem resolução de 1520 X 1120 *pixels* cada.

As imagens obtidas foram divididas em 4 conjuntos. O conjunto 1 apresenta imagens capturadas com o microscópio B que possuía um filtro de visualização. Esse filtro é similar a uma lente que não apresenta nenhum aumento, porém apresenta uma coloração. Essa coloração influencia na imagem a ser capturada e, por esse motivo, a imagem obtém a cor do filtro usado. Esse filtro é posicionado em cima da luz do microscópio, por esse motivo ele consegue refletir uniformemente a coloração em toda a amostra. Este conjunto possui imagens com filtro azul. Similar ao conjunto 1, o conjunto 2 foi obtido utilizando um filtro roxo.

O terceiro conjunto é formado por imagens capturadas através do microscópio B, essas imagens foram capturadas diretamente do grão de pólen. O quarto conjunto é representado por imagens obtidas através do microscópio A, essas imagens foram capturadas diretamente do mel. O conjunto 1 possuía 4 classes e os conjuntos 2, 3 e 4 possuíam 6 classes.

Para realizar experimentos, os algoritmos de extração de atributos foram aplicados e atributos das classes foram obtidas. Com essas informações, os algoritmos de classificação foram treinados e testados. Os experimentos realizados utilizaram a validação cruzada com 10 dobras presente no WEKA¹.

Com o resultado da classificação, o teste de Friedman foi aplicado para verificar se há diferença entre os classificadores. Para a utilização do teste de Friedman, um programa chamado R foi utilizado [8]. Um valor, chamado de *p* é calculado e, com base no resultado que ele apresenta, a hipótese nula pode ser rejeitada ou não. Se o valor de *p* for menor que 5%, valor de significância utilizado nos experimentos, a hipótese nula é rejeitada, ou seja, pode-se afirmar que os classificadores são diferentes, caso contrário a hipótese nula é aceita e pode-se dizer que os classificadores são iguais. Se a hipótese nula foi aceita, o teste de pos-Hoc será aplicado para verificar qual classificador apresenta melhor resultado. Também foram avaliadas informações de taxa de acerto, *recall*, *precision* e *f-score*.

O conjunto de atributos extraídos foram a média dos canais H, S, B, R, G e B e atributos de forma (todos os atributos de forma foram utilizados), matriz de coocorrência (entropia, diferença inversa, correlação, dissimilaridade, momento da diferença inverso, contraste e uniformidade) e mapas de iteração (entropia, diferença inversa, correlação, dissimilaridade, momento da diferença inverso, contraste e uniformidade). Os valores dos parâmetros utilizados nos ex-

¹ www.cs.waikato.ac.nz/ml/weka

perimentos para o algoritmo matriz de coocorrência foi 1 *pixel* de distância e ângulos de 180°, 135°, 90° e 45°. Para mapas de interação foram usados os ângulos 180°, 135°, 90° e 45° para 2, 1 e 0 *pixels* de distância, respectivamente.

A fim de se obter os atributos mais relevantes, foi aplicado um algoritmo de seleção de atributos, denominado *best-first*. O algoritmo *best-first* foi escolhido por apresentar os resultados de forma rápida e por resultar atributos que apresentaram, para esses experimentos, melhorias na taxa de acerto do classificador. Após a seleção de atributos foi obtido um conjunto com 9 atributos, sendo eles: a média dos canais H, S, B, G, B, *form factor*, dissimilaridade 180, entropia 135 e dissimilaridade 45). Os valores presentes após os atributos de textura referem-se ao ângulo de cada atributo.

Os experimentos visam verificar quais conjuntos de atributos são mais relevantes para a classificação de grãos de pólen e qual é o resultado do OPF em relação à classificação e tempo de treinamento quando comparado aos demais classificadores selecionados.

4. Resultados e Análises

Nesta seção serão descritos os experimentos realizados e análises dos resultados. Para cada experimento as métricas *recall*, *precision*, *f-score* e taxa de acerto foram analisadas e os algoritmos de teste de hipótese foram aplicados.

O WEKA calcula o teste T-STUDENTS, por isso quando o resultado de um experimento é obtido, um círculo pode estar presente ao lado do resultado de algum classificador. Este círculo revela que o classificador obteve performance melhor ou pior que o classificador posicionado na primeira coluna das Tabelas de resultado. Quando o círculo estiver vazio significa que o classificador obteve melhor resultado do que o primeiro algoritmo.

No primeiro experimento, foi possível concluir que os outros classificadores apresentaram resultados mais satisfatórios quando comparados ao OPF. Por esse motivo, foram inseridos atributos de textura e um segundo experimento foi realizado. Com esse experimento foi possível observar que o OPF apresentou resultados tão satisfatórios quanto os outros classificadores. Usando os parâmetros obtidos com o segundo experimento, o terceiro experimento foi realizado com os quatro conjuntos de imagens presentes no projeto.

Os resultados apresentados na Tabela 1, são referentes ao experimento 3. Esses resultados foram analisados através do teste de Friedman para verificar se há ou não distinção entre os classificadores. Foi considerado 5% de significância. Ao submeter o resultado do experimento presente na Tabela 1 ao teste de Friedman o resultado de p obtido foi igual a 5.7%. De acordo com o resultado obtido a hipótese nula pode ser aceita. Porém quando o resultado do tempo de trei-

Dataset	(OPF)	(SVM)	(IBK)	(C4.5)
Conjunto1	74.43	67.69	87.00 ○	71.45
Conjunto2	45.83	67.37 ○	73.13 ○	75.27 ○
Conjunto3	74.85	80.68	87.03 ○	81.95
Conjunto4	43.35	47.35	52.00	59.15

○, ● statistically significant improvement or degradation

Tabela 1. Porcentagem de acerto de todas as bases de aprendizagem e todos os classificadores

Dataset	(OPF)	(SVM)	(IBK)	(C4.5)
Conjunto1	0.78	0.78	0.70	0.70
Conjunto2	0.54	0.90 ○	0.64	0.68
Conjunto3	0.84	0.85	0.87	0.89
Conjunto4	0.59	1.00 ○	0.63	0.66

○, ● statistically significant improvement or degradation

Tabela 2. Resultado do recall de cada classificador

namento dos classificadores foi analisado, obteve-se um valor de p igual a 0.7%. A Tabela 5 apresenta o tempo utilizado no treinamento de cada classificador. De acordo com o resultado a hipótese nula pode ser rejeitada, porém precisa saber qual é o classificador que apresenta maior tempo de treinamento. Para isso, o teste de pos-Hoc foi aplicado. De acordo com o resultado pode-se afirmar que o SVM apresenta maior tempo de treinamento quando comparado aos outros classificadores.

Ao analisar, através do teste de Friedman o *recall*, *precision* e o *f-score* das Tabelas 4, 2 e 3 respectivamente, pode-se confirmar o resultado obtido através da taxa de acerto dos classificadores. De acordo com o experimentos, analisando o *recall*, o valor de p foi igual a 20.54%. Para o *precision*, o valor de p foi igual a 55.20% e para o *f-score* o valor obtido foi de 99.42%. Analisando os dados supracitado, os 3 valores obtiveram resultado superior à 5%, por isso também pode-se rejeitar a hipótese nula para os 3 itens supracitados. Analisando apenas o resultado do teste T-Students para a métrica da taxa de acerto, pode-se perceber que o IBK apresentou melhores resultados quando comparado ao OPF

Dataset	(OPF)	(SVM)	(IBK)	(C4.5)
Conjunto1	0.74	0.55	0.64	0.57
Conjunto2	0.52	0.57	0.68	0.75
Conjunto3	0.93	0.87	0.90	0.86
Conjunto4	0.50	0.45	0.44	0.57

○, ● statistically significant improvement or degradation

Tabela 3. Resultado do precision de cada classificador

Dataset	(OPF)	(SVM)	(IBK)	(C4.5)
Conjunto1	0.73	0.62	0.64	0.60
Conjunto2	0.50	0.67	0.63	0.67
Conjunto3	0.86	0.84	0.87	0.86
Conjunto4	0.51	0.60	0.48	0.58

◦, • statistically significant improvement or degradation

Tabela 4. Resultado do *f-score* de cada classificador

Dataset	(OPF)	(SVM)	(IBK)	(C4.5)
Conjunto1	0.002980	0.172150	0.000000	0.001090
Conjunto2	0.002030	0.109370	0.000000	0.002030
Conjunto3	0.013280	0.104380	0.000000	0.003600
Conjunto4	0.000460	0.146240	0.000000	0.001080

◦, • statistically significant improvement or degradation

Tabela 5. Tempo utilizado para realizar o treinamento dos classificadores

em 3 conjuntos.

5. Conclusões

Três experimentos foram realizados e quando comparados, percebeu-se que houve uma melhora nos resultados das métricas (taxa de acerto, *recall*, *precision* e *f-score*) com a inserção de atributos de textura. Para classificar os grãos de pólen, técnicas de extração de atributos foram aplicadas e teste de classificação foram realizados, utilizando os classificadores OPF, SVM, KNN e C4.5. O resultado foi analisado através do teste de Friedman, utilizando 5% de significância. Com o valor de *p* obtido pelo experimento a hipótese nula pode ser aceita, pois o resultado do *p* foi igual a 5.7%. Desse modo, pode-se afirmar que, de acordo com o teste de Friedman, os classificadores não apresentam diferença estatística na classificação dos grãos.

Como trabalhos futuros, testes usando diferentes extractores de atributos serão realizados com o intuito de encontrar um conjunto de informações capaz de distinguir as classes existentes no projeto. Novas imagens serão obtidas com o intuito de enriquecer o banco de imagens, e a geração de novos experimentos serão realizados, utilizando novos algoritmos de seleção de atributos e novos algoritmos de aprendizagem e classificação de dados.

Agradecimentos

Este trabalho recebeu apoio financeiro da Universidade Católica Dom Bosco, UCDB, da Fundação de Apoio ao Desenvolvimento do Ensino, Ciência e Tecnologia do Estado

de Mato Grosso do Sul, FUNDECT e do Conselho Nacional de Desenvolvimento Científico e Tecnológico, CNPq.

Referências

- [1] O. M. Barth. *O Pólen no mel Brasileiro*. 1989.
- [2] A. X. Falcao, J. Stolfi, and de Alencar. The image foresting transform: theory, algorithms, and applications. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(1):19–29, 2004.
- [3] L. C. S. K. S. VENTURINI, M. F. SARCINELLI. Características do mel. Pró-Reitoria de Extensão - Programa Institucional de Extensão. Boletim Técnico - PIE-UFES: 01107 - Editado: 18.08.2007.
- [4] P. e. A. Ministério da Agricultura. Instrução normativa número 11, de 20 de outubro de 2000 e regulamento técnico de identidade e qualidade do mel. Diário Oficial da União de 23/10/2000, Seção 1, Página 23, Fevereiro 2010.
- [5] J. A. Montoya-Zegarra, J. P. Papa, N. J. Leite, R. S. Torres, and A. X. Falcão. Learning how to extract rotation-invariant and scale-invariant features from texture images. *In Journal on Advances in Signal Processing, volume 2008*, pages 1–16, 2008.
- [6] J. A. Montoya-Zegarra, J. P. Papa, N. J. Leite, R. S. Torres, and F. A. X. Rotation invariant texture recognition. *In 3rd International Symposium on Visual Computing, volume Part II, LNCS 4842, Lake Tahoe, Nevada, CA, USA*, pages 193–204, 2007.
- [7] J. P. Papa. *Classificação Supervisionada de Padrões Utilizando Florestas de Caminhos Ótimos*. PhD thesis, UNICAMP, novembro 2008.
- [8] L. A. Peternelli and M. P. Mello. *Conhecendo o R: uma visão estatística*. Série Didática. Editora UFV, 1 edition, March 2011.
- [9] T. Vargas. Avaliação da qualidade do mel produzido na região dos campos gerais do paran. Master's thesis, Universidade Estadual de Ponta Grossa, 2006. Mestrado em cincia e tecnologia de alimentos.