

MINERAÇÃO DE DADOS ESPECTRAIS PARA MODELAGEM DE OCORRÊNCIA DA BROCA DO CAFÉ¹

Margarete M. L. Volpato²; Helena Maria R. Alves³; Tatiana G. C. Vieira⁴; Victor G.Boell⁵; Júlio Cesar de Souza⁶; Miler G. Boell⁷; Wilian L. Soares⁸

¹ Trabalho financiado pelo Consórcio Brasileiro de Pesquisa e Desenvolvimento do Café – Consórcio Pesquisa Café

² Pesquisadora, Doutora, EPAMIG/URESM, Lavras-MG, BIPDT/ FAPEMIG, margarete@epamig.ufla.br

³ Pesquisadora, PhD, Embrapa Café, Brasília-DF, helena.alves@embrapa.br

⁴ Pesquisadora, Bolsista Consórcio Pesquisa Café, tatianagevieira@gmail.com

⁵ Graduando em Ciência da Computação/UFLA, BIC FAPEMIG/EPAMIG, grudtnerv@gmail.com

⁶ Pesquisador, Doutor, EPAMIG/URESM, Lavras-MG, BIPDT/ FAPEMIG, jcsouza@navinet.com.br

⁷ Bolsista Consórcio Pesquisa Café, DCI-2, milergrudtner@gmail.com

⁸ Professor, Dep. Ciência da Computação/UFLA, Lavras-MG, lacerda@dcc.ufla.br

RESUMO: O monitoramento fitossanitário possibilita prever o aparecimento ou aumento da ocorrência da Broca do Café (*Hypothenemus hampei*). Tradicionalmente esse monitoramento é baseado em observação de períodos críticos ocorridos. Entretanto uma das maiores dificuldades para se utilizar esse tipo de monitoramento é a aquisição de dados climáticos. Uma alternativa para superar este problema é utilizar dados e produtos de imagens de satélites, em função da cobertura espacial e temporal, e de sua relação com as variações do clima e da vegetação de uma região. Uma das dificuldades para realização desse estudo é o grande número de dados gerados, por isso optou-se pela metodologia de mineração de dados, etapa principal do processo de Descoberta de Conhecimento em Bancos de Dados (*Knowledge Discovery in Databases* - KDD). O presente estudo objetivou aplicar técnicas de mineração de dados para encontrar modelos de dados climáticos e espectrais associados à ocorrência *H. hampei* em frutos de café. As coletas de dados de campo foram realizadas na fazenda experimental da EPAMIG, em São Sebastião de Paraíso, MG, e os dados espectrais foram adquiridos pelo sensor MODIS do satélite Terra. Os modelos gerados mostraram que a temperatura média foi o atributo de maior separabilidade na totalidade dos dados climático estudados com taxa de acerto de 67% e 73%.

PALAVRAS-CHAVE: Sensoriamento remoto, *Hypothenemus hampei*, café, agrometeorologia.

DATA MINING SPECTRAL TO MODELING OF OCCURRENCE OF COFFEE BERRY BORER

ABSTRACT: The phytosanitary monitoring enables to predict the onset or increase of intensity of *Hypothenemus hampei*. Traditionally this monitoring is based on the observation of critical periods of occurred. However one of the major difficulties in using this type of monitoring is to acquire weather data. An alternative to overcome this problem is to use data and satellite imagery products, depending on the spatial and temporal coverage, and its relationship to changes in climate and vegetation of an area. One of the difficulties in carrying out this study is the large number of generated data, so we opted for the methodology of data mining, the main stage of the process of Knowledge Discovery in Databases (KDD). This study aims to apply data mining techniques to find models of climate data and spectral associated with the occurrence of *H. hampei* on coffee. The collected field data were conducted at the experimental farm of EPAMIG in São Sebastião do Paraíso, MG, and the spectral data were acquired by the MODIS sensor Terra satellite. The models showed that the average temperature was higher separability with hit rate of 67% and 73%.

KEYWORDS: Remote sensing, *Hypothenemus hampei*, coffee, agrometeorology.

INTRODUÇÃO

O estado de Minas Gerais é responsável por aproximadamente metade da produção nacional de café. Por causa de sua importância econômica, é necessário realizar o monitoramento dessas áreas cafeeiras visando o controle de pragas e doenças que afetam a produção. Uma das principais pragas que afeta a lavoura é a broca do café (*Hypothenemus hampei*). Este inseto ataca os frutos em todos os estágios de crescimento (SOUZA e REIS, 1997). Monitorar as lavouras permite acompanhar seu desenvolvimento e as pragas que as atacam. Permite também a descoberta dos principais fatores meteorológicos relacionados ao ciclo de vida do inseto, tais como: umidade relativa do ar, temperatura do ar e precipitação (SOUZA e REIS, 1997). O projeto desenvolvido tem como principal objeto monitorar a broca do café. Para tanto, foram usados dados referentes a infestação da broca na amostras coletadas. Devido à dificuldade de se conseguir dados climáticos de boa qualidade, utilizaram-se também dados de séries temporais de índices de vegetação espectrais, oriundos de satélites orbitais, que surgem como alternativa, uma vez que apresentam correlação com o vigor vegetativo dos dosséis das vegetações e seus status hídrico. Para o desenvolvimento de modelos de ocorrência da

broca do café, optou-se pelo processo de classificação através do uso de mineração de dados, que combina análises estatísticas em grandes quantidades de dados, obtendo assim novas informações ou padrões (WITTEN et al., 2011). Portanto o objetivo deste estudo foi realizar a modelagem dos dados no sistema computacional de mineração de dados, de forma a alcançar melhora da acurácia dos resultados e do modelo visível através de árvore de decisão. Os dados climáticos e de ocorrência da broca do café foram coletados na fazenda experimental da EPAMIG, na região de São Sebastião do Paraíso, MG. A modelagem consistiu na quantidade de dados usados no processo de mineração de dados. Os resultados obtidos foram avaliados por especialistas para averiguar a conformidade dos mesmos com o comportamento real da praga.

MATERIAL E MÉTODOS

Os dados utilizados para mineração são referentes à ocorrência da broca do café, e contém informações sobre a umidade relativa do ar, temperatura do ar, precipitação, mês do ano e EVI2. O EVI2 (Enhanced Vegetation Index) é um índice usado para estimar o vigor vegetativo do dossel das plantas sendo oriundo do sensor MODIS ((MODerate resolution Imaging Spectroradiometer) do satélite Terra, disponibilizado pelo INPE (Instituto Nacional de Pesquisas Espaciais - <http://www.dsr.inpe.br/laf/series>). A preparação dos dados consistiu na discretização, em que os mesmos são ajustados em intervalos de valores, permitindo a sua classificação pela árvore de indução, além de sua disposição em uma tabela para melhor visualização (Tabela 1). Foi realizada a avaliação e organização dos dados de índices de vegetação (EVI2) para inserção no Banco de Dados Geográfico (BDG) e em seguida a limpeza dos dados. Valores identificados como erros foram removidos da base de dados. Os dados da ocorrência da Broca do Café, climáticos e espectrais foram organizados e padronizados quanto a tamanho, formato e domínio, em seguida foi iniciada a etapa de pré-processamento dos dados espectrais EVI2, com a discretização da série temporal de dados EVI2. Esta discretização foi realizada segundo as seguintes classes: EVI > 0,5 (dossel vigoroso); EVI 0,5 - 0,4 (dossel com vigor vegetativo intermediário); EVI 0,4 - 0,3 (dossel com vigor vegetativo baixo); EVI < 0,3 (dossel com vigor vegetativo muito baixo). Após a discretização os dados foram encaminhados para a fase de mineração de dados. Para mineração de dados foi usado o software WEKA que contém um conjunto de algoritmos implementados para estudos nas subáreas de inteligência artificial e aprendizagem de máquina. Dentre os algoritmos, escolheu-se o J48, que serve para indução de modelos de classificação, mais conhecidos por árvores de decisão, tendo sua elaboração baseada em sistemas de inferência e em conceitos de sistemas de aprendizagem. Ele constrói árvores de decisão a partir de um dado conjunto de exemplos, sendo a árvore resultante usada para classificar amostras (MITCHELL, 1997).

Tabela 1. Exemplo do Banco de Dados discretizado, prontos para mineração, onde: DataMes é o número correspondente ao mês de coleta, PrecipitaçãoSoma é dado em milímetros mensais, EVI é o Enhanced Vegetation Index, TempMedia é a temperatura média em graus Celsius, Umidade é a média mensal dada em %, infestação é a ocorrência da Broca.

DataMes	PrecipitaçãoSoma	TempMedia	Umidade	EVI	infestação
1	>100	>22	<80	>=0.5	BAIXA
2	>100	>22	>80	0.4-0.5	BAIXA
3	>100	>22	>80	0.4-0.5	BAIXA
4	>100	>22	<80	0.4-0.5	BAIXA
5	<50	18-22	<80	0.4-0.5	BAIXA
6	<50	18-22	<80	0.3-0.4	BAIXA
7	<50	18-22	<80	<0.3	BAIXA
8	<50	<18	<80	<0.3	BAIXA
9	<50	18-22	<80	0.3-0.4	BAIXA
10	>100	18-22	<80	0.3-0.4	BAIXA
11	<50	>22	<80	0.4-0.5	BAIXA
12	>100	>22	<80	0.4-0.5	BAIXA
1	>100	>22	<80	0.4-0.5	BAIXA
2	>100	>22	<80	0.4-0.5	BAIXA
3	50-100	>22	<80	>=0.5	BAIXA
4	>100	>22	<80	0.4-0.5	BAIXA

RESULTADOS E DISCUSSÃO

Dentre os modelos gerados dois foram escolhidos por demonstrarem melhores resultados. A leitura da árvore faz-se a partir do topo da árvore (nó raiz) em direção aos nós folhas, passando pelos seus ramos, de acordo com os testes nos valores dos atributos. Os nós folhas representam o objeto de estudo, ou seja, a ocorrência da broca do café. Cada ramo na árvore é uma conjunção de condições. Assim, o percurso na árvore (da raiz à folha) corresponde a uma regra de classificação (WITTEN et al., 2011). O modelo da Figura 1 foi gerado com os dados completos, em que a principal variável do modelo foi o mês de coleta, observa-se que os meses de julho a dezembro, foram os de maior infestação da praga. A segunda variável foi a temperatura do ar, entre 18 e 22°C, e a terceira variável foi a precipitação, menor que 50 mm. A taxa de acertos na classificação das amostras foi de 73%. No segundo modelo gerado (Figura 2), a principal variável de influência foi a temperatura do ar, entre 18 e 22°C, diretamente ligado ao tempo de ciclo de vida do inseto (fase larval a adulta), seguido do EVI2, entre 0,4 e 0,5 (dossel com vigor vegetativo intermediário) e a precipitação, menor que 50 mm. A taxa de acertos na classificação das amostras foi de 67%. Apesar da semelhança entre os modelos o acerto é consideravelmente maior, quando é avaliada a existência de frutos nos cafeeiros.

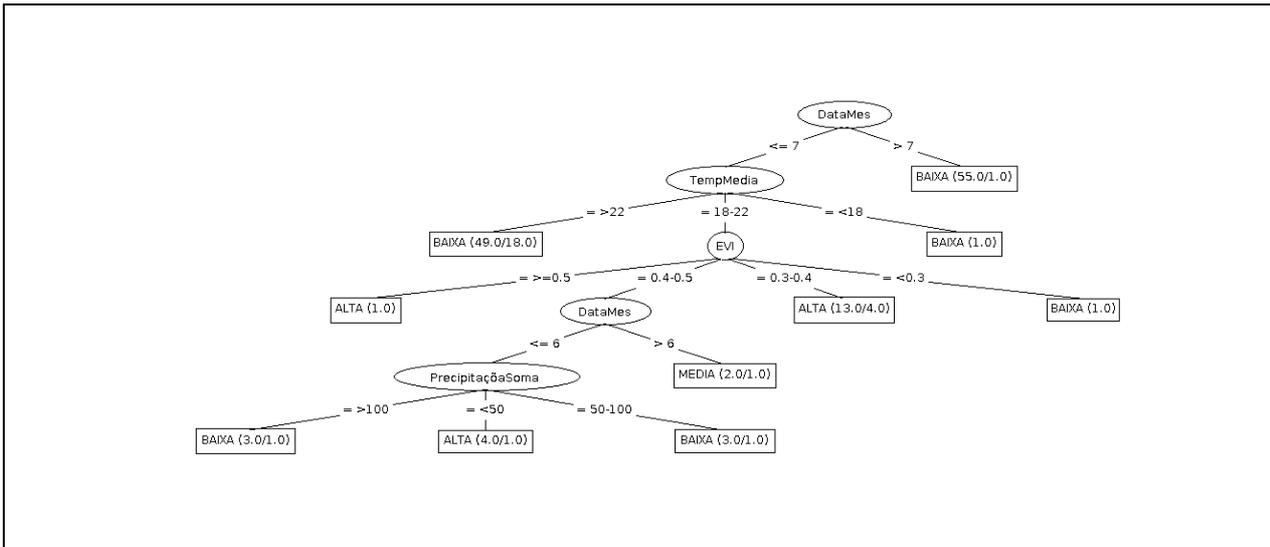


Figura 1. Árvore de decisão gerada a partir dos dados coletados, utilizando os meses da coleta.

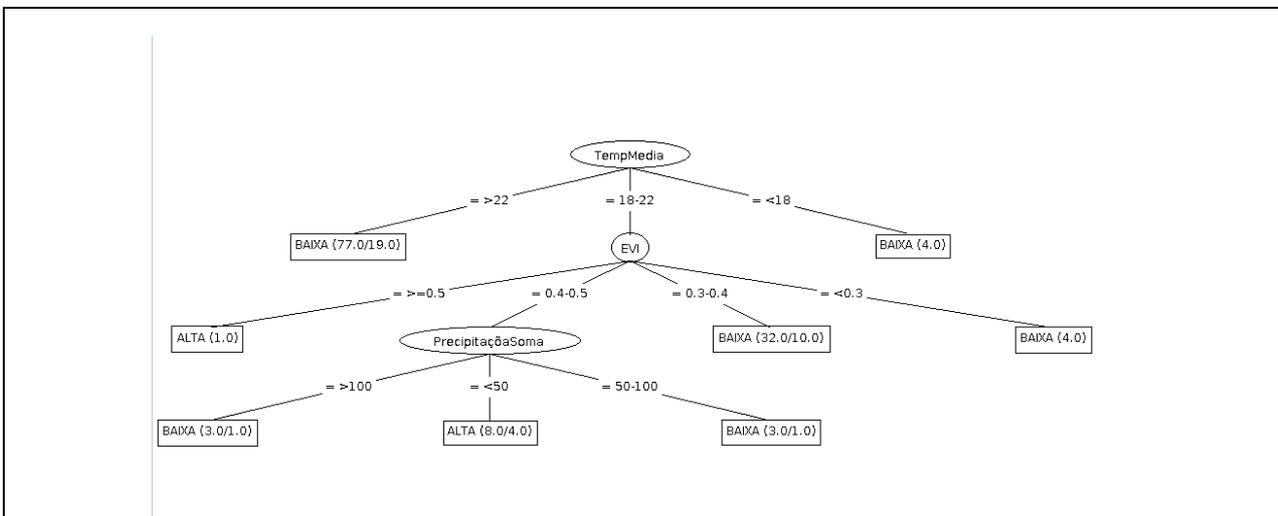


Figura 2. Árvore de decisão gerada a partir dos dados coletados, sem utilizar informações sobre os meses de coleta.

CONCLUSÕES

A classificação baseada em indução de árvores de decisão no desenvolvimento de modelos para explicar a ocorrência da Broca do Café, no sul de Minas Gerais, utilizando dados espectrais do satélite MODIS e dados climáticos, mostrou-se satisfatório, com taxa de acerto de 73%.

AGRADECIMENTOS

Ao Consórcio Pesquisa Café (CBP&D Café) pelo financiamento do projeto e bolsas concedidas e à Fundação de Amparo à Pesquisa do Estado de Minas Gerais (Fapemig), pelas bolsas concedidas.

REFERÊNCIAS BIBLIOGRÁFICAS

- MITCHELL, T. **Machine Learning**. Aprendizado de máquina: árvore de decisão indutiva. Disponível em: <http://www.iceb.ufop.br/decom/prof/guarda/BCC740/ArvoreDecisaoIndutiva.pdf>. Acesso em 12 de junho 2013.
- SOUZA, J.C.; REIS, P. R. **Broca-do-café**: histórico, reconhecimento, biologia, prejuízos, monitoramento e controle. 2.ed. Rev. Aum – Belo Horizonte: Epamig, 1997.
- WITTEN I. H.; FRANK, E.; HALL, M. A. **Data Mining: Practical Machine Learning Tools and Techniques**, 3ed. Elsevier: Burlington, 2011, 665p.
- INPE. **Visualização das séries temporais MODIS para análise de mudanças de uso e cobertura da terra**. Disponível em: www.dsr.inpe.br/laf/series. Acesso em: junho de 2011.