

MINERAÇÃO DE DADOS CLIMÁTICOS PARA PREVISÃO DE GEADA E DEFICIÊNCIA HÍDRICA PARA AS CULTURAS DO CAFÉ E CANA-DE-AÇÚCAR PARA O ESTADO DE SÃO PAULO

Luciana Corpas Bucene¹
Luiz Henrique Antunes Rodrigues²
Carlos Alberto Alves Meira³

RESUMO

As perdas que ocorrem na agricultura são grandes, devido, principalmente, à ocorrência de sinistros às plantações por eventos climáticos. Como exemplo, a produção de café e a de cana-de-açúcar no Estado de São Paulo sofrem alternâncias motivadas por eventos climáticos adversos e, em especial, as geadas e as secas, que reduzem drasticamente as produções.

Este estudo propõe identificar relações entre parâmetros climáticos, como temperatura máxima, temperatura mínima, precipitação, radiação, etc., visando descobrir eventuais novos conhecimentos, possibilitando a previsão de geada para a cultura de café e a previsão de *déficit* hídrico para as culturas de cana-de-açúcar e café no Estado de São Paulo, com maior grau de confiança e num intervalo de tempo satisfatório, com o objetivo de auxiliar os produtores na tomada de decisões. Para isso, serão aplicadas técnicas de mineração de dados nos grandes bancos de dados climáticos.

Mineração de dados (“data mining”) pode ser definida como o processo de extração de informação nova, não trivial e útil de repositórios de dados, ou seja, descobrir, automaticamente, conhecimento interessante, mas escondido, em grandes bancos de dados – usar os computadores para descobrir novas informações úteis. É uma tecnologia para explorar grandes bases de dados e descobrir relações e padrões existentes nas informações.

Este trabalho é um projeto de Doutorado da FEAGRI/UNICAMP que está vinculado ao projeto "Desenvolvimento e Evolução de um Sistema de Monitoramento Agroclimatológico para o Estado de São Paulo", desenvolvido pela EMBRAPA em conjunto com o Instituto Agrônomo de Campinas (IAC), ao qual se pretende adicionar novos produtos para incorporação no sistema de monitoramento, além de propor uma nova infra-estrutura para o sistema já existente e evoluções nos modelos operacionais disponíveis.

PALAVRAS-CHAVE: Inteligência artificial; geada; deficiência hídrica; café; cana-de-açúcar.

DATA MINING CLIMATIC FOR FORECAST OF FROST AND DEFICIT HIDRIC FOR COFFEE AND CANE SUGAR FOR STATE SÃO PAULO

ABSTRACT

The losses that occur in agriculture are great, because, mainly, to the occurrence of accidents to the plantations for climatic events. For example, the production of coffee and of cane sugar in the State of São Paulo suffers alternations motivated for adverse climatic events e, in special, the frosts and the droughts, that reduce the productions.

This study it considers to identify relations between climatic parameters, as maximum temperature,

¹ Eng. Agrícola, Mestre em Geoprocessamento, doutoranda da Faculdade de Engenharia Agrícola / UNICAMP, Pesquisadora colaboradora da Embrapa Informática Agropecuária. (email: lucianac@cnptia.embrapa.br)

² Eng. Agrícola, Prof. Dr. Faculdade de Engenharia Agrícola/UNICAMP. (email: lique@agr.unicamp.br)

³ Mestre em Ciências da Computação. Embrapa Informática Agropecuária. Caixa Postal 6041 – CEP 13083-970 – Campinas, SP.(email: carlos@cnptia.embrapa.br)

minimum temperature, precipitation, radiation, etc., aiming at to discover eventual new knowledge, making possible the forecast of frost for the coffee culture and the forecast of hídric deficit for the cane sugar cultures and coffee in the State of São Paulo, with bigger reliable degree and in an interval of satisfactory time, with the objective of assisting the producers in the taking of decisions. For this, techniques of data mining in the great climatic data bases will be applied.

Data mining can be defined as the process of extration of new, not trivial and useful information of repositories of data, or either, to discover, automatically, interesting, but hidden knowledge, in great data bases to use the computers to discover new useful information. It is a technology to explore great databases and to discover existing relations and standards in these information.

This work is a project of PhD of the FEAGRI/UNICAMP that is tied with the project "Development and Evolution of a System of Monitoring Agroclimatológico for the State of São Paulo", developed for the EMBRAPA in set with the Institute Agrônômico de Campinas (IAC), to which if it intends to add new products for incorporation in the monitoring system, besides considering a new infrastructure for the existing system already and evolutions in the available operational models.

KEYWORDS: Artificial intelligence, frost, deficit hidric, coffee, cane sugar.

1. INTRODUÇÃO

A Descoberta de Conhecimento em Bancos de Dados é um processo não trivial de identificar padrões válidos, não conhecidos, potencialmente úteis e interpretáveis, consistindo, basicamente, em descobrir conhecimento nos dados armazenados, a partir da aplicação de técnicas de mineração de dados, da aplicação dos padrões obtidos e da interpretação dos resultados (Fayyad et al., 1996).

Fayyad et al. (1996) mostram um processo típico de mineração de dados. Ele possui três passos, o pré-processamento, a mineração de dados e o pós-processamento.

Segundo Baranauskas & Monard (2000), no início do desenvolvimento, há a necessidade de preparação dos dados, fase considerada na literatura como a que consome mais tempo. O primeiro passo é pré-processar os dados para aprontá-los para a análise. Usualmente os dados têm que ser formatados, amostrados, adaptados e, algumas vezes, transformados para que possam ser usados pelo algoritmo de mineração. Ocorre então, o desenvolvimento do entendimento do domínio da aplicação, avaliação do hardware e software disponíveis, seleção, limpeza e transformação dos dados.

Após, os dados estão prontos para serem minerados por um algoritmo. É a fase conhecida como a própria mineração de dados. É definida a escolha da tarefa e das técnicas a serem utilizadas, identificação da ferramenta que satisfaça a essas condições e aplicação das ferramentas nos dados. Segundo Mendonça Neto (2001), as tarefas de mineração podem ser classificadas em predição, classificação, descoberta de associações, visualização de dados, entre outras. As técnicas para a execução dessas tarefas são variadas, entre elas (Mendonça Neto, 2001): árvores de decisão, regras de indução, redes neurais artificiais, mineração visual de dados e clusterização.

O último passo do processo de mineração de dados é assimilar a informação minerada, chamado pós-processamento. É a interpretação dos resultados e incorporação do conhecimento adquirido.

A qualidade do conhecimento descoberto no final é dependente da qualidade do dado, do pré-processamento, do algoritmo de mineração e do processo de assimilação. Mais do que isso, a qualidade é altamente dependente de como o processo é montado como um todo.

2. MATERIAL E MÉTODOS

O trabalho está sendo desenvolvido na Embrapa Informática Agropecuária, juntamente com a equipe do projeto "Desenvolvimento e Evolução de um Sistema de Monitoramento Agroclimatológico para o Estado de São Paulo" (EMBRAPA, 2001).

Os dados em estudo compreendem dados climáticos históricos do Estado de São Paulo, como temperatura máxima, temperatura mínima, precipitação diária, radiação e balanço hídrico, para um longo período de tempo, disponibilizados pelo Instituto Agrônomo de Campinas (IAC) e CEPAGRI. Contém dados de 136 estações climatológicas, coletados num total de 12 anos, no período de 1991 a 2002.

As seguintes etapas do processo de mineração de dados estão sendo aplicadas no conjunto dos dados presentes.

- a) pré-processamento. Os dados estão sendo preparados para que, futuramente, possam ser analisados pelo algoritmo de mineração de dados. Está sendo feito então, a seleção, limpeza e transformação dos dados. A partir dos dados de temperatura máxima, temperatura mínima, precipitação, evapotranspiração, disponibilidade de água no solo e radiação solar incidente e refletida, obteve-se outros atributos, como amplitude de temperatura, precipitação acumulada, dias acumulados com precipitação, dias acumulados sem precipitação, temperatura mínima de 1, 2, 3, 4 e 5 dias anteriores, comportamento da temperatura mínima, até 5 dias anteriores, amplitude de temperatura mínima de 1, 2, 3, 4 e 5 dias anteriores, temperatura máxima de 1, 2, 3, 4 e 5 dias anteriores, comportamento da temperatura máxima, até 5 dias anteriores, amplitude de temperatura máxima de 1, 2, 3, 4 e 5 dias anteriores, ocorrência de geada para temperatura mínima menor ou igual à -4, -3, -2, -1, 0, 1, 2, 2.5, 3 e 4 °C, radiação incidente total diária, radiação refletida total diária e ocorrência de geada e de deficiência hídrica.

Todos os dados foram preparados em planilhas eletrônicas no software Excel.

- b) mineração de dados. No presente caso, utilizam-se as técnicas de mineração visual de dados e indução de regras, já que o intuito é classificar o banco de dados em ocorrência ou não de geada e de deficiência hídrica.

A mineração visual de dados é uma técnica que mapea volumes de dados para a tela de um computador. Visualização é uma ferramenta importante para mineração de dados porque seres humanos são muito bons em processar informação visual e muito ruins em processar informação numérica e/ou tabular. Essa técnica foi trabalhada no software Eureka 1.2, podendo-se mapear o banco de dados envolvendo múltiplas variáveis simultaneamente, o que permitiu a exploração inteligente dos dados.

A indução de regras é o processo de se obter uma hipótese a partir dos dados e fatos já existentes. Objetos semelhantes são agrupados em classes e regras formuladas por meio das quais é possível prever a classe de objetos não vistos. Este processo de classificação identifica grupos nos quais cada qual tem um padrão único de valores que constitui a descrição da classe. A natureza do ambiente é dinâmica e por isso o modelo deve ser adaptável, isto é, deve ser capaz de aprender (UNESP, 1998). O software utilizado para indução de regras foi o software CBA 2.0.

- c) pós-processamento. Como o trabalho está na fase de mineração de dados, esta etapa de pós-processamento encontra-se não desenvolvida, pois os resultados adquiridos ainda não foram analisados. A partir de análises e avaliações das regras geradas poderão surgir novos conhecimentos através de evidências não detectadas anteriormente, indicando-se a necessidade de estudos aprofundados acerca dessas relações.

Os conhecimentos adquiridos com a utilização de técnicas de mineração de dados deverão ser verificados junto a especialistas humanos na área climatológica, ao longo de todo o desenvolvimento do projeto. A validação do modelo gerado deve ser realizada através da consulta a outros especialistas que não tenham participado do seu desenvolvimento. Devem ser apresentados aos especialistas diversos cenários e comparadas as suas previsões com as obtidas com o modelo gerado.

3. RESULTADOS E DISCUSSÃO

O atual estudo ainda não se encontra em fase final. O projeto apresenta-se na fase de pré-processamento dando início a fase de mineração de dados, utilizando-se um subconjunto de dados da estação climatológica de Campinas.

Dentre os resultados esperados, no fim desse projeto, pretende-se, a partir da mineração de dados, identificar novos conhecimentos entre os parâmetros climáticos (temperatura máxima, temperatura mínima, precipitação diária, entre outros), permitindo a previsão de geada para a cultura do café e previsão de déficit hídrico para as culturas do café e da cana-de-açúcar, visando a prevenção contra déficit hídrico e geada para o Estado de São Paulo.

4. CONCLUSÕES

Como qualquer iniciativa de mineração de dados, este trabalho parte da hipótese de que é possível descobrir conhecimento novo "escondido" no grande volume de dados climáticos e ainda, a partir do comportamento conhecido dos atributos climáticos, aumentar as chances de se descobrir padrões que podem explicar e ajudar a prever o comportamento futuro dos mesmos. Estas hipóteses são reforçadas pela percepção dos especialistas em climatologia que possuem um claro sentimento de que é possível extrair informação nova e útil e prever o comportamento futuro dos parâmetros climáticos. Para isso, as técnicas de mineração de dados que serão aplicadas nos grandes bancos de dados climáticos do Estado de São Paulo, possibilitarão a previsão de geada para cultura do café e a previsão de deficiência hídrica no solo tanto para as culturas de café como cana-de-açúcar no Estado de São Paulo, com maior grau de confiança e em um intervalo de tempo satisfatório, podendo, então, auxiliar os produtores na tomada de decisões visando a proteção contra essas ocorrências, reduzindo os impactos causados.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- BARANAUSKAS, J.A.; MONARD, M.C. Reviewing Some Machine Learning Concepts and Methods. ICMC-USP, São Carlos, SP, 2000. (Relatório Técnico).
- EMBRAPA. Empresa Brasileira de Pesquisa Agropecuária. Desenvolvimento e evolução de um sistema de monitoramento agroclimático para o Estado de São Paulo. Embrapa Informática Agropecuária: Campinas, 2001. 15p.
- FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. American Association for Artificial Intelligence. 738-4602-1996. p. 37-54.
- MENDONÇA NETO, M. G.de Mineração de dados. In: ESCOLA REGIONAL DE INFORMÁTICA DA SBC REGIONAL DE SÃO PAULO, 6., 2001, São Carlos. Minicursos: coletânea de textos: anais. São Carlos: USP-ICMC, 2001. p. 189-218.
- UNESP. Data mining. Universidade Estadual de São Paulo. Rio Claro. 1998 Disponível em <http://black.rc.unesp.br/IA/cintiab/datamine/teoria.html>. Capturado em 05/2002.