

USO DE REGRAS DE ASSOCIAÇÃO PARA ENRIQUECER O SUPORTE DO PLANEJAMENTO AGRÍCOLA MUNICIPAL: ESTUDO DE CASO

FERNANDO ATTIQUE MAXIMO¹
STANLEY ROBSON DE MEDEIROS OLIVEIRA²
MARINA JORGE DE MIRANDA³

RESUMO: As técnicas de mineração de dados são úteis para explorar padrões em grandes banco de dados, subsidiando novos conhecimentos. Essas técnicas tem sido usadas com sucesso em diversas áreas, da comercial a científica. O objetivo deste trabalho é apresentar um estudo de caso em Taquarituba-SP, onde os dados das propriedades rurais foram coletados e submetidos à regras de associação, uma das tarefa da mineração de dados . A eficiência desta tarefa evidenciará novas informações para os municípios, destacando pontos fortes e fracos e enriquecerá os subsídios para o planejamento rural municipal.

PALAVRAS CHAVES: *Planejamento Agrícola, Associação, Mineração de Dados*

USE OF ASSOCIATION RULES IN AIDING COUNTY AGRICULTURAL PLANNING: CASE STUDY

ABSTRACT: Data mining techniques are useful to search for patterns in huge data banks, providing new knowledge. These techniques have been used with success in several realms, like commercial and scientific. The goal of the this work is to present a case study conveyed in Taquarituba-SP county, where data from farms were collected and submitted to the association rules for data mining. This task helped to highlight new information to the county, emphasizing strong and weak points, and, this way, contributing to the county agricultural planning.

KEY-WORDS: *Agricultural Planning, Association, Data Mining*

I - INTRODUÇÃO

A constituição brasileira de 1988 permitiu que os municípios assumissem novos papéis no desenvolvimento agrário nacional. Esses municípios tiveram aumento nas suas receitas, criaram suas secretárias para se adequar as mudanças e também, permitiram a organização das comunidades de diversas formas como conselhos, comitês etc (OLIVEIRA, 2006).

Dentro desse contexto de mudanças para os municípios, o Projeto Governança e Planejamento Agrícola dos Municípios no Estado de São Paulo - Fase 1, 2006, desenvolveu estudos de microclima e solos para criar um sistema de informações com opções de desenvolvimento regional de municípios do Estado com Índice de Desenvolvimento Humano (IDH) menores que 0,5. Os trabalhos envolveram a coletividade de agricultores e as prefeituras municipais, a Universidade Estadual de Campinas, através do Centro de Pesquisas Meteorológicas e Climáticas Aplicadas à Agricultura (CEPAGRI), a Embrapa Informática Agropecuária e a Coordenadoria de Assistência Técnica Integrada da Secretaria da Agricultura (CATI) (MIRANDA, 2007). Esse projeto foi finalizado com um sistema que recebe dados de socioeconomia e agrícola e foi testado com informações de uma das cidades envolvidas na pesquisa (Taquarituba-SP).

1 Bacharel em Matemática Aplicada e Computacional, pesquisador da Embrapa Informática Agropecuária, Campinas, SP, e-mail: fernando@cnptia.embrapa.br.

2 PhD em Ciência da Computação, pesquisador da Embrapa Informática Agropecuária, Campinas, SP, e-mail: stanley@cnptia.embrapa.br.

3 Geógrafa, estagiária do CEPAGRI/Unicamp, E-mail: marina@cpa.unicamp.br

Na atual etapa, esse sistema incluiu os municípios de todo Brasil, evoluiu a interface e incorporou novas informações pedidas pelos clientes que fizeram os treinamentos. Além disso, está em evolução o uso de mapas que mostrará a localização do município e das propriedades. Para usar esse sistema, a Embrapa Informática treina os interessados e os cadastra para acessá-lo. Na sequência, o objetivo é que as pessoas treinadas voltem as suas cidades, façam coleta de dados em cada propriedade por meio de um formulário de campo e transfiram as informações para o sistema. Dessa maneira, o município enxerga esses dados organizados através de relatórios, dando uma visão geral da área rural e oferecendo uma perspectiva de como construir um plano agrícola municipal.

Nesse contexto, os dados coletados na cidade de Taquarituba e cadastrados no sistema, é o objeto de estudo para verificar se as tarefas de mineração de dados (MD) podem ser úteis para acrescentar novos conhecimentos e agregar no suporte do planejamento municipal.

Uma das tarefas da MD é o uso de regras de associação, as quais representam um padrão de relacionamento entre os dados do domínio da aplicação. Uma regra de associação demonstra a presença de um determinado conjunto que implica na presença de algum outro conjunto distinto de itens (AGRAWAL et al., 1994).

II - OBJETIVO

O objetivo desse trabalho é descobrir regras (relações) através da tarefa de associação para enriquecer o suporte ao planejamento agrícola de municípios.

III - MATERIAL E MÉTODOS

Os *conjuntos de dados* escolhidos trazem dados sobre os investimentos da propriedade rural em tecnologia, tanto do ponto de vista da agricultura como em termos da tecnologia da informação. Foram analisados 8 conjuntos de dados e escolhidos 3 para o trabalho de mineração de dados. Dos 5 não escolhidos, um deles foi útil para para alimentar os outros (informações gerais da propriedade), em outro conjunto, os dados eram insuficientes e os outros 3 restantes, tinham um processo semelhante de análise aos outros escolhidos.

Os conjuntos de dados escolhidos foram: a) Benfeitorias praticadas na propriedade (4 atributos, 3112 instâncias); b) Sistemas de produção (9 atributos, 1908 instâncias); c) Informações gerais sobre vários aspectos como ser cooperado ou não, buscar assistência técnica ou não, usar ou não computadores etc (34 atributos, 882 instâncias).

E para delimitar esse trabalho, foi selecionado o conjunto de Informações Gerais porque contém dados sobre a área de tecnologia de informação, entre outras, área de atuação da Embrapa Informática Agropecuária.

A *metodologia* utilizada para o trabalho de mineração de dados seguiu essa abordagem:

a) Entender o negócio: a importância de criar um suporte para o planejamento rural dos municípios;

b) Entender os dados: entender quais dados do sistema era possível usar as tarefas de MD e que pode ser útil para o suporte do planejamento municipal;

c) Preparar os dados: a preparação dos dados é muito importante para fazer a mineração. Um dos passos foi verificar se havia dados ausentes no conjunto “Informações Adicionais”. Nesses dados pesquisados e cadastrados, os proprietários respondem “sim” ou “não” para diversos itens. Por exemplo, tem ou não computador na propriedade, tem ou não rádio amador (nesse caso, nenhuma propriedade tinha), é ou não cooperado, é ou não associado, usa ou não a internet etc. Portanto, essa tabela foi preenchida com 0 (não) e 1 (sim).

Nesse conjunto, houve um cuidado na *aplicação* da tarefa da mineração de dados porque, além de ser uma tabela com “sim” e “não”, a presença de “não” (ter, usar, é etc) foi muito expressiva nos dados desse município. Portanto, transformar as respostas “não” em

valores nulos, descaracterizaria os resultados. Além disso, existem inúmeros trabalhos que trabalham com regras de associação negativas (GONÇALVES, 2005), onde não será explorado o seu potencial nesse trabalho.

Para ajudar na limpeza dos dados e na sua organização, foi usado o Calc, planilha de cálculos do BrOffice, versão 2.4.

d) Escolher as tarefas adequadas: o objetivo inicial era usar duas tarefas diferentes de MD, *associação* e *classificação*, para poder *comparar e avaliar os resultados* das duas e poder ter visões diferentes. Para realizar a tarefa de classificação, era preciso de um atributo chave para caracterizar um resultado preditivo, que não foi encontrado. Então, foi usado apenas a tarefa de associação na busca de conhecimento.

Uma regra de associação é do tipo $X \rightarrow Y$ (se X, então Y), onde X e Y são conjuntos de atributos independentes, presentes numa linha de um banco de dados. Alberto Filho (2009) define:

Suporte corresponde à frequência com que todos os itens presentes na regra aparecem juntos em uma mesma transação da base de dados, mesmo que eles apareçam com outros itens não presentes na regra. A confiança expressa o percentual de transações em que, ocorrendo o antecedente, ocorre o conseqüente da regra.

Para auxiliar a realizar essa tarefa, foi usado o algoritmo de associação Apriori (Agrawal, 1994). O Apriori foi usado no ambiente Weka 3.5.6 (Witten & Frank, 2005).

Na aplicação desse algoritmo, inicialmente, foram usados todos os atributos do conjunto de dados, e depois foram escolhidos manualmente alguns atributos cuja ideia foi focar o acesso a tecnologia da informação. E num terceiro momento, foram escolhidas as linhas da tabela, considerando o tamanho das propriedades. E por último, escolher as regras com resultados mais importantes para essa análise.

IV - RESULTADOS E DISCUSSÕES

Os resultados das regras de associação são em grande quantidade, portanto foi preciso escolhê-los, excluindo as repetições e observando quais seriam os mais úteis. Para analisar nesse requisito de utilidade, foi definido para esse trabalho, clientes potenciais do sistema, que podem se valer desta análise. Os clientes escolhidos foram: i) o gestor municipal, ii) da Embrapa, em particular, a Embrapa Informática Agropecuária, iii) do investidor e iv) do próprio proprietário.

Na abordagem dos testes *foi destacado* o que cada proprietário tem em termos de *relacionamento* como pertencer a uma cooperativa, sindicato, associação; informações em *Tecnologia de Informação* quanto ao uso de computadores e internet que é um meio interessante para se atualizar e acessar novas possibilidades; se buscam informações em *órgãos de extensão*, se fazem uso de algumas técnicas importantes na *agropecuária* como inseminação artificial.

No primeiro teste, usando os dados completos (34 atributos mais o atributo sobre tamanho das propriedades, 882 instâncias), considerando o suporte de 95%, e confiança de 100%, mostrou que quando não há uso de tecnologia de informação (TI) e também há ausência de técnicas que podem ser úteis na pecuária como inseminação artificial e cultivo em estufa na agricultura. É comum muitos ainda não utilizarem a TI, mas para quem vive da agricultura, seria muito importante usar técnicas da mesma para alavancar os lucros. Nesse teste, mostra que se usa pouco tanto a TI como as técnicas da área rural. Ver Figura 1.

No segundo teste, aproveitando o resultado anterior e tentando aprofundar o seu entendimento, foi feita uma seleção de atributos manual, escolhendo atributos de tecnologia de informação e locais onde os proprietários poderiam acessá-la, reduzindo-se a 12 atributos. Nesse caso, observou-se que além de não usar a TI, alguns locais onde eles poderiam acessá-la como uma cooperativa, uma associação ou um sindicato, percebe-se que a maioria, mais de

70% não está ligado nenhuma dessas instituições. Aqui o suporte foi de 70 % e confiança de 90%.

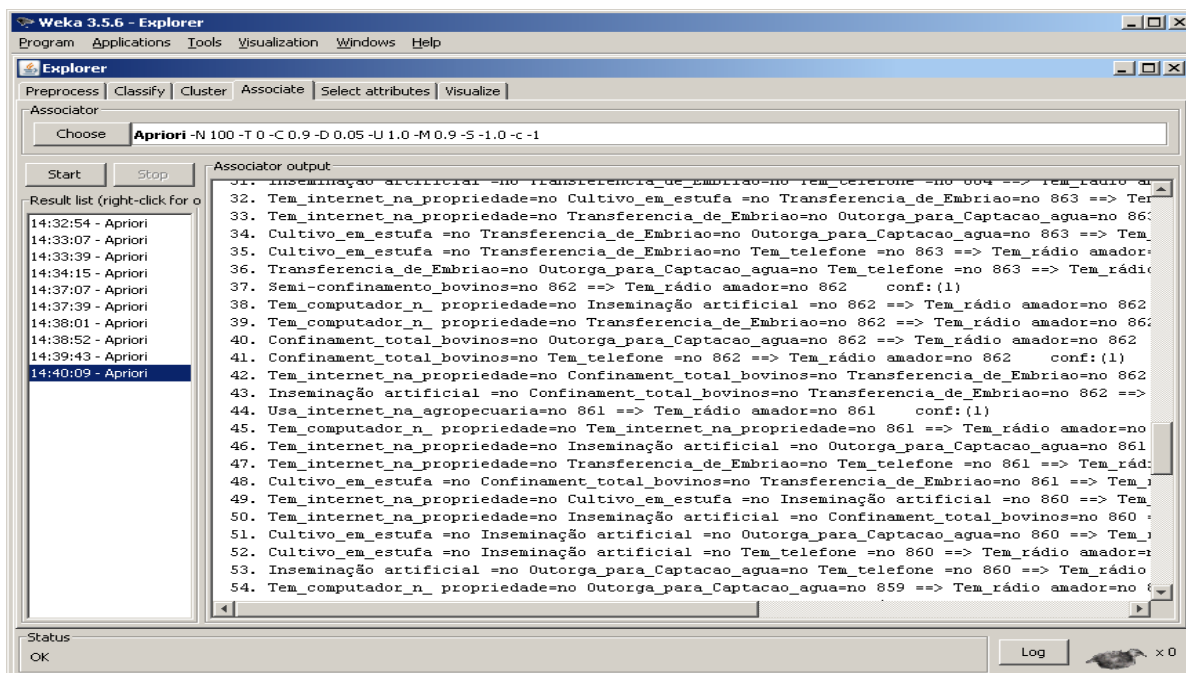


Figura 1 – Tela do Weka sobre o teste 1

No terceiro teste, mantendo os 12 atributos, mas considerando o suporte de 50%, confiança de 90%, foi verificado que não se usa assistência técnica privada, mas usa a assistência do estado e tem energia elétrica na propriedade. Aqui dá algumas informações por onde o gestor municipal poderia ajudar os proprietários: usar a assistência técnica do estado e aumentar o acesso a eletricidade.

No quarto teste, foram mantidos os mesmos 12 atributos anteriores, tendo apenas as propriedades rurais com até 50 hectares (dos 882 registros, 709 registros tem propriedade com esse tamanho). Considerando o suporte de 55% e confiança de 90%, mostrou que as propriedades de até 20 hectares, não tem computadores e não são associados.

Os resultados destes testes, mostram que o fato de não usar TI pode não tem relação com o desenvolvimento no agronegócio, a princípio, porque pode-se buscar informações de outras maneiras. Mas quando há ausência de contato com associações, cooperativas, percebe-se que esses proprietários perdem chances de desenvolver por falta de integração e conhecimento que esses tipos de agrupamentos podem gerar. E que poderiam acessar a internet.

Esses resultados indicam que *para os proprietários*, existe um espaço para modernizarem em termos de TI e seria bom buscarem contatos através de organizações. *Para o investidor*, esses resultados mostram que, tanto prefeitura como proprietários, são clientes em potencial e devem buscar formas atrativas para empréstimo para haver uma modernização dessas propriedades. *Para o gestor municipal* seria interessante viabilizar o intercâmbio entre os proprietários como incentivar o uso de tecnologias, como permitir o acesso a internet; incentivar a participação (ou criação) de cooperativas e associações. *Para a Embrapa*, o ideal para chegar suas tecnologias, é considerar uma aproximação com órgãos de extensão rural. A **Embrapa Informática Agropecuária** deve seguir o mesmo caminho, principalmente porque seus produtos são voltados para internet.

V - CONCLUSÕES

O trabalho procurou mostrar que a mineração de dados é um instrumento para realçar

alguns conhecimentos e apresentar outros novos. A análise foi feita em mais tabelas e para escopo desse trabalho, foi usada apenas uma, podendo se fazer testes variados.

Em segundo lugar, o teste foi feito em apenas um município. Os resultados podem ser interessantes para analisar municípios de uma região ou de uma bacia hidrográfica, porque pode-se adotar soluções conjuntas, ampliando o leque de opções de desenvolvimento regional. Seguindo esse raciocínio, um estado pode gerar um política estadual.

Em terceiro lugar, os municípios podem aumentar a sua participação no agronegócio brasileiro, beneficiando proprietários e consumidores, em termos de qualidade de produtos e preços.

Em quarto lugar, mostra que há uma aproximação que deve ser intensificada por pesquisadores e extensão rural para chegar os resultados de pesquisa no campo. No caso da cidade em questão, o órgão de extensão rural do estado é um elo importante que a Embrapa deve explorar.

Em quinto lugar, a ênfase em TI nas análises se deve pelo interesse em viabilizar recursos na forma de sistemas de informação para internet, produzidos pela Embrapa Informática, como esse sistema citado nesse trabalho.

Como trabalho futuro, as sugestões que podem ser abordadas: a) Explorar as outras tabelas, completando melhor a análise e expor em encontro de prefeitos junto com o sistema; b) Em termos de mineração de dados, na tarefa de associação, explorar os padrões das regras negativas e positivas; c) Fazer um contato com os clientes citados e evoluir essa análise.

V - BIBLIOGRAFIA

AGRAWAL, R. et al. (1994) Fast algorithms for mining association rules in large databases. In: International Conference on Very Large Data Bases, VLDB, 20., 1994, Santiago. Proceedings... Hove: Morgan Kaufmann, pages. 478-499.

ALBERTO FILHO, Luiz, Favero, Eloi L.. Aplicação de Técnicas de Mineração de Dados e Textos no Apoio à Tomada de Decisão em Segurança Pública. Disponível em <http://www.ufpa.br/ppgcc/ppgcc/files/File/Seminario_Andamento/2008/SI/artigo_dissertacao_mestrado%20luiz.pdf>. Acesso em 10/06/2009.

GONÇALVES, E. C., Regras de Associação e suas Medidas de Interesse Objetivas e Subjetivas. UFF - Universidade Federal Fluminense, 2005.

MIRANDA, Marina Jorge de ; PINTO, H. S. ; MEDEIROS, Stanley R de ; ROMANI, Luciana S . e-Sisplam - Sistema de Suporte ao Planejamento e Levantamento Agrícola Municipal. In: XV Congresso Brasileiro de Agrometeorologia - CBAgro, 2007, Aracaju, Se. Anais do XV Congresso Brasileiro de Agrometeorologia - CBAgro. Campinas, SP. : SBAgro, 2007. v. 1. p. 45-49.

OLIVEIRA, Stanley R de M.; ROMANI, Luciana Alvim Santos ; PINTO, H. S. ; ZULLO JR, Jurandir ; MIRANDA, Marina Jorge de ; PEREIRA, Vanildo H ; SOUSA, Antonio Carlos de . Tecnologia da Informação para Automatizar o Processo de Planejamento de Atividades Agropecuárias da Bacia do Alto Paraguai. In: I Simpósio de Geotecnologias no Pantanal - GeoPantanal., 2006, Campo Grande, MS. Anais - I Simpósio de Geotecnologias no Pantanal - GeoPantanal.. Campo Grande, MS : Embrapa, 2006. v. 1. p. 645-653.

WITTEN, I. H.; FRANK, E. Data mining: practical machine learning tools and techniques. 2nd ed. San Francisco: Morgan Kaufmann, 2005.