

Resumos Expandidos

XI Mostra de estagiários e bolsistas da Embrapa Informática Agropecuária

Campinas, 10 a 12 de novembro, 2015

>>> Material e Métodos

Foram utilizados dados espaciais do IBGE e de ANA (BRASIL, 2008), bem como dados estatísticos do Censo Agropecuário IBGE 2008 (IBGE, 2008). Utilizando o software ArcGIS, foram criadas mapas de uso do solo para o uso de pastagem, lavouras e matas, na bacia do rio Parapuá.

>>> Resultados e Discussão

Os principais produtos do trabalho foram os mapas de uso do solo para os municípios de Itapetininga e Itapicuru no Censo Agropecuário 2008. Área total de pastagens, área total de lavouras, áreas de matas ou florestas, rios e áreas de preservação ambiental. É possível observar, a partir da criação dos mapas, as áreas de maior atividade agropecuária da bacia do rio Parapuá, que está localizada nos estados Mato Grosso do Sul e Mato Grosso.

A parte mais central da bacia, onde se localiza o pantanal, apresenta muita mata e áreas de preservação ambiental. Assim como as áreas de matas ou florestas. A prática de agricultura já não é tão intensa, talvez pela maior dificuldade de se preparar as terras de planície para o plantio.



Figura 2. Área total de lavouras. Fonte: IBGE

>>> Conclusão

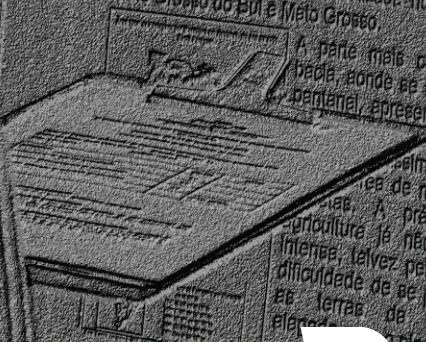
Chegamos ao fim da bacia do rio Parapuá e seus municípios. A área de pastagem e lavouras foram suas principais atividades. São equívocos e erros, áreas localizadas em grandes áreas desmatadas e florestas.

>>> Agradecimentos

Meus agradecimentos à Embrapa Informática Agropecuária, pela disponibilização do trabalho, bem como ao Laboratório de Geoprocessamento sempre que precisei de informações e não às atividades de planejamento.

>>> Referências

BRASIL. Ministério do Meio Ambiente. Instituto Brasileiro de Geografia e Estatística. IBGE. Instituto Brasileiro de Geografia e Estatística. Censo Agropecuário 2008. Brasília, 2008.



*Empresa Brasileira de Pesquisa Agropecuária
Embrapa Informática Agropecuária
Ministério da Agricultura, Pecuária e Abastecimento*

Resumos Expandidos

XI Mostra de Estagiários e Bolsistas da Embrapa Informática Agropecuária

Campinas, 10 a 12 de novembro, 2015

*Giampaolo Queiroz Pellegrino
Maria Giulia Croce
Poliana Fernanda Giachetto
Stanley Robson de Medeiros Oliveira*

Editores Técnicos

Embrapa
Brasília, DF
2015

Embrapa Informática Agropecuária

Av. André Tosello, 209 - Barão Geraldo
Caixa Postal 6041 - 13083-886 - Campinas, SP
Fone: (19) 3211-5700
www.embrapa.br/informatica-agropecuaria
SAC: www.embrapa.br/fale-conosco/sac/

Unidade responsável pelo conteúdo e edição

Embrapa Informática Agropecuária

Comitê de Publicações da Embrapa Informática Agropecuária

Presidente: *Giampaolo Queiroz Pellegrino*

Secretária: *Carla Cristiane Osawa*

Membros: *Adhemar Zerlotini Neto, Stanley Robson de Medeiros Oliveira, Thiago Teixeira Santos, Maria Goretti Gurgel Praxedes, Adriana Farah Gonzalez, Neide Makiko Furukawa, Carla Cristiane Osawa*

Membros suplentes: *Felipe Rodrigues da Silva, José Ruy Porto de Carvalho, Eduardo Delgado Assad, Fábio César da Silva*

Supervisor editorial: *Stanley Robson de Medeiros Oliveira, Neide Makiko Furukawa*

Revisor de texto: *Adriana Farah Gonzalez*

Normalização bibliográfica: *Maria Goretti Gurgel Praxedes*

Editoração eletrônica/Arte capa: *Neide Makiko Furukawa*

Foto capa: *Rosangene Santos*

1ª edição

Publicação digitalizada (2015)

Todos os direitos reservados.

A reprodução não autorizada desta publicação, no todo ou em parte, constitui violação dos direitos autorais (Lei nº 9.610).

Dados Internacionais de Catalogação na Publicação (CIP)

Embrapa Informática Agropecuária

Mostra de Estagiários e Bolsistas da Embrapa Informática Agropecuária (11. : 2015 : Campinas, SP).

Resumos Expandidos : XI Mostra de Estagiários e Bolsistas da Embrapa Informática Agropecuária : Campinas, 10 a 12 de novembro, 2015 / Giampaolo Queiroz Pellegrino ... [et al.], Editores técnicos. -- Brasília, DF : Embrapa, 2015.

124 p.

ISBN 978-85-7035-544-7

1. Agroinformática. 2. Tecnologia da informação. 3. Iniciação científica. 4. Mudanças climáticas. I. Pellegrino, Giampaolo Queiroz. II. Embrapa Informática Agropecuária. III. Título.

CDD 630.285

© Embrapa 2015

Editores Técnicos

Giampaolo Queiroz Pellegrino

Engenheiro-florestal, doutor em Engenharia Agrícola
Pesquisador da Embrapa Informática Agropecuária, Campinas, SP

Maria Giulia Croce

Relações públicas, especialista em Comunicação Mercadológica
Analista da Embrapa Informática Agropecuária

Poliana Fernanda Giachetto

Zootecnista, doutora em Produção Animal
Pesquisadora da Embrapa Informática Agropecuária, Campinas, SP

Stanley Robson de Medeiros Oliveira

Bacharel em Ciência da Computação, doutor em Ciência da Computação
Pesquisador da Embrapa Informática Agropecuária

Apresentação

É com muita satisfação que a Embrapa Informática Agropecuária apresenta os anais da décima primeira edição da Mostra de Estagiários e Bolsistas ocorrida em novembro de 2015 nas instalações da Unidade em Campinas - SP. A Mostra, que acontece anualmente na Embrapa Informática Agropecuária, tem como objetivo divulgar os trabalhos desenvolvidos pelos estudantes na Unidade, além permitir que os mesmos vivenciem a experiência de um evento científico. Os trabalhos apresentados na Mostra, na forma de pôster ou oral, são submetidos à avaliação por um comitê científico composto por pesquisadores e analistas da Unidade e obedecem a regras e critérios previamente estabelecidos e divulgados aos autores. Além da área de Pesquisa e Desenvolvimento, empregados de outras áreas como o Núcleo de Comunicação Organizacional, Núcleo de Tecnologia de Informação, Núcleo de Apoio à Pesquisa, Setor de Gestão de Pessoas e demais áreas da Administração, integram esforços para a realização do evento.

A Mostra de 2015 contou com a apresentação de 21 trabalhos, distribuídos nas seguintes categorias: Apoio à Pesquisa, Pesquisa - Pós-graduação e Pesquisa - Graduação, sendo apresentados oralmente 10 trabalhos e 11 sob a forma de pôsteres.

Além do aspecto científico, a Mostra de Estagiários e Bolsistas é uma oportunidade de integração dos estudantes e empregados das diferentes áreas da Unidade, que visitam a sessão de pôsteres e a sessão oral nos dias do evento. A Chefia da Embrapa Informática Agropecuária parabeniza e agradece aos estudantes e seus orientadores pela participação na Mostra e agradece o empenho e comprometimento de todos os envolvidos, em especial o apoio do Comitê Local de Publicações e do Comitê Organizador pelo sucesso da XI Mostra de Estagiários e Bolsistas da Embrapa Informática Agropecuária.

Giampaolo Queiroz Pellegrino

Chefe Adjunto de Pesquisa e Desenvolvimento
Embrapa Informática Agropecuária

Sumário

Uma ferramenta para expansão do vocabulário com base em coocorrência

Exupério Lédo Silva Júnior; Roberta Akemi Sinoara; Solange Oliveira Rezende;
Ricardo Marcondes Marcacini; Maria Fernanda Moura 11

I-Preproc: uma ferramenta para pré-processamento e indexação incremental de documentos

Renan Gomes Pereira; Maria Fernanda Moura..... 17

Extração de hierarquias de tópicos em textos para apoiar a construção de portfólios tecnológicos

Carolina Tavares de Oliveira; Luisa Miyashiro Tápias;
Stanley Robson de Medeiros Oliveira; Maria Fernanda Moura 24

Variabilidade interanual das áreas com potencial de risco de desertificação da Bahia

Camila da Silva Dourado; Stanley Robson de Medeiros Oliveira;
Ana Maria Heuminski de Avila..... 31

Identificação de Pontos Perceptualmente Importantes (PIP) em séries temporais de tópicos extraídos de dados textuais

Lucas Santiago Rodrigues; Roberta Akemi Sinoara; Solange Oliveira Rezende;
Ricardo Marcondes Marcacini; Maria Fernanda Moura 38

Levantamento de conteúdo para o sistema WebAmbiente

Meiry Helen Bordim; Alan Massaru Nakai 45

Extração de regras de associação em textos para apoiar a construção de portfólios tecnológicos

Luisa Miyashiro Tápias; Carolina Tavares de Oliveira;
Stanley Robson Medeiros Oliveira; Maria Fernanda Moura..... 50

Processo de produção de imagens para comunicação visual	
Giovana Roggeri Affonso; Deise Rocha Martins dos Santos Oliveira.....	56
Processo de produção em audiovisual das tecnologias da Embrapa Informática Agropecuária	
Mariana Moreti Vieira; Deise Rocha Martins dos Santos Oliveira	62
Protótipo de sistema de segurança criptográfica via porta lógica quântica	
Jacomo Giovanetti Minto Neto; Alexandre de Castro; Adauto Luiz Mancini	68
Automação de experimentos científicos utilizando KnowledgeFlow	
Nádia Vieira Ribeiro; Luiz Manoel Silva Cunha	72
Implementação do Best Linear Unbiased Prediction (BLUP) em Python para avaliação genética animal	
Caio Augusto C. Volpato; Roberto Hiroshi Higa.....	78
Interpretação e classificação de imagens de satélite na extração de informações sobre o uso da terra e cobertura vegetal na Bacia Hidrográfica do Rio Paracatu, MG	
Paula Guimarães; João dos Santos Vila da Silva	85
Interpretação e classificação de imagens de satélite na extração de informações sobre o uso da terra e cobertura vegetal na bacia hidrográfica do rio Teles Pires/MT-PA	
Jéssica da Silva Rodrigues Cecim; João dos Santos Vila da Silva	89
Mecanismo para verificação de conteúdo de temas disponíveis em banco de dados geográficos (PostGIS)	
João Luís dos Santos; João dos Santos Vila da Silva.....	94
Disponibilização de dados geoespaciais no SIAGEO Amazônia	
Edson Antonio Mengatto Junior; João dos Santos Vila da Silva	98
Uso da terra na bacia hidrográfica do alto rio Paraguai no Brasil	
Cezar Freitas Barros; João dos Santos Vila da Silva	102

**Sistema Interativo de Análise Geoespacial da Amazônia Legal:
análise da distribuição e localização de dados**

Daniel Dias de Andrade; João dos Santos Vila da Silva 107

**Explorando o uso da Teoria Fuzzy no software R: estudo de caso
para avaliar a força de infecção da Anemia Infecciosa Equina (AIE)**

Gabriela Servidone; Sônia Ternes; Marcelo Rossi 112

**Disseminando a aplicação do R-Shiny em métodos quantitativos e
computação científica na web**

Bernardo França Rocha; Sônia Ternes; Marcelo Rossi 119

Uma ferramenta para expansão do vocabulário com base em coocorrência

Exupério Lédo Silva Júnior¹

Roberta Akemi Sinoara²

Solange Oliveira Rezende³

Ricardo Marcondes Marcacini⁴

Maria Fernanda Moura⁵

Resumo: Neste trabalho é apresentado um módulo desenvolvido para a experimentação de algumas técnicas de pré-processamento visando uma boa representação de coleções de documentos. As técnicas experimentadas são voltadas à expansão de vocabulário do domínio por meio da inclusão de termos coocorrentes. Um módulo, chamado DATool, foi desenvolvido em Java e experimentos estão sendo realizados. Caso os resultados sejam positivos, as técnicas serão transformadas em filtros de pré-processamento e indexação a serem incorporados ao arcabouço de ferramentas do projeto CRITIC@.

Palavras-chave: mineração de textos, pré-processamento de textos, coocorrência de termos.

¹ Estudante de Ciências de Computação da Universidade de São Paulo (ICMC-USP), estagiário da Embrapa Informática Agropecuária, Campinas, SP.

² Bacharel em Informática, doutoranda em Ciências da Computação e Matemática Computacional no Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo (ICMC-USP), São Carlos, SP.

³ Graduada em Licenciatura em Ciências Habilitação Matemática, doutora em Engenharia Mecânica, professora associada da Universidade de São Paulo (ICMC-USP), São Carlos, SP.

⁴ Bacharel em Informática, doutor em Ciências da Computação e Matemática Computacional, docente e pesquisador da Universidade Federal de Mato Grosso do Sul (UFMS), Três Lagoas, MS.

⁵ Estatística, doutora em Ciências Matemáticas e da Computação, pesquisadora da Embrapa Informática Agropecuária, Campinas, SP.

Introdução

Técnicas de Mineração de Textos auxiliam a extrair e organizar conhecimento de grandes coleções de documentos textuais, uma atividade cada vez mais importante dada a quantidade crescente de documentos textuais no meio digital (ROSSI, 2011). Tais técnicas baseiam-se na busca de padrões, tendências e regularidades em documentos escritos em língua natural.

O processo de Mineração de Textos pode ser dividido em cinco etapas: identificação do problema, pré-processamento, extração de padrões, pós-processamento e utilização do conhecimento (REZENDE et al., 2003). Na identificação do problema são definidos os objetivos do processo de Mineração de Textos. Uma vez definidos o escopo do problema e o objetivo da aplicação, e selecionados documentos que representem o domínio do problema, partimos para a próxima etapa, o pré-processamento. Na etapa de pré-processamento, é realizado um tratamento dos dados. Tal tratamento engloba atividades como: eliminação de termos sem valor terminológico para o domínio em questão (as *stopwords*), normalização das palavras, identificação dos termos do domínio e seleção de atributos (LAGUNA, 2015; MOURA, 2009). Outra atividade importante realizada na etapa de pré-processamento é a estruturação dos dados, que devem ser representados em um formato apropriado para a extração de conhecimento. As tarefas a serem realizadas em seguida, na extração de padrões, consistem na utilização de algoritmos de aprendizado de máquina, que podem rotular, categorizar, ou detectar comportamentos intrínsecos da coleção de dados. O pós-processamento, por sua vez, consiste na avaliação dos padrões extraídos, verificando sua validade e aplicabilidade para que, no fim, o conhecimento extraído possa ser utilizado na etapa de utilização do conhecimento. Caso as atividades realizadas no pós-processamento indiquem que o conhecimento não atinge os objetivos estabelecidos, um novo ciclo do processo de Mineração de Textos é iniciado, com uma nova execução das etapas anteriores.

As atividades realizadas no pré-processamento são essenciais para o sucesso da Mineração de Textos, visto que a representação dos textos obtida nesta etapa deve manter os padrões a serem descobertos na extração de padrões. O módulo abordado neste artigo, chamado de DATool, é uma experimentação de técnicas para auxiliar as atividades de pré-processamento e indexação incremental de bases de dados textuais. Com a aplicação dessas técnicas, espera-se obter um conjunto maior de termos relacionados

ao domínio em questão e melhorar os resultados da Mineração de Textos. Na próxima seção são descritos os requisitos e funcionalidades do módulo desenvolvido. Logo após são apresentados os resultados obtidos até o momento e, por fim, as considerações finais.

Materiais e Métodos

O módulo DATool foi desenvolvido com o objetivo de permitir a experimentação de técnicas de pré-processamento e indexação de bases textuais no formato XML no contexto do projeto CRITIC@, que visa melhorar a gestão do conhecimento técnico-científico na área de recursos hídricos, por meio de análises cruzadas das informações, bem como subsidiar ações de investigação e disseminação do conhecimento na rede de pesquisa (SILVA; MOURA, 2014). Para o desenvolvimento desse módulo foram realizadas atividades de especificação dos requisitos, estudo das técnicas envolvidas (tanto do processo de Mineração de Textos, quanto de programação), implementação do módulo, documentação e testes. O módulo foi desenvolvido com base na ferramenta Torch (MARCACINI; REZENDE, 2010), permitindo o processamento incremental e a sua integração com outros métodos de Mineração de Textos.

Na Figura 1 são apresentados os tratamentos realizados nos textos de entrada para gerar a representação desses textos com base em um vocabulário controlado e termos coocorrentes, que pode ser incrementalmente atualizada a medida em que novos textos são adicionados à coleção.

O módulo recebe como entrada um arquivo de configuração fornecido pelo usuário que contém os parâmetros para execução do módulo. Nesse arquivo o usuário define: a) idioma da coleção de textos: sendo que o módulo suporta os idiomas português, inglês e espanhol; b) diretório de entrada: endereço do diretório contendo a coleção de textos; c) diretório de saída: endereço do diretório no qual são salvos os arquivos de saída; d) vocabulário inicial: endereço do vocabulário contendo as palavras que formam o vocabulário inicial; e) *stopwords*: endereço do arquivo contendo a lista de *stopwords*; f) *stemmização*: define se os termos são stemmizados ou não (*true* ou *false*); g) *tags*: nome das *tags* cujo conteúdo são considerados no pré-processamento; h) coocorrência: define os parâmetros para o cálculo dos coocorrentes: método e coeficiente de corte.

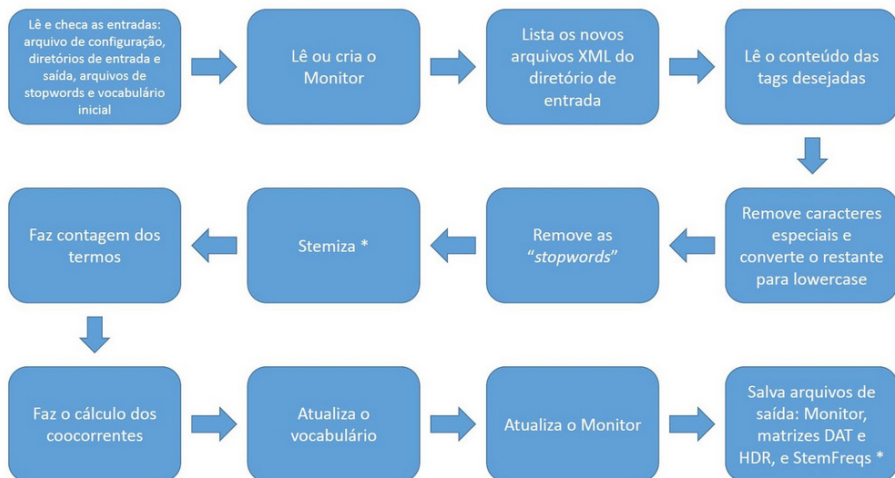


Figura 1. Fluxograma do módulo DATool.

Inicialmente o módulo lê e verifica todos os parâmetros passados pelo arquivo de configuração. Para garantir o aspecto incremental do módulo, utiliza-se um arquivo que guarda as informações dos processamentos anteriores, o Monitor. Então, caso seja a primeira execução, cria-se um novo Monitor; caso contrário, lê-se o Monitor existente. Os novos arquivos são listados e, para fins de normalização, o módulo realiza uma limpeza dos dados, selecionando apenas o conteúdo das tags desejadas, removendo os caracteres especiais e as stopwords, e facultativamente realizando a stemmização das palavras. A stemmização conta com o apoio da API Lucene⁶ 5.2.0 da Apache, além da ferramenta Torch (MARCACINI; REZENDE, 2010). E o cálculo da coocorrência utiliza a API Commons Math⁷ 3.5, também da Apache.

Posteriormente realiza-se a contagem dos termos por arquivo. Aqueles que também aparecem no arquivo de Vocabulário Inicial entram para o vocabulário. E então são calculados os coocorrentes por meio do coeficiente de correlação de Pearson. No caso da DATool, este coeficiente mede o grau de relação entre pares de termos do vocabulário e termos que estão fora do vocabulário. O cálculo leva em consideração a frequência destes termos

⁶ Disponível em: <<https://lucene.apache.org/>>. Acesso em: 28 set. 2015.

⁷ Disponível em: <<http://commons.apache.org/>>. Acesso em: 28 set. 2015.

em cada texto da coleção, e o resultado para o coeficiente varia entre [-1, 1]. Os termos que formam pares com o coeficiente maior, em módulo, que o coeficiente de corte definido no arquivo de configuração também são adicionados ao vocabulário.

No caso deste módulo, a saída é uma matriz atributo-valor, cujas linhas correspondem a cada documento da coleção de textos e as colunas a cada atributo selecionado, respeitando os padrões utilizados no projeto CRITIC@. Tal padrão consiste no arquivo texto HDR, que contém a lista dos documentos e a lista dos atributos, e o arquivo texto DAT, que informa a frequência dos atributos nos arquivos.

Após o desenvolvimento, foram realizados testes de unidade para validar as funcionalidades e os parâmetros do módulo. A ferramenta foi executada com uma coleção de 956 documentos em português no formato XML. A funcionalidade de *stemmização* também foi testada individualmente com outros textos escritos em espanhol e inglês.

Resultados e Discussão

Durante os testes de unidade, foi verificado que as funcionalidades implementadas estão funcionando conforme esperado, porém não foi feita uma análise sobre os resultados obtidos pela expansão do vocabulário. Atualmente, o módulo está em fase de teste e validação desses resultados. Os resultados esperados são que, dado um conjunto de termos inicial, outros termos também relacionados ao domínio sejam encontrados. Desta forma, é possível um grupo maior de termos relacionados ao domínio em questão, aumentando a representatividade da coleção de textos por meio do vocabulário.

Considerações Finais

Neste trabalho é apresentado o módulo DATool, que é uma experimentação de técnicas para expansão de vocabulário com base em coocorrência de termos para representação incremental de coleções textuais. Caso os experimentos que estão sendo realizados no momento tenham resultados

positivos, as técnicas validadas serão transformadas em filtros de pré-processamento e indexação a serem incorporados ao sistema do projeto CRITIC@.

Referências

LAGUNA, M. da S. C. **Extração automática de termos simples baseada em aprendizado de máquina**. 2014. Tese (Doutorado) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, SP. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/55/55134/tde-11082014-103430/pt-br.php>>. Acesso em: 28 set. 2015.

MARCACINI, R. M.; REZENDE, S. O. Torch: a tool for building topic hierarchies from growing text collections. In: WORKSHOP ON TOOLS AND APPLICATIONS, 9.; BRAZILIAN SYMPOSIUM ON MULTIMEDIA AND THE WEB, 2010, Belo Horizonte. **Proceedings...** [S.l.: s.n.], 2010. p. 1-3. WFA'2010; Webmedia 2010.

MOURA, M. F. **Contribuições para a construção de taxonomias de tópicos em domínios restritos utilizando aprendizado estatístico**. 2009. Tese (Doutorado) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, SP. Disponível em: <http://ainfo.cnptia.embrapa.br/digital/bitstream/item/17875/1/MFM_Tese_5318963-2.pdf>. Acesso em: 28 set. 2015.

REZENDE, S. O.; PUGLIESI, J. B.; MELANDA, E. A.; de PAULA, M. F. Mineração de dados. In: REZENDE, S. O. (Ed.). **Sistemas inteligentes: fundamentos e aplicações**. Manole, 2003. p. 307-335.

ROSSI, R. G. **Representação de coleções textuais por meio de regras de associação**. 2011. Dissertação (Mestrado) - ICMC - USP - São Carlos. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/55/55134/tde-31082011-125648/pt-br.php>>. Acesso em: 28 set. 2015.

SILVA, L. E. A.; MOURA, M. F. Componentes para a integração e extração de padrões em textos para versão 1.0 do ambiente CRITIC@. In: MOSTRA DE ESTAGIÁRIOS E BOLSISTAS DA EMBRAPA INFORMÁTICA AGROPECUÁRIA, 10., 2014, Campinas. **Resumos...** Brasília, DF: Embrapa, 2014. p. 17-19. Disponível em: <<http://ainfo.cnptia.embrapa.br/digital/bitstream/item/118438/1/043-14.pdf>>. Acesso em: 28 set. 2015.

I-Preproc: uma ferramenta para pré-processamento e indexação incremental de documentos

Renan Gomes Pereira¹

Maria Fernanda Moura²

Resumo: O objetivo deste trabalho é apresentar a ferramenta I-Preproc, utilizada no pré-processamento e indexação incremental de documentos. A ferramenta foi implementada em Java utilizando a biblioteca open source Apache Lucene. Embora a ferramenta esteja em desenvolvimento, os resultados parciais obtidos têm sido bastante satisfatórios, como mostrado no experimento realizado.

Palavras-chave: mineração de textos, Apache Lucene, máquina de busca.

Introdução

I-Preproc é uma ferramenta flexível e expansível desenvolvida em Java utilizando a biblioteca open source Apache Lucene (APACHE SOFTWARE FOUNDATION, 2015), para pré-processamento, indexação incremental e busca de documentos. O usuário especifica os parâmetros da indexação por meio de um arquivo de configuração em XML, no qual podem ser especificados quais filtros serão aplicados nos documentos na fase de indexação e, caso o documento seja um arquivo XML, as tags do documento que terão seus conteúdos indexados. Além disso, é possível informar os parâmetros desejados durante a extração dos resultados na forma de matrizes atributo-valor. Essas matrizes são utilizadas em processos de mineração de textos,

¹ Engenharia da Computação, Universidade Estadual de Campinas; Estagiário da Embrapa Informática Agropecuária, Campinas, SP.

² Estatística, doutora em Ciências Matemáticas e da Computação, pesquisadora da Embrapa Informática Agropecuária, Campinas, SP.

por ferramentas de aprendizado de máquina. Nessas matrizes, as linhas correspondem aos documentos (observações, instâncias), as colunas aos atributos (palavras, composições de palavras, frases, etc.) e cada célula ao grau de importância do atributo para o documento correspondente; por exemplo, o valor da célula pode representar a frequência de uma palavra em um texto.

A possibilidade de realizar a indexação incremental de uma coleção extensa de documentos é crucial. Neste tipo de indexação, a ferramenta insere novos documentos em um índice já existente sem a necessidade de reindexar toda a coleção. Esse processo economiza tempo e recursos computacionais que seriam gastos na reindexação de documentos que já estavam indexados.

Por fim, com a I-Preproc também é possível fazer buscas nos índices por palavras, N-gramas, frases, termos exatos, e, também excluir termos dos resultados. Caso o número de documentos retornados na busca seja menor que um valor determinado pelo usuário, a ferramenta realiza uma outra busca por termos similares utilizando o conceito de edit distance (Konchady, 2008).

Dessa forma, a I-Preproc é uma ferramenta flexível para ser usada em indexação, busca e geradora de dados para ferramentas de aprendizado de máquina, que pode e está sendo evoluída (e desenvolvida) de forma incremental, com uma boa performance, conforme apresentado neste trabalho.

Materiais e Métodos

Nesta seção é mostrado o experimento de performance realizado e um exemplo de um arquivo de configuração para a indexação de uma base de textos, bem como a arquitetura de alto nível da I-Preproc, com a explicação da função de cada módulo constituinte.

- **Experimento:** foi realizado um primeiro experimento para avaliar o tempo de indexação, com uma base exemplo de 2054 textos de tamanhos variáveis. As opções para a indexação, nesse experimento, foram: a) utilizar um vocabulário controlado (o disponível para testes contém apenas 33178 unigramas); b) indexar e filtrar o texto completo removendo acentuação, convertendo letras maiúsculas para minúsculas e removendo caracteres

especiais. Foram indexados apenas os termos stemmizados; quanto ao idioma, o escolhido foi o português do Brasil, e os testes foram feitos em uma máquina Intel(R) Core(TM) i3-2120 CPU @ 3.30GHz com 8GB de RAM rodando no sistema operacional Ubuntu 14.04.

- **Arquivo de configuração:** neste arquivo é possível fornecer todos os parâmetros desejados na indexação, na extração das matrizes atributo-valor e na busca. Um exemplo de arquivo de configuração para a indexação é mostrado na Figura 1, onde:

```
<IncPreProc>
  <Paths indexdir = "../indexdir/"
    textbase = "../textbase/"
    incrementaltextbase = "../textbase/newtexts/"
    domainVoc = "../vocabularies/domainvoc.txt"
    stoplist = "../stopwords_bra.txt"/>
  <Tags
    xmlTags = "author, location, content"
    filteredXmlTags = "content" />
  <Ngram ngrams = "false" maxNGrams = "4"/>
  <Stemimization language = "BRA" stemming = "STEM_NO_STEM"/>
  <Filters domainVoc = "true"
    stoplist = "false"
    toLowercase = "true"
    removeAccents = "true"
    removeSpecials = "true"/>
</IncPreProc>
```

Figura 1. Exemplo de um arquivo de configuração.

<Paths>: especifica os endereços dos arquivos e diretórios necessários para a indexação.

<Tags>: quais serão as tags de um arquivo XML a serem indexadas e quais dessas tags serão filtradas. Caso <Tags> seja removida do arquivo, a ferramenta indexa e filtra o arquivo de texto completo.

<Ngrams>: opções para a indexação de N-gramas, se serão indexados e qual o tamanho máximo de cada n-grama.

<Stemimization>: escolher dentre as opções de indexação. As opções podem ser: a) indexar apenas as palavras sem stemmizar; b) indexar apenas as palavras stemmizadas; c) indexar ambas as formas. Também

é fornecido o idioma de stemmização. As opções disponíveis no momento são inglês “ENG” e português brasileiro “BRA”. Pretende-se incluir novos idiomas como o francês e o espanhol.

<Filters>: são escolhidos os filtros que se deseja usar. Os filtros disponíveis até o momento, estão listados abaixo:

- Vocabulário controlado: especifica-se o arquivo que contém um vocabulário controlado. O vocabulário controlado é escolhido pelo usuário.
- Stopwords: se o usuário optar por não usar um vocabulário controlado, é possível fornecer um arquivo com stopwords que são palavras que o índice deve desconsiderar. As stopwords são escolhidas e especificadas pelo usuário.
- Indexar em letras minúsculas: desconsidera a diferença entre letras maiúsculas e minúsculas.
- Remover acentuação: remove os acentos das palavras dos textos antes de indexar.
- Remover caracteres especiais: remove todos os caracteres diferentes de letra, número e hífen.

Arquitetura da I-Preproc: a arquitetura de alto nível do projeto está presente na Figura 2. A I-Preproc é composta por 4 módulos e um índice criado pela biblioteca Lucene:

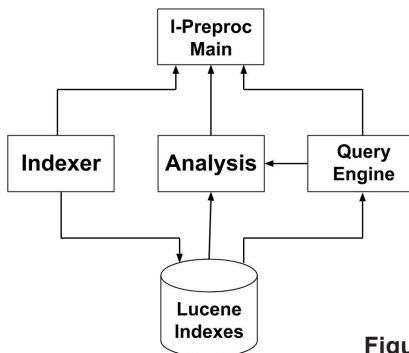


Figura 2. Arquitetura da I-Preproc.

I-Preproc Main é o módulo responsável pela comunicação entre os módulos da ferramenta e pelo controle do fluxo de execução do programa. A comunicação é feita por meio de um arquivo de comunicação que é salvo quando um índice é criado. Este arquivo contém todos os parâmetros que foram usa-

dos para criar este índice, como os filtros utilizados, tamanho dos N-gramas e opções de stemmização. Quando algum módulo opera sobre este índice, ele carrega o arquivo de comunicação deste índice. Deste modo, é possível manter a coerência do índice e das operações efetuadas nele.

O **Módulo Indexer** realiza a indexação incremental de novos textos sem a necessidade de reindexar toda a coleção. A I-Preproc cria um índice de acordo com as opções escolhidas pelo usuário no arquivo de configuração. Após a criação desse índice, quando o usuário deseja indexar novos textos, ele fornece o diretório onde estão os novos textos (este diretório deve ser diferente do diretório da coleção inicial). A ferramenta carrega o arquivo de comunicação, que contém as informações do índice, e indexa todos os textos contidos neste novo diretório movendo-os para o diretório da coleção total. Como, em geral, o número de textos a serem adicionados incrementalmente no índice é relativamente bem menor do que a coleção total, esse processo é muito mais rápido do que a reindexação da coleção completa.

O **Módulo Analysis** extrai as matrizes atributo-valor que contém informações estatísticas importantes sobre o índice e tem formato adequado para ferramentas de aprendizado de máquina. O usuário pode escolher, pelo arquivo de configuração, quais as medidas que ele deseja extrair, como TF ou TF-IDF. Também é possível filtrar os resultados a serem colocados na matriz utilizando filtros de frequência.

O **Módulo Query Engine** é responsável pelas buscas no índice. As buscas podem ser feitas por palavras, N-gramas, frases, termos exatos, e, também excluir termos dos resultados. Caso o número de documentos retornados na busca seja menor do que um valor determinado pelo usuário, a ferramenta realiza uma outra busca por termos similares utilizando o conceito de edit distance.

Resultados e Discussão

Os intervalos em segundos para 10 execuções do programa se encontram na Tabela 1. Ao indexar incrementalmente 78 textos em temas agrícolas de tamanhos variados no índice descrito acima, o tempo médio em 10 execuções foi de aproximadamente 2 segundos (2,183 segundos).

Tabela 1.

Media	10,8125
Desvio padrão	0,1849133431
Intervalo de confiança	0,1146083714
Intervalo inferior	10,697891629
Intervalo superior	10,927108371

Tabela 2.

Termos processados	7153
Media	20,2748
Desvio padrão	0,1742455483
Intervalo de confiança	0,1079965253
Intervalo inferior	20,166803475
Intervalo superior	20,382796525

Na Tabela 2 são mostrados os tempos obtidos para 10 execuções do módulo Analysis sobre o índice criado acima para a extração de uma matriz atributo-valor utilizando a medida estatística TF-IDF.

Foram feitas algumas buscas no índice criado com termos que estão presentes no vocabulário controlado e com alguns termos que não estão no vocabulário controlado. Como esperado, os termos não presentes no

vocabulário controlado não foram retornados como resultado de uma busca, pois não foram indexados. Para os termos pesquisados que estavam presentes no vocabulário controlado, alguns exemplos de resultados estão ilustrados na Figura 3.

Como ilustrado na Figura 3, a busca encontrou rapidamente os termos que estavam presentes no vocabulário controlado. O módulo Query Engine aplica os mesmos filtros que o Indexer aplicou na fase de criação do índice e neste caso é indiferente pesquisar com acentos, ou letras maiúsculas ou

```

Insira a busca: leite
Sua busca retornou 205 resultados.
Tempo: 2 ms
Insira a busca: abacaxi
Sua busca retornou 15 resultados.
Tempo: 2 ms
Insira a busca: ABÁCÂXÍ
Sua busca retornou 15 resultados.
Tempo: 1 ms
Insira a busca: Irrigação
Sua busca retornou 0 resultados.
Tempo: 12 ms

```

Figura 3. Resultado da busca no índice.

minúsculas. Quando a busca não encontra resultados, o tempo é um pouco maior pois ela tenta pesquisar termos similares utilizando o conceito de edit distance.

Considerações Finais

Com uma coleção relativamente grande de textos, os tempos de execução para uma única máquina com capacidade de processamento de um computador residencial atual foram bem baixos. Como trabalhos futuros, serão implementadas novas funcionalidades, tais como, o tratamento de sinônimos e relações taxonômicas entre termos para indexação e busca, bem como novas formas de extração da matriz atributo-valor, tais como agrupamentos de atributos por classes pré-determinadas e agrupamentos de documentos, por exemplo, por publicações referentes a um mesmo tema. Além disso, pretende-se melhorar o planejamento experimental para avaliar a performance da ferramenta.

Referências

APACHE SOFTWARE FOUNDATION. **Apache Lucene Core**. Disponível em: <<https://lucene.apache.org/core/>>. Acesso em: out. 2015.

KONCHADY, M. **Building search applications**: Lucene, LingPipe, and Gate. Oakton: Mustru Pub., 2008. 430 p. ill.

Extração de hierarquias de tópicos em textos para apoiar a construção de portfólios tecnológicos

Carolina Tavares de Oliveira¹

Luisa Miyashiro Tápias²

Stanley Robson de Medeiros Oliveira³

Maria Fernanda Moura⁴

Resumo: Neste trabalho são relatadas etapas para se extrair hierarquias de tópicos em coleções de documentos, compostas por publicações científicas, com o objetivo de auxiliar a construção de portfólios de tecnologias agrícolas diretamente relacionadas ao uso de recursos hídricos. Com base nas análises ao longo do processo, que é interativo, foi sendo construído um vocabulário controlado, que melhor representa o tema dos portfólios. Com esse vocabulário, que é utilizado para agrupar os documentos, os tópicos encontrados apresentaram um papel relevante na construção de tais portfólios.

Palavras-chave: mineração de textos, agrupamento de textos, recuperação de informação.

Introdução

O foco deste trabalho é construir portfólios de tecnologias para agricultura, nos quais são relatados os recursos tecnológicos empregados para viabili-

¹ Estudante de Engenharia Agrícola da Universidade Estadual de Campinas (Unicamp) bolsista da Embrapa Informática Agropecuária, Campinas, SP.

² Estudante de Engenharia Agrícola da Universidade Estadual de Campinas (Unicamp), estagiário da Embrapa Informática Agropecuária, Campinas, SP.

³ Doutor Ciências da Computação, Pesquisador Embrapa Informática Agropecuária, Campinas, SP.

⁴ Doutora em Ciências da Computação, Pesquisadora Embrapa Informática Agropecuária, Campinas, SP.

zar a produção de culturas de forma sustentável, voltando-se a atenção para a questão do uso da água no setor agrícola brasileiro; dado que, este é o principal usuário consuntivo do recurso se comparado ao setor industrial e ao de consumo doméstico. Os portfólios são planilhas elaboradas contendo as informações coletadas e organizadas a partir de uma coleção delimitada de textos. O trabalho aqui proposto visa automatizar ou auxiliar futuros processos de construção de portfólios a partir da literatura, utilizando técnicas de mineração de textos.

O processo de Mineração de Textos pode ser dividido em cinco etapas (MOURA, 2009): a) identificação do problema; b) pré-processamento; c) extração de padrões; d) pós-processamento; e) utilização do conhecimento. Na identificação do problema são definidos os objetivos do processo de Mineração de Textos. Na etapa de pré-processamento, os dados são manipulados a fim de se obter uma representação que possa ser lida por ferramentas de extração de padrões (ferramentas de aprendizado de máquina). Na extração de padrões, técnicas específicas são usadas de acordo com os resultados esperados. No caso dos tópicos, são analisados todos os ramos da hierarquia, procurando identificar os tópicos de maior interesse, que auxiliem a identificação de tecnologias e relações associadas a elas, tais como, tipo de solo, geolocalidade, etc. Ainda, com o arsenal de ferramentas utilizado, também pode-se observar a distribuição temporal de cada tópico em análise, ou seja, em que épocas aquelas tecnologias e suas características associadas estiveram em evidência. No pós-processamento, os resultados encontrados são analisados e validados. Se o processo não resultar em resultados usáveis, repetem-se as etapas, isto é, desde o pré-processamento.

O objetivo deste trabalho é utilizar técnicas de mineração de textos para extrair tópicos de hierarquias em textos visando à construção de portfólios de tecnologias agrícolas relacionadas ao uso de recursos hídricos.

Materiais e Métodos

A metodologia empregada foi composta de quatro fases: a) busca; b) pré-processamento; c) geração de hierarquia; d) pós-processamento (Figura 1), resultando em um processo retroalimentável.

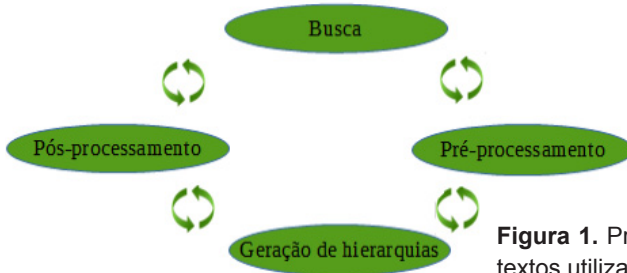


Figura 1. Processo de mineração de textos utilizado.

Busca: Os textos foram selecionados do Sistema Integrado e Aberto de Informação em Agricultura (SABIIA) (VACARI et al., 2011), este é um mecanismo de busca automatizado, que coleta metadados de provedores de dados científicos.

Pré-processamento: utilizou-se a ferramenta I-PreProc, em desenvolvimento na Empresa Brasileira de Pesquisa Agropecuária (Embrapa), para gerar uma matriz de termos (colunas) por documentos (linhas); considerando-se os termos de uma lista de vocábulos previamente fixados. Cada célula da matriz contém a frequência de ocorrência do vocábulo no texto. São gerados dois arquivos: o de extensão DAT com os valores das células (grau de importância de cada termo/palavra em cada documento) e o de extensão HDR com a descrição dos textos (nomes) e vocábulos (termos) utilizados.

Geração de hierarquias: geram-se dendrogramas, que são representações de agrupamentos hierárquicos, por meio do software torch⁵ que transforma a matriz em uma hierarquia (arquivo xml). Cada nó corresponde a um tópico com seus descritores, e, associados a cada tópico, os gráficos de distribuição temporal destes. Na sequência, utiliza-se a ferramenta topicVis (em desenvolvimento na Embrapa), que lê o xml gerado pela torch e cria arquivos nos padrões jsons e html, que podem ser visualizados utilizando-se um navegador.

Pós-processamento: analisam-se as hierarquias, verificando se os resultados são descartáveis ou utilizáveis, neste caso, na construção dos portfólios. Se o resultado ainda for insuficiente, dada esta validação subjetiva, então verifica-se a qual etapa retornar.

⁵ Disponível em: <<http://sites.labc.icmc.usp.br/torch/>>. Acesso em: 20 out. 2015.

A partir do uso deste processo, são relatados alguns resultados iniciais (portfólios tecnológicos) a partir dos quais fez-se uma primeira análise. Depois dessa primeira análise, verificou-se a necessidade de se buscar por novas fontes de dados (selecionar uma nova base de textos e melhorar os vocábulos e expressões de busca) para que fossem geradas novas hierarquias.

Resultados e Discussão

A partir da SABIIA foram reunidos 643 documentos e metadados, gerou-se dendrogramas a partir da coleção conforme a literatura. A partir das hierarquias realizou-se uma análise exploratória nos tópicos identificados, utilizando as informações apresentadas para a elaboração manual de uma primeira aproximação do portfólio de tecnologias, mostrado na Tabela 1.

Tabela 1. Representação de portfólios tecnológicos.

Portfólio de tecnologias								
Gráfico	Documento	Tópico	Localidade	Ano	Título	Cultura	Tecnologia	Tecnologia associada
Entre 09/05/1995 a 08/09/2001	0083.190.txt	Condutividade	Sete Lagoas MG	2013	Irrigação de pastagens	Forrageira	Irrigação por micro- aspersão	Fertirrigação
		Hidráulica						Localizada
		Censo						Aspersão
		Crédito						Capsula porosa
		Aviaria						Gotejamento
		Cevada						Infiltração
		Contrato						
...	

Além das hierarquias, verificavam-se os conteúdos dos textos em cada tópico e a distribuição temporal do tópico – picos desta. Uma parte da hierarquia é mostrada na Figura 2.

Depois da análise dos dados, verificou-se que poderia ser mais interessante ampliar o vocabulário do domínio, para que os descritores dos tópicos fossem mais significativos e abrangessem temas relacionados às tecnologias agrícolas associadas a recursos hídricos. Selecionou-se terminologias nas subáreas de geo localidades, tipos de solo, e tecnologias. Esses termos foram obtidos da planilha e expandidos com as relações do

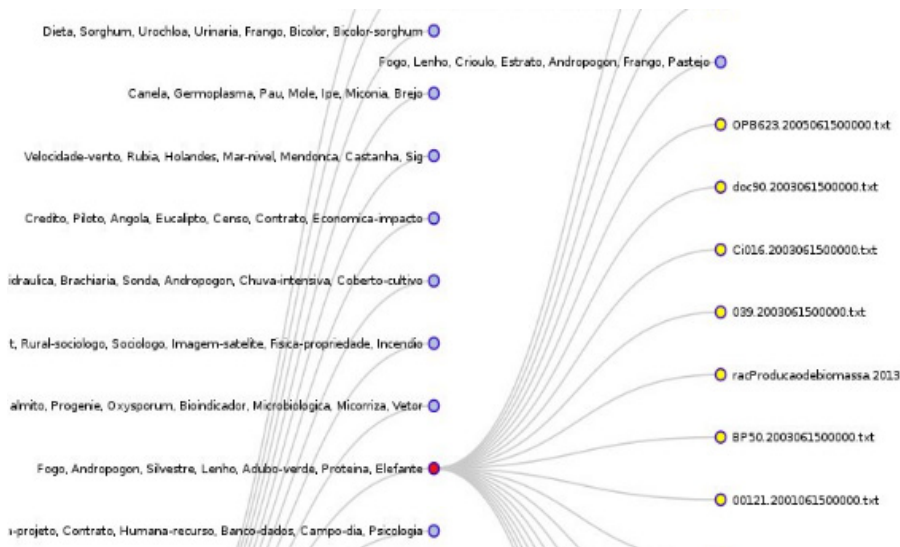


Figura 2. Tópicos construídos para os primeiros 643 documentos.

Thesagro⁶. Alguns termos obtidos são Irrigação por aspersão, Irrigação por microaspersão, Irrigação pivô central, Irrigação subterrânea, Irrigação por sulco, Irrigação por pote de barro, pivô central etc. De geo localidades, foram utilizadas as presentes nos textos, como Teresina, Salvador, Sete Lagoas, Campinas, etc. E, também foram incluídos só tipos de solos, como Latossolo, Neossolo, Organossolo, etc. Construiu-se um novo vocabulário, buscaram-se por novos textos e metadados no SABIIA, novas hierarquias foram geradas a partir da mudança do vocabulário. O resultado é mostrado na Figura 3.

Nesses diagramas cada ramo da árvore corresponde às palavras que funcionam como descritores dos tópicos, agrupando um conjunto de textos similares. Um possível tópico é apresentado na Figura 3. Na Figura 3 está em evidência o tópico descrito pelas palavras Melancia, Irrigação, Fertirrigação, Água, Adubo, Parnaíba; ao qual está associada uma série temporal, que indica a distribuição no temporal desse tópico em relação aos textos agrupados. Os resultados apresentam uma melhora subjetiva, como

⁶ Disponível em: <http://snida.agricultura.gov.br:81/binagri/html/Cen_Thes1.html>. Acesso em: 20 out. 2015.

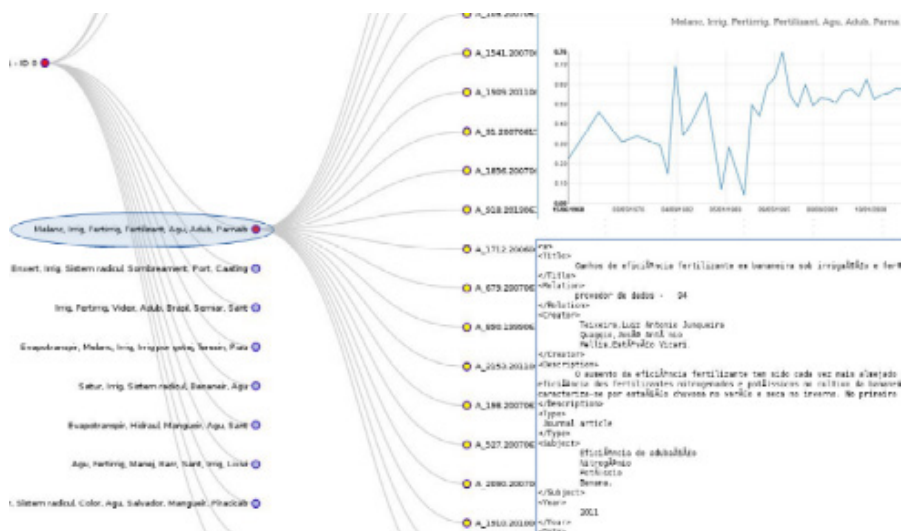


Figura 3. Distribuição temporal de um tópico e exemplo de documento.

se pode observar, na Figura 2, os tópicos contém termos, tais como nomes de tecnologias utilizadas no setor agrícola brasileiro, diretamente associadas a recursos hídricos.

Considerações Finais

O processo interativo utilizado está possibilitando que se obtenham melhores resultados, e mais especificidade na construção dos portfólios, isso foi observado repetindo o processo para se gerar novas hierarquias de tópicos.

Referências

MOURA, M. F. **Contribuições para a construção de taxonomias de tópicos em domínios restritos utilizando aprendizado estatístico**. 2009. 137 f. Tese (Doutorado em Ciências de Computação e Matemática Computacional) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, SP. Disponível em: <http://ainfo.cnptia.embrapa.br/digital/bitstream/item/17875/1/MFM_Tese_5318963-2.pdf>. Acesso em: 20 out. 2015.

VACARI, I.; VISOLI, M. C.; GONZALES, L. E. Acesso aberto a informação científica agropecuária na internet: caso do sistema aberto e integrado de informação em agricultura (Sabíia). In: CONGRESSO BRASILEIRO DE AGROINFORMÁTICA, 8., 2011. Bento Gonçalves. **Anais...** Florianópolis: UFSC; Pelotas: UFPel, 2011. Não paginado. Disponível em: <<http://ainfo.cnptia.embrapa.br/digital/bitstream/item/46253/1/89811-1.pdf>>. Acesso em: 20 out. 2015.

Variabilidade interanual das áreas com potencial de risco de desertificação da Bahia

Camila da Silva Dourado¹

Stanley Robson de Medeiros Oliveira²

Ana Maria Heuminski de Avila³

Resumo: O objetivo deste trabalho foi analisar a expansão ou contração das áreas susceptíveis à desertificação (ASDs) na Bahia. Para isto, foram utilizados dados de precipitação e temperatura, para o período de 2000 a 2014, de 25 estações meteorológicas. A partir disto, foram calculadas a Evapotranspiração Potencial (ETP), pelo método de Thornthwaite e o Índice de Aridez (IA) para cada ano. As ASDs foram determinadas a partir de valores do IA. Para mapear a variabilidade interanual das ASDs, foi aplicado o método geostatístico Inverse Distance Weighted (IDW). Os resultados evidenciaram que os baixos volumes de chuva influenciaram diretamente na variabilidade das ASDs.

Palavras-chave: degradação de terras, indicadores climáticos, índice de aridez, geotecnologias.

Introdução

A desertificação é definida pela Convenção das Nações Unidas de Combate à Desertificação como a degradação de terras nas zonas áridas, semiáridas

¹ Eng^a Agrônoma, Doutoranda pela Faculdade de Engenharia Agrícola/Unicamp, Campinas, SP.

² Bacharel em Ciência da Computação, PhD em Ciências da Computação, pesquisador da Embrapa Informática Agropecuária, Campinas, SP.

³ Bacharel em Meteorologia, Doutora em Engenharia Agrícola, pesquisadora do Centro de Pesquisas Meteorológicas e Climáticas Aplicadas à Agricultura/Unicamp, Campinas, SP.

e subúmidas a seca, resultante das variações climáticas e ações antrópicas (BRASIL, 1994). Tal fenômeno transforma terras agricultáveis em terras improdutivas e inférteis, causa impactos ambientais como a destruição da biodiversidade, diminuição da disponibilidade de recursos hídricos e provoca a perda física e química dos solos.

O critério internacionalmente estabelecido para delimitação das áreas susceptíveis à desertificação é o IA definido por Thornthwaite (1948), o qual é dado pela razão entre a precipitação e a evapotranspiração potencial. No Brasil, segundo o Programa de Ação Nacional de Combate à Desertificação e Mitigação dos Efeitos da Seca (PAN - Brasil) (BRASIL, 2007), as áreas consideradas com potencial de risco de desertificação concentram-se na região Nordeste do país. A semiaridez existente nestes estados é resultante de fenômenos puramente naturais, mas está se intensificando por causas antrópicas, devido uma exploração mais intensa nos recursos naturais, com modelos de desenvolvimento imediatistas (SILVA, 2013).

Em particular, o estado da Bahia possui 258 municípios pertencentes a uma região semiárida, compreendendo uma área de 388.274 km², o que corresponde a 68% do seu território. Os estudos de Dourado (2013) e Santos et al. (2010) indicam que esta região vem sofrendo com fortes variabilidades climáticas, o que favorece a expansão das áreas com risco de desertificação. Este cenário se torna um desafio para diversos setores econômicos e sociais da região, principalmente o setor agropecuário.

Neste contexto, o objetivo deste trabalho foi identificar áreas com potencial de risco de desertificação no estado da Bahia, por meio do índice de aridez, bem como, analisar a variabilidade interanual dessas áreas para o período de 2000 a 2014.

Materiais e Métodos

A área de estudo corresponde ao estado da Bahia, localizado entre as latitudes 8°32'00"S e 18°20'07"S e as longitudes 37°20'37"W a 46°36'54"W. Dividido geopoliticamente em sete mesorregiões: Centro Norte Baiano (CNB), Centro Sul Baiano (CSB), Vale São-Franciscano da Bahia (VSFB), Nordeste Baiano (NB), Sul Baiano (SB), Extremo Oeste Baiano (EOB) e Metropolitana de Salvador (MS), conforme Figura 1.

Área de estudo: estado da Bahia

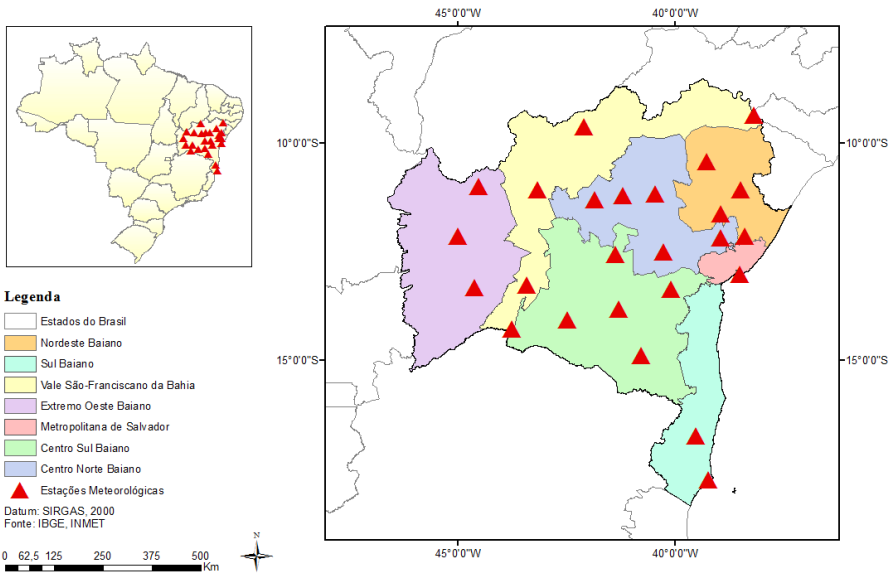


Figura 1. Espacialização das estações meteorológicas nas sete mesorregiões da Bahia.

Os dados climatológicos (precipitação, temperatura máxima e mínima) foram obtidos de 25 postos meteorológicos, fornecidos pelo Instituto Nacional de Meteorologia (Inmet) (INMET, 2015). Para espacialização dos dados, foi adotado o sistema de coordenadas geográficas Latitude e Longitude, Datum SIRGAS 2000, e sistema de projeção Universal Transversa de Mercator (UTM), conforme Figura 1.

As médias anuais de temperatura média e precipitação, entre 2000 e 2014, foram utilizadas na determinação da EPT, pelo método de Thornthwaite (1948), pelo programa “BHnorm” desenvolvido por Rolim et al. (1998). Para identificar os diferentes níveis de risco de desertificação foi adotado o critério do Índice de Aridez, estabelecido por Thornthwaite (1948) e utilizado pelas Nações Unidas para aplicação do Plano de Combate à Desertificação (PACD). O IA é definido como: $IA = P/ETP$, em que P representa as precipitações médias anuais e a ETP, a evapotranspiração potencial média anual. Quanto menor for o valor da relação, maior é a aridez, conforme a Tabela 1.

Tabela 1. Faixas climáticas relacionadas ao índice de aridez.

Classificação	Índice de Aridez (IA)
Hiperárido	< 0,05
Árido	0,051 a 0,20
Semiárido	0,21 a 0,50
Subúmido a seco	0,51 a 0,65
Subúmido úmido a úmido	> 0,65

Para confecção dos mapas das áreas de risco de desertificação da Bahia, foram interpolados dados do IA, por meio do IDW, método geoestatístico, Ponderação do Inverso da Distância. Este método, de acordo com Chen et al. (2012) e Dourado et al. (2013) é uma das técnicas de interpolação mais usadas para pontos espalhados espacialmente, apresentando como vantagem a sua utilidade quando a distribuição dos parâmetros estimados não é uma distribuição normal.

Resultados e Discussão

Verifica-se, na Figura 2, a distribuição do IA nas mesorregiões do território baiano, delimitando as áreas com potencial de risco de desertificação no estado. A partir da análise desses índices, pode-se inferir que as áreas de risco do estado estão classificadas em Árido (0,05 a 0,20), Semiárido (0,21 a 0,50), Subúmido a seco (0,51 a 0,65) e Subúmido úmido a úmido (> 0,65).

Observa-se, em todos os anos estudados (2000 a 2014), que grande parte da extensão territorial é classificada como área de risco de desertificação, apresentando variações de potencial de risco em alto, moderado, baixo e nenhum. Os índices variaram de 0,1, registrado no ano de 2012, a 1,7 registrado no ano de 2006.

A Figura 2 mostra que as mesorregiões com os maiores IA foram a MS e a SB. Estas áreas são classificadas como subúmido úmido a úmido, não apresentando risco de desertificação. A maioria das localidades dessas mesorregiões encontra-se na costa litorânea e apresentam altos índices pluviométricos.

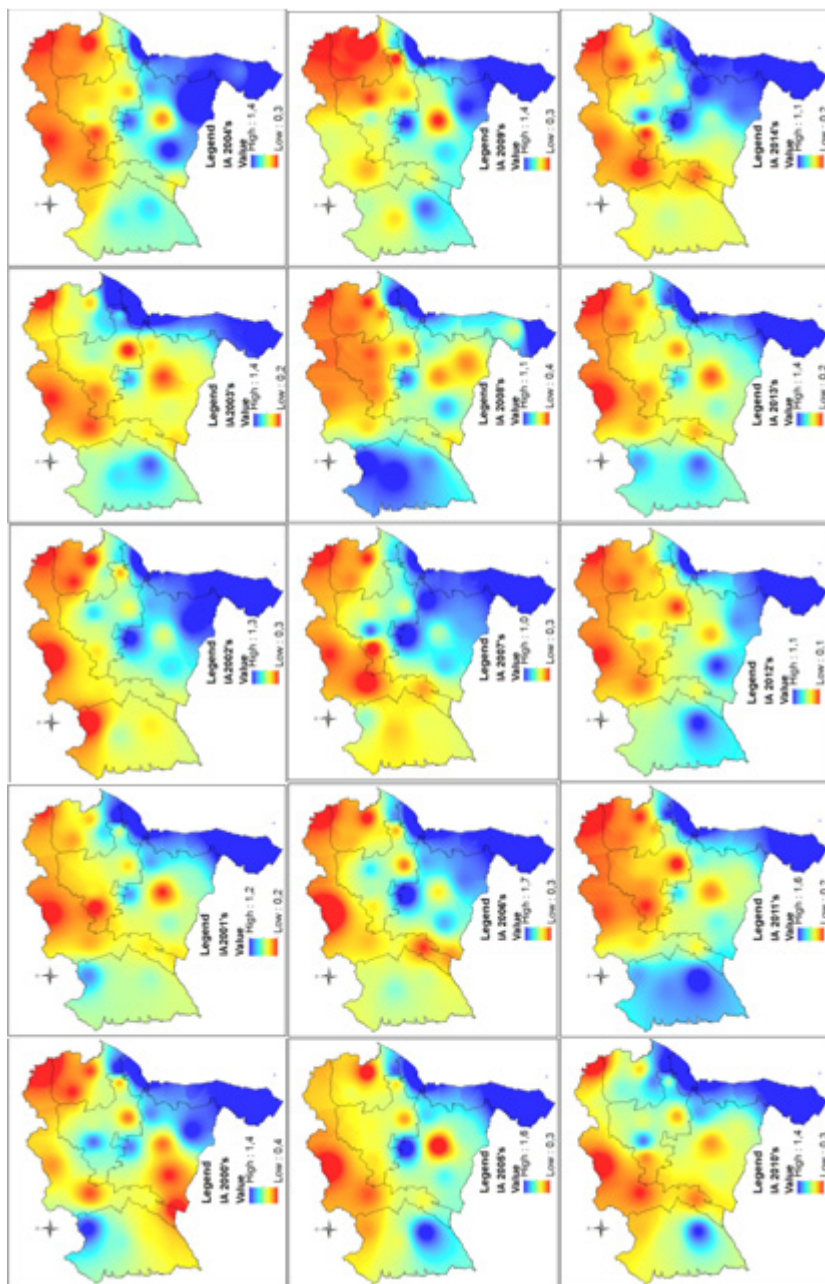


Figura 2. Interpolação dos índices de aridez e delimitação das áreas com potencial de risco de desertificação do estado da Bahia para os anos de 2000 a 2014.

Os menores índices de aridez, variando entre 0,1 a 0,61, estão nas mesorregiões do VSFB, NB e CNB, que apresentam os menores volumes de chuva do estado (DOURADO et al., 2013). Dessas mesorregiões, a VSFB é a que apresenta o menor IA em todos os anos da série. O norte da VSFB é responsável pelo polo da fruticultura do estado, sendo responsável por 10% da produção da fruticultura nacional (IBGE, 2015).

A mesorregião EOB, formada por localidades do polo agrícola do estado (CONAB, 2013; IBGE, 2015), apresentou nos anos de 2002, 2007, 2012 e 2014 IA mais baixos, sendo considerada, nestes anos, como área de risco de desertificação moderado. Isso se deve aos baixos volumes pluviométricos ocorridos nestes anos.

Considerações Finais

Os menores índices de aridez são referentes aos anos com baixos volumes de chuva. Com destaque para o ano de 2012, cujo índice variou de 0,1 a 1,1, classificando o norte da mesorregião do VSFB como área de risco de desertificação muito alta.

As mesorregiões Metropolitana de Salvador e Sul Baiano não apresentaram variabilidade em sua classificação de área de baixo ou nenhum risco de desertificação.

A mesorregião Extremo Oeste, grande polo agrícola do estado, variou da classe de risco moderado para alto risco.

Referências

BRASIL. Ministério do Meio Ambiente. **Atlas das áreas susceptíveis à desertificação do Brasil**. Organizador do texto: Marcos Oliveira Santana. Brasília, 2007. 134 p.

BRASIL. Ministério do Meio Ambiente. **Convenção Internacional de Combate à Desertificação nos países afetados por seca e/ou desertificação**. Brasília, 1994. Disponível em: <http://www.mma.gov.br/port/sdi/ea/documentos/convs/conv_desertificacao>. Acesso em: mar. 2015.

CHEN, F. W.; LIU, C. W. Estimation of the spatial rainfall distribution using inverse distance weighting (IDW) in the middle of Taiwan. **Paddy and Water Environment**, v. 10, p. 209-222, Feb. 2012.

CONAB (Brasil). **Acompanhamento da safra brasileira de grãos - safra 2011/2012**. Disponível em: <http://www.conab.gov.br/boletim_marco_2012.pdf>. Acesso em: fev. 2013.

DOURADO, C. da S. **Mineração de dados climáticos para análise de eventos extremos de precipitação**. 2013, 131 p. Dissertação (Mestrado em Engenharia Agrícola), Faculdade de Engenharia Agrícola, Campinas, SP.

DOURADO, C. da S.; OLIVEIRA, S. R. M. de; AVILA, A. M. H. de. Análise de zonas homogêneas em séries temporais de precipitação no Estado da Bahia. **Bragantia**, v. 72, n. 2, p. 192-198, abr./jun. 2013.

IBGE. **Clima da Bahia**. Disponível em: <www.ibge.gov.br>. Acesso em: 15 mar. 2015.

INMET (Brasil). **Inmet**. [2015]. Disponível em: <<http://www.inmet.gov.br>>. Acesso em: 20 fev. 2015.

ROLIM, G. S.; SENTELHAS, P. C.; BARBIERI, V. Planilhas no ambiente Excel para os cálculos de balanços hídricos: normal, sequencial, de cultura e de produtividade real e potencial. **Revista Brasileira de Agrometeorologia**, v. 6, n. 1, p. 133-137, 1998.

SANTOS, D. N. dos; SILVA, V. de P. R. da; SOUSA, F. de A. S.; SILVA, R. A. Estudo de alguns cenários climáticos para o Nordeste do Brasil. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 14, n. 5, 2010.

THORNTHWAITE, C. W. An approach toward a rational classification of climate. **Geographical Review**, v. 38, p. 55-94, Jan. 1948.

UNITED NATIONS ENVIRONMENT PROGRAMME. **World atlas of desertification**. London: Edward Arnold, 1992. 69 p.

Identificação de Pontos Perceptualmente Importantes (PIP) em séries temporais de tópicos extraídos de dados textuais

Lucas Santiago Rodrigues¹

Roberta Akemi Sinoara²

Solange Oliveira Rezende³

Ricardo Marcondes Marcacini⁴

Maria Fernanda Moura⁵

Resumo: Neste trabalho é apresentado um módulo computacional denominado PIPC (PIP Classification) que permite identificar Pontos Perceptualmente Importantes (PIP) em séries temporais. O módulo foi desenvolvido para apoiar o projeto Compilação e Recuperação de Informações Técnico-científicas e Indução ao Conhecimento (CRITIC@), permitindo identificar os pontos relevantes da evolução temporal de um tópico extraído dos textos, identificar documentos textuais que possam auxiliar a interpretar tais pontos, bem como classificar a formação de próximos PIPs nas séries temporais. Foram realizados testes do módulo a partir de notícias sobre produção de milho no Brasil, e os resultados preliminares de avaliação do módulo são promissores.

Palavras-chave: séries temporais, extração de tópicos, classificação.

¹ Estudante de Sistemas de Informação da Universidade Federal de Mato Grosso do Sul - Campus de Três Lagoas (UFMS), estagiário da Embrapa Informática Agropecuária, Campinas, SP.

² Bacharel em Informática, doutoranda em Ciências da Computação e Matemática Computacional no Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo (ICMC-USP), São Carlos, SP.

³ Graduada em Licenciatura em Ciências Habilitação Matemática, doutora em Engenharia Mecânica, professora associada da Universidade de São Paulo (ICMC-USP), São Carlos, SP.

⁴ Bacharel em Informática, doutor em Ciências da Computação e Matemática Computacional, docente e pesquisador da Universidade Federal de Mato Grosso do Sul (UFMS), Três Lagoas, MS.

⁵ Estatística, doutora em Ciências Matemáticas e da Computação, pesquisadora da Embrapa Informática Agropecuária, Campinas, SP.

Introdução

A extração automática de tópicos a partir de dados textuais tem recebido grande atenção nos últimos anos, uma vez que os tópicos permitem sumarizar, de forma inteligente, o conhecimento embutido em grandes conjuntos de dados textuais (REZENDE et al., 2003). Assim, é possível que usuários explorem coleções textuais, que são inerentemente não estruturadas, de uma forma organizada e intuitiva (MOURA, 2009). Uma aplicação promissora para esse tipo de tarefa é demonstrada no projeto CRITIC@, em que os resultados retornados por motores de busca sobre informação técnico-científicas são organizados por meio de tópicos. Além disso, um diferencial do projeto CRITIC@ é analisar a “evolução temporal” do tópico, ou seja, computar uma série temporal que permite visualizar os períodos em que um tópico é mais ou menos ativo (MARCACINI; REZENDE, 2010).

Nesse trabalho foi desenvolvido um módulo em linguagem de programação JAVA para identificação e classificação de PIP em séries temporais de tópicos extraídos a partir de dados textuais, denominada PIPC (PIP Classification). Os PIP's são definidos como pontos críticos de determinada série temporal, uma vez que representam pontos com valores muito diferentes de seus vizinhos, geralmente picos e vales de uma série. No módulo PIPC, foi implementado um algoritmo tradicional da literatura para identificação dos PIP's, baseada em distância Euclidiana para selecionar os pontos de nosso interesse. Além de identificar os PIP's, o módulo desenvolvido possui duas funções: a) identificar um documento textual mais relevante associado ao PIP de uma série; e b) prever a probabilidade de ocorrência de um PIP nos próximos pontos da série por meio de algoritmos de classificação. Dessa forma, no contexto do projeto CRITIC@, os usuários podem analisar as séries temporais com um nível mais rico de detalhes e obter a tendência de formação de ponto perceptualmente importante no futuro. Os experimentos de avaliação da ferramenta foram realizados em tópicos extraídos de textos sobre a produção de Milho no Brasil e se mostraram promissores.

Materiais e Métodos

O módulo PIPC (PIP Classification) desenvolvido durante o estágio é ilustrado na Figura 1. Inicialmente, recebe, como entrada, uma série temporal de

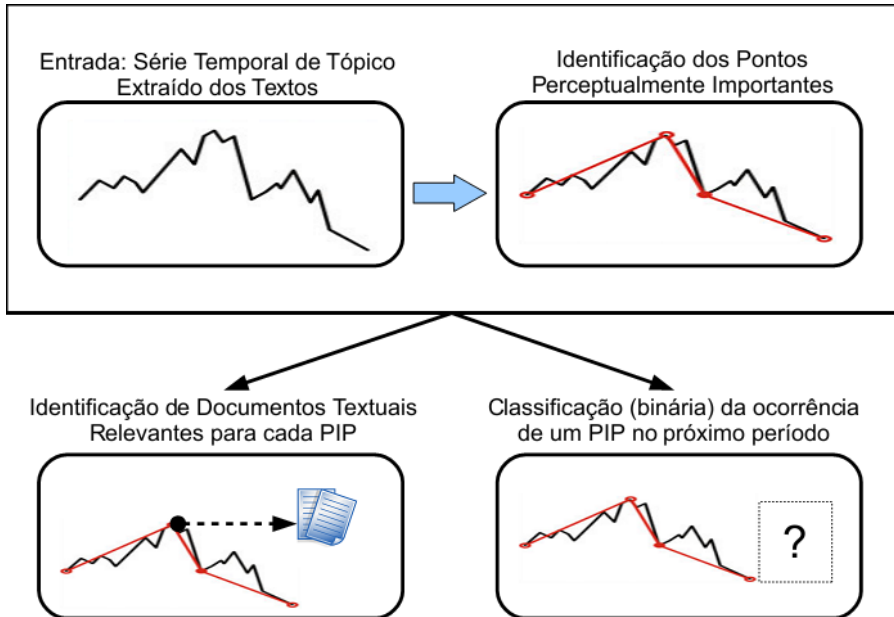


Figura 1. Fluxograma do módulo DATool.

tópico extraído dos textos. Esta série de entrada é um formato XML utilizado pelo CRITIC@, com os descritores dos tópicos, os documentos associados e a série temporal. No próximo passo é executado um algoritmo para identificação dos PIP's, conforme descrito em Sanches (2006). A ideia geral deste algoritmo utiliza a distância euclidiana entre os pontos da série temporal, inicializando-se com o primeiro e o último ponto da série. A partir desses dois pontos iniciais, calcula-se o ponto mais distante entre os dois e este é definido como um PIP. Este mesmo procedimento é executado recursivamente para cada dois PIPs anteriores da série. Um exemplo da execução desse algoritmo é apresentado na Figura 2.

Após a identificação dos PIPs da série temporal, o módulo PIPC permite a execução de duas funcionalidades, conforme descritas a seguir.

A primeira funcionalidade é a Identificação de documentos textuais relevantes para cada PIP, na qual o usuário pode selecionar um PIP na série temporal e visualizar qual documento textual representa aquele período. Para obter este resultado, foi utilizado o conceito de centroide, geralmente empregado na área de agrupamento de dados. Para tal, é obtido o conjunto

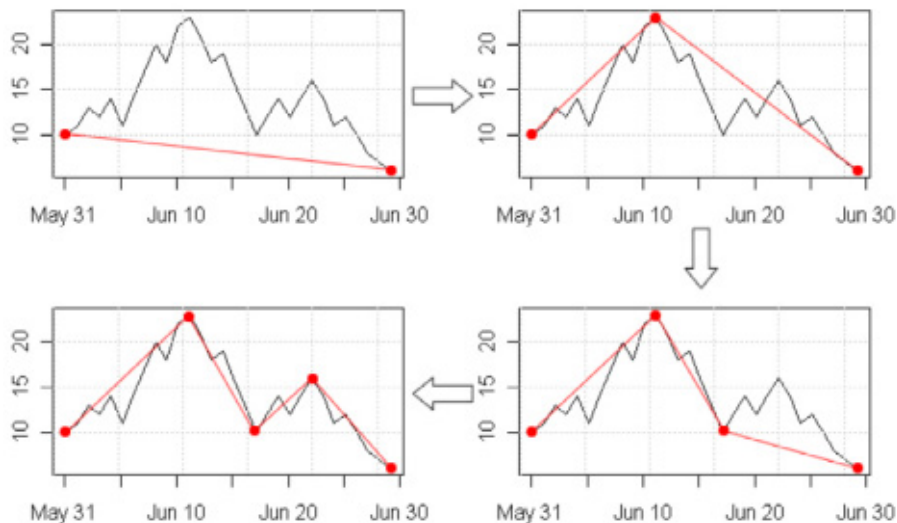


Figura 2. Exemplo de Identificação de Pontos Perceptualmente Importantes. No primeiro quadro são selecionados os pontos iniciais. Nas próximas iterações são selecionados os pontos mais distantes para cada dois pontos consecutivos.

Fonte: Sanches (2006).

de documentos publicados no período temporal de um PIP de interesse a partir do XML de entrada. A partir desse conjunto, é calculado o vetor centroide, que é um vetor médio que sumariza todos os documentos do conjunto. Em seguida, calcula-se a similaridade de cada documento do período a este centroide por meio da similaridade cosseno. O documento textual que apresenta maior similaridade ao centroide é então considerado o mais relevante do período. Desse modo, o usuário pode interagir com cada PIP e utilizar esses documentos mais relevantes como uma opção para interpretar e entender o motivo dos picos e vales existentes na série temporal.

A segunda funcionalidade é Classificação (binária) da ocorrência de um PIP no próximo período da série temporal do tópico. Nesse caso, a série temporal é modelada em uma nova tabela atributo-valor por meio de suas subsequências. Assim, definido o tamanho da subsequência, busca-se todas as subsequências da série temporal em que o último ponto da subsequência é um PIP. Essas subsequências são definidas como classe POSITIVA, ou seja, que representam a ocorrência de um PIP. Todas as outras subsequências da série temporal que não possuem PIP são definidas como classe

NEGATIVA. Por meio dessa estrutura, é possível treinar classificadores binários para que, dada uma nova subsequência, retorne à classe (positiva ou negativa) e ao respectivo valor de confiança da classificação. Essas informações são utilizadas para identificar uma possível tendência do próximo ponto para apoiar os usuários na análise das séries temporais de tópicos extraídos dos textos. Para a análise dos PIP's e dos documentos utilizamos a biblioteca Weka, que possui algoritmos de análises e foi incorporada aos códigos desenvolvidos durante a pesquisa.

Resultados e Discussão

A avaliação do módulo PIPC foi feita no domínio “Produtividade de Milho no Brasil”. Este domínio foi escolhido devido à sua crescente demanda com o setor de avicultura, suinocultura e consumo industrial. Os meios de comunicação do Brasil divulgam com grande destaque o setor agrícola, dessa forma, temos incontáveis quantidades de dados que foram publicados durante décadas até a atualidade. Atualmente, com os avanços da tecnologia, esses dados podem ser usados por instituições e empresas para gerar conhecimento, sendo possível analisar vários indicadores de uma safra. No caso particular desse estudo, foi analisada a safra do milho.

Foram utilizados 5674 documentos textuais sobre produtividade de milho coletados de diversos portais especializados. Para os experimentos, foi selecionado um tópico em que os descritores “[aumento, produção, milho]” (e suas variações morfológicas) estão presentes com maior frequência, no período de 2008 à 2010, conforme ilustrado na Figura 3. Na figura também há um PIP selecionado para analisar a funcionalidade sobre identificação e notícias relevantes para o PIP. Neste período (entre maio e junho de 2009), o referido tópico contém notícias sobre expectativas e tendências de alta de produtividade no setor, e a notícia identificada mais próximo ao centroide é



Figura 3. Evolução temporal do tópico [aumento, produção, milho] no período de 2008 à 2010.

“Tendência é de alta nas exportações de milho”⁶. Esta análise é subjetiva e uma validação mais robusta depende de especialistas de domínio da área.

Para avaliar a classificação de ocorrência de um PIP no próximo período foi utilizado o algoritmo KNN, com a seguinte configuração: a) correlação de Spearman como critério de proximidade; b) número de vizinhos mais próximos (k) igual a 1; c) tamanho da subsequência (w) igual a 15 dias. O conjunto de treinamento foi construído utilizando 50 exemplos da classe positiva e 50 da classe negativa. A estratégia de avaliação selecionada foi a *leave-one-out*, obtendo uma taxa de acerto em torno de 78%. Para a série temporal em questão, a classe retornada foi negativa.

Considerações Finais

Neste trabalho foi apresentado o módulo PIPC (PIP Classification) desenvolvido durante o trabalho de estágio. É importante ressaltar que durante o estágio foram estudados e implementados algoritmos existentes na literatura e adaptados ao contexto do projeto CRITIC@, para fins de pesquisa e desenvolvimento. Outra observação é que os algoritmos foram empregados para a extração e análise computacional de séries de valores brutos, sem qualquer tratamento, a fim de que fosse possível utilizar o módulo para qualquer série temporal a ser escolhida.

Os trabalhos futuros incluem uma avaliação mais robusta considerando diferentes domínios de aplicação com apoio de especialistas de domínio.

Referências

MARCACINI, R. M.; REZENDE, S. O. Torch: a tool for building topic hierarchies from growing text collection. In: WORKSHOP ON TOOLS AND APPLICATIONS, 9.; BRAZILIAN SYMPOSIUM ON MULTIMEDIA AND THE WEB, 8., 2010, Belo Horizonte. **Proceedings...** Belo Horizonte: UFMG, 2010. p. 133-135. Webmedia.

⁶ Disponível em: <<http://www.abramilho.org.br/noticias.php?cod=463>>.

MOURA, M. F. **Contribuições para a construção de taxonomias de tópicos em domínios restritos utilizando aprendizado estatístico**. 2009. 137 f. Tese (Doutorado em Ciências de Computação e Matemática Computacional) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, SP.

REZENDE, S. O.; PUGLIESI, J. B.; MELANDA, E. A.; PAULA, M. F. de. *Mineração de dados*. In: REZENDE, S. O. (Ed.). **Sistemas inteligentes: fundamentos e aplicações**: Manole, 2003. p. 307-335.

SANCHES, R. A. **Redução de dimensionalidade em séries temporais**. 2006. 92 p. Dissertação (Mestrado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Carlos, SP.

Levantamento de conteúdo para o sistema WebAmbiente

Meiry Helen Bordim¹

Alan Massaru Nakai²

Resumo: O WebAmbiente é um sistema que está sendo desenvolvido pela Empresa Brasileira de Pesquisa Agropecuária (Embrapa), em cooperação com o Ministério do Meio Ambiente (MMA), cujo objetivo é auxiliar o produtor rural na recuperação de seus passivos ambientais por meio do acesso a informações relacionadas a espécies nativas e métodos de recuperação. O objetivo deste trabalho foi realizar levantamento bibliográfico e elaborar conteúdo didático e simplificado para auxiliar os usuários a interagirem com o sistema.

Palavras-chave: código florestal, recomposição de áreas de preservação permanente e reserva legal, preservação e recuperação de florestas.

Introdução

O Novo Código Florestal - Lei nº 12.651 de 25/05/2012 - estabelece limites de uso das áreas dos imóveis rurais para que se mantenha o equilíbrio entre as dimensões ambiental e econômica na exploração agropecuária. A lei refere-se à proteção e preservação de florestas, matas ciliares, Áreas de Preservação Permanente (APP) e Reserva Legal (INSERNHAGEN, 2009 ; NOVO CÓDIGO FLORESTAL, 2012). Além disso, regulamenta a exploração do uso da terra no Brasil, apoiado no fato de que este é um bem comum que deve ser devidamente utilizado e protegido.

¹ Estudante de Engenharia Ambiental e Sanitária da Pontifícia Universidade Católica de Campinas (PUC-Campinas), estagiário da Embrapa Informática Agropecuária, Campinas, SP.

² Doutor em Ciência da Computação, analista da Embrapa Informática Agropecuária, Campinas, SP.

Neste contexto, a Embrapa, em cooperação com MMA, está desenvolvendo o sistema WebAmbiente que tem o objetivo de apoiar o produtor rural na recuperação de seus passivos ambientais, provendo acesso a informações relativas às espécies nativas e aos métodos de recuperação mais adequados para a situação do passivo em questão.

O projeto do WebAmbiente prevê um módulo de sistema especialista por meio do qual o usuário deverá informar a situação de sua propriedade. Com base nisto, o sistema retornará às informações pertinentes para a situação relatada. Entretanto, o usuário pode não ter o conhecimento necessário para responder aos questionamentos do sistema especialista, sendo necessário prover conteúdo explicativo que o auxilie nesta tarefa. O objetivo deste estágio é apoiar o levantamento do conteúdo que subsidiará o desenvolvimento do sistema.

Materiais e Métodos

O levantamento de conteúdo foi realizado a partir de uma lista de termos obtidos de esquemas organizacionais fornecidos pela equipe do projeto. A realização do estágio envolveu as seguintes atividades:

- Pesquisa bibliográfica.
- Elaboração de conteúdo conciso a respeito dos termos pesquisados.
- Busca de imagens ilustrativas e catalogação de suas fontes.

As principais fontes de consulta para levantamento das definições dos termos foram:

- Manual técnico da vegetação brasileira (MANUAL..., 2012): reúne informações sobre o sistema fitogeográfico, inventário das formações florestais e campestres, técnicas de manejo de coleções botânicas e procedimentos para mapeamento da vegetação em escala regional.
- Pacto pela restauração da Mata Atlântica (INSERNHAGEN, 2009): estudo aprofundado dos termos referentes à restauração florestal, para recuperação de ecossistemas degradados.
- Técnicas de restauração de áreas degradadas (SOARES, 2009): trabalho acadêmico que descreve a problemática do desmatamento e projetos de restauração de um ecossistema degradado.

- Plano de restauro florestal (INSTITUTO BRASILEIRO DE FLORESTAS, 2011): apresenta os passos necessários para restauração florestal, desde o reconhecimento da área até o pós-plantio.
- Práticas de conservação do solo e recuperação de áreas degradadas (WASDT, 2003): trata da importância da fertilidade do solo e conceitua processos de degradação para definir práticas de conservação de solos.
- Atributos físicos de solos e distribuição das fitofisionomias de Cerrado na Bacia Hidrográfica do Rio Jardim - (SPERA et al., 2005): características físico-hídricas de cada tipo de solo e suas respectivas relações com as principais fitofisionomias correspondentes.

Resultados e Discussão

Até o momento, o trabalho gerou conteúdo sobre aproximadamente 70 termos. Os termos pesquisados estão relacionados a diversos temas, tais como: a) adubação; b) degradação do solo; c) métodos de reflorestamento; d) biomas e suas fisionomias, entre outros. Além disso, foram realizadas apresentações para a equipe de tecnologia da informação do projeto com o objetivo de contextualizar os seus integrantes.

A Figura 1 apresenta exemplos de telas do sistema WebAmbiente que foram criadas a partir do levantamento de termos realizado pela estagiária:

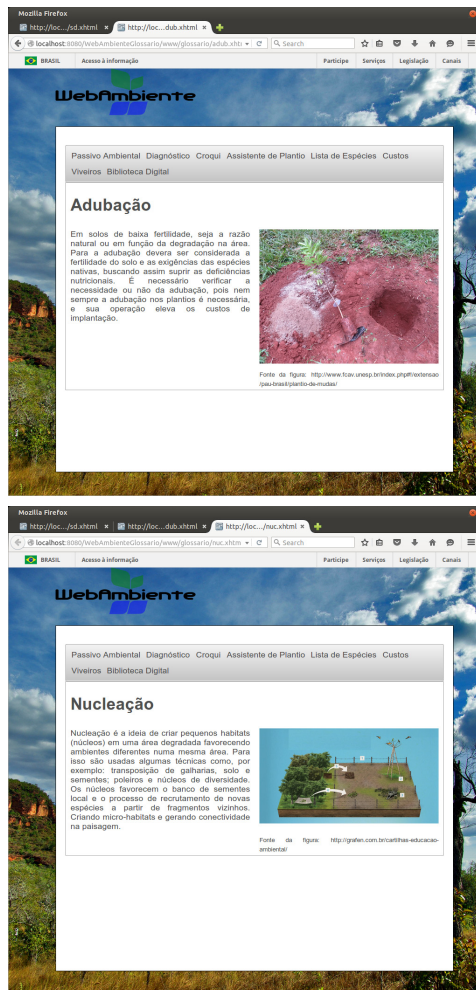


Figura 1. Exemplos de telas geradas a partir do levantamento de termos.

Considerações Finais

Os termos pesquisados estão sendo incluídos no sistema, cuja elaboração ainda está em desenvolvimento e em processo de ajustes e mudanças. Em uma próxima etapa, especialistas da área ambiental serão convidados a validar o conteúdo gerado. Ainda há muitos termos a serem pesquisados e o trabalho continua em andamento.

Referências

INSERNHAGEN, I.; BRANCALION, P. H. S.; RODRIGUES, R. R.; GANDOLFI, S.; RODRIGUES, R. T. **Pacto pela restauração da Mata Atlântica**: Referencial dos conceitos e ações de restauração florestal. São Paulo: Instituto BioAtlântica, 2009. Disponível em: <<http://www.pactomataatlantica.org.br/pdf/referencial-teorico.pdf>>. Acesso em: 01 out. 2015.

INSTITUTO BRASILEIRO DE FLORESTAS. **Plano de restauro florestal**. 2011. Disponível em: <<http://www.brasil.diplo.de/contentblob/3285776/Daten/1609154/PlanoRestauro.pdf>>. Acesso em: 19 out. 2015.

MANUAL técnico da vegetação brasileira. 2. ed. rev. ampl. Rio de Janeiro: IBGE, 2012. 274 p. il. (Manuais técnicos em geociências, n. 1). Disponível em: <ftp://geoftp.ibge.gov.br/documentos/recursos_naturais/manuais_tecnicos/manual_tecnico_vegetacao_brasileira.pdf>. Acesso em: 29 set. 2015.

NOVO Código Florestal. Curitiba: Sistema Faep, 2012. 83 p. il. Disponível em: <<http://codigoflorestal.sistemafaep.org.br/wp-content/uploads/2012/11/novo-codigo-florestal.pdf>>. Acesso em: 1 out. 2015.

SOARES, S. M. P. **Técnicas de restauração de áreas degradadas**. Juiz de Fora: UFJF, 2009. Disponível em: <http://www.ufjf.br/ecologia/files/2009/11/estagio_Silvia_Soares1.pdf>. Acesso em: 22 out. 2015.

SPERA, S. T.; REATTO, A.; MARTINS, E. de S.; CORREIA, J. R. **Atributos físicos de solos e distribuição das fitofisionomias de cerrado na bacia hidrográfica do Rio Jardim, DF**. Planaltina, DF: Embrapa Cerrados, 2005. 18 p. (Embrapa Cerrados. Boletim de Pesquisa e Desenvolvimento, 146). Disponível em: <http://bbeletronica.cpac.embrapa.br/versaomodelo/html/2005/bolpd/bolpd_146.shtml>. Acesso em: 19 out. 2015.

WADT, P. G. S. (Ed.). **Práticas de conservação do solo e recuperação de áreas degradadas**. Rio Branco, AC: Embrapa Acre, 2003. 29 p. (Embrapa Acre. Documentos, 90). Disponível em: <<http://ainfo.cnptia.embrapa.br/digital/bitstream/CPAF-AC/7611/1/doc90.pdf>>. Acesso em 19 out. 2015.

Extração de regras de associação em textos para apoiar a construção de portfólios tecnológicos

Luisa Miyashiro Tápias¹

Carolina Tavares de Oliveira²

Stanley Robson Medeiros Oliveira³

Maria Fernanda Moura⁴

Resumo: O objetivo deste trabalho foi explorar o uso de técnicas para extração de regras de associação em um conjunto de textos para auxiliar a construção de portfólios tecnológicos. A metodologia adotada foi constituída de um processo de busca de documentos na literatura e um processo de mineração de textos, com a intenção de obter regras que indiquem relações entre tecnologias existentes, geolocalidades, tipos de solos, entre outras. Os resultados dos experimentos revelaram a necessidade de mudanças no processo de geração de regras, pois o formato e a quantidade dos dados utilizados resultou em matrizes altamente esparsas, com muitas classes e poucos atributos, o que não nos permitiu obter inferências com um nível de precisão aceitável. Nos próximos passos serão realizados ajustes nos experimentos para que as regras de associação geradas apresentem um nível de precisão aceitável, em termos de suporte e confiança.

Palavras-chave: expressão de busca, vocabulário controlado, Apriori.

¹ Estudante de Engenharia Agrícola da Universidade Estadual de Campinas (Unicamp), estagiária da Embrapa Informática Agropecuária, Campinas, SP.

² Estudante de Engenharia Agrícola da Universidade Estadual de Campinas (Unicamp), estagiária da Embrapa Informática Agropecuária, Campinas, SP.

³ Bacharel em Ciência da Computação, doutor em Ciências da Computação, pesquisador da Embrapa Informática Agropecuária, Campinas, SP.

⁴ Estatística, doutora em Ciências Matemáticas e da Computação, pesquisadora da Embrapa Informática Agropecuária, Campinas, SP.

Introdução

Neste trabalho pretendeu-se construir um portfólio para selecionar tecnologias para o uso sustentável da água na agricultura, a partir da literatura. Os portfólios são planilhas com a relação de tecnologias, locais, época, culturas associadas e possíveis tecnologias associadas. Após a construção dos portfólios, os especialistas de domínio verificarão se essas tecnologias são soluções para problemas decorrentes da avaliação e da adaptação de tecnologias para os biomas brasileiros. A princípio, buscou-se obter, manualmente, informações nos textos, referentes à data, à localidade, às tecnologias presentes e ausentes, e às informações adicionais relevantes, para que seja possível comparar com o método automático. Como o conjunto de textos é muito grande, é interessante automatizar ou semiautomatizar os processos de construção de portfólios utilizando-se técnicas de mineração de textos.

O processo de Mineração de Textos pode ser dividido em cinco etapas: a) identificação do problema; b) pré-processamento; c) extração de padrões; d) pós-processamento; e) utilização do conhecimento. Na identificação do problema são definidos os objetivos do processo de Mineração de Textos. Uma vez definidos o escopo do problema e o objetivo da aplicação, e selecionados os documentos que representem o domínio do problema, pode-se definir a forma de representação dos dados e extração de conhecimento.

Na etapa de pré-processamento, a coleção de textos é manipulada a fim de se obter uma representação que possa ser lida por ferramentas de extração de padrões. Os textos são transformados em uma matriz atributo valor, na qual as linhas correspondem a cada texto e as colunas, às palavras (ou composições de palavras) presentes na coleção de textos, e, cada célula corresponde a uma medida da importância da palavra no texto.

Neste trabalho, na etapa de extrações de padrões, o foco é a experimentação de técnicas de extração de regras de associação, a fim de verificar relações entre tecnologias e outras palavras (composições de palavras) nos textos, tais como condições de aplicação das tecnologias, por exemplo: em presença de seca aplica-se irrigação, seria uma regra do tipo “seca->irrigação”. As regras de associação buscam encontrar o relacionamento entre itens de dados que ocorram com uma certa frequência, ou seja, identificar padrões em dados históricos (AGRAWAL et al., 1994).

Na última etapa, do processo de mineração de textos, realiza-se o pós-processamento em que é analisado se os dados obtidos nas regras de associação aproximam-se ao esperado. No caso das regras, várias métricas de qualidade podem ser utilizadas para uma avaliação objetiva, bem como a avaliação subjetiva. Se o processo não alcançar resultados usáveis, repete-se o processo, isto é, desde o pré-processamento até que os resultados avaliados sejam compatíveis de acordo com o objetivo do trabalho.

Assim, o objetivo deste trabalho é avaliar o uso de técnicas de mineração de dados para auxiliar a construção de portfólios tecnológicos.

Materiais e Métodos

Foi realizada uma busca de textos no Sistema Aberto e Integrado de Informação em Agricultura (SABIIA) (VACARI et al., 2011) e assim, com as expressões e combinações das palavras fornecidas pelos especialistas, se obteve 2.210 textos. A expressão de busca foi uma combinação de termos fornecidos pelos especialistas do domínio, mais ou menos na forma: *“(bacia hidrográfica) OR (gestão hídrica) OR (modelagem hidrológica) OR (hidrossedimentometria) OR (sistemas de informação geográfica) OR (sensoriamento remoto) OR (disponibilidade hídrica) OR (balanço energético) OR (qualidade da água) OR (indicadores de sustentabilidade) OR (pegada hídrica ecológica) OR (águas cinzas)) AND ((videira) OR (bananeira) OR (goiabeira) OR (mangueira) OR (coqueiro) OR (melão) OR (melancia) OR (tomate) OR (cebola)”*.

O resultado foi convertido para uma base de textos, na qual cada resposta de busca corresponde a um texto. Para isso, escreveu-se um script em python, que usa os metadados da busca e baixa os pdfs, quando disponíveis, convertendo-os para textos.

Aplicou-se um pré-processamento à base textual utilizando a ferramenta I-PreProc (Incremental Pre-Processing), em desenvolvimento no projeto CRITIC@, com um vocabulário controlado. O vocabulário foi determinado por um estudo de resultados anteriores de busca. No vocabulário constam: tipos de solos, geo localidades, todos os termos da expressão de busca e a especificação de tecnologias agrícolas relacionadas ao uso de água, de acordo com o Thesagro (BRASIL, 2015). I-PreProc gera uma representação

de dados com dois arquivos: HDR de descritores e DAT com a matriz atributo valor. Para poder utilizar os resultados como input do algoritmo Apriori (LIU et al., 1998), disponível no ambiente Weka (HALL et al., 2009), foram escritos dois scripts em Python, que leem DAT e HDR e criam: a) um arquivo para o Apriori de Christian Borgelt⁵, cujas linhas, que correspondem a cada texto (ou transação), são preenchidas com os vocábulos presentes no texto; e, b) um arquivo do Apriori implementado no Weka, que é uma matriz não esparsa, representada em formato CSV, sendo a primeira linha o nome de todos os vocábulos utilizados na coleção e as demais indicando presença (S) ou ausência (N) do vocábulo no texto, e, a última coluna as variáveis de interesse para o consequente da regra (classes). Sendo que o Apriori de Christian Borgelt apenas calcula as regras de associação. O Apriori do Weka, com matriz não esparsa e classes, gera regras do tipo: *vocábulo x e vocábulo y* → *tecnologia*.

Uma regra de associação é uma implicação da forma $X \rightarrow Y$, onde X e Y são conjuntos frequentes, suporte é a frequência em que X e Y aparecem no conjunto de dados e confiança mede a frequência de itens em Y que aparece nas transações que contém X, equação (2).

Resultados e Discussão

No primeiro experimento, o algoritmo Apriori do Weka foi avaliado para diferentes valores de suporte, mantendo-se a confiança em 90%. A partir da Tabela 1 é possível verificar que para um valor de suporte muito baixo, o número de regras cresce exponencialmente, sendo a maior parte delas regras redundantes, bem como regras que não apresentam nenhuma novidade, como, por exemplo, “aspersão e chuva ==> água”.

Tabela 1. Número de regras geradas pelo Apriori mantendo a confiança de 90% e variação do suporte.

Suporte (%)	Número de regras
1	100.000
3	2.512
5	174
10	13
15	3
20	0

⁵ Disponível em <http://www.borgelt.net/apriori.html>, consultado em setembro de 2015.

Além disso, a maioria das regras encontradas foram constituídas de poucos vocábulos, pois a matriz de termos submetida ao algoritmo Apriori era muito esparsa (grande número de textos e ausência do vocábulo no texto). Com isso, a frequência de muitos termos era muito baixa (próxima de zero) resultando em regras fracas, ou seja, formadas por termos com um baixo valor de suporte.

Considerando que a análise de regras de associação não apresentou os resultados esperados, outros algoritmos de aprendizado de máquina foram utilizados para tentar identificar a relação entre as palavras-chave e as tecnologias.

O primeiro algoritmo de aprendizado utilizado foi o C.4.5 (QUILAN, 1993) para a indução de regras de classificação por meio de árvores de decisão. A escolha desse algoritmo foi influenciada pela facilidade de uso do algoritmo e pela forma simbólica em que o conhecimento gerado pode ser representado por meio de regras. No entanto, devido à matriz de dados ser tão esparsa, a árvore de decisão não foi gerada completamente, pois a maioria dos atributos possuía valores faltantes.

Em seguida, tentou-se outra abordagem baseada em redes Bayesianas (JOHN; LANGLEY, 1995). Novamente, observou-se que a matriz esparsa como input não favoreceu o uso dessa abordagem. Como consequência, a rede Bayesiana não foi gerada e a relação entre os termos e as tecnologias não foi identificada.

O que pode ser observado é que essa forma de matriz não esparsa, com presença e ausência dos vocábulos e utilizando-se a última coluna com todas as classes de interesse, não permite que se possa inferir algum modelo. Foram utilizados 134 vocábulos, com 114 diferentes classes (tecnologias) em 1.702 textos, sendo que várias classes aparecem nos mesmos textos. Assim, uma opção que se apresenta é fixar uma tecnologia como classe positiva e todas as demais como atributos comuns, ou seja, se a tecnologia em questão não aparecer no texto, a classe será negativa.

Considerações Finais

Os experimentos iniciais mostram que o formato dos dados, como colocado, está levando a matrizes altamente esparsas, com muitas classes e poucos

atributos, o que não nos permite obter inferências com um nível de precisão aceitável. Assim, como trabalho futuro, deve-se procurar diminuir a esparsidade da matriz, tanto considerando as classes como binárias uma a uma, quanto ampliando o vocabulário utilizado.

Referências

AGRAWAL, R.; SRIKANT, R. Fast algorithms for mining association rules in large databases. In: INTERNATIONAL CONFERENCE ON VERY LARGE DATA BASES, 20., 1994, Santiago. **Proceedings**... San Francisco: Morgan Kaufmann, 1994. p. 478-499.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. **Thesagro**. Disponível em: <http://snida.agricultura.gov.br:81/binagri/html/Gen_Thes1.html>. Acesso em: 15 out. 2015.

HALL, M.; FRANK, E.; HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P.; WITTEN, I. H. The WEKA data mining software: an update. **ACM SIGKDD Explorations Newsletter**, v. 11, n. 1, p. 10-18, June, 2009. DOI: 10.1145/1656274.1656278.

JOHN, G. H.; LANGLEY, P. Estimating continuous distributions in Bayesian classifiers. In: CONFERENCE ON UNCERTAINTY IN ARTIFICIAL INTELLIGENCE, 11., San Francisco. **Proceedings**... San Francisco: Morgan Kaufmann, 1995. p. 338-345.

LIU, B.; HSU, W.; MA, Y. Integrating classification and association rule mining. In: Fourth INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 4., 1998, New York. **Proceedings**... Menlo Park: AAAI Press, 1998. p. 80-86.

QUILAN, R. **C4.5**: programs for machine learning. San Mateo: Morgan Kaufmann, 1993. 302 p. ill.

VACARI, I.; VISOLI, M. C.; GONZALES, L. E. Acesso aberto a informação científica agropecuária na internet: caso do sistema aberto e integrado de informação em agricultura (Sabíia). In: CONGRESSO BRASILEIRO DE AGROINFORMÁTICA, 8., 2011, Bento Gonçalves. **Anais**... Florianópolis: UFSC; Pelotas: UFPel, 2011. Não paginado.

Processo de produção de imagens para comunicação visual

Giovana Roggeri Affonso¹

Deise Rocha Martins dos Santos Oliveira²

Resumo: Neste trabalho é apresentado o processo de produção de imagens, com foco em ilustrações, utilizadas nos folders, palestras e vídeos da Embrapa Informática Agropecuária. Esse processo inclui etapas desde o briefing e planejamento, rascunho, finalização até o fechamento da ilustração em softwares de criação de imagens vetoriais. Em seguida, discute-se vantagens e desvantagens das imagens produzidas por meio desta metodologia. Como resultado, foram produzidas um total de 102 imagens, sendo 68 delas vetoriais. Com o trabalho, concluiu-se que o método é eficiente, melhora a comunicação visual dos materiais em que é aplicado e terá continuidade de uso até dezembro de 2015.

Palavras-chave: ilustração, Adobe Illustrator, imagens vetoriais, design.

Introdução

A utilização de imagens para a comunicação visual tem a função primeira de aproximar público e serviço, pois promovem uma identificação e familiarização com o usuário. As imagens tornam a leitura mais agradável e dinâmica, e permitem a fácil associação pelo leitor. Segundo Dondis (2007), a experiência visual no design está na interação de polaridades duplas: primeiro, as forças do conteúdo (mensagem e significado) e da forma (design, meio e ordenação); em segundo lugar, o efeito recíproco do articulador (design,

¹ Estudante de Artes Visuais da Universidade Estadual de Campinas (Unicamp), estagiário da Embrapa Informática Agropecuária, Campinas, SP.

² Mestre em Engenharia da Produção, analista da Embrapa Informática Agropecuária, Campinas, SP.

artista, artesão) e do receptor (público), onde a mensagem é emitida pelo criador e modificada e assimilada pelo observador. As vantagens da utilização de ilustrações produzidas pela empresa são o alto grau de personalização, definido entre o cliente e o ilustrador - para que forma e conteúdo sejam bem específicos - e o baixo custo em relação a imagens compradas de banco de imagens.

Na Empresa Brasileira de Pesquisa Agropecuária (Embrapa), as ilustrações são utilizadas para potencializar a comunicação de vídeos, palestras, folhetos e torná-los mais diferenciados e chamativos. Para suprir esta demanda, usavam-se imagens feitas e coloridas à mão sobre papel e depois escaneadas, porém essas ilustrações possuíam diversas limitações quanto: a) edição; b) ampliação; c) qualidade. Além disso, as imagens utilizadas em material áudio-visual precisavam ser exportadas em partes, para depois serem animadas no software de edição de vídeo, outra dificuldade das imagens feitas de forma tradicional. A utilização do software Adobe Illustrator (ADOBE, 2010) para a composição dessas imagens solucionou essas limitações.

Este trabalho tem por objetivo apresentar o processo de produção dessas imagens vetoriais, utilizadas para atender esta demanda.

Materiais e Métodos

O processo para produção das imagens é realizado em várias etapas, a saber: briefing, planejamento, digitalização, vetorização, finalização e exportação.

A etapa de briefing consiste em uma reunião em que o cliente expõe sua demanda. Nessa reunião, ilustrador e cliente definem quais são as ilustrações necessárias para seu projeto e onde as imagens potencializam o entendimento e a comunicação deste. O estilo da Figura - se é mais cartoon ou realista - prazos, público-alvo e paleta de cores também são definidos nessa reunião. A etapa seguinte é a do planejamento, onde é feito um esboço a lápis, um rascunho de como a ilustração será e um retorno com o cliente, para aprovação. A Figura 1 mostra dois esboços feitos para a palestra “As Cinco Linguagens de Valorização no Ambiente de Trabalho”, apresentada na Unidade. A escolha por realizar o rascunho diretamente no papel vem da organicidade e espontaneidade das formas obtidas, características não



Figura 1. Planejamento em papel e lápis de ilustrações para a palestra “As Cinco Linguagens de Valorização no Ambiente de Trabalho”.

permitidas pelo software utilizado, que exigem mais precisão. Além disso, por se tratarem de rascunhos, a velocidade de criação, experimentação e descarte é muito maior. Após a concordância entre as partes, é realizada a etapa de digitalização, onde o rascunho em papel é escaneado e importado no software Adobe Illustrator CS5.5.

Esse software é muito utilizado nas áreas de design e ilustração por trabalhar com imagens vetoriais. Diferente das imagens bitmaps (mapa de bits), que são compostas por pixels, cada um contendo uma informação de cor, as imagens vetoriais são formadas por fórmulas matemáticas interpretadas pelo software e transformadas em elementos visuais (JORDÃO, 2010). A Figura 2 demonstra a diferença entre a imagem bitmap e a vetorial. As vantagens da imagem vetorial são a precisão das formas - principalmente ao se considerar o uso do mouse para sua manipulação, a ampliação dos elementos sem prejuízo da qualidade da imagem e a facilidade de edição e alteração, caso necessário, diferente do desenho feito e colorido à mão, em que a alteração é limitada ou inviável.

No Adobe Illustrator, é executada a etapa da vetorização, que consiste em transformar manualmente o rascunho importado em formas vetorizadas. Cada ilustração é constituída por um conjunto de formas geométricas de cores chapadas. As formas vetorizadas são então coloridas conforme conversado na reunião de briefing e aplica-se alguns efeitos - como desfoque, transparência, sombreamento e, degradê - para melhor apresentação da

Imagem bitmap e detalhe ampliado

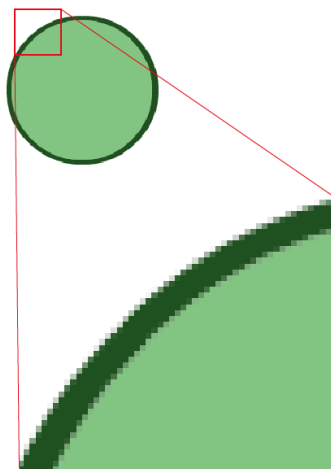


Imagem vetorial e detalhe ampliado

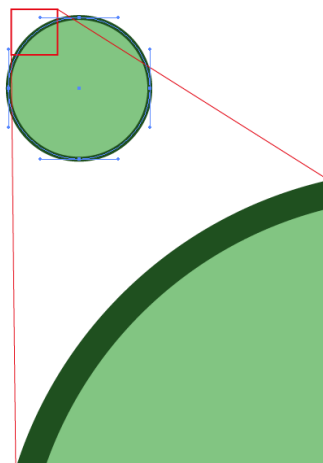


Figura 2. Exemplo de diferenças entre a estrutura da imagem bitmap e da imagem vetorial.

imagem, consistindo na etapa da finalização. Por fim, ela é exportada em formato adequado ao seu uso. Para o caso de algumas imagens utilizadas em vídeos, ela é exportada em partes, para depois ser animada no software.

Resultados e Discussão

As imagens produzidas no software Adobe Illustrator, a partir de um rascunho feito de forma tradicional, têm um resultado muito satisfatório e atendem às necessidades de maneira eficiente. A Figura 3 mostra a versão final das imagens produzidas a partir das imagens da Figura 1. Somente entre o período de outubro de 2014 a setembro de 2015, foram produzidas 102 ilustrações para materiais variados. Dessas 102, 68 são ilustrações produzidas com o uso do software.

A desvantagem em comparação a imagens não-vetoriais é que a imagem fica mais lisa, com texturização limitada, e as nuâncias de transição de cores, como em uma pintura tradicional, também são limitadas pois a cor deve ser única para cada forma vetorizada. Porém as vantagens de qualidade de ampliação e fácil edição o tornam a melhor alternativa. Por exemplo, se



Figura 3. Ilustrações finalizadas para a palestra “As cinco Linguagens de Valorização no Ambiente de Trabalho”.

compararmos as Figuras 1 e 3, nota-se que a expressão dos personagens e o tamanho da caixa de presente sofreram alterações entre o rascunho e a versão final, correções resolvidas em poucos minutos com um simples redimensionamento de formas. Uma solução utilizada para essas desvantagens de textura e variação de cor é uma mistura de softwares de edição de imagem vetorial e de imagens bitmap.

Considerações Finais

As solicitações de ilustrações têm sido atendidas por meio desta metodologia, principalmente em relação à produção de vídeos. Quando às imagens, apresentam necessidade de alteração, estas são resolvidas de uma forma mais ágil em função da utilização de imagens vetoriais. Esta metodologia apresentou melhoria na comunicação visual dos materiais em que ela foi aplicada e terá continuidade de uso nos vídeos produzidos pela Unidade, até dezembro de 2015 e em outros materiais de comunicação, como folders, palestras e cartões comemorativos.

Agradecimentos

Agradecimentos a minha companheira de mesa Mariana Moreti Vieira, que sempre me apoia e me ajuda com as ilustrações.

Referências

ADOBE Illustrator CS5. [Berkeley]: Adobe Press, 2010. 463 p. ill.

DONDIS, D. A. **Sintaxe da linguagem visual**. Tradução Jefferson Luiz Camargo. 3. ed. São Paulo: Martins Fontes, 2007. 236 p. il.

JORDÃO, F. **Quais as diferenças entre pixels e vetores**. 2010. Disponível em: <<http://www.tecmundo.com.br/6135-quais-as-diferencas-entre-pixels-e-vetores-.htm>>. Acesso em: 1 out. 2015.

Processo de produção em audiovisual das tecnologias da Embrapa Informática Agropecuária

Mariana Moreti Vieira¹

Deise Rocha Martins dos Santos Oliveira²

Resumo: Este trabalho consiste em descrever a metodologia aplicada no desenvolvimento de vídeos de divulgação sobre as tecnologias produzidas pela Embrapa Informática Agropecuária, incluindo a produção do roteiro, a narração, o desenvolvimento de imagens, e a criação e edição dos vídeos. Estes vídeos são curtos e elaborados para demonstrar e divulgar as tecnologias de maneira instrutiva e concisa, divulgando-as em feiras, eventos e redes sociais, facilitando a transferência de tecnologia e a absorção do conteúdo por diversos tipos de públicos. São elaboradas a partir dos programas VideoScribe, Adobe Illustrator CS5, Gravador de Voz e Audacity. Os resultados obtidos foram a produção de quatro vídeos no primeiro semestre de 2015 e dessa forma pode-se agregar a produção de imagens digitais, em substituição às imagens tradicionais.

Palavras-chave: comunicação visual, vídeo, roteiro, narração, design.

Introdução

A comunicação visual é uma ferramenta indispensável ao apelo do consumidor. Visualmente é possível estabelecer vínculos e facilitar a leitura de informações complexas, possibilitando maior acesso a estas informações.

¹ Estudante de Artes Visuais da Universidade Estadual de Campinas (Unicamp), estagiária/bolsista da Embrapa Informática Agropecuária, Campinas, SP.

² Mestre em Engenharia da Produção, analista da Embrapa Informática Agropecuária.

Esta produção visual, com objetivo de transmitir conhecimento, torna-se uma aliada na disseminação das tecnologias.

A produção dos vídeos supre a necessidade de inovação nas apresentações em feiras e eventos, bem como a conexão com redes sociais, rompendo com a busca de vídeos externos, e criando conteúdo original para uso da empresa.

Segundo Dondis (2007), a imagem facilita a compreensão, promovendo a comunicação, diminuindo o uso de textos e fixando melhor as informações na memória do público.

O nível representacional da inteligência visual é fortemente governado pela experiência direta que ultrapassa a percepção. Aprendemos sobre as coisas das quais não podemos ter experiência direta através dos meios visuais de demonstrações e de exemplos em forma de modelo. [...] Não se faz necessária a intervenção de nenhum sistema de códigos para facilitar a compreensão, e de nenhuma decodificação que retarde o entendimento. (DONDIS, 2007, p. 20).

A visualidade é pensada para abranger um grande público e busca se fazer o mais acessível possível. Esta proposta que prima pela comunicação visual busca construir novos caminhos dentro da Empresa Brasileira de Pesquisa Agropecuária (Embrapa), para auxiliar o processo de comunicação em si e também a transferência de tecnologia. Os processos são desenvolvidos em acordo com critérios estabelecidos, utilizando o design.

Materiais e Métodos

A metodologia utilizada nesse processo conduz o ritmo da produção dos vídeos. Depois de escolhida a tecnologia, é feito um projeto do roteiro baseado em informações contidas nos sites, repositórios da Embrapa, textos, artigos e contato com o responsável pelo desenvolvimento da tecnologia. A estrutura do texto é desenvolvida de maneira a: a) conceituar a tecnologia (o que?); b) apontar suas características, vantagens e benefícios (para que?); c) seu público-alvo (para quem?); d) informações para contato (onde?). O texto não deve ser muito longo, pois os vídeos devem ser curtos, com duração de 2 a 4 minutos. Em seguida, o texto é escrito, usando uma linguagem simples, de fácil entendimento. O texto é revisado pela supervisão e por um respon-

sável pelo desenvolvimento da tecnologia, e, em seguida, encaminhado ao Comitê Local de Publicações (CLP). Após o parecer do CLP, inicia-se o processo de criação das imagens para o vídeo.

São produzidas em torno de 20 imagens por vídeo, utilizando meios variados, desde imagens digitais usando o programa Adobe Illustrator³ até imagens tradicionais como lápis e canetas, como pode ser observado na Figura 1.

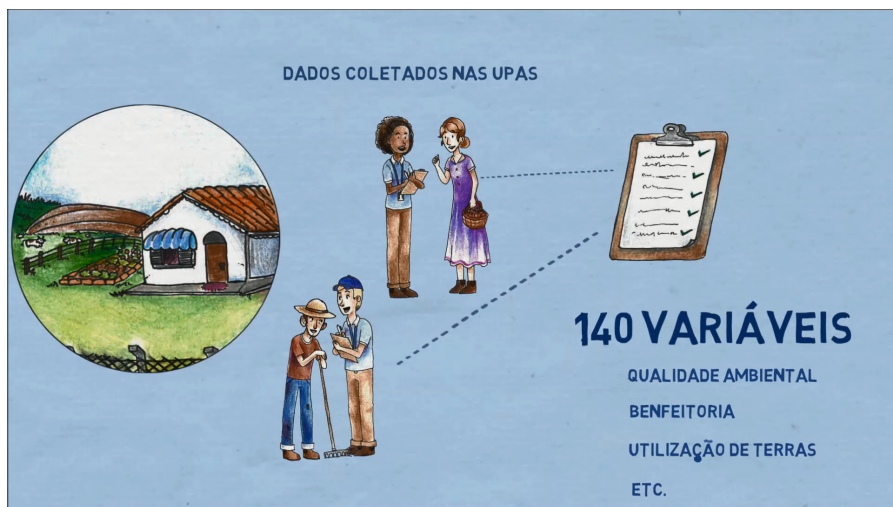


Figura 1. Frame do Vídeo “Conheça o Planeja”.

Ao se concluir a fase de produção de imagens, estas são incluídas no roteiro. O responsável pela tecnologia verifica novamente esta versão com as imagens e, após aprovada, se inicia o processo da gravação da narração. A gravação é composta pela totalidade do texto e é usado o Gravador de Som do Windows. Já o áudio é feito em partes, sendo editado apropriadamente no Audacity 2.0.6 (MAZZONI; DANNENBERG, 2007). Cada parcela é tratada de maneira a remover possíveis ruídos e deixar a fala o mais clara possível.

Foi observado que o modo mais eficiente para incluir a narração no vídeo era fazê-la anteriormente à sua produção, pois facilita o posicionamento

³ Disponível em: <<http://www.adobe.com>>. Acesso em: 15 out. 2015.

das informações no programa escolhido para edição, o VideoScribe⁴. Este programa permite a criação de vídeos com funções de inclusão animada de texto, imagens, além de inclusão de música e narração. As sequências são desenvolvidas de acordo com o roteiro prévio e visando à construção de uma identidade visual que se relacione com a Embrapa e os sistemas tecnológicos. As paletas de cores e propostas de designs utilizadas desde o início da elaboração do vídeo são retomadas.

É essencial que se estabeleça uma relação entre as palavras ditas e as imagens animadas no programa. Evidencia-se este processo na Figura 2, onde as cores do logo da tecnologia são retomadas nas setas indicativas e nas cores da fonte. Também fica claro, a tipografia usada para enfatizar determinadas palavras, a fim de destacar suas informações e conectá-las com a visualidade do vídeo.



Figura 1. Frame do Vídeo “Agropedia brasilis”.

Resultados e Discussão

Desde outubro de 2014, foram produzidos quatro vídeos de tecnologias, sendo estes: a) Planeja; b) Agritempo; c) Agropedia brasilis; c) Invernada. Atualmente, o vídeo das Bases de acesso aberto e Agritempo mobile estão em andamento.

⁴ Disponível em: <<http://www.videoscribe.co/>>. Acesso em: 15 out. 2015.

Sobre os vídeos, a Tabela 1 evidencia que conforme houve acesso ao programa Adobe Illustrator, a produção de imagens tradicionais foi substituída pelas imagens digitais. Isso aumenta a qualidade estética do vídeo e facilita a execução de alterações, caso sejam necessárias. A quantidade de ilustrações também tende a aumentar, promovendo maior ênfase nas imagens do que nas palavras e tornando a elaboração e a aplicação destas mais complexa.

Tabela 1. Informações sobre os vídeos produzidos na Embrapa Informática Agropecuária - primeiro semestre de 2015.

Vídeos	Tempo	Imagens	
		Quantidade	Tipo
Conheça o Planeja	02'04"	13	Tradicional
Agritempo	04'03"	23	Tradicional/Digital
Agropedia brasilis	04'22"	15	Digital
Invernada	02'55"	20	Digital

A aplicação é dificultada pelo próprio programa de criação de vídeos, o VideoScribe, tendo em vista que não possui tanto suporte a grandes quantidades de elementos vinculados ao projeto, provocando assim uma tendência à redução no tempo dos vídeos.

Considerações Finais

Durante este processo, evidenciou-se a necessidade da comunicação visual dentro da Embrapa como um elemento inovador na transferência de tecnologia e criou-se, além disso, um novo atrativo para as tecnologias desenvolvidas na Embrapa Informática Agropecuária. As imagens produzidas tornam-se parte do acervo da Embrapa, possibilitando seu uso em outros projetos.

Este trabalho demonstra como a produção de vídeos coloca a empresa num patamar diferenciado, por conta da intersecção entre as áreas de Comunicação, Transferência de Tecnologia e Artes Visuais, revelando uma importância maior dentro da sua relação com o público em geral, e em particular.

Agradecimentos

A Giovana Roggeri Affonso, pela colaboração na elaboração das imagens e apoio na produção dos vídeos; e a Junia Rodrigues de Alencar, pelo suporte na finalização deste trabalho.

Referências

DONDIS, A. D. **Sintaxe da linguagem visual**. Tradução Jefferson Luiz Camargo. 3. ed. São Paulo: Martins Fontes, 2007. 236 p.

MAZZONI, D.; DANNENBERG, D. **Audacity 2.0.6**. 2007. Disponível em: <<http://audacityteam.org/>>. Acesso em: 15 out. 2015.

VIDEOSCRIBE. 2014. Disponível em: <<http://www.videoscribe.co/>>. Acesso em: 15 out. 2015.

Literatura recomendada

MATEUS, S. **Publicidade e consumo nas sociedades contemporâneas**. Covilhã: Livros LabCom, 2011. 227 p. (Estudos em comunicação).

Protótipo de sistema de segurança criptográfica via porta lógica quântica

Jacomo Giovanetti Minto Neto¹

Alexandre de Castro²

Adauto Luiz Mancini³

Resumo: O cifrador XOR não é considerado completamente seguro, pois não há chaves de criptografia verdadeiramente aleatórias, dado que as chaves são geradas em geradores de números pseudorrandômicos a partir de sementes que não podem ser negligenciadas. No entanto, uma chave que destrói a própria semente que a gerou, pode ser considerada verdadeiramente aleatória, pois o seu estado inicial é definitivamente ignorado. Neste trabalho, nós apresentamos um modelo de chaves verdadeiramente aleatórias que podem ser obtidas a partir da porta quântica denominada controlled-NOT (CNOT) usada para emaranhar estados EPR.

Palavras-chave: porta lógica quântica, criptografia, OTP, cifrador XOR, campo de galois.

Introdução

Em um trabalho anterior, Bayer (2006) sugeriu que a porta CNOT é assimétrica. Nesta mesma linha, recentemente mostramos que a porta CNOT se torna irreversível com restrições adiabáticas (CASTRO, 2014), uma vez que o seu circuito quântico só pode ser completado se uma operação de

¹ Estudante de Análise e Desenvolvimento de Sistemas da Faculdade de Tecnologia de Americana (Fatec, Americana, SP), estagiário da Embrapa Informática Agropecuária, Campinas, SP.

² Doutor em Ciências (Biomatemática), pesquisador da Embrapa Informática Agropecuária, Campinas, SP.

³ Mestre em Ciência da Computação, pesquisador da Embrapa Informática Agropecuária, Campinas, SP.

disjunção exclusiva no seu qubit ancilla ganha uma informação extra igual a $\text{Log}(2)$. Aqui, mostramos que, se a chave de criptografia é obtida por uma função quadrática módulo 2 da mensagem (plaintext), o resultado é um qubit ancilla perfeitamente emaranhado, que produz um cifrador XOR com comportamento de um one-time pad(OTP).

Materiais e Métodos

O operador unitário U_{CNOT} pode ser escrito sobre dois qubits, operacionalmente, $|a\rangle$ e $|b\rangle \in GF_2$, onde o primeiro é o qubit de controle que representa cada bit da chave criptográfica e o último é o qubit ancilla que representa cada bit da mensagem original, e GF_2 é o campo de Galois (MULLEN; PLANARIO, 2013) de dois elementos, $F_2 = \{0,1\}$:

$$U_{\text{CNOT}} |a \otimes b\rangle = |a\rangle \otimes |a \oplus b\rangle$$

onde $a \oplus b = (a+b) \bmod 2$ representa a mensagem cifrada. Na operação CNOT, o primeiro qubit é conservado, ao passo que o segundo qubit é o resultado de uma operação XOR entre o primeiro e o segundo qubit (NIELSEN; CHUANG, 2000).

A matricial para esta transformação é:

$$U(0,0) = (0,0) \Rightarrow \begin{pmatrix} U_{11} & U_{12} & U_{13} & U_{14} \\ U_{21} & U_{22} & U_{23} & U_{24} \\ U_{31} & U_{32} & U_{33} & U_{34} \\ U_{41} & U_{42} & U_{43} & U_{44} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \therefore \begin{pmatrix} U_{11} \\ U_{21} \\ U_{31} \\ U_{41} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\Rightarrow U_{11} = 1, U_{21} = U_{31} = U_{41} = 0$$

Da mesma forma, para $U(0,1) = (0,1)$: $U_{22} = 1, U_{12} = U_{32} = U_{42} = 0$

Para:

$$U(1,0) = (1,1) \Rightarrow \begin{pmatrix} 1 & 0 & U_{13} & U_{14} \\ 0 & 1 & U_{23} & U_{24} \\ 0 & 0 & U_{33} & U_{34} \\ 0 & 0 & U_{43} & U_{44} \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \therefore \begin{pmatrix} U_{13} \\ U_{23} \\ U_{33} \\ U_{43} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

$$\Rightarrow U_{43} = 1, U_{13} = U_{23} = U_{33} = 0$$

Para , $U(1,1) = (1,0)$, $U_{34} = 1$, $U_{14} = U_{24} = U_{34} = 0$.

Logo,

$$U = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \text{ onde } U^2 = I_d \text{ é uma permutação.}$$

Note-se que na transformação:

	(0,0)	(0,1)	(1,0)	(1,1)
(0,0)	1	0	0	0
(0,1)	0	1	0	0
(1,0)	0	0	0	1
(1,1)	0	0	1	0

O estado $U_{\text{CNOT}} |1,0\rangle \rightarrow |1,1\rangle$ pode ser substituído por $U_{\text{CNOT}} |1,b\rangle \rightarrow |1, \text{NOT}(b) \oplus b\rangle$, onde $|\text{NOT}(b) \oplus b\rangle$ representa uma função irreversível.

Resultados e Discussão

A partir da construção acima, mesmo considerando a porta CNOT unitária (permutação), somente poderá ocorrer a decifração se a chave gerada aleatoriamente por meio da função irreversível for aplicada, executando, assim, uma transformação unitária que é a sua própria inversa, de modo que exista uma involução. (ver aspectos termodinâmicos (CASTRO, 2014).

Considerações Finais

Neste trabalho, apresentamos um protocolo de criptografia via porta CNOT, que representa um modelo OTP seguro, pois a chave de criptografia não é obtida a partir de um gerador pseudorrandômico e, sim, através do próprio protocolo de criptografia.

Referências

BAYER, G. W. Quantum computation violates mirror symmetry. **Quantum Information Processing**, v. 5, n. 1, p. 25-30, Feb. 2006. DOI: 10.1007/s11128-005-0010-1.

CASTRO, A. One-way-ness in the input-saving (Turing) machine. **Physica A: Statistical Mechanics and its Applications**, v. 415, n.1, p. 473-478, Dec. 2014. DOI: 10.1016/j.physa.2014.08.021.

MULLEN, G. L.; PANARIO, D. **Handbook of finite fields**. Boca Raton: CRC Press, 2013. 1033 p.

NIELSEN, M. A.; CHUANG, I. L. **Quantum computation and quantum information**. Cambridge: New York: Cambridge University Press, 2000. 676 p. ill.

Automação de experimentos científicos utilizando KnowledgeFlow

Nádia Vieira Ribeiro¹
Luiz Manoel Silva Cunha²

Resumo: Este trabalho apresenta um estudo de caso referente à utilização da tecnologia KnowledgeFlow, para automação de experimento científico, em Mineração de Dados, na busca por novos conhecimento visando ao aperfeiçoamento dos processos de caracterização e classificação de solos tipo Bruno, classificados nas classes Latossolo Bruno e Nitossolo Bruno. O KnowledgeFlow, mesmo com algumas limitações pontuais, trouxe um retorno positivo. Sua inserção no Sistema Brasileiro de Classificação de Solos (SiBCS) contribuirá para tornar a classificação dos solos mais ágil e de maneira mais otimizada.

Palavras-chave: workflow científico, automação de processos, mineração de dados, weka.

Introdução

Os processos de elaboração de conhecimento científico baseiam-se em constantes repetições de experimentos e análises de grandes conjuntos de dados, que podem tornar a atividade exaustiva ou até mesmo inviável. Diante disso, a busca por metodologias e ferramentas que agilizem os procedimentos são indispensáveis para garantir melhor estruturação, flexibilidade e sucesso do trabalho em questão.

¹ Estudante de Engenharia de Agrícola da universidade Estadual de Campinas (Unicamp), estagiária da Embrapa Informática Agropecuária, Campinas, SP.

² Estatístico, mestre em Engenharia de Software, analista da Embrapa Informática Agropecuária, Campinas, SP.

Uma das metodologias utilizada atualmente é a adoção de fluxos de trabalho, mais conhecidos como *workflows*. Segundo Cuevas-Vicentitin et al. (2012) e Yu; Buyya (2005), *workflow* científico é uma abordagem para automação de um experimento ou de um processo científico, expressa em termos das atividades a serem executadas e, principalmente, das dependências dos dados manipulados.

Este trabalho refere-se ao desenvolvimento de modelos de classificação por meio de Árvores de Decisão, utilizando a ferramenta KnowledgeFlow do software Weka, com intuito de aprimorar, automatizar e facilitar os processos de mineração de dados - como seleção de atributos, balanceamento das classes e modelagem - e suas análises, resultando em melhorias na caracterização e na classificação mais acurada dos solos Brunos no escopo do SiBCS.

Materiais e Métodos

O estudo foi dirigido no Laboratório de Inteligência Computacional (LabIC), situado na Embrapa Informática Agropecuária³, Unidade Descentralizada da Empresa Brasileira de Pesquisa Agropecuária (Embrapa).

Os dados usados foram extraídos do Sistema de Informação de Solos Brasileiros⁴ e de documentos - como ficha para descrição morfológica dos solos no campo, teses, dissertações de mestrado e artigos científicos. Inicialmente, 182 observações contendo 28 atributos, sendo o último denominado "classe" - para o estudo de caso. As observações obtidas são oriundas de levantamentos pedológicos realizados nos estados do Paraná, de Santa Catarina e do Rio Grande do Sul. Elas foram agrupadas nas classes: Latossolo Bruno (LB), Nitossolo Bruno (NB) e Outros Latossolos, Nitossolos e Cambissolos (OLNC). A Figura 1 ilustra os processos incluídos no experimento em questão e as informações de parametrização.

O KnowledgeFlow surgiu como uma alternativa para a ferramenta Explorer do software Weka (HALL; REUTEMANN, 2008), embora algumas funções sejam específicas de cada uma. A ferramenta consiste no fluxo de dados

³ Embrapa Informática Agropecuária: <https://www.embrapa.br/informatica-agropecuaria>

⁴ Sistema de Informação de Solos Brasileiros: <http://www.sisolos.cnptia.embrapa.br>

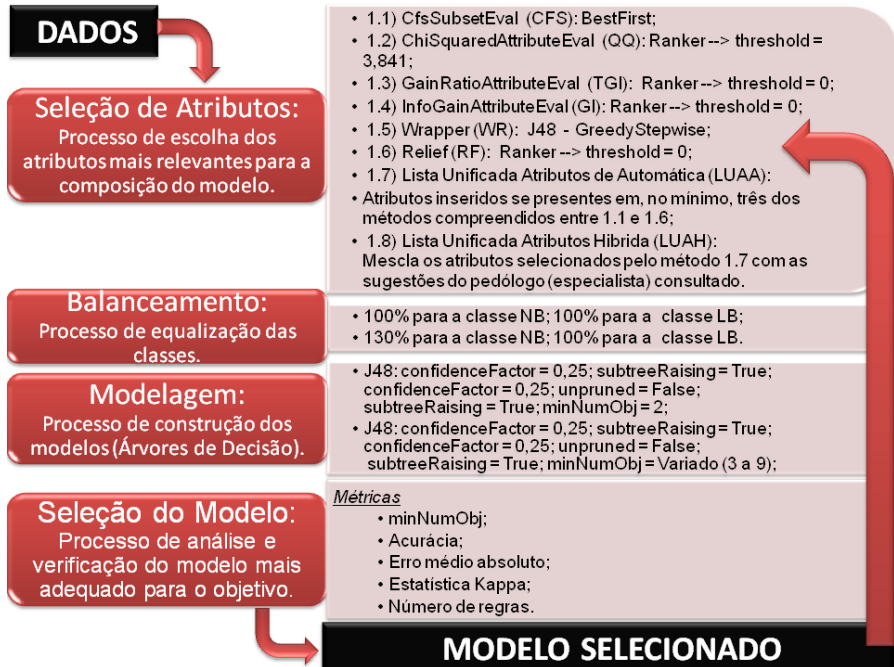


Figura 1. Processos do experimento de solos Brunos.

conectados de maneira a formar o processo desejado, onde o usuário seleciona o componente da barra de ferramentas e o dispõe da forma que lhe for conveniente. Estes são parametrizados, conectados em um grafo direcionados para processar e analisar os dados. Dessa forma, é possível visualizar em termos de como os dados fluem através do sistema.

Resultados e Discussão

Os processos metodológicos descritos acima foram estruturados e automatizados com a utilização do KnowledgeFlow, estando inseridos no âmbito da atividade “Descoberta de conhecimentos em bases de dados de solos: uma contribuição à classificação dos solos Brunos”, incorporada no projeto “Pesquisa e Inovação para o aprimoramento da taxionomia de Solos Brasileiros”, este conduzido pela Embrapa Solos. Apenas com o intuito de distinção de tarefas, o fluxo de seleção de atributos foi feito separadamente

do fluxo usado para balanceamento e modelagem, conforme representa a Figura 2.

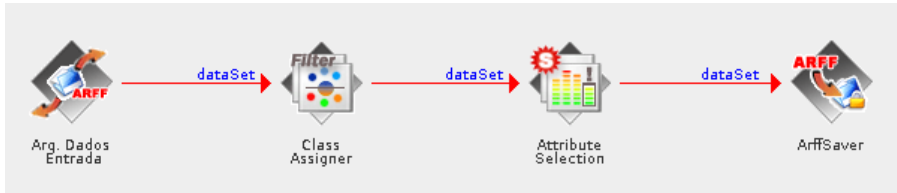


Figura 2. KnowledgeFlow utilizado para a seleção de atributos.

Para o processo de seleção de atributos, o uso da ferramenta foi relevante para a redução e a simplificação das etapas construtivas, visto que com um mesmo fluxo, apenas alterando o algoritmo de seleção e o nome do arquivo de saída, foi possível usar 6 métodos distintos (CFS, QQ, TGI, GI, WR e RF) com um despendimento de tempo mínimo.

Para determinar o melhor nível de balanceamento de classes, utilizou-se o fluxo apresentado na Figura 3, onde o ramo superior refere-se ao balanceamento de 100% para a classe NB (1º componente SMOTE) e 100% para a classe LB (2º componente SMOTE) e o ramo inferior refere-se ao balanceamento de 130% para a classe NB e de 100% para a classe LB.

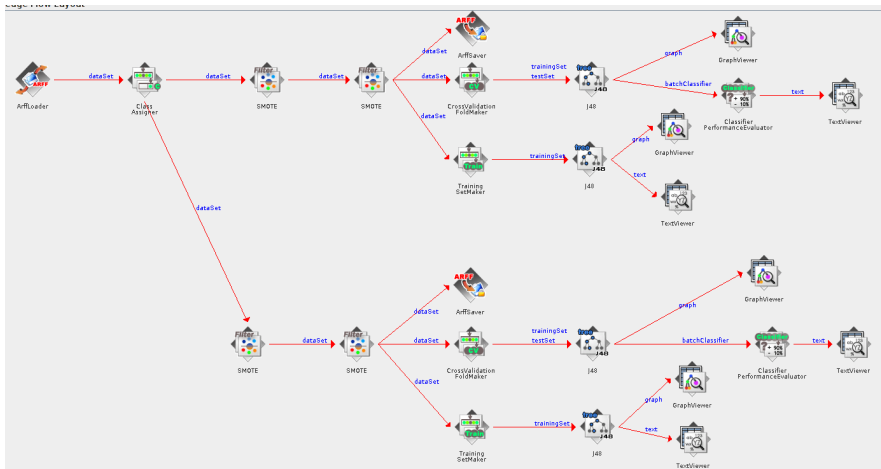


Figura 3. KnowledgeFlow utilizado para o balanceamento e modelagem com número mínimo de objetos fixo.

O KnowledgeFlow permitiu, a partir de diversas combinações, encontrar os melhores níveis de balanceamento pois, dada uma porcentagem, um modelo era gerado automaticamente; assim, selecionando o modelo de melhor desempenho, determinava-se o melhor nível de balanceamento.

Analogamente ao procedimento anterior, para o número mínimo de objetos por folha fixo variado, utilizou-se o fluxo apresentado na Figura 4, em que os ramos à direita originavam os resultados e parâmetros da melhor árvore, usando o método de amostragem de dados Cross Validation 10 folds para cada número mínimo de objetos e os ramos à esquerda originavam as árvores - impressas em formato de texto ou gráfico, para cada número mínimo de objetos correspondente.

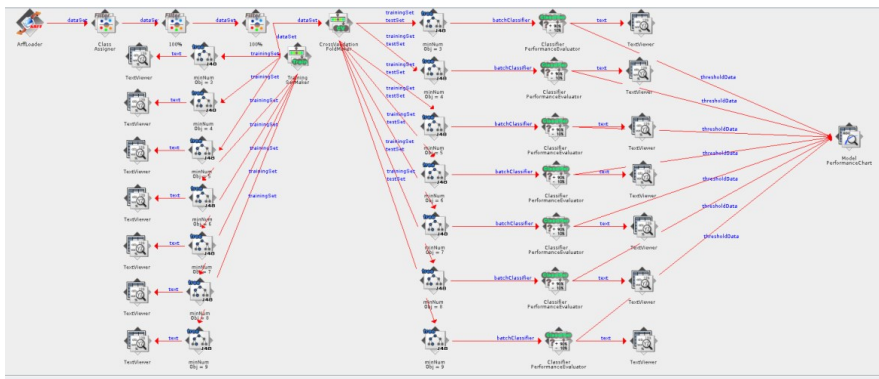


Figura 4. KnowledgeFlow utilizado para o balanceamento e modelagem com número mínimo de objetos variado.

Dentre as vantagens e benefícios agregados pelo uso da ferramenta, destacam-se a estruturação e organização dos procedimentos realizados (ILKAY, 2006) - o que implica na facilidade de exposição do trabalho para especialistas de outras áreas, a redução no tempo gasto para execução do trabalho proposto e a viabilidade de correções e adequações pontuais sem que seja necessário o retrabalho completo. KnowledgeFlow permite o projeto e a execução de configurações para processamento de dados de forma contínua, recurso este que a ferramenta Explorer, do software Weka, não suporta.

As limitações encontradas referem-se à ausência da possibilidade de salvar os resultados dos buffers de maneira automatizada, com a inserção de um ícone saver, por exemplo; a ferramenta permite armazenar os resultados de

forma individual. Na versão utilizada (3.6.10), a ferramenta não permitiu salvar as representações gráficas das árvores, fazendo-se necessário a utilização de meios alternativos como Print Screen. O arquivo do KnowledgeFlow gerado na extensão “.kf”, não foi compatível com outras versões testadas (3.6.9 e 3.7.12); a solução encontrada foi salvar o arquivo utilizando a extensão “.kfml”.

Considerações Finais

Concluído esse estudo, foi possível averiguar que o uso do KnowledgeFlow, apesar de ainda apresentar algumas limitações pontuais, trás um retorno positivo para a implementação da experimentação científica automatizada. A continuação desse trabalho pode se dar avaliando a utilização da ferramenta em rede.

Referências

ALTINTAS, I.; BARNEY, O.; CHENG, Z.; CRITCHLOW, T.; LUDAESCHER, B.; PARKER, S.; SHOSHANI, A.; VOUK, M. Accelerating the scientific exploration process with scientific workflows. **Journal of Physics**, p. 468-477, 2006.

BOUCKAERT, R. R.; FRANK, E.; HALL, M.; KIRKBY, R.; REUTEMANN, P.; SEEWALD, A.; SCUSE, D. **Weka manual for Version 3-6-10**. Hamilton: University of Waikato, 2013. Disponível em: <<http://facweb.cs.depaul.edu/mobasher/classes/csc478/Notes/WekaManual-3-6-10-1.pdf>>. Acesso em: 15 out. 2015.

CUEVAS-VICENTTÍN, V.; DEY, S.; KÖHLER, S.; RIDDLE, S.; LUDÄSCHER, B. Scientific workflows and provenance: Introduction and research opportunities. **Datenbank-Spektrum**, 12, n. 3, p. 193-203, Oct. 2012.

HALL, M.; REUTEMANN, P. **WEKA knowledgeflow tutorial for version 3-5-8**. University of Waikato, 2008. 13 p. Disponível em: <<http://software.ucv.ro/~eganea/AIR/KnowledgeFlowTutorial-3-5-8.pdf>>. Acesso em: 15 out. 2015.

YU, J.; BUYYA, R. A taxonomy of scientific workflow systems for grid computing. **Sigmod Record**, v. 34, n. 3, p. 44-49, 2005.

Implementação do Best Linear Unbiased Prediction (BLUP) em Python para avaliação genética animal

Caio Augusto C. Volpato¹
Roberto Hiroshi Higa²

Resumo: Valores genéticos para animais em programas de melhoramento de grandes populações são obtidos pela resolução de Best Linear Unbiased Prediction (BLUP) correspondente ao modelo de avaliação utilizado. O sistema de equações lineares que resulta da resolução de BLUP contém centenas de milhares ou até milhões de equações para serem resolvidas. Neste trabalho, pretende-se avaliar a estratégia de representação desse sistema de equações lineares em memória Random Access Memory (RAM), baseada em matrizes esparsas, e sua resolução por meio do algoritmo de gradiente conjugado preconditionado (WATKINS, 2010), para avaliar sua aplicabilidade em grandes populações de animais. Até o presente momento, uma solução baseada em matrizes densas foi implementada e sua adaptação para utilizar matrizes esparsas encontra-se em curso. Este trabalho integra as atividades do Plano de Ação 4 do projeto “Desenvolvimento e implementação de metodologias genético-estatísticas em avaliações genéticas de gado de corte” (MENEZES, 2015) que pretende desenvolver um software Embrapa para avaliação genética de grandes populações de animais, livrando os programas de melhoramento genético coordenado pela Empresa Brasileira de Pesquisa Agropecuária (Embrapa) do pagamento de licenças de uso.

Palavras-chave: Melhoramento genético, grandes populações, BLUP, gradiente conjugado, Python.

¹ Estudante de Matemática Aplicada e Computacional da Unicamp, estagiário da Embrapa Informática Agropecuária, Campinas, SP.

² Engenheiro Eletricista, pesquisador da Embrapa Informática Agropecuária, Campinas, SP.

Introdução

Programas de melhoramento genético de animais utilizam medidas de características e de genealogia de dezenas de milhares, ou ainda milhões, de animais para avaliar a contribuição do fator genético na manifestação de características de interesse, em gerais econômicos, da cadeia pecuária (BOURDON, 2000; ROSA et al., 2013).

Para inferir a contribuição da genética na manifestação de características de interesse (valor genético) dos animais que participam de um programa de melhoramento utilizam-se modelos estatísticos conhecidos como modelos lineares mistos (BOURDON, 2000). Esses modelos acomodam tanto fatores fixos quanto aleatórios (JIANG, 2007). Nesses modelos, enquanto fatores como sexo e grupo de manejo são modelados como fatores fixos, o valor genético é modelado como um fator aleatório e todos eles são inferidos a partir da utilização dos dados observados para ajuste do modelo considerado (JIANG, 2007).

Uma das metodologias mais utilizadas para resolução de modelos lineares mistos é o BLUP, um estimador que minimiza a variância do erro de predição (JIANG, 2007). Na prática, a resolução do BLUP implica a solução de um sistema de equações lineares que, na prática, é obtida pela utilização de soluções computacionais (MRODE; THOMPSON, 2014).

No caso de programas de melhoramento genético de grandes populações de animais, a resolução do BLUP é ainda dificultada pelas dimensões do sistema de equações lineares que precisa ser resolvido (centenas de milhares ou milhões de equações) (MRODE; THOMPSON, 2014).

Em geral, a solução de sistemas de equações lineares de grande dimensão utiliza algoritmos iterativos, tais como Jacobi, Gauss-Seidel, sobre-relaxação sucessiva, gradiente conjugado e gradiente conjugado preconditionado (MISZTAL, 2011).

Em particular, no caso da resolução do BLUP aplicado a programas de melhoramento genético de grandes populações de animais, duas estratégias são possíveis. A primeira delas, bastante utilizada, consiste na modelagem da solução iterativa de sistemas de equações lineares com iteração nos dados (do inglês Interaction on Data - IOT), em que se mantém em memória apenas a solução corrente e para cada iteração do algoritmo lê-se o conjunto de dados de arquivos, adequadamente formatados (MISZTAL, 2011).

A segunda estratégia consiste em aproveitar-se do fato de que, o sistema de equações lineares para resolução do BLUP utiliza matrizes esparsas (matrizes esparsas são matrizes onde temos muitos elementos nulos, assim são armazenados apenas os elementos não nulos) para realizar toda a computação com o sistema de equações armazenado na memória RAM. Obviamente, essa estratégia é limitada pela quantidade de memória disponível e, a priori, não se presta a grandes populações de animais. A presente proposta de trabalho pretende explorar exatamente essa estratégia de solução do BLUP.

O objetivo do presente trabalho é: a) implementar a montagem do sistema de equações lineares em memória, utilizando estruturas de matrizes esparsas; b) implementar a resolução do BLUP em memória, utilizando o método do gradiente conjugado preconditionado (método iterativo para resolver sistemas lineares definidos positivos, estamos utilizando o método com suas matrizes preconditionadas, fazendo assim que o método convirja muito menos iterações); c) avaliar a escalabilidade dessa solução em termos de memória e tempo de processamento. Espera-se, dessa forma, obter informações sobre o tipo de população (dimensões) para as quais esse tipo de solução é aplicável, bem como a correspondente necessidade em termos de hardware (quantidade de memória requerida). Este trabalho integra as atividades do Plano de Ação 4 do projeto “Desenvolvimento e implementação de metodologias genético-estatísticas em avaliações genéticas de gado de corte” (MENEZES, 2015), que pretende desenvolver um software Embrapa para avaliação genética de grandes populações de animais, livrando os programas de melhoramento genético coordenado pela Embrapa do pagamento de licenças de uso.

Materiais e Métodos

Para realização da presente proposta de trabalho, a linguagem Python (PYTHON SOFTWARE FOUNDATION, 2015) será utilizada para integrar algoritmos e estruturas de dados disponíveis em diferentes bibliotecas, incluindo:

- SciPy: contém implementações do método do gradiente conjugado e matrizes esparsas.

- Numpy: contém funcionalidades para manipulações de vetores.
- PyPedal: contém a implementação de algoritmos necessários para a montagem do sistema de equações lineares para solução do BLUP, tais como a construção da matriz de parentesco (ou, equivalentemente, do inglês, numerator relationship matrix – NRM), e os algoritmos para obtenção dos elementos de sua inversa em que não seja necessário obter a matriz de parentesco, especificamente os métodos conhecidos como de Henderson e de Quaas.

Inicialmente, o pacote PyPedal (COLE, 2012) será utilizado diretamente na construção do sistema de equações lineares para solução do BLUP. Uma vez que ele não utiliza recursos de matrizes esparsas, num segundo passo, suas implementações da construção da matriz de parentesco e obtenção das matrizes inversas serão estudadas e adaptadas utilizando recursos de matrizes esparsas disponíveis na biblioteca SciPy.

Uma vez tendo construído a solução de BLUP, utilizando o algoritmo de gradiente conjugado preconditionado com o sistema representado em memória RAM por meio de matrizes esparsas, ele será resolvido considerando diferentes tamanhos de populações, sendo medidos o consumo de memória RAM e o tempo de processamento.

Para esse trabalho, será considerado o modelo animal simples, ilustrado na equação matricial $y = Xb + Za + e$, onde y é o vetor com as observações, X é matriz de design dos efeitos fixos, Z é matriz de design dos efeitos aleatórios, a é o vetor com os efeitos animais aleatórios, b é o vetor com os efeitos fixos e e é o vetor com os efeitos residuais extensamente explorado em exemplos em Mrode e Thompson (2014). Os conjuntos de dados para teste serão obtidos de três fontes: dados dos exemplos apresentados em Mrode e Thompson (2014), dados simulados e partes de conjuntos de dados reais, obtidos junto ao programa de melhoramento genético de gado de corte GenePlus-Embrapa.

Para realização dos testes apresentados na seção “Resultados e Discussão” foi utilizado um computador de mesa, com 8GB de memória RAM, com CPU:AMD Phenom(tm) II X4 B95 Processor, utilizando o Ubuntu versão 14.04. Para avaliação do comportamento da implementação baseada em matrizes esparsas, os testes serão realizados em um servidor com maior capacidade de processamento e, em particular, disponibilidade de memória RAM.

Resultados e Discussão

Para implementação da versão da solução de BLUP baseada em matrizes densas, utilizou-se a biblioteca PyPedal tanto para gerar a matriz de parentesco (NRM) quanto sua inversa; enquanto que para montagem do sistema de equações lineares e sua resolução, utilizou-se as bibliotecas SciPy e NumPy (SCIPY, 2015).

Como prova de conceito, essa implementação foi testada utilizando-se o exemplo 3.1 apresentado na referencia (MRODE; THOMPSON, 2014), onde a característica analisada é o ganho de peso à desmama, do inglês Pre-weaning gain – WWG (Tabela 1), assumido-se que a variância genética é dada por $\alpha = 2$. Esses dados são apresentados na Tabela 1.

Tabela 1. Pedigree e ganho de peso à desmama.

Bezerro	Sexo	Pai	Mãe	WWG (kg)
4	Macho	1	Desconhecido	4,5
5	Fêmea	3	2	2,9
6	Fêmea	1	2	3,9
7	Macho	4	5	3,5
8	Macho	3	6	5,0

Fonte: Mrode e Thompson (2014).

O modelo misto utilizado neste exemplo pode ser descrito pelo seguinte conjunto de equações $y_{ij} = p_i + a_j + e_{ij}$, onde y_{ij} é o WWG do j -ésimo bezerro do i -ésimo sexo, p_i é o efeito fixo do i -ésimo sexo, a_j é o efeito aleatório do j -ésimo bezerro e e_{ij} é o erro do efeito aleatório.

O sistema de equações lineares utilizado para estimar os valores dos efeitos estudados é dado por:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + A^{-1}\alpha \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

onde A é a matriz de parentesco.

Inicialmente, o programa implementado recebe o Pedigree e a variância genética α , em seguida, monta as matrizes X e Z e, por meio do software PyPedal, obtém a inversa da matriz de parentesco. Por fim, todas essas matrizes são juntadas para formar o sistema de equações lineares acima. Para obter a solução do sistema foi utilizado a implementação do método do gradiente conjugado presente na biblioteca SciPy.

A solução obtida é apresentada na Tabela 2, ilustrando que o resultado coincide com aquele apresentado em Mrode e Thompson (2014).

Considerações Finais

Até o presente momento, utilizamos Python, suas bibliotecas (SCIPY, 2015) e o pacote PyPedal para construção do sistema linear para solução de BLUP baseado nas matrizes densas geradas por PyPedal

(NRM e suas inversas). Essa implementação foi validada por meio de um estudo de caso (dados do exemplo 3.1) em Mrode e Thompson (2014).

No momento, estamos estudando as implementações de PyPedal para geração da NRM e suas inversas para reimplementá-las com base em matrizes esparsas. Em seguida, reimplementaremos a solução do BLUP com base nessas matrizes e avaliaremos a escalabilidade dessa solução em um servidor com grande capacidade de memória RAM.

Tabela 2. Solução BLUP (valores estimados dos efeitos genéticos) do problema apresentado na Tabela 1.

Efeitos	Solução
..... Sexo.....	
1 = Macho	4,358
2 = Fêmea	3,404
..... Bezerro.....	
1	-0,041
2	-0,019
3	0,098
4	-0,009
5	-0,186
6	0,177
7	-0,249
8	0,183

Referências

BOURDON, R. M. **Understanding animal breeding**. Upper Saddle River: Prentice Hall, 2000. 538 p. ill.

COLE, J. B. **A manual for use of PyPedal**: a software package for pedigree analysis. 2012. Disponível em: <<http://pypedal.sourceforge.net/manual/index.html>>. Acesso em: 15 out. 2015.

JIANG, J. **Linear and generalized linear mixed models and their applications**. New York: Springer, 2007. 268 p.

MENEZES, G. R. O. **Desenvolvimento e implementação de metodologias genético-estatísticas em avaliações genéticas de gado de corte**. Campo Grande: Embrapa Gado de Corte, 2015. (Embrapa. Macroprograma 2). Código SEG: 02.13.14.004.00.00. Projeto em andamento.

MISZTAL, I. **Computational techniques in animal breeding**. Athens: University of Georgia, 2011. 188 p. ill. Disponível em: <<http://nce.ads.uga.edu/~ignacy/course11.pdf>>. Acesso em: 15 out. 2015.

MRODE, R. A.; THOMPSON, R. **Linear models for the prediction of animal breeding values**. 3rd ed. Boston: CABI, 2014. 360 p.

PYTHON SOFTWARE FOUNDATION. **Python**. 2015. Disponível em: <<https://www.python.org/>>. Acesso em: 15 out. 2015.

ROSA, A. do N.; MARTINS, E. N.; MENEZES, G. R. de O.; SILVA, L. O. C. da (Ed.). **Melhoramento genético aplicado em gado de corte**: Programa Geneplus-Embrapa. Brasília, DF: Embrapa; Campo Grande, MS: Embrapa Gado de Corte, 2013. 241 p.

SCIPY. **Numpy and Scipy documentation**. Disponível em: <<http://docs.scipy.org/doc/>>. Acesso em: 15 out. 2015.

WATKINS, D. S. **Fundamentals of matrix computations**. 3rd ed. Hoboken: Wiley, 2010. 644 p.

Interpretação e classificação de imagens de satélite na extração de informações sobre o uso da terra e cobertura vegetal na bacia hidrográfica do Rio Paracatu, MG

Paula Guimarães¹

João dos Santos Vila da Silva²

Resumo: Este trabalho, que está inserido no projeto da Rede AgroHidro, tem por objetivo apresentar resultados preliminares do mapeamento do uso da terra e da cobertura vegetal da bacia hidrográfica do Rio Paracatu, MG em 2002, para fins de planejamento. Baseado na segmentação por região e interpretação visual de imagens de satélite, foram identificadas sete classes de usos e quatro de vegetação natural.

Palavras-chave: Planejamento ambiental, gestão ambiental, fronteira agropecuária.

Introdução

O mapeamento de bacias hidrográficas se mostra importante no processo de monitoramento do uso da terra e da cobertura vegetal. Por meio dele é possível avaliar as diferentes formas de atividade que são executadas em uma determinada área, viabilizando o planejamento eficaz das formas de apropriação e exploração dos recursos naturais.

¹ Estudante de Geografia na Universidade Estadual de Campinas (Unicamp), estagiária da Embrapa Informática Agropecuária, Campinas, SP.

² Licenciado em Matemática pela Universidade Federal do Mato Grosso do Sul (UFMS), Mestre em Sensoriamento Remoto pelo Inpe, Doutor em Engenharia Agrícola (Unicamp), pesquisador da Embrapa Informática Agropecuária, Campinas, SP.

Assim sendo, estudos desta natureza podem servir de instrumento orientador nas ações de gestão ambiental, pois como salienta Yassuda (1993) a bacia hidrográfica é a área única de interação das águas com o meio físico, o meio biótico e o meio social, econômico e cultural. Portanto, a elaboração de constantes estudos em áreas de bacias hidrográficas é relevante para analisar e avaliar a ocupação antrópica frente ao avanço das fronteiras agropecuárias nos diferentes biomas brasileiros.

A bacia do Rio Paracatu está inserida no bioma cerrado, no domínio das savanas, compreendida entre os paralelos 15°30' e 19°30' de latitude sul e os meridianos 45°10' e 47°30' de longitude oeste³. O Rio Paracatu é um afluente do Rio São Francisco e drena uma área de aproximadamente 46.600 km², passando pelos estados de Minas Gerais (92%), Goiás (5%) e Distrito Federal (3%), envolvendo 16 municípios, de acordo com Bragança (2012). A figura 1 apresenta a localização da bacia estudada.

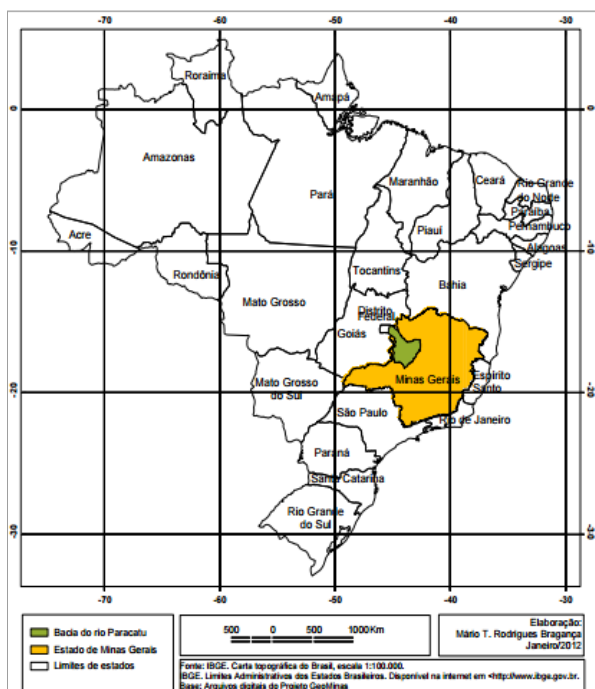


Figura 1. Localização da bacia do Rio Paracatu em relação ao Estado de Minas Gerais e ao Brasil.

Fonte: Bragança (2012).

³ Disponível em: <http://www.atlasdasaguas.ufv.br/paracatu/resumo_paracatu.html/>.

Materiais e Métodos

A classificação do uso e ocupação da terra está sendo feita por meio da interpretação das imagens de satélites Landsat 5 TM (bandas 3, 4 e 5), ano 2002, adquiridas no site Earth Explorer⁴. As imagens foram manipuladas no software Sistema de Processamento de Informações Georreferenciadas (SPRING)⁵ na versão 5.2.6. Foram usadas diversas ferramentas de edição nesse ambiente para efetuar a correção de possíveis erros nos limites dos polígonos, tais como, junção ou separação polígonos para readequação da classe mapeada, linhas não poligonalizadas e linhas duplicadas.

As classes mapeadas foram baseadas na proposta do projeto AgroHidro (REDE AGROHIDRO, 2015) e no Manual Técnico da Vegetação Brasileira (IBGE, 2012), usando como ferramentas de auxílio, as imagens do Google Earth e as séries temporais do EVI2-MODIS⁶ - Plataforma séries Views.

Resultados e Discussão

O projeto ainda se encontra em desenvolvimento. No estágio atual estão sendo feitas revisões e edições dos polígonos em toda a dimensão da bacia, para que na etapa seguinte seja finalizada a classificação dos alvos. Contudo, já foi possível identificar 12 classes, tais como, Pastagem Plantada, Agricultura Anual, Agricultura Perene, Agropecuária, Área Urbana, Solo Nu e Áreas Indiscriminadas (de difícil identificação). No que se refere às áreas naturais identificou-se Floresta, Savana Florestada, Savana Arborizada, Savana Parque e Gramíneo-Lenhosa, e Água.

Considerações Finais

A análise e classificação das imagens do satélite Landsat TM5 no ambiente SPRING e ferramentas como o Google Earth e a Plataforma Series View têm

⁴ Disponível em: <<http://www.earthexplorer.usgs.gov/>>

⁵ Disponível em: <<http://www.dpi.inpe.br/spring/>>

⁶ Disponível em: <<https://www.dsr.inpe.br/laf/series/index.php>>

se mostrado eficaz para o projeto proposto. Com o uso de tais ferramentas será possível gerar um diagnóstico a respeito de uso e ocupação da terra e cobertura vegetal alcançando, deste modo, os objetivos propostos pelo Projeto AgroHidro.

Referências

BRAGANÇA, M. T. R. **Superfícies de erosão do setor centro - oriental da bacia do rio Paracatu no Estado de Minas Gerais**. 2012. 109 f. Dissertação (Mestrado em Geografia) - Faculdade de Filosofia e Ciências Humanas, Universidade de São Paulo, São Paulo.

IBGE. **Manual técnico da vegetação brasileira**. Rio de Janeiro, 2012. 271 p.

LOEBMANN, D. G. dos S. W.; MAÇORANO, R. P.; SILVA, G. B. S. da; VICENTE, L. E.; VICTORIA, D. de C. **Interpretação de alvos a partir de imagens de satélite de média resolução espacial**. Campinas: Embrapa Monitoramento por Satélite, 2012. 24 p. (Embrapa Monitoramento por Satélite. Circular técnica, 21). Disponível em: <<http://ainfo.cnptia.embrapa.br/digital/bitstream/item/85516/1/048-12.pdf>>. Acesso em: 29 set. 2015.

REDE AGROHIDRO. **Agropedia brasilis**. Disponível em: <<https://www.agropediabrasilis.cnptia.embrapa.br/web/agrohidro>>. Acesso em: 29 set. 2015.

YASSUDA, E. R. Gestão de recursos hídricos: fundamentos e aspectos institucionais. **Revista de Administração Pública**, v. 27, n. 2, p. 5-18, 1993.

Interpretação e classificação de imagens de satélite na extração de informações sobre o uso da terra e cobertura vegetal na bacia hidrográfica do rio Teles Pires/MT-PA

Jéssica da Silva Rodrigues Cecim¹

João dos Santos Vila da Silva²

Resumo: Este trabalho é uma das atividades do projeto AgroHidro. O objetivo é mapear o uso da terra e cobertura vegetal na bacia hidrográfica do Rio Teles Pires, em 2003. Utiliza-se imagens do satélite Landsat 5TM (R5G4B3) e o software SPRING 5.3, bem como outras ferramentas, como Google Earth e plataforma Séries View, para a interpretação e classificação dos usos da terra. Mesmo que ainda não classificadas, já foram identificadas sete classes de áreas antrópicas e cinco classes de áreas naturais.

Palavras-chave: análise de imagem, sensoriamento remoto, recursos hídricos, Amazônia, Cerrado.

Introdução

O mapeamento do uso da terra tem grande importância, sobretudo, para o planejamento territorial, pois seus produtos servem de subsídio para a determinação da capacidade da terra para diferentes usos, bem como para a elaboração de uma melhor política de uso da terra (BRITO; PRUDENTE, 2005).

¹ Estudante de Geografia na Universidade Estadual de Campinas (Unicamp), estagiária na Embrapa Informática Agropecuária (Campinas, SP).

² Licenciado em Matemática pela Universidade Federal do Mato Grosso do Sul (UFMS), Mestre em Sensoriamento Remoto pelo Inpe, Doutor em Engenharia Agrícola (Unicamp), pesquisador da Embrapa Informática Agropecuária, Campinas, SP.

O projeto AgroHidro, busca, por meio de estudos inter e multidisciplinares, analisar a interação dos recursos hídricos com as cadeias produtivas agropecuárias e florestais pelo uso de tecnologias, buscando a sustentabilidade e aumento da qualidade de vida dos produtores rurais (REDE AGROHIDRO, 2015).

A bacia do Rio Teles Pires (142.626,9 km²) se configura como uma das grandes bacias hidrográficas brasileiras, abarcando 35 municípios - 2 situados no estado do Pará e 33 no Estado de Mato Grosso. É composta pelos biomas Amazônico e Cerrado, além do contato Floresta-Cerrado, apresentando significativa riqueza mineral e hídrica. (EMPRESA DE PESQUISA ENERGÉTICA, 2009, p. 205).

O objetivo deste trabalho é apresentar o mapeamento que vem sendo desenvolvido na Bacia Teles Pires por meio da interpretação e classificação de imagens de satélite, expondo as etapas já desenvolvidas e resultados alcançados.

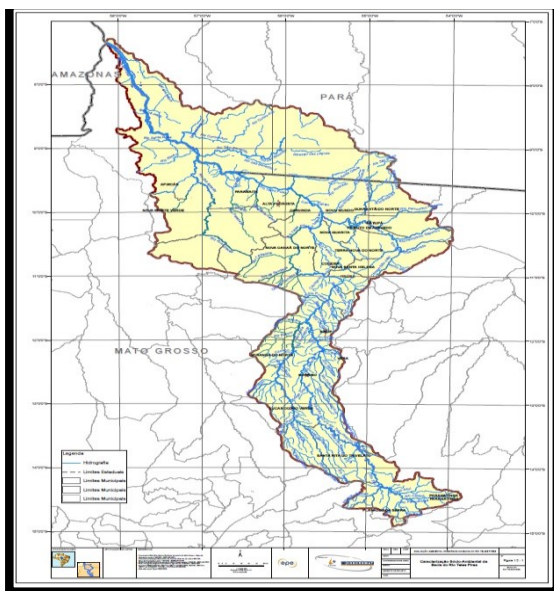


Figura 1. Mapa de Localização da Bacia Hidrográfica do Rio Teles Pires.

Fonte: Empresa de Pesquisa Energética (2009).

Materiais e Métodos

A classificação está sendo realizada por meio da interpretação das imagens de satélites Landsat 5 TM na composição colorida (bandas R5G4B3, correspondendo ao infravermelho médio, infravermelho próximo e vermelho) referentes ao ano de 2003. As imagens foram adquiridas no site Earth Explorer³ já georreferenciadas (localização em um sistema de referência) e ortoretificadas (correção geométrica das imagens). A manipulação das imagens é feita no Sistema de Processamento de Informações Georreferenciadas (SPRING 5.3), desenvolvido pelo Instituto Nacional de Pesquisas Espaciais (Inpe).

As principais ferramentas utilizadas nesse software dizem respeito à edição das linhas (supressão ou adição) que permite ajuste dos polígonos às feições identificadas no terreno, das linhas não poligonalizadas ou linhas duplicadas, por exemplo. A não correção desses erros dificultará a posterior classificação dos polígonos.

A classificação - em tipos de uso e vegetação natural - se dá com base na proposta do projeto AgroHidro e do Manual Técnico da Vegetação Brasileira (IBGE, 2012). Utiliza-se, como referência, o projeto Probio de 2002, o Google Earth e as séries temporais do EVI2-MODIS⁴ - Plataforma Series Views. Busca-se nessas plataformas, por imagens que se aproximem do ano selecionado para essa etapa do projeto (2002/2003). As interpretações são realizadas visualmente.

Resultados e Discussão

O projeto ainda se encontra em andamento. A atual fase diz respeito à revisão e edição dos polígonos para a sua posterior classificação. Polígonos não devidamente editados configuram erros de classificação dos usos da terra.

As correções são realizadas a partir das ferramentas de edição do software SPRING 5.3, com o auxílio do Google Earth e da Plataforma Series Views,

³ Disponível em: <<http://www.earthexplorer.usgs.gov/>>. Acesso em: 28 set. 2015.

⁴ Disponível em: <<https://www.dsr.inpe.br/taf/series/index.php>>. Acesso em: 28 set. 2015.

que permitem uma visualização mais clara das áreas trabalhadas, facilitando, dessa forma, as interpretações acerca dos usos da terra.

Até o presente momento os erros já foram corrigidos em toda a porção norte da bacia, restando apenas a porção sul, menos extensa. Em uma pré-classificação já pode-se identificar no tema áreas antrópicas, as classes de Pastagem Plantada, Agricultura Anual, Agricultura Perene, Agropecuária, Área Urbana, Solo Nu e Áreas Indiscriminadas (de difícil identificação). No que se refere às áreas naturais identificou-se: a) floresta; b) savana florestada; c) savana arborizada; d) savana parque e gramíneo-Lenhosa; e) Água.

Considerações Finais

A utilização do geoprocessamento para a análise e classificação de imagens de satélite Landsat TM5 e o trabalho em ambiente SPRING e demais ferramentas, tem se mostrado eficaz para o projeto proposto. O emprego dessas ferramentas auxilia na obtenção de informações e permite que os pesquisadores realizem análises capazes de compreender e formular um diagnóstico acerca do uso e ocupação da terra e cobertura vegetal na bacia hidrográfica do rio Teles Pires.

Referências

BRITO, J. L. S.; PRUDENTE, T. D. **Mapeamento do uso da terra e cobertura vegetal do município de Uberlândia - MG, utilizando imagens ccd/cbers 2**. Caminho de Geografia. p. 144-153, 2005.

EMPRESA DE PESQUISA ENERGÉTICA (Brasil). **Avaliação ambiental integrada da bacia hidrográfica do Rio Teles Pires**: avaliação ambiental distribuída e conflitos: relatório Parcial 2. Brasília, DF, 2009. Disponível em: <<http://www.epe.gov.br/MeioAmbiente/Documents/AA1%20Teles%20Pires/AA1%20Teles%20Pires%20-%20AAD%20e%20conflitos.pdf>>. Acesso em: 28 set. 2015.

IBGE. Manual técnico da vegetação brasileira. Rio de Janeiro, 2012. 271 p.

REDE AGROHIDRO. **Agropedia brasilis**. Disponível em: <<https://www.agropediabrasilis.cnptia.embrapa.br/web/agrohidro>>. Acesso em: 29 set. 2015.

SOARES, E. R.; SILVA, J. dos S. V. da. Mapeamento da cobertura vegetal e uso da terra na bacia hidrográfica do rio Teles Pires. In: MOSTRA DE ESTAGIÁRIOS E BOLSISTAS DA EMBRAPA INFORMÁTICA AGROPECUÁRIA, 10., 2014, Campinas. **Resumos...** Brasília, DF: Embrapa, 2014. p. 168-169. Disponível em: <<http://ainfo.cnptia.embrapa.br/digital/bitstream/item/118512/1/resumos-mapeamento.pdf>>. Acesso em: 28 set. 2015.

Mecanismo para verificação de conteúdo de temas disponíveis em banco de dados geográficos (PostGIS)

João Luís dos Santos¹
João dos Santos Vila da Silva²

Resumo: A criação de um banco de dados geográficos e sua disponibilização via web demanda atividades de preparação, organização e padronização em cada um dos seus temas. Essas etapas, geralmente executadas de forma manual, podem ser automatizadas e este trabalho apresenta um mecanismo de recuperação de dados a partir do catálogo do sistema gerenciador de banco de dados PostgreSQL, criando subsídios para validação do seu conteúdo pela geração de uma planilha de análise de forma automática.

Palavras-chave: publicação de banco de dados geográficos, Siageo Amazônia, projeto UZEE da Amazônia Legal.

Introdução

O projeto de Uniformização do Zoneamento Ecológico-Econômico da Amazônia Legal (UZEE) vem executando atividades de apoio técnico, financeiro e institucional aos estados que compõem a Amazônia Legal. Um dos principais objetivos é a disponibilização de uma Infraestrutura de Dados Espaciais (IDE) que permita obter um conjunto de serviços que ofereçam uma série de funcionalidades úteis e interessantes aos usuários de dados

¹ Tecnólogo em Análise de Sistemas de Informação, bolsista da Embrapa Informática Agropecuária, Campinas, SP.

² Matemático, doutor em Engenharia Agrícola, pesquisador da Embrapa Informática Agropecuária, Campinas, SP.

geoespaciais. Para isso, etapas preliminares de organização das informações dos Zoneamentos Ecológico-Econômico (ZEEs) dos estados que integram a Amazônia Legal resultaram na criação de um banco de informações geográficas no software Quantum GIS (QGIS). Posteriormente, esse conteúdo foi carregado em um banco de dados no Sistema Gerenciador de Banco de Dados (SGBD) objeto-relacional PostgreSQL e para manipulação de dados geográficos foi utilizado o PostGIS, que é sua extensão de banco de dados espaciais. Para disponibilização dos dados via web foi utilizada a plataforma (BRASIL 2015), onde são utilizadas técnicas que possibilitam a visualização de dados e sua análise pela geração de relatórios espaciais por meio da comparação de áreas de terra com todos os dados do catálogo do i3Geo referentes à Amazônia Legal (POSSAN et. al., 2014). Essas funcionalidades resultaram na criação da ferramenta WebGIS do Sistema Interativo de Análise Geoespacial da Amazônia Legal (Siageo Amazônia). Para a integração de todos esses recursos tecnológicos são realizadas atividades de organização e padronização em cada um dos seus temas e essas etapas, antes executadas de forma manual, foram otimizadas no decorrer do projeto pela automatização de rotinas e integração de recursos. Nesse sentido e para garantir o funcionamento correto da ferramenta de geração de relatórios espaciais, o conteúdo do banco de dados deve ser verificado e validado. Essa tarefa é o objeto de estudo deste trabalho, pela criação de um mecanismo que simplifique o processo de verificação do conteúdo dos temas publicados no banco PostGIS para a criação de um banco de dados consistente.

Materiais e Métodos

O processo de criação e alimentação de um banco de dados geográficos produz como resultado a geração de diversos objetos, de acordo com a arquitetura do SGBD PostgreSQL. Os dados são organizados em grupos de informações em um objeto denominado Schemas, que por sua vez contém todos os objetos relacionados a esse grupo como tabelas, funções, índices, entre outros. Do processo de alimentação do banco PostgreSQL, destacam-se a criação das tabelas que representam cada um dos temas, seus respectivos índices, além da alimentação da estrutura de tabelas do catálogo responsáveis pela organização interna do banco de dados. Uma vez

concluído esse processo, os objetos criados no catálogo ficam disponíveis para consulta, podendo ser acessados e manipulados de forma que sejam coletadas informações sobre os temas, entre elas o conteúdo das colunas que compõem a estrutura das tabelas. Para isso foram construídos *scripts* na linguagem Structured Query Language (SQL), onde é gerado automaticamente o código necessário para recuperação dos dados do catálogo do PostgreSQL referentes ao conteúdo de uma coluna desejada. A execução desse *script* cria na estrutura do banco de dados uma função chamada *getNameColumn* que ao ser executada, exibirá ao usuário o código para recuperação dos dados de cada tabela do banco. Com isso, é possível gerar uma planilha dos dados a serem analisados.

Resultados e Discussão

Os códigos SQL gerados pela função *getNameColumn* permitem a execução de uma nova consulta que terá como resultado a geração de uma planilha com as informações contidas em uma coluna específica de cada tema. No escopo do projeto UZEE, foi definido que cada tema deve conter uma coluna denominada “nome”, que será usada para composição do relatório. Esse conteúdo pode variar em função de cada tema e por isso a planilha é analisada pelo grupo de geotecnologias do projeto, com capacidade para determinar a validade das informações disponíveis. O resultado da recuperação do conteúdo dessa coluna pode ser conferido na planilha exibida na Figura 1. Ela foi estruturada de acordo com os requisitos do projeto UZEE e do catálogo do i3Geo, auxiliando, dessa forma, sua análise pela equipe responsável. Nesse caso, o conteúdo da coluna “nome” do tema Aeroportos, inserido na categoria Meio Socioeconômico do conjunto dos Dados Básicos Estaduais do estado do Acre, referente à iniciativa de Zoneamento Ecológico-Econômico (ZEE) Fase 2 executado na escala de 1:250.000, pode ser confirmado pelos especialistas do projeto e seu con-

Menu	Grupo	Subgrupo	Tema	Código (tema)	Conteúdo (coluna “NOME”)
Dados Básicos Estaduais	Acre - ZEE Fase II (1:250.000)	Meio Socioeconômico	Aeroportos	a_aeroportosirgas2000	Tarauacá
Dados Básicos Estaduais	Acre - ZEE Fase II (1:250.000)	Meio Socioeconômico	Aeroportos	a_aeroportosirgas2000	Sena Madureira
Dados Básicos Estaduais	Acre - ZEE Fase II (1:250.000)	Meio Socioeconômico	Aeroportos	a_aeroportosirgas2000	Cruzeiro do Sul

Figura 1. Planilha com o resultado da consulta pela execução dos códigos de recuperação.

teúdo pode ser validado rapidamente, sem a necessidade de uma consulta direta ao banco de dados.

Considerações Finais

Este trabalho trouxe contribuições importantes ao projeto UZEE no tocante à disponibilização de uma IDE consistente e aderente ao trabalho realizado pelos pesquisadores e especialistas que compõem o grupo de trabalho do projeto. Como resultado principal, pode-se destacar sua integração às ferramentas desenvolvidas no decorrer do projeto que visam à automatização de rotinas e integração de recursos. Em trabalhos futuros, podem ser incorporadas novas funcionalidades que permitam a seleção das colunas de interesse em cada tema, além da otimização do algoritmo e do processo de execução do *script* para geração da planilha.

Referências

BRASIL. Ministério do Meio Ambiente. **Software I3Geo ajuda no processamento de dados geográficos via web**. Disponível em: <<http://www.softwarelivre.gov.br/noticias/software-i3geo-ajuda-no-processamento-de-dados-geograficos-via-web>>. Acesso em: 3 ago. 2015.

POSSAN, L. H. J.; SANTOS, J. L. dos; SILVA, J. dos S. V. da. Uma ferramenta de análise espacial e geração de relatórios dinâmicos para auxílio à gestão ambiental. In: SIMPÓSIO DE GEOTECNOLOGIAS NO PANTANAL, 5., 2014, Campo Grande, MS. **Anais...** São José dos Campos: INPE, 2014. p. 905-914. 1 CD-ROM. Geopantanal 2014. Disponível em: <<http://ainfo.cnptia.embrapa.br/digital/bitstream/item/112155/1/ferramenta-analise.pdf>>. Acesso em: 28 set. 2015.

Disponibilização de dados geoespaciais no SIAGEO Amazônia

Edson Antonio Mengatto Junior¹
João dos Santos Vila da Silva²

Resumo: O Zoneamento Ecológico-Econômico (ZEE) é um recurso técnico de planejamento estratégico direcionado ao ordenamento territorial. Tem a incumbência de estabelecer medidas e padrões que possibilitem a proteção ambiental, a conservação da biodiversidade, o uso integrado e sustentado dos recursos naturais e a melhoria das condições de vida das populações. Tendo em vista a grande quantidade de informações produzidas na elaboração de um ZEE, sua organização por meio de um Sistema de Informações Geográficas (SIG) deu origem ao Sistema Interativo de Análise Geoespacial da Amazônia Legal (Siageo) Amazônia, contendo 634 dados. Possibilita a elaboração de estudos e relatórios técnicos que auxiliem no desenvolvimento de suas áreas de forma sustentável.

Palavras-chave: QGIS, banco de dados, SIG, i3Geo.

Introdução

O planejamento do meio ambiente atualmente considera bases sustentáveis, ao qual possa se fazer uso de seus recursos naturais sem prejudicá-lo e garantindo a existência destes recursos para gerações futuras. Neste contexto, surge o ZEE, um instrumento de planejamento e organização do território, cuja finalidade última é aperfeiçoar o uso do espaço e as políticas públicas (BECKER; EGLER, 1997).

¹ Geógrafo, Doutorando em Geografia - Universidade Estadual de Campinas (Unicamp), bolsista da Embrapa Informática Agropecuária, Campinas, SP.

² Matemático, Doutor em Engenharia Agrícola, pesquisador da Embrapa Informática Agropecuária, Campinas, SP.

O uso de SIG permite a incorporação das informações que podem ser organizadas a partir de um sistema gerenciador de banco de dados para posteriormente serem replicados e disponibilizados via web.

O Siageo Amazônia reúne, de forma sistematizada, as informações geoespaciais utilizadas e produzidas nas diversas iniciativas de ZEE da região. Possibilita a visualização de mapas georreferenciados por meio da manipulação de diferentes níveis de informação de acordo com o interesse e a necessidade do usuário.

Desenvolvido pela Embrapa Informática Agropecuária, em parceria com o Ministério do Meio Ambiente (MMA), os governos dos estados da Amazônia Legal e a Embrapa Amazônia Oriental, com recursos da Financiadora de Estudos e Projetos (Finep) e Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), o Siageo Amazônia fornece, assim, uma base de informações territoriais integradas que permite maior conhecimento do seu território, subsidiando a formulação e a espacialização de planos, programas e políticas, de forma a contribuir para a superação da desarticulação e da sobreposição de ações que incidem sobre a região.

O objetivo deste trabalho foi criar um banco de dados respeitando os parâmetros definidos no projeto, para replicação em um sistema WebGIS que permita a disponibilização dos dados geoespaciais selecionados no projeto.

Materiais e Métodos

Para a construção do banco de dados, foram recebidos dados repassados pelo MMA, que posteriormente foram avaliados e consolidados para a inserção no banco local a partir do uso do SIG QGIS, versão 1.8 (Lisboa), além de se utilizar do Sistema Gerenciador de Banco de Dados (SGBD) PostGreSQL + PostGIS (extensão espacial), atendendo todos os pré-requisitos necessários para o trabalho proposto.

Os dados vetoriais foram recebidos e processados em SIG para atender as necessidades de padronização, aferição e validação. Isto significa definir o sistema de coordenadas e projeção (Sistema de Coordenadas Geográficas, datum Sirgas 2000), o sistema de representação dos dados (RGB) e ajustar a sistemática de nomes e a estrutura de diretórios para o PostGIS (QGIS) e o i3Geo.

Foram disponibilizados dados referentes a nove estados e do Macrozoneamento Amazônia Legal; dados do projeto TerraClass e também dados derivados de estudos que propõe a uniformização das legendas estaduais dos ZEE.

Resultados e Discussão

A recente liberação do Siageo Amazônia para o uso externo à Empresa Brasileira de Pesquisa Agropecuária (Embrapa) torna-se um dos principais resultados apresentados frente ao projeto proposto.

No banco de dados local, replicado e disponibilizado na web, foram inseridos 634 dados geoespaciais. Os dados estão armazenados e agrupados entre os seguintes temas: a) aspectos Institucionais; b) meio Biológico; c) meio Físico; d) meio socioeconômico; e) estudos Integrados. Estes temas foram definidos a partir de reuniões junto ao MMA e podem ser acessados pela estrutura “WebGIS”, de acordo com a Figura 1. A página principal do Siageo Amazônia apresenta informações sobre a construção do sistema.

Outra estrutura disponível no Siageo Amazônia é “Publicações”, que disponibiliza todos os documentos oficiais e legais referentes ao ZEE de cada es-

The image shows the initial presentation of the WebGIS Siageo Amazônia. At the top, there is a navigation bar with a Brazilian flag icon and the text 'BRASIL'. Below this, a secondary navigation bar contains the text 'Acesso à informação' and several menu items: 'Participe', 'Serviços', 'Legislação', and 'Canais'. The main content area features a logo for 'SIAGEO Amazônia' (Sistema Interativo de Análise Geoespacial da Amazônia Legal) on the left. To the right of the logo, the title 'SIAGEO Amazônia' is displayed in large green letters, with the subtitle 'Sistema Interativo de Análise Geoespacial da Amazônia Legal' below it. Below the title, there is a horizontal navigation menu with buttons for 'Home', 'Atlas', 'WebGIS', 'Relatório', 'Publicações', and 'Projeto UZEE'. The 'Publicações' button is highlighted, and a dropdown menu is open, listing the following states: Acre, Amapá, Amazonas, Maranhão, Mato Grosso, Pará, Rondônia, Roraima, and Tocantins. The main content area below the navigation menu displays the title 'Sobre o SIAGEO' and a paragraph of text: 'Instrumento da Política Nacional do Meio Ambiente, o zoneamento ecológico-econômico (ZEE) busca orientar os tomadores de decisão na adoção de ações convergentes com a conservação ambiental, de acordo com as vulnerabilidades e potencialidades do território. Nesse sentido, o Sistema Interativo de Análise Geoespacial da Amazônia Legal - SIAGEO disponibiliza informações geoespaciais utilizadas e produzidas nas diversas iniciativas de ZEE da região amazônica, por meio de mapas georreferenciados e de informações tabulares de modo interativo, por meio da interface de usuário. Do mesmo modo, permite efetuar consultas sobre os mapas temáticos inseridos no catálogo, além de disponibilizar aos usuários o acesso aos documentos técnicos e legais referentes a cada iniciativa de zoneamento.'

Figura 1. Apresentação inicial do WebGIS Siageo Amazônia.

tado. Há ainda uma ferramenta de análise espacial na estrutura “Relatório”. A estrutura “Atlas” mostra um modelo de atlas construído e gerenciado pelo MMA. Já a estrutura “projeto UZEE” apresenta o histórico, os parceiros e outros produtos derivados do projeto em questão.

Considerações Finais

A correta interpretação das informações podem ser úteis para a elaboração de políticas públicas mais adequadas. Seus dados permitem a elaboração de estudos que fomentem um desenvolvimento sustentável.

Referências

BECKER, B. K.; EGLER, C. A. G. **Detalhamento da metodologia para execução do zoneamento ecológico-econômico pelos Estados da Amazônia Legal**. Brasília. DF: Ministério do Meio Ambiente, dos Recursos Hídricos e da Amazônia

Uso da terra na bacia hidrográfica do alto rio Paraguai no Brasil

Cezar Freitas Barros¹
João dos Santos Vila da Silva²

Resumo: Busca-se neste trabalho mapear os principais usos da terra na bacia hidrográfica do alto rio Paraguai, uma das 12 grandes bacias delimitadas pela Agência Nacional de Águas (BRASIL, 2003). O estudo representa um recorte do trabalho que está sendo realizado em cada uma dessas bacias. Inserido na rede de pesquisa AgroHidro, esta pesquisa faz uso de dados do Censo Agropecuário 2006 (IBGE), e também de dados espaciais do território brasileiro. Os mapas foram criados utilizando o software ArcGIS, permitindo assim, a identificação dos municípios com as maiores áreas de lavouras, pastos e florestas na área de estudo.

Palavras-chave: censo agropecuário, Pantanal, regiões produtivas.

Introdução

A rede de pesquisa AgroHidro³ formada por 45 pesquisadores, 12 centros de pesquisa da Empresa Brasileira de Pesquisa Agropecuária (Embrapa) e oito universidades brasileiras, além de outros parceiros internos e externos, tem como objetivo o estudo das interações entre os recursos hídricos e as cadeias produtivas agropecuárias e florestais. Este trabalho se insere na

¹ Bacharel em Geografia e estudante de licenciatura em Geografia pela Universidade Estadual de Campinas (Unicamp), estagiário da Embrapa Informática Agropecuária, Campinas, SP.

² Licenciado em Matemática pela Universidade Federal do Mato Grosso do Sul (UFMS), Mestre em Sensoriamento Remoto pelo INPE, Doutor em Engenharia Agrícola (Unicamp), pesquisador da Embrapa Informática Agropecuária, Campinas, SP.

³ Disponível em: <<http://www.AgroHidro.www.agropediabrasilis.cnptia.embrapa.br>>. Acesso em: 25 set. 2015.

rede de pesquisa com o objetivo de mapear o uso da terra nas grandes bacias brasileiras.

Nesse contexto, o presente trabalho tem, como produtos, mapas do uso da terra para agricultura, pecuária e florestas na Bacia Hidrográfica do Rio Paraguai. A partir destes mapas, pode-se fazer uma comparação com outras bacias hidrográficas delimitadas pela Agência Nacional de Águas (ANA), bem como analisar todo o território brasileiro, identificando portanto as regiões em que há uma maior área destinada à agropecuária, uma vez que é crescente a importância do manejo consciente dos recursos hídricos e de estudos que facilitem a gestão hídrica.

Materiais e Métodos

Os dados utilizados são provenientes, basicamente, do Instituto Brasileiro de Geografia e Estatística (IBGE) e da ANA que disponibiliza, em seu site, dados espaciais vetorizados do território brasileiro, contendo as 12 bacias hidrográficas determinadas pela agência (BRASIL, 2003). No site do IBGE, foram adquiridos dados espaciais do Brasil com a divisão municipal brasileira, sendo que dados quantitativos foram obtidos por meio de tabulação especial.

Para este estudo, os dados obtidos foram filtrados para estabelecimentos que possuem gado de corte como principal atividade econômica. Na bacia analisada, uma das menores, foram identificados 74 municípios. O que determina a presença destas localidades em sua respectiva bacia é o fato de possuírem a maior parte da sua área na região de drenagem do Rio Paraguai.

A bacia do Rio Paraguai, em sua porção brasileira, abrange uma área de 363.446 km². Abriga cerca de 1% da população brasileira e se divide entre os estados do Mato Grosso e Mato Grosso do Sul. É nesta bacia que se localiza o Pantanal, uma das maiores extensões úmidas contínuas do mundo (AGÊNCIA NACIONAL DE ÁGUAS, 2006).

Para executar o mapeamento, foi utilizado o software ArcGIS, especialmente a ferramenta de junção de tabelas de acordo com a localização espacial, ou *spatial join*, que liga tabelas ou elementos no mapa de acordo com as coordenadas, sendo necessário o georreferenciamento dos vetores. Uma vez

relacionadas as tabelas, foi necessário determinar em qual bacia os municípios localizados nos limites entre duas ou três unidades hídricas estavam, o que foi feito de forma manual ou visual, observando em qual bacia estava a parte de maior área de cada município. Posteriormente, foram criados os mapas de acordo com os usos escolhidos: a) área total de lavouras; b) área total de pastagem; c) área total de matas e florestas. A Bacia do Rio Paraguai, com um pequeno número de municípios é, no entanto, uma importante região agropecuária brasileira.

Resultados e Discussão

Os principais produtos gerados neste trabalho foram os três mapas das respectivas classes escolhidas para o estudo: área total de lavouras; área total de pastagens; e área total de matas ou florestas. A determinação da bacia em que cada município está localizado leva em conta os limites municipais, e não das bacias, resultando, portanto, na incompatibilidade entre os limites, visível nos três mapas gerados.

A Figura 1 representa o mapa da área total de lavouras, variando entre até 500 hectares e mais de 10.000 hectares. Entre os usos específicos contidos nesta classe de uso, estão as áreas de lavouras permanentes; lavouras não permanentes ou temporárias; áreas de forrageiras para corte e área com cultivo de flores e mudas.

O município com maior área destinada às lavouras é Itiquira, com cerca de 83.600 ha para a agricultura. Há um total de 648 mil hectares destinados à agricultura na bacia. A Figura 2 representa o mapa da área total de pastagens por município, sendo este provavelmente a principal atividade agrope-

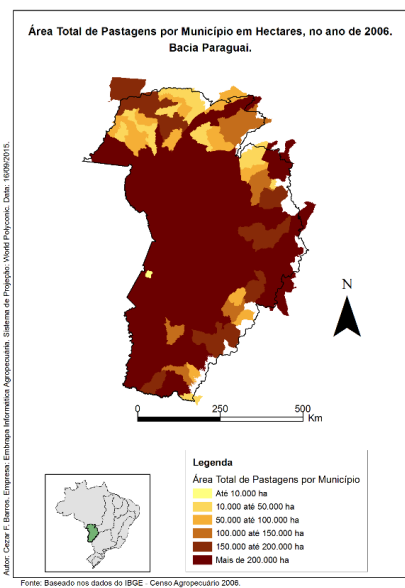


Figura 1. Área total de Lavouras.

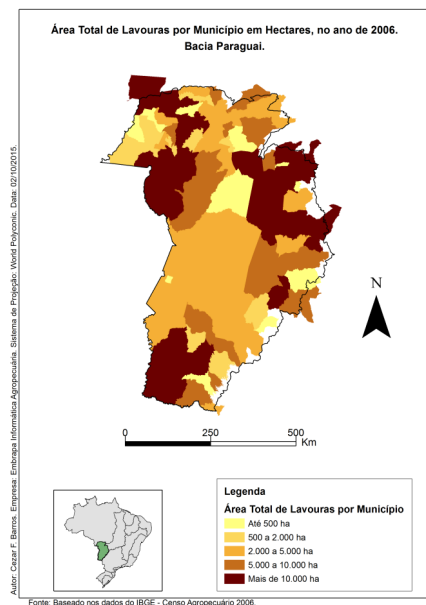


Figura 2. Área total de pastagens.

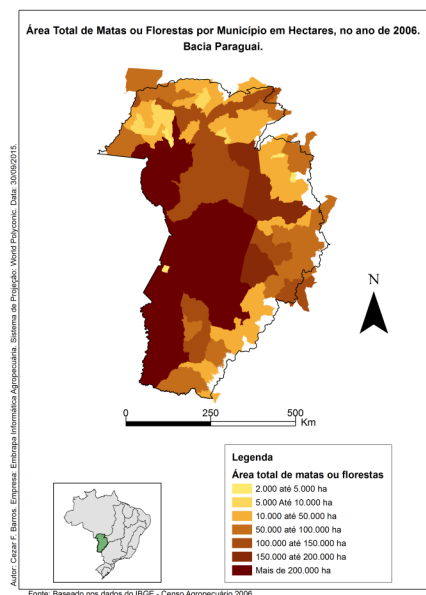


Figura 3. Área total de matas e florestas.

cuária desenvolvida na bacia. Entre os usos específicos estão as áreas de pastagem natural e de pastagem plantada, degradada e não degradada. Grande parte dos municípios possuem área de mais de 200 mil hectares destinados a pastagens, chegando também a casos de menos de 10 mil hectares. Destaca-se o Município de Corumbá, com cerca de 3,6 milhões de hectares. A bacia possui um total de aproximadamente 15,9 milhões de hectares utilizados para pastos.

O mapa de matas e florestas está representado pela Figura 3. Possui classes que vão de 2.000 a mais de 200.000 hectares. Os tipos recenseados nesta categoria foram o de áreas ou florestas naturais destinadas a preservação permanente ou reserva legal; matas ou florestas naturais existentes nos estabelecimentos; área de silvicultura e área de sistemas florestais. O Município de Corumbá, no sudoeste da bacia, também se destaca, com quase 1 milhão de hectares. A área recenseada nesta classe foi de 5,3 milhões de hectares.

Observando a bacia como um todo, foram recenseados cerca de 21,8 milhões de hectares, distribuídos entre estas três grandes categorias de uso, revelando o predomínio da extensão territorial do uso de pastagem.

Considerações Finais

Tendo em vista os mapas e valores pesquisados e citados, pode-se dizer que a bacia Paraguai, apesar de possuir uma pequena quantidade de municípios, além de uma extensão territorial não tão expressiva como algumas outras bacias, demonstra grande importância em relação à sua agropecuária extremamente expressiva, sendo assim uma das regiões produtivas mais importantes do Brasil.

Referências

AGÊNCIA NACIONAL DE ÁGUAS (Brasil). **Região Hidrográfica do Paraguai**. 2006. Disponível em: <<http://www2.ana.gov.br/Paginas/portais/bacias/paraguai.aspx>>. Acesso em: 25 set.. 2015.

CONSELHO NACIONAL DE RECURSOS HÍDRICOS (Brasil). Resolução n° 32 de 15 de outubro de 2003. Institui a Divisão Hidrográfica Nacional. **Diário Oficial da União**, Brasília, DF, 17 dez. 2003. Seção 1, p. 142-14.

IBGE. **Censo Agropecuário 2006**. Disponível em: <www.ibge.gov.br>. Acesso em: 25 set. 2015.

Sistema Interativo de Análise Geoespacial da Amazônia Legal: análise da distribuição e localização de dados

Daniel Dias de Andrade¹
João dos Santos Vila da Silva²

Resumo: O uso de operações estatísticas baseadas em distância são a base para a Análise Espacial e para a tomada de decisão com a utilização dos Sistemas de Informações Geográficas. O módulo de Análise Espacial da ferramenta “Relatório”, integrante do Sistema Interativo de Análise Geoespacial da Amazônia Legal (Siageo Amazônia) possibilita identificar em um raio de até 15 quilômetros as feições de interesse inseridas no sistema e selecionadas conforme necessidade do usuário, cabendo a este a correta interpretação e análise dos resultados.

Palavras-chave: gestão do território, análise espacial, distância entre pontos.

Introdução

Os Zoneamentos Ecológicos-econômicos (ZEE) brasileiros encontram-se em diversas fases de execução e implantação, necessitando de apoio técnico, financeiro e institucional, e têm como atribuição fornecer subsídios para a gestão do território, tendo como eixo norteador os princípios do desenvolvimento sustentável.

Dentre outros objetivos, o projeto “Uniformização do Zoneamento Ecológico-econômico da Amazônia Legal e integração com zoneamentos agroecoló-

¹ Geógrafo pela Universidade de São Paulo (USP), Mestre em Geografia pela Universidade Estadual de Campinas (Unicamp), Bolsista CNPq - DTI-B da Embrapa Informática Agropecuária, Campinas, SP.

² Doutor em Engenharia Agrícola Pesquisador da Embrapa Informática Agropecuária, Campinas, SP.

gicos da região (UZEE)” reúne mapeamentos temáticos diversos que são replicados em uma plataforma WebGIS i3Geo³ para disponibilização via web, denominada Siageo⁴.

As operações estatísticas baseadas em distância são tipos de descrição de padrões que utilizam uma abordagem espacial, baseada no espaçamento das distâncias entre os pontos e/ou de seus valores. Tomando-se cada ponto por sua vez, o ponto mais próximo a ele é encontrado, e então a distância é calculada ou medida. O ponto correto mais próximo é frequentemente óbvio, mas para um trabalho preciso, é mais seguro calcular distâncias entre as coordenadas (UNWIN, 1981).

A ferramenta “Relatório” foi projetada para gerar informações a partir da comparação de pontos, linhas ou polígonos localizados na Amazônia Legal selecionados pelo usuário. No “Módulo Análise Espacial” é gerado um relatório com distâncias, sobreposições e porcentagem destas, entre o tema que se quer analisar e os dados geográficos do Siageo.

Para restringir a busca, é necessário definir um raio máximo, e os temas considerados na análise serão aqueles dentro deste raio. Em cada linha ligando dois pontos são inseridos atributos que indicam a distância e o identificador dos pontos. Esses identificadores são escolhidos com base nas colunas de atributos dos temas envolvidos na análise⁵.

Desta forma, objetiva-se a demonstração e análise da utilização do Módulo Análise Espacial do Siageo e o uso de operações estatísticas baseadas em distância, por meio da escolha do mapeamento das “áreas, com alteração da cobertura vegetal, antropizadas por ocupação rural”⁶, presentes em um raio de 15 quilômetros em relação a um ponto aleatório à Noroeste do município de Boca do Acre, no Amazonas.

³ Interface Integrada para Internet de Ferramentas de Geoprocessamento, desenvolvida pelo Ministério do Meio Ambiente, versão 5.0 SP3.

⁴ Disponível em: <<http://www.amazonia.cnptia.embrapa.br/>>. Acesso em 28 set. 2015.

⁵ Disponível em: <http://www.amazonia.cnptia.embrapa.br/i3geo/ajuda_usuario.php>. Acesso em: 20 set. 2015.

⁶ Definidas como “aquelas nas quais ocorreu supressão vegetal com a consequente exposição dos solos locais [...] Áreas onde ocorre exploração econômica do potencial madeireiro de modo planejado e limitado. Áreas onde ocorre exploração da cobertura vegetal para instalação de atividades/ocupação antrópica características de zonas rurais e que requerem monitoramento ambiental e assistência técnica regular”. Foram utilizados os polígonos de desmatamento acumulado de 1997 a 2007 (Prodes/Inpe) (AMAZONAS, 2008).

Materiais e Métodos

A ferramenta relatório possibilita realizar análises de pontos, linhas ou territórios localizados na área da Amazônia Legal a partir da submissão de qualquer tema e da seleção de distintos temas que compõem o Siageo, como rios, limites municipais, unidades de conservação, dados socioeconômicos, etc. O sistema identifica: a) a porcentagem do tema submetido que sobrepõe cada classe dos temas selecionados; b) a porcentagem de cada classe dos temas que se sobrepõe ao tema submetido; c) as classes dos temas selecionados que estão contidas no tema submetido; d) as classes dos temas selecionados onde o tema submetido está contido; e) as classes dos temas selecionados distantes até 15 quilômetros do tema submetido; f) o(s) Bioma(s) presente(s) no tema submetido.

Quando necessário, é efetuado a reprojeção do tema submetido. Para que o resultado dos cálculos a serem exibidos no Relatório seja obtido em quilômetros, é utilizado o sistema de projeção plana Universal Transversa de Mercator (UTM) (POSSAN et al., 2014).

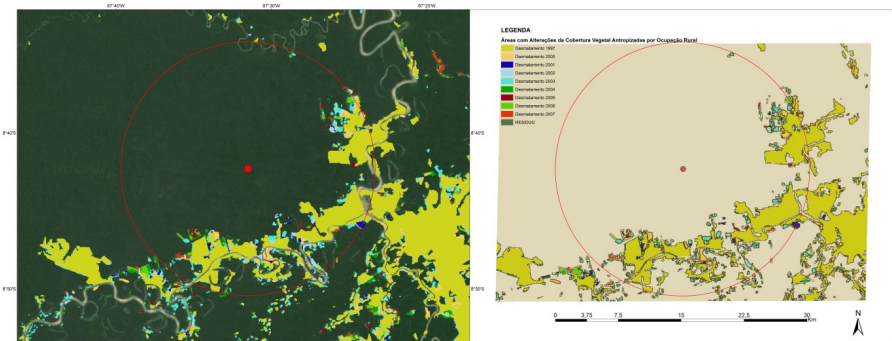


Figura 1. Pontos aleatórios e polígonos da classe “áreas, com alteração da cobertura vegetal, antropizadas por ocupação rural” inscritos por representação do raio de análise sobre imagem orbital (à esquerda) e figura gerada pela ferramenta relatório (à direita).

Resultados e Discussão

Os resultados mostraram que, em até 15 quilômetros do ponto aleatório utilizado, foram detectadas 37 áreas alteradas no ano de 1997, 39 áreas no ano de 2000, 14 no ano de 2001, 46 no ano de 2002, 158 no ano de 2003, 30 no ano de 2004, 10 em 2005, 26 em 2006 e 5 em 2007. A eficácia de políticas públicas que venham a ser adotadas, no entanto, dependerão da natureza dos dados utilizados e de sua correta interpretação e análise dos resultados.

O problema em utilizar medidas de distância entre dois pontos em linha reta pode derivar, primeiramente, por assumirmos que uma geometria euclidiana é apropriada em determinadas análises. Algumas informações, como destaca Unwin (1981), podem resultar em uma melhor análise pela reafirmação da distância como custo, tempo, ou mesmo distância percebida, que não são suscetíveis de serem euclidianas.

Em segundo lugar, encontrar apenas a distância em linha reta pode não ser adequado para algum problema de pesquisa em particular. Pode-se, por exemplo, estar mais interessado na distância rodoviária, na distância a um rio, e assim por diante (UNWIN, 1981).

Também como adverte Gatrell (1983), não necessariamente deve-se objetivar as propriedades de distância somente pela propriedade “métrica”. Isto nos leva ao conceito de espaço relativo, que concebe o espaço como uma relação definida em um conjunto de objetos, mais geral e empiricamente mais útil. Quando nos referimos a uma “relação”, estamos prevendo muitas maneiras em que a separação espacial dos objetos pode ser descrita. Existe um número quase infinito de relações definidas em conjuntos de objetos, onde a distância física é apenas uma delas. Estas, conclui, nem sempre irão produzir uma representação gráfica ou visual (GATRELL, 1983).

Considerações Finais

Considerando o objetivo de localizar e identificar o padrão de distribuição das áreas alteradas por ocupação rural em um raio de 15 quilômetros a partir de um ponto aleatório, a ferramenta obteve êxito na execução.

O módulo de análise espacial é passível de customização em relação aos atributos do banco de dados que são exibidos no relatório, possibilitando

análises distintas entre os diferentes dados levantados em cada um dos mais de 600 mapeamentos temáticos disponibilizados.

Dessa forma, a possibilidade de diversos cruzamentos entre dados na geração do relatório e da customização dos atributos exibidos podem constituir importante contribuição na análise do meio e nas diversas etapas dos processos de tomada de decisão e execução de políticas públicas.

Referências

AMAZONAS. Secretaria de Estado do Meio Ambiente e Desenvolvimento Sustentável. Macrozee - **Macrozoneamento econômico-ecológico do Estado do Amazonas**: resumo executivo, 2008. 78 p.

GATRELL, A. Spatial analysis: a critical perspective. In: A. GATRELL, **Distance and Space**: a geographical perspectives. Oxford: Claredon, 1983. cap. 1. Contemporary problems in geography).

POSSAN, L. H. J.; SANTOS, J. L. dos.; SILVA, J. dos S. V. da. Uma ferramenta de análise espacial e geração de relatórios dinâmicos para auxílio à gestão ambiental. In: SIMPÓSIO DE GEOTECNOLOGIAS NO PANTANAL, 5., 2014, Campo Grande, MS. **Anais...** São José dos Campos: Inpe, 2014. p. 905-914. 1 CD-ROM.

UNWIN, D. Point on maps. In: UNWIN, D. **Introductory spatial analysis**. London: Methuen, 1981. p. 29-65.

Explorando o uso da Teoria Fuzzy no software R: estudo de caso para avaliar a força de infecção da Anemia Infecciosa Equina (AIE)

Gabriela Servidone¹

Sônia Ternes²

Marcelo Rossi³

Resumo: A Anemia Infecciosa Equina (AIE) é uma doença grave no Pantanal Sul-Mato-Grossense, e uma das possíveis vias de infecção é a iatrogênica. A obtenção de dados do processo infeccioso no campo que permitam estimar a força de infecção da transmissão da AIE é uma etapa onerosa, pois envolve altos custos e o bem estar animal. Para contornar a escassez de tais dados, implementou-se um sistema baseado em regras fuzzy para mimetizar a força de infecção correspondente ao reúso de possíveis agulhas contaminadas nos cavalos. Este parâmetro epidemiológico foi gerado por diferentes t-normas, verificando-se sua importância nos respectivos resultados do modelo matemático.

Palavras-chave: teoria de conjuntos fuzzy, força de infecção, epidemiologia.

Introdução

O cenário científico mundial mostra o crescimento vertiginoso da capacidade de aquisição de dados de diferentes naturezas. Com relação à pesquisa agropecuária, seus diversos setores não podem ser tratados de maneira

¹ Estudante de Matemática Aplicada e Computacional na Universidade Estadual de Campinas (Unicamp), bolsista PIBIC da Embrapa Informática Agropecuária, Campinas, SP.

² Pós-doutor do Laboratoire d'Ingénierie pour les Systèmes Complexes (LISC-Cemagref), Pesquisadora da Embrapa Informática Agropecuária, Campinas, SP.

³ Doutor do Programa Inter unidades em Biotecnologia da Universidade de São Paulo (USP), bolsista PNPd/CNPq da Embrapa Informática Agropecuária, Campinas, SP.

homogênea no que diz respeito à medição de variáveis e parâmetros de interesse.

Ao trabalhar com dados com incerteza, a informação deixa de ser representada por um valor determinado e passa a ser representada por um conjunto de dados imprecisos. Desta forma, o uso da teoria clássica dos conjuntos torna-se inviável devido à sua ineficiência no tratamento dessas informações. Entretanto, essas incertezas podem ser estudadas e modeladas, de forma mais robusta, utilizando a teoria dos conjuntos nebulosos, também conhecida como teoria dos conjuntos fuzzy (MEIRELLES et al., 2007).

Conforme Bönisch et al. (2004), a lógica fuzzy é capaz de modelar, de maneira aproximada, o raciocínio humano, visando manipular informações em um ambiente de incerteza e imprecisão. Esse ferramental pode fornecer uma resposta aproximada, baseada em um conhecimento inexato, incompleto ou não totalmente confiável.

Este trabalho tem como objetivo realizar um experimento utilizando lógica fuzzy, buscando avaliar quais t-normas e t-conormas são mais adequadas para representar a força de infecção do possível uso de agulhas contaminadas no modelo determinístico da dinâmica de propagação da AIE (TERNES; VILAMIU, 2013). Esse modelo é representado pelo seguinte sistema de equações diferenciais ordinárias:

$$\left\{ \begin{array}{l} \frac{dS_h}{dt} = \Phi - (\lambda_h + \lambda_n + \mu)S_h \\ \frac{dI_h}{dt} = (\lambda_h + \lambda_n)S_h - \mu I_h \\ \frac{dS_v}{dt} = \varphi(S_v + I_v) + \varepsilon I_v - (\lambda_v - \zeta)S_v \\ \frac{dI_v}{dt} = \lambda_v S_v - (\varepsilon + \zeta)I_v \end{array} \right. \quad \text{onde} \quad \lambda_h = \frac{\alpha N_v I_v}{N_h}, \quad \lambda_v = \beta I_h, \quad \lambda_n = \gamma I_v$$

No modelo acima, as variáveis de interesse representam a população de animais suscetíveis (S_h), equinos infectados (I_h), mutucas infectivas (S_v) e mutucas não infectivas (I_v). Os parâmetros λ_h , λ_v e λ_n representam, respectivamente, a força de infecção da AIE em equinos transmitida por mutucas; força de infecção em mutucas, devido à alimentação do sangue contaminado de equinos e a força de infecção em equinos, a partir do possível reúso de agulhas contaminadas, sendo esta última o objetivo de interesse deste estudo.

Materiais e Métodos

Para criar um sistema baseado em regras fuzzy, são necessárias as seguintes informações: a) funções de pertinência das variáveis de entrada e saída; b) base de regras fuzzy; c) definição da t-norma, t-conorma e implicação; d) modelo de inferência; e) processador de saída (defuzzificação).

Considerando o conceito biológico de λ_n , utilizou-se, como variáveis dependentes de entrada, a taxa de injeções aplicadas por ano e o volume morto da agulha, como mostrado nas Figuras 1a e 1b.

Para representar a função de pertinência de saída, a variável independente escolhida é a força de infecção dada por $\gamma = \alpha * \beta$, onde β é a taxa de injeções e α o volume morto da agulha (Figura 1c).

O próximo passo é a definição da base de regras. Neste momento é despendido tempo na verificação das pertinências, no estudo do modelo biológico, para que as regras sejam as mais realistas possíveis.

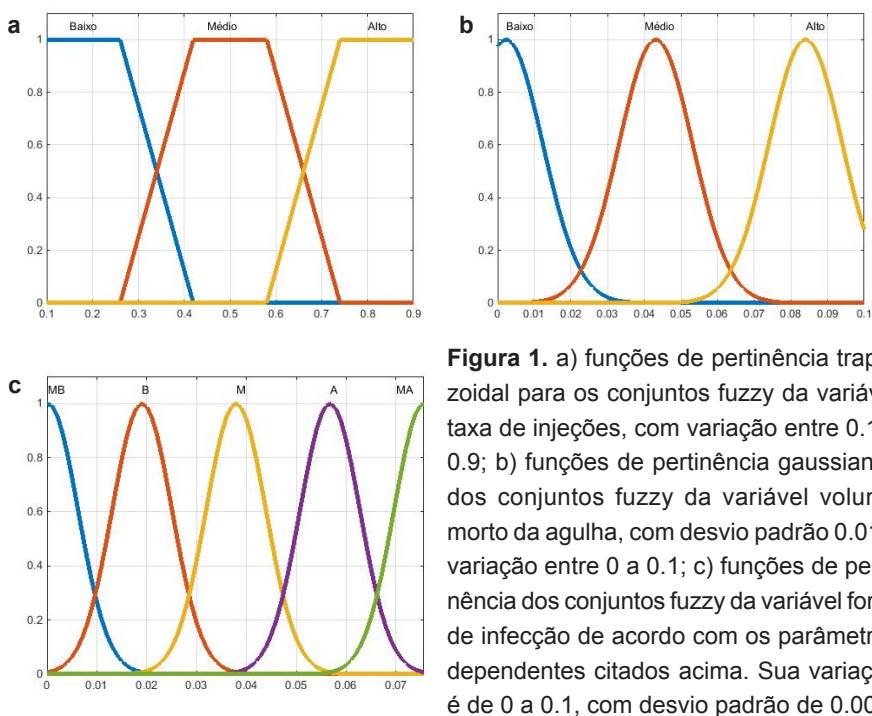


Figura 1. a) funções de pertinência trapezoidal para os conjuntos fuzzy da variável taxa de injeções, com variação entre 0.1 a 0.9; b) funções de pertinência gaussianas dos conjuntos fuzzy da variável volume morto da agulha, com desvio padrão 0.01 e variação entre 0 a 0.1; c) funções de pertinência dos conjuntos fuzzy da variável força de infecção de acordo com os parâmetros dependentes citados acima. Sua variação é de 0 a 0.1, com desvio padrão de 0.006.

Baseado em dados da literatura sobre a infecção pelo Human Immuno deficiency Virus (HIV), que é da mesma família do vírus transmissor da AIE, verificou-se que para agulhas de pequeno volume morto, a possibilidade de infecção é menor. Portanto, tal possibilidade é dependente do volume do conjunto agulha/seringa utilizada por pessoas usuárias de drogas (ABDALA et al., 1999; BOBASHEV; ZULE, 2010). Esta realidade direcionou a escolha das regras para a força de infecção, descritas na Figura 2.

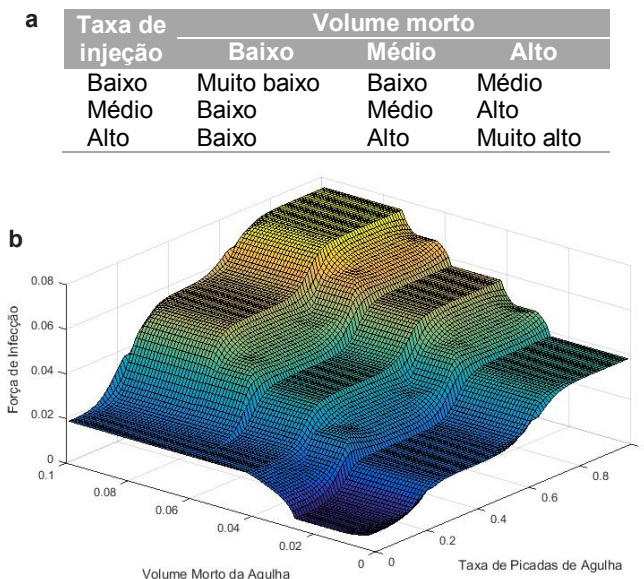


Figura 2. a) base de regras definida na Tabela de acordo com o modelo biológico estudado: Anemia Infecciosa Equina; b) relação gráfica dos valores obtidos aplicados a essa base de regras com t-norma MIN e t-conorma MAX.

O software R (R DEVELOPMENT CORE TEAM, 2015) possui o pacote Fuzzy Rule-Based Systems (FRBS) (RIZA et al., 2015), o qual possui em sua estrutura as diferentes t-normas e t-conormas que foram testadas para o sistema fuzzy que modela a força de infecção do modelo compartimental determinístico. Estas normas são comentadas, de forma reduzida, na Tabela 1.

Importante observar que cada t-norma tem sua respectiva t-conorma para o cálculo. Por fim, utilizando o modelo de inferência de Mamdani e o cálculo do centroide como método de defuzzificação, os conjuntos fuzzy resultantes do

Tabela 1. Listagem das t-normas e t-conormas utilizadas no experimento.

	t-norma	t-conorma
Min/Max	$\min(x_1, x_2)$	$\max(x_1, x_2)$
Hamacher	$(x_1 \cdot x_2) / (x_1 + x_2 - x_1 \cdot x_2)$	$(x_1 + x_2 - 2x_1 \cdot x_2) / (1 - x_1 \cdot x_2)$
Yager	$1 - \min(1, ((1 - x_1) + (1 - x_2)))$	$\min(1, (x_1 + x_2))$
Product	$(x_1 \cdot x_2)$	$(x_1 + x_2 - x_1 \cdot x_2)$
Bounded	$\max(0, x_1 + x_2 - 1)$	$\min(1, x_1 + x_2)$

processo de implicação da base de regras são defuzzificados para a variável de saída “força de infecção”.

Resultados e Discussão

Após o teste das cinco diferentes combinações de t-normas e t-conormas, encontramos a variação descrita na Figura 3. Uma matriz 10x10 com parâmetros igualmente espaçados foi construída com valores estimados de força de infecção e alimentado o respectivo parâmetro nas simulações do modelo. O gráfico demonstra a prevalência dos cavalos infectados no período de 4 anos.

Claramente, quando analisamos entre YAGER e MIN/MAX, vemos que a t-norma de MIN é biologicamente mais aceitável por refletir uma variação

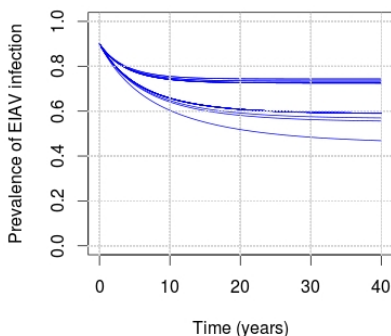
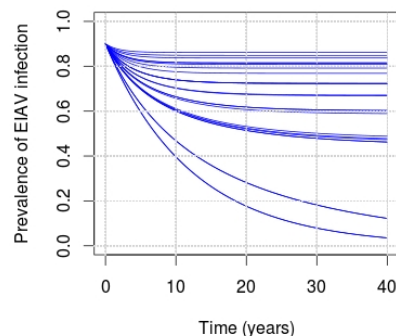
a Prevalence of infected horse - YAGER**b** Prevalence of infected horse - MIN

Figura 3. Gráfico da porcentagem de cavalos infectados por anos. (a) t-norma e t-conorma YAGER. (b) t-norma MIN e t-conorma MAX. A simulação foi feita para 100 diferentes valores de entrada (força de infecção).

da porcentagem de cavalos infectados em relação às diferentes forças de infecção utilizadas. Resultados semelhantes foram obtidos quando testadas as normas PRODUCT e HAMACHER. Já a BOUNDED, não teve um comportamento esperado, pois todos os valores retornados pelos diferentes inputs foram 0.04.

Considerações Finais

Para a organização do experimento, foi primeiramente avaliado se as definições de t-normas, t-conormas, implicação, defuzzificação e tipo do modelo eram as preferíveis para a utilização. Desse modo, foi possível fazer um levantamento do que o pacote FRBS do software R podia oferecer como recurso. Inicializando os parâmetros do modelo como constantes, foi analisado como essas definições se comportavam quando eram variadas. O resultado obtido foi o esperado: mesmo sendo a norma do mínimo e a conorma do máximo as mais simples, elas foram as que retornaram uma abrangência de valores os quais são biologicamente aceitáveis. Outras normas como a do produto e a Hamacher retornaram resultados iguais, mas como seus cálculos possuem grau de complexidade maior, mais uma vez o uso das normas de mínimo e máximo se mostra mais adequado.

Como trabalho futuro, visualiza-se testar o efeito da quantidade de conjuntos fuzzy (granularidade) no particionamento das variáveis, bem como diferentes formatos das funções de pertinência.

Agradecimentos

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pelo suporte financeiro (Processo 160212/2015-4).

Referências

ABDALA, N.; STEPHENS, P. C.; GRIFFITH, B. P.; HEIMER, R. Survival of HIV-1 in syringes. *Journal of Acquired Immune Deficiency Syndromes*, v. 20, n. 1, p. 73-80, Jan.1999.

BOBASHEV, G. V.; ZULE, W. A. Modeling the effect of high dead space syringes on the human immunodeficiency virus (HIV) epidemic among injecting drug users. **Addiction**, v. 105, n. 8, p. 1439-47, Aug. 2010. DOI: 10.1111/j.1360-0443.2010.02976.x.

BÖNISCH, S.; ASSAD, M. L. L.; CÂMARA, G.; MONTEIRO, A. M. V. Representação e propagação de incertezas em dados de solos. I. Atributos categóricos. **Revista Brasileira de Ciências do Solo**, v. 28, n. 1, p. 21-32, jan./fev. 2004.

MEIRELLES, M. S. P.; MOREIRA, F. R.; CAMARA, G. Técnicas de inferência espacial. In: MEIRELLES, M. S. P.; CAMARA, G.; ALMEIDA, C. M. de (Ed). **Geomática: modelos e aplicações ambientais**. Brasília, DF : Embrapa Informação Tecnológica, 2007. Cap. 3.

R DEVELOPMENT CORE TEAM. **R: a language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing, 2015. Disponível em: <<http://www.R-project.org/>>. Acesso em: 15 out. 2015.

RIZA, L. S.; BERGMEIR, C.; HERRERA, F.; BENÍTEZ, J. M. Frbs: fuzzy rule-based systems for classification and regression in R. **Journal of Statistical Software**, v. 65, n. 6, p. 1-30, May 2015. Disponível em: <<http://www.jstatsoft.org/v65/i06/>>. Acesso em: 15 out. 2015.

TERNES, S.; VILAMIU, R. **Epidemiologia matemática da Anemia Infecciosa Equina**. In: WORKSHOP PROJETO, 1., 2013, Corumbá. Corumbá: [s.n.], 2013. Slides. Trabalho não publicado.

Disseminando a aplicação do R-Shiny em métodos quantitativos e computação científica na web

Bernardo França Rocha¹
Sônia Ternes²
Marcelo Rossi³

Resumo: Neste trabalho é apresentada uma alternativa para a disseminação via web de aplicações que envolvem técnicas de métodos quantitativos e computação científica, a ser usada por profissionais que não estão completamente familiarizados com linguagens de programação em seu cotidiano. Para tal, utilizando o software R e o pacote Shiny, ambas ferramentas open source, foi criada uma aplicação web para mostrar os resultados obtidos na avaliação da dinâmica de propagação da Anemia Infecciosa Equina (AIE), expressa por um sistema de equações diferenciais ordinárias previamente desenvolvidas. A aplicação web ofereceu melhor interação e maior facilidade no manuseio e na interpretação do modelo determinístico por profissionais de diversas formações, em especial os das áreas biológicas, que são os especialistas do domínio em estudo, quando comparado ao formato tradicional, baseado no uso de linguagens de programação clássicas. **Palavras-chave:** Software R, Shiny, anemia infecciosa equina, aplicação web.

¹ Estudante de Matemática Aplicada e Computacional da Universidade Estadual de Campinas (UNICAMP), bolsista PIBIC da Embrapa Informática Agropecuária, Campinas, SP.

² Pós-doutor do Laboratoire d'Ingénierie pour les Systèmes Complexes (LISC-Cemagref), Pesquisadora da Embrapa Informática Agropecuária, Campinas, SP.

³ Doutor do Programa Inter unidades em Biotecnologia da Universidade de São Paulo (USP), bolsista PNPd/CNPq da Embrapa Informática Agropecuária, Campinas, SP.

Introdução

No Laboratório de Matemática Computacional (LabMac) da Embrapa Informática Agropecuária existe a necessidade de soluções que permitam aos usuários externos testar e reusar os sistemas de simulação e algoritmos desenvolvidos em projetos de pesquisa pelo laboratório. Dentro desse contexto, desenvolveu-se um protótipo de aplicação que permite a interação, via web, do usuário final na escolha dos parâmetros de simulação de um modelo matemático consolidado e implementado para avaliar a dinâmica de propagação da AIE no Pantanal brasileiro (TERNES; VILAMIU, 2013; VILAMIU et al., 2013). Esse protótipo utiliza como base o software de análise estatística e de computação científica R (R DEVELOPMENT CORE TEAM, 2015) e seu framework de aplicações web conhecido como Shiny (CHANG et al., 2015), ambos open source.

Assim, o objetivo deste trabalho é oferecer uma alternativa para a disseminação do uso de modelos matemáticos e de simulação entre usuários diversos, aumentando a abrangência e facilitando a compreensão dos métodos quantitativos e de computação científica fornecidos pela Empresa Brasileira de Pesquisa Agropecuária (Embrapa) no suporte à tomada de decisão para problemas complexos do setor agropecuário.

Materiais e Métodos

O desenvolvimento de ferramenta para uso via web, visando à simulação e ao teste de modelos quantitativos é possibilitado pelo uso do framework de aplicações Shiny (R package version 0.12.2), que permite acessar todo o conteúdo do software R (version 3.2.2) via interface web.

O Sistema de Equações Diferenciais Ordinárias (EDO) que descreve o modelo matemático compartimental utilizado como teste (TERNES; VILAMIU, 2013; VILAMIU et al., 2013) foi implementado por meio dos recursos de resolução de EDO do software R, mais especificamente pelo uso dos pacotes deSolve (R PACKAGE VERSION 3.2.2) (SOETAERT et al., 2010) e PRACMA (R PACKAGE VERSION 1.8.6) (BORCHERS, 2015). Utilizou-se bibliotecas de interface gráfica existentes no Shiny para que os usuários pudessem definir os valores dos parâmetros de entrada do modelo.

Adotou-se o modelo para a AIE supracitado para melhor compreensão das etapas de elaboração do protótipo, assim como das hipóteses que nortearam as escolhas das variáveis de controle. Note que o usuário final do protótipo, ou seja, especialista do domínio da aplicação, não necessita de conhecimento prévio de programação, de modelagem matemática ou de qualquer método de computação científica para testar e obter novos resultados da simulação computacional, bastando apenas conhecimento sobre o problema representado pelo modelo, significado biológico de seus parâmetros e acesso à web.

Resultados e Discussão

A partir de rotinas implementadas usando o ambiente de programação do software R, denominado RStudio (RSTUDIO TEAM, 2015), foi desenvolvido um aplicativo que simula o sistema de EDO do modelo representando a dinâmica de infecção da AIE, a ser executado via web. Ele é totalmente multiplataforma e é executado localmente a partir de qualquer ambiente R (R CONSOLE, RGui, RStudio). Também pode ser hospedado em um servidor web remoto, com a possibilidade de ser acessado por qualquer usuário a partir de um navegador web instalado em seu computador. O protótipo desenvolvido pode, ainda, ser customizado ou estendido utilizando as tecnologias HTML, CSS e Java Script.

A Figura 1 mostra a interface do RStudio acessada pelo programador ou especialista na modelagem matemática, onde estão inseridas as equações do modelo e a programação da interface do protótipo.

A interface do protótipo acessado via web é mostrada na Figura 2. A partir dos botões do tipo “slide” apresentados na interface, o usuário pode modificar os parâmetros do modelo de forma a adaptá-lo à análise que pretende realizar. Após a alteração dos valores dos parâmetros, a simulação ocorre automaticamente em um servidor remoto utilizando os pacotes do software R, e o resultado da nova simulação é visualizado em tempo real pelos gráficos disponibilizados na interface criada. A escolha de quais parâmetros o usuário poderá modificar, assim como a interface a ser visualizada, é opção do criador ou do desenvolvedor do modelo matemático.

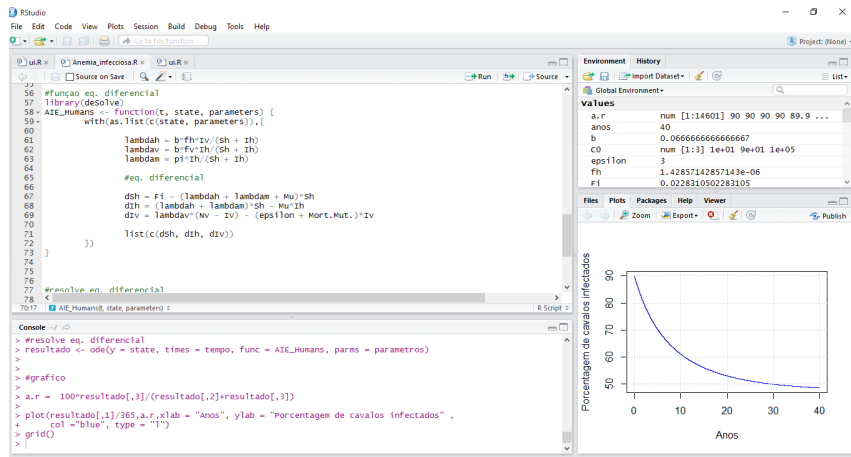


Figura 1. Ambiente RStudio, com trecho do código implementado em linguagem do software R.

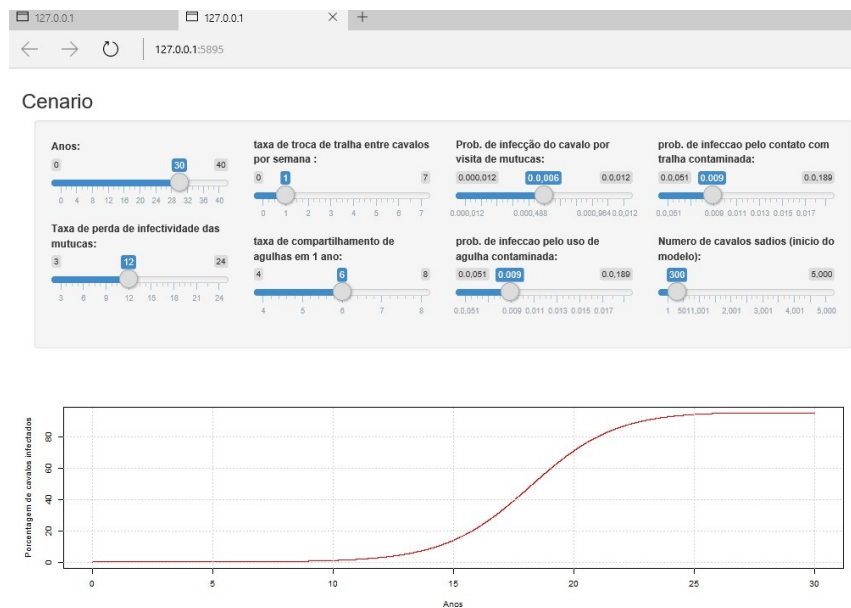


Figura 2. Interface do protótipo visível ao usuário. No quadro superior são apresentados os botões “slide” para escolha dos valores dos parâmetros de entrada e no quadro inferior é mostrado o gráfico gerado em tempo real pela execução automática após a movimentação de qualquer um dos botões.

Conclusões

O Shiny combina o poder computacional do software R com a interatividade da web moderna, o que o torna uma opção interessante para uso em computação científica.

O protótipo criado facilita o uso de modelos matemáticos e de simulação por usuários diversos, possibilitando a disseminação dos resultados das simulações e algoritmos desenvolvidos em projetos de pesquisa.

A criação do aplicativo permite uma modularização no processo de desenvolvimento de sistemas web. Após o término do protótipo, o desenvolvedor pode estabelecer um processo de adaptação a novos modelos matemáticos, necessitando apenas de modificações nas equações utilizadas em modelos anteriores e seus respectivos parâmetros.

Como trabalho futuro, pretende-se testar diferentes layouts de interfaces gráficas, além de replicar o processo de construção de interface web em outros projetos de pesquisa que possuem como foco o desenvolvimento de modelos quantitativos.

Agradecimentos

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pelo suporte financeiro (Processo 113993/2015-3).

Referências

BORCHERS, H. W. **Pracma**: practical numerical math functions. 2015. Disponível em: <<https://cran.r-project.org/web/packages/pracma/index.html>>. Acesso em: 15 out. 2015.

CHANG, W.; CHENG, J.; ALLAIRE, J. J.; XIE, Y.; MCPHERSON, J. **Shiny**: Web application framework for R. 2015. Disponível em: <<http://CRAN.R-project.org/package=shiny>>. Acesso em: 15 out. 2015.

R DEVELOPMENT CORE TEAM. **R**: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2015. Disponível em: <<http://www.R-project.org/>>. Acesso em: 15 out. 2015.

RSTUDIO TEAM. **RStudio**: integrated development for R. Boston, 2015. Disponível em: <<http://www.rstudio.com>>. Acesso em: 10 out. 2015.

SOETAERT, K.; PETZOLDT, T.; SETZER, W. R. Solving differential equations in R: Package deSolve. **Journal of Statistical Software**, v. 33, n. 9, p. 1-25, Feb. 2010. Disponível em: <<http://www.jstatsoft.org/v33/i09/>>.

TERNES, S.; VILAMIU, R. **Epidemiologia matemática da Anemia Infecciosa Equina**. In: WORKSHOP PROJETO, 1., 2013, Corumbá. Corumbá: [s.n.], 2013. Slides. Trabalho não publicado.

VILAMIU, R.; TERNES, S.; MARQUES, A . P. D.; NOGUEIRA, M. F. The role of horseflies in the prevalence of Equine Infectious Anemia (EIA) in the Brazilian Pantanal. In: INTERNATIONAL CONFERENCE ON MATHEMATICAL METHODS AND MODELS IN BIOSCIENCES, Sofia. **Conference book**. [S.l.: s.n.], 2013. Não paginado. BIOMATH 2013. Disponível em: <<http://ainfo.cnptia.embrapa.br/digital/bitstream/item/96829/1/role-horseflies.pdf>>. Acesso em: 15 out. 2015.

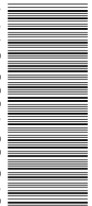
Embrapa

Informática Agropecuária

Ministério da
Agricultura, Pecuária
e Abastecimento

GOVERNO FEDERAL
BRASIL
PÁTRIA EDUCADORA

978-85-7035-544-7



CGPE 12445