WORLD CLASS RESEARCH WITH REAL LIFE IMPACT



www.fapesp.br/convenios/microsoft

AGRODATAMINE: DEVELOPMENT OF ALGORITHMS AND METHODS OF DATA MINING TO SUPPORT RESEARCHES ON CLIMATE CHANGES RECORDING AGROMETEOROLOGY

Agma Juci Machado Traina

São Carlos Institute of Mathematics and Computer Sciences / University of São Paulo (USP) FAPESP Process 2009/53153-3 | Term: Dec 2009 to Jul 2012

This research project aims at investigating and developing techniques and computational methods to evaluate complex data sets used by meteorologists and agro-meteorologists to assist in their research on the impacts of climate change in the Agriculture. Therefore, this project proposes: 1) developing new methods to analyze the outputs of regional models of climate change prediction; and 2) to compare them with real measurements collected by surface stations and satellites in order to assess the quality of the generated data and make allowances for calibrating the models. We intend to develop new methods to filter, analyze and extract association patterns between researchers from the Database and Images Group (GBDI) at ICMC/USP, the Embrapa Agriculture Informatics, the Cepagri (Unicamp), the Databases Groups at UFSCar, UFABC and CPTEC-Inpe.

Step 3: Time series mining Association Quantization Rules patterns presentation process generation Step 1: Image georeferencing **Radiometric Calibration** Step 2: Time series extraction One series **Geometric Correction** for each Extraction pixel Maximum Cross Correlation by area Series for Multitemporal average Images value **Products Generation** Extraction Series for by point pixel Images Climate 21 Database Database Surface Cloud Cover Temperature Databases

MAIN RESULTS

1. A large database of the Agrodatamine project.

The database was designed considering data requirements identified by the experts in the agrometeorology area members of the project. We have developed the conceptual and logical designs for the database schema, respectively using the Entity-Relationship Model and Relational Model. In order to build the database, we have chosen the DBMS (Databse Management System) PostGreSQL 8.4 (http://www.postgresql.org), as it is an open source and provides the resources to meet the project needs. This database is only for the project use, since the data are part of an agreement between Cepagri-Unicamp and CPTEC-Inpe. 2. The first version of the software SatimageExplorer.

The software has been developed aimed at automatically generating time series from satellite image sequences. The specialist can define a region of interest (ROI) to be analyzed and the system provides the time series of the region regarding a given index or measurement. New indexes and measurements can be designed and introduced into the system in order to provide fast creation of new time series and their analysis. The system is a valuable tool to aid agrometeorologists to evaluate a crop development and climate changes. Versions for MS Windows as well as Linux Ubuntu are now available. It is already available as a beta test version.



PRODUCTS/PUBLICATIONS

JOURNALS

Vespa TG, Traina AJM, Traina Jr. C. Efficient bulk-loading on dynamic metric access methods. *Information Systems Journal (IS)*. Elsevier, **35(5)**: 557-569. July 2010.

Vieira MR, Chino FJT, Traina Jr. C, Traina AJM. A visual framework to understand similarity queries and explore data in metric access methods. Special issue on Beyond Multimedia and XML Streams Querying and Mining. *International Journal of Business Intelligence and Data Mining (IJBIDM)*. **5(4)**: 370-397. 2010.

Kaster D, Bugatti PH, Traina AJM, Traina Jr.C. FMI-SiR: A flexible and efficient module for similarity searching on oracle database. *Journal of Information and Data Management (JIDM)*. **1(2)**: 229-244. Sep 2010.

Romani LAS, Avila AMH, Zullo Jr. J, Traina Jr.C, Traina AJM. Mining relevant and extreme patterns on climate time series with CLIPSMiner. *Journal of Information and Data Management* (*JIDM*). **1(3)**: 245-260, Sep 2010.

Traina Jr. C, Traina AJM, Leejay W, Faloutsos C. Fast feature selection using fractal dimension. *Journal of Information and Data Management (JIDM)*. **1(1)**: 3-16. June 2010.

Rodrigues Jr. JF, Tong H, Traina AJM, Faloutsos C. Large graph visualization with GMine. *Journal of Visual Languages & Computing*. Elsevier, 28 pags. (*accepted*)

INTERNATIONAL CONFERENCES

Cordeiro RLF, Guo F, Haverkamp DS, Horne JH, Hughes EK, Kim G, Traina AJM, Traina Jr. C, Faloutsos C. QMAS: querying, mining and summarizing multi-modal databases. *10th IEEE International Conference on Data Mining (ICDM 2010)*. December 14-17, 2010. Sydney, Australia.

Bueno R, Ribeiro MX, Traina AJM, Traina Jr. C. Improving medical image retrieval through multi-descriptor similarity functions and association rules. 23rd International Symposium on Computer-Based Medical Systems (CBMS 2010), pp. 309-314. October 12-15, 2010. Perth, Australia.

Porto Ferreira MR, Ponciano Silva M, Amo S, Pereira FSF, Traina AJM, Traina Jr. C, Chbeir R. Integrating user preference to similarity queries over medical images datasets. 23rd International Symposium on Computer-Based Medical Systems (CBMS 2010), pp. 486-491. October 12-15, 2010. Perth, Australia.

Silva SF, Brandoli B, Eler DM, Batista Neto JES, Traina AJM. Silhouette-based feature selection for classification of medical images. 23rd International Symposium on Computer-Based Medical Systems (CBMS 2010). pp 315-320. October 12-15, 2010. Perth, Australia.



NATIONAL CONFERENCES

Chino DYT, Romani LAS, Traina AJM. Extração de séries temporais de imagens de satélite para monitoramento de safras agrícolas e de dados climáticos. *CTIC - XXIX Concurso de Trabalhos de Iniciação Científica da SBC*, pp. 08 pags. 20-23 de julho de 2010, Belo Horizonte, MG. Second place in the context.

Nunes SA, Romani LAS, Avila AMH, Traina Jr. C, Sousa EPM, Traina AJM. Análise baseada em fractais para identificação de mudanças de tendências em múltiplas séries climáticas. 25 Brazilian Symposium on Databases. Belo Horizonte, MG. Proceedings of the XXV Brazilian Symposium on Databases - Short Paper Session, 2010. pp. 65-72.

Vieira MR, Chino FJT, Traina Jr. C, Traina AJM. MAMView: a framework for visualization of metric trees. *Simpósio Brasileiro de Banco de Dados (SBBD 2010)*. Belo Horizonte, MG. October 5-8, 2010. 16 – *Demo Session, 2010*. pp. 1-6.

POSTERS

Traina AJM, Romani LAS, Cordeiro R, Sousa EPM, Ribeiro MX, Ávila AMH, Zullo Jr. J, Traina Jr. C. How to find relevant patterns in climate data: an efficient and effective framework to mine climate time series and remote sensing images. *2010 SIAM Annual Meeting (AN10)*. Pittsburgh. July 12-16, 2010.

Agma Juci Machado Traina

Instituto de Ciências Matemáticas e Computação de São Carlos – Universidade de São Paulo (USP)

Av. do Trabalhador Sãocarlense, 400 Caixa Postal 668 – Centro 13560-950 – São Carlos, SP – Brazil

agma@icmc.usp.br 55.16.3373-9674