NEWS and EVENTS

Submit your news and announcements up to two weeks prior to publication for inclusion in the current issue. Submissions should be relevant to the SIGKDD community and should not be advertisements for products or services. Success stories from Data Mining vendors are welcome.

News Items

SIGKDD Explorations publishes news-oriented articles as submitted without review. News articles can be up to 2 pages long and cover important timely topics in the area. The Editor reserves the right to reject any submissions at his discretion.

Announcements Policy

SIGKDD Explorations publishes announcements that are submitted as is without review. Announcements cannot be advertisements and should be of general interest to the wider community. The Editor reserves the right to reject any requests for announcements at his discretion.

NEWS

ACM NOTICE to Past Authors of ACM Published Articles.

submitted by: Julie Goetz, ACM
URL: http://www.acm.org/

ACM intends to create a complete electronic archive of all articles and/or other material previously published by ACM.

If you have written a work that was previously published by ACM in any journal or conference proceedings prior to 1978, or any SIG Newsletter at any time, and you do NOT want this work to appear in the ACM Digital Library, please inform permissions@acm.org, stating the title of the work, the author(s), and where and when published.

If you have any questions concerning this statement, please contact permissions@acm.org, as well.

Gartner Group Report on Data Mining.

submitted by: Beverly Brown Stockstill, SAS Institute
URL:
www.gartnerweb.com/public/axl/reprints/sas/00082107.html

LAKE BUENA VISTA, Fla. -- During a presentation titled "Data Mining in the E-Commerce Age" at GartnerGroup's ITxpo99 in October, Gartner analyst Alexander Linden identified IBM, Oracle, and SAS Institute as vendors best equipped to address emerging market needs, such as e-business, customer relationship management and more.

In the 1999 Data Mining Workbench Magic Quadrant, SAS Enterprise Miner from SAS Institute was singled out as the product that best addresses what Gartner identifed as the "most common needs of most data mining projects, solving the 'average' project faster than any other approach."

To see the full report, visit the web site above.

New CMU Masters Program in Knowledge Discovery and Data Mining

submitted by: Diane Stilde
URL: http://www.cs.cmu.edu/~kdd

The extraordinary spread of computers and online data is changing forever the way that important decisions are made in many organizations. Hospitals analyze online medical records to decide which treatments to apply to future patients. Banks routinely analyze past financial records to learn to spot future fraud. Today's demand for data mining expertise far exceeds the supply, and this imbalance will become more severe over the coming decade.

To educate the next generation of experts in this important area, Carnegie Mellon University offers a new Master's program in Knowledge Discovery and Data Mining (KDD). This new inter-disciplinary program trains students to become tomorrow's leaders in the rapidly growing area of Knowledge Discovery and Data Mining. It is offered by Carnegie Mellon's Center for Automated Learning and Discovery (CALD), which has assembled a large multi-disciplinary team of faculty and students across several academic departments.

KDD candidates will be trained in all important areas related to scientific data mining and knowledge discovery. The Master's program balances interdisciplinary course work, hands-on project work, and cutting-edge research carried out under direct faculty supervision. The curriculum addresses areas such as advanced machine learning algorithms, statistical principles and foundations, database and data warehousing methods, complexity analysis, approaches to data visualization, privacy and security, and specific application areas such as business, marketing, finance, and public policy. Our graduates are uniquely positioned to pioneer new data mining and knowledge discovery efforts, and to pursue top notch research on the next generation of data mining tools, algorithms, and systems.

CALD received $100,000 from Microsoft for Graduate Fellowships for this program. "We are pleased to support Carnegie Mellon in establishing a new academic program in this area of computing," said Usama Fayyad, Head of Microsoft Research's Data Mining and Exploration (DMX) Group. "The potential for data mining to positively impact education, business, and even scientific data analysis is enormous. The future depends on laying academic foundations and scientific principles to develop the needed innovative algorithms and solutions."

Carnegie Mellon invites applications of qualified individuals. Admission is highly competitive. A limited number of fellowships are available, which will be provided on a competitive basis. The application deadline is February 5, 2000.

For more details about the program or to apply please vist the web site above.

Microsoft To Release Public Beta Version of OLE DB for Data Mining Specification.

submitted by: Z. Tang
URL: http://www.microsoft.com/data

The Analysis Services Group in Microsoft's Data Warehousing Product Unit, in collaboration with Microsoft Research's Data Mining & Exploration (DMX) Group are working on enhancing the SQL Server platform by extending it with data mining components. Data Mining is a relatively young and promising area. However, the data mining industry today is highly fragmented, making it difficult for application software vendors and corporate developers to integrate different knowledge-discovery tools.

OLE DB for Data Ming (DM) is an extension to OLE DB supporting data mining operations over OLE DB data providers. There are more 40 ISVs in the business intelligence fields that contributed to this specification. Its goal is to provide an industry standard for data mining so that different data mining algorithms from various data mining ISVs can be easily plug into user applications. Those software packages that provide data mining algorithms are called Data Mining Provider, those applications that use data mining features are called Data Mining Consumer. OLE DB for DM specifies the common interface between Data Mining Consumer and Data Mining Provider.

OLE DB for DM doesn't add new OLE DB interfaces, rather it defines a query language with which consumer applications can communicate with the data mining providers. The syntax of this new data mining query language is similar to SQL syntax, which keeps it familiar to application developers. The data mining objects (such as models and input data sets) and data mining operations (such as prediction, browsing, and so forth) are mapped to analogous familiar entities in the relational world. For example, a data mining model is similar to a relational table, which is composed with a number of column attributes. Some columns are used for training while others are used to do prediction. Creating and training a data mining model is similar to creating and populating a relational table.

We are also working closely with the DMG Group, and industry consortium supporting the PMML format for persisting models. We hope to coordinate with the PMML Board (Dr. Usama Fayyad of MSR being the Microsoft representative) on using the PMML standard whenever possible.

OLE DB for DM has also defined a number of schema rowset which allow consumer application to discover information such as mining services, mining model, mining columns and model contents. Consumer applications can access these schema rowsets through ADO.

Today most industry data mining packages requires data to be stored in some proprietary format. OLE DB for DM allows data mining providers to access to any tabular data source through OLE DB provider, which means data mining analysis can be performed directly on a relation database.

OLE DB for DM is designed to support most popular data mining algorithms. The current specification is a public beta version and its contents are continuously being enriched. We appreciate readers to send feedbacks to **oledbdm@microsoft.com.**

New UCI KDD Archive of Databases.

submitted by: Padhraic Smyth
URL: http://www.ics.uci.edu/~kdd/

The University of California, Irvine KDD Archive is a new online repository of large data sets encompassing a wide variety of data types, analysis tasks, and application areas. The primary role of this repository is to serve as a benchmark testbed to enable researchers in knowledge discovery and data mining to scale existing and future data analysis algorithms to very large data sets.

This repository is currently under construction and is still in a preliminary form. This work is supported by a grant from the Information and Data Management Program at the National Science Foundation and is intended to extend the existing UCI Machine Learning Database Repository by several orders of magnitude. Visit http://www.ics.uci.edu/~kdd/ for more details.

New book on Self-organizing Data Mining.

submitted by: Frank Lemke
URL: http://www.knowledgeminer.net

A PDF version of the new book by Mueller/Lemke "Self-Organizing Data Mining" can be downloaded from the Web. The concept of an inductive, evolutionary data mining using parametric and nonparametric approaches is described. Parametric algorithms like GMDH Neural Networks are able to self-select relevant input variables while evolving an optimal network that does not overfit the design data and that can be analyzed analytically on the fly. Many sample applications from economy, ecology, sociology are included and can be downloaded free along with a demo version of the KnowledgeMiner software.

EVENTS

PADD 2000: The Practical Application of Knowledge Discovery and Data Mining

submitted by: info@pap.com
URL www.practical-applications.co.uk/PADD2000/

Announcing PADD 2000, The Practical Application of Knowledge Discovery and Data Mining 11th-14th April 2000, Manchester, UK For more Details visit web site above.

KDD-2000 Submissions Deadline March 3rd, 2000

submitted by: Paul Bradley
URL http://www.acm.org/sigs/sigkdd/kdd2000/

Reminder tat deadlines for KDD-2000 submissions (papers, tutorial proposals, workshop proposals, etc is fast approaching.

GECCO-2000 Workshop on Data Mining with Evolutionary Algorithms

submitted by: Alex A. Freitas
URL http://www.ppgia.pucpr.br/~dmea/

GECCO-2000 Workshop on Data Mining with Evolutionary Algorithms. July 8, 2000. Las Vegas, NV, USA. The workshop is part of GECCO-2000 (Genetic and Evolutionary Computation Conf.), to be held in Las Vegas, NV, USA, July 8-12, 2000. Workshop Chairman: Alex A. Freitas - alex@ppgia.pucpr.br

URL for GECCO-2000: http://www.genetic-algorithm.org/ GECCO2000 /gecco2000mainpage.htm

Microsoft Corporation Data Mining Positions

Contact Information: *Kristine Ingraham, Recruiting* **e-mail: datajobs@microsoft.com**

We have several open development positions in Microsoft Research (DMX Group) and product groups in Data Mining, Databases, and Data Warehousing. This is an exciting opportunity to contribute to data mining products or to help with analysis and warehousing of data. Top notch development skills are a must.

Job Title: SOFTWARE DESIGN ENGINEER (Research or Products): build data mining components at Microsoft. Primary responsibilities include the design, implementation and debugging of server, interface or client components of data mining platform. Designing, developing, and maintaining algorithms for data mining, data analysis, pattern recognition, and data visualization from databases, heavy emphasis on database considerations. Qualifications should include a minimum of 2 years full-time C++ development experience under Windows and experience building database applications with SQL Server, preferably using OLEDB, ADO, or ODBC. COM experience strongly desired. Proficiency in SQL and/or OLAP strongly preferred. Experience in Visual Basic would be a strong asset.

Job Title: DEVELOPMENT MANAGER (Products): This is a Development Manager role as part of a team responsible for delivering a data warehouse and reporting service (iDSS) that collects, processes, and manages the usage metrics for the MSN online properties. The group is focused on delivering and maintaining a world-wide reporting system that will allow CG online properties to make informed business decisions across markets and CG products. Primary responsibilities will include: Lead the teams which are building the Data Warehouse Infrastructure (data collection, monitoring, transformation, Modeling, OLAP). Provide architecture leadership in designing next generation real-time reporting systems and Data Mining and profiling stores. Become a visionary and advocate for iDSS technology. Retain and motivate a high power development team. Translate Architecture to specifications. Develop and maintain schedules and allocate development resources. Partner with other groups within Microsoft on leveraging SQL/OLAP/OWC technology Qualifications: Experience in building large scale service oriented systems, experience in C/C++ and SQL, Excellent planning, leadership, and communication skills. Experience in internet technologies and measurement technologies. Ability to identify and balance quality, time, costs, risks of development efforts. BS degree in Computer Science or related technical field preferred.

Job Title: SOFTWARE DESIGN ENGINEER (Products): Responsibilities: analyzing business problem, transforming data, developing models, deploying and measuring models; and providing and maintaining a software service environment to data mining analyst and customers. Required qualifications: demonstrated programming experience with Perl, C/C++, SQL, and scripting languages; demonstrated experience and knowledge of commercial data mining, data transformation, and visualization tools; thorough technical understanding of Internet technologies; a track record of successfully exploiting large, dirty, real-world data sets, particularly for internet applications; working knowledge of relational databases; ability to prototype ideas rapidly based on existing software components from Windows and UNIX environment; and technical problem solving. Preferred qualifications include experience with large data warehousing systems and an understanding of business issues in data mining design and deployment.

Job Title: DATA ANALYST/WEB MINER (Products): Projects include methodology research, tools evaluation, model development, prototyping, and deployment. Required qualifications: at least MS degree (Ph.D. highly preferred) in Computer Science, Statistics, or Mathematics or equivalent job experience. Broad experience applying data mining techniques, neural networks, statistics, pattern recognition, to real-world data. Ability to work independently and research innovative solutions to challenging technical problems. Working knowledge of relational databases. Familiarity with data mining tools and statistical packages. Fluent in one of the following: C/C++, SQL, Perl. Must possess good communications skills. Preferred qualifications include significant experience solving business problems involving internet data mining. Substantial expertise with large business data sets. Good scripting and rapid prototyping skill. We offer excellent compensation plus a complete benefits package including an employee stock option plan.

Microsoft is an Equal Opportunity Employer and supports workforce Diversity

DBMiner 2.0 (Enterprise)

DBMiner 2.0 (Enterprise) that incorporates many years of research and development on data mining is a powerful data mining tool for data warehouses and relational databases.

It consists of the following function modules: (1) data warehouse construction (using Microsoft SQLServer 7.0 OLAP Manager), (2) table/chart driven OLAP operations (using Microsoft Excel'2000), (3) statistics-based exploration and marking of cube cells at both current and drill-down levels, (4) 3-D cube view of the data warehouse, (5) boxplot view of the data cube, (6) multi-level association and correlation analysis, (7) multi-level decision tree analysis, and (8) multi-level clustering analysis. Filtering, drilling, slicing, and dicing can be performed on any part of the data cube in both OLAP and data mining analyses. It can also be linked to S-PLUS, a statistical analysis software, for statistical analysis of OLAP and data mining results. Time-series analysis will be included in the future release.

DBMiner 2.0 runs on Windows/NT. It uses Microsoft SQLServer 7.0 (required) and Office'2000 (required) to perform data warehousing and OLAP. It communicates with various relational database systems, via SQLServer 7.0 OLAP Manager. We are also working on a new, light-weighted version which runs on Microfosft Office'2000 without using Microsoft SQLServer7.0.

A 90-day free evaluation copy of DBMiner 2.0 can be downloaded at http://www.dbminer.com. A single-user license of DBMiner 2.0 (Enterprise) can be purchased at the price of $999.00 (U.S.) plus tax (when applicable). This also includes free upgrades up to one year. As an academic discount, one user license may cover the use of it in teaching in one university. We will be pleased to take contracts for customized development of particular data mining applications based on the DBMiner system. For more information, please visit the Web page http://www.dbminer.com or http://db.cs.sfu.ca/DBMiner, or by e-mails to: sales@dbminer.com.