

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

KDD-99 to include an Industrial Track.

submitted by: Ronny Kohavi and Jim Gray

URL: <http://research.microsoft.com/datamine/kdd99>

For the first time, the Knowledge Discovery & Data Mining conference (KDD-99) will feature an industrial track. The track, co-chaired by Jim Gray from Microsoft Research and Ronny Kohavi from Blue Martini Software, includes nine talks covering

- Three case studies at DaimlerChrysler and Amdocs
- An integrated mining system for finding interesting patterns
- IBM's Intelligent Miner's text mining system
- A system for monitoring newsfeeds used at Lexis-Nexis
- A visualization system for link analysis
- New SQL primitives for knowledge discovery that were implemented at Compaq/Tandem
- An SQL extension that supports efficient complex queries.

These talks were selected from a field of 27 submissions. The track will provide a great opportunity to meet people who are designing and implementing knowledge discovery and data mining systems in industry. Come to share your experiences and learn from others!

New UCI Academic Program: MS Specialization in Knowledge Discovery in Data.

submitted by: Padhraic Smyth

URL: <http://www.ics.uci.edu/~gcounsel/masterreqs.html>

The Information and Computer Science Department at the University of California at Irvine is pleased to announce the

introduction of a Masters degree in computer science with a KDD specialization. Students can choose from a curriculum including courses in data mining, database modeling, machine learning, information retrieval, probabilistic learning, and multivariate statistics.

Participating faculty in this program include Padhraic Smyth, Sharad Mehrotra, Michael Pazzani, Dennis Kibler, and Bob Newcomb.

The goal of this MS specialization is to educate students in both the fundamental principles of computational methods for modeling data as well as providing a broad foundation in emerging methods for knowledge discovery and data mining. In addition to the coursework, numerous opportunities exist for students to participate in research projects with participating faculty and internships with local companies. Visit <http://www.ics.uci.edu/~gcounsel/masterreqs.html> for further details.

Microsoft Research Grant Establishes

UW Data Mining Institute

submitted by: R. Ramakrishnan & O. Mangasarian

URL: <http://www.news.wisc.edu/thisweek/Research/Engr/Y99/datamine.html>

Data mining at the University of Wisconsin-Madison received a boost this month from Microsoft Corporation. The research division of the company based in Redmond, Washington, awarded the Computer Sciences Department a four-year grant, valued at approximately \$720,000, to establish a Data Mining Institute to study the hidden potential of huge databases. "This grant is part of our overall commitment to collaborating with major academic institutions and fostering the growth of important new cross-disciplinary areas like data mining," said Jim Gray, senior researcher for scalable servers at Microsoft Research. "UW-Madison is one of the top academic database research groups in the world, and we're delighted to be working with this premiere group of scientists."

"Data mining has received a great deal of attention among large corporations and industrial research labs for years," said Usama Fayyad, senior researcher in data mining and exploration for Microsoft Research. "I'm excited to see one of the top universities in database systems form a Data Mining Institute bridging several disciplines. I hope to see many more computer science departments establish formal academic programs in data mining."

The DMI will be directed by Olvi Mangasarian and Raghu Ramakrishnan and will also have Jeff Naughton and Michael Ferris as principal investigators. Around six PhD students will be supported while doing their research in data mining. It is anticipated that the unrestricted nature of the grant will enable

DMI researchers to investigate higher-risk problems and to work with a broader range of application experts.

The goals of the UW-Madison DMI are to bring together the powerful tools of the database and the mathematical programming communities to harness and extract knowledge from the vast store of data that is being accumulated by industrial, research and internet organizations. Ramakrishnan said the university is especially equipped to work on real-world applications with enormous databases.

Very large-scale applications already exist. Companies specializing in credit card fraud reduction have programs that can analyze millions of daily credit transactions. The search programs are trained to flag peculiarities that might suggest theft, such as changes in location or types of purchase.

The World Wide Web will be the catalyst for much broader data mining applications, said Ramakrishnan, since it provides the ultimate publicly accessible database. Mining the Web is different from conventional key word searching, since the programs are designed to find patterns or trends across different subjects.

One exciting example at UW-Madison is a breast cancer diagnosis and prognosis tool developed by UW-Madison computer sciences professor Olvi Mangasarian and Medical School colleagues. The data mining program analyzes tumor size and fine-needle aspirate samples to estimate cancer-free periods for patients. The program recently mined a National Cancer Institute database of more than 40,000 breast cancer patients, helping the program achieve more reliable results. The goal is to provide a non-invasive option for prognosis. Patients currently have to undergo a painful removal of lymph nodes under their arm to receive an accurate prognosis, said Mangasarian.

Ramakrishnan plans to exploit systematic data evolution to achieve significant performance improvements in incremental maintenance of data mining models, and to "monitor" the changes in data characteristics as the database evolves. As organizations collect and maintain ever-larger data sets over time, describing operations conducted at several locations, they want to know how trends evolve over time and how trends vary by location.

Ramakrishnan will investigate efficient algorithms for the answering these questions. In developing algorithms, the approach will be to consider how these algorithms impact the underlying data management systems, and how they can benefit from existing query processing capabilities.

Naughton's principal interest in data mining is to study how relational database systems such as Microsoft SQL-Server and OLAP tools such as Microsoft Plato can be more tightly integrated with data mining techniques. Among the tasks to be considered are (a) how can the existing capabilities of systems like SQL-server and Plato be exploited for data mining purposes, (b) how can systems like SQL-server and Plato be extended to better support data Mining, and (c) how can the results of a data mining session be incorporated back into these systems to enable further interactive analysis.

Mangasarian plans to work on generalized support vector machines (GSVMs) which are key to data discrimination in very high dimensional feature spaces that cannot easily be handled by other techniques. GSVMs separate data by a very general nonlinear surface induced by an arbitrary kernel and simplified by

parameter suppression. Linear programming chunking and successive overrelaxation algorithms for massive data discrimination will be further extended beyond the 10-million datasets currently processed. These approaches will be applied to various datasets and particularly to medical ones such as the National Cancer Institute SEER Database of over 40,000 breast cancer cases which has already been mined for practical survival curves.

Ferris plans to efficiently use vast amounts of computing resources to solve very large-scale mathematical programs generated by data mining problems. The principal aim of his research has been the development of tools that enable applications experts to formulate and solve such optimization problems on a metacomputer. In order to make many optimization techniques practical, they need to be carried out in parallel.

Rather than require a large parallel computer, Ferris will utilize metacomputer, a confederation of heterogeneous computing resources including, but not limited to, supercomputers, workstations, and specialized machines connected through a network. The tool developed for concurrently solving optimization problems exploits one vast, and largely untapped, part of this metacomputer, a pool of pre-existing, "off-the-shelf" workstations via Condor. Ferris also plans to work on modeling languages such as GAMS and AMPL in conjunction with the Condor distributed computing system to solve extremely large data mining problems.

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.

EVENTS

DBMINER Educational version E1.1

Released + Related Data Mining Ph.D.

Theses Available

submitted by: Jiawei Han

URL [http:// www.dbminer.com](http://www.dbminer.com)

DBMiner Technology Inc. and Intelligent Database Systems Research Laboratory of Simon Fraser University (<http://db.cs.sfu.ca>) have released DBMiner Educational Version

E1.1 on the Web for demo downloading and/or adoption. Research publications and Ph.D./M.Sc. theses can be downloaded from there as well.

Cvis Updated Release on AlphaWorks

submitted by: Scott Spangler

URL [http:// www.alphaWorks.ibm.com/formula/CViz](http://www.alphaWorks.ibm.com/formula/CViz)

The CViz tool for cluster visualization and analysis has recently been updated on the Alphaworks web site. CViz allows the analyst to visualize high dimension data sets by clustering the examples drawing scatter plots along the planes defined by the cluster centroids. These scatter plots can then be dynamically rotated. This version of CViz can take raw text input files and cluster each line of the file as a text example. To download CViz go to: <http://www.alphaWorks.ibm.com/formula/CViz>

Special Session on Biomedical Applications of KDD at ACM Symposium on Applied Computing

submitted by: Warren Jones

URL <http://www.acm.org/conferences/sac/sac00>

Special Session on Biomedical Applications of Knowledge Discovery and Data Mining at the ACM Symposium on Applied Computing, March 19-21, 2000, Villa Como, Italy. This Session will be part of the Biomedical Computing Track which is in cooperation with ACM SIGBIO. Contact: Warren T. Jones (jones@cis.uab.edu). The deadline for papers is September 1, 1999. For additional information about the Symposium and submission procedures visit the home page above.

Special Issue of ACM SIGBIO Newsletter on Biomedical Applications of KDD

submitted by: Warren Jones (jones@cis.uab.edu)

URL <http://www.cis.uab.edu/info/kdrg/biomedical.html>

A Special Issue of ACM SIGBIO Newsletter on Biomedical Applications of KDD (Editor: Warren T. Jones) Vol. 18, No. 3 (December 1998) contains twenty-five brief research project descriptions representing twelve countries. A web site has been created by the Knowledge Discovery Research Group of the University of Alabama at Birmingham (UAB) which contains the URL's for these and other similar research groups. Groups not included at the site listed above are welcome to send information and requests to above e-mail.

Data Mining Club Established

submitted by: Amit Seth

URL <http://clubs.yahoo.com/clubs/datamining>

Data Mining Club - a vendor neutral place to discuss data mining related issues (technologies, methodologies and applications).

This club was recently bestowed with the datawarehouse.com/DM Review Magazine Editor's Choice Award.

AAAI-99 & GECCO-99 Workshop on Data Mining - Evolutionary Algorithms

submitted by: Alex Alves Freitas

URL <http://www.ppgia.pucpr.br/~dmea/>

AAAI-99 & GECCO-99 Workshop on Data Mining with Evolutionary Algorithms: Research Directions. July 18, Orlando, FL, USA. Chairman: Alex A. Freitas (alex@ppgia.pucpr.br)

Advanced Course on Machine Learning and Applications at ACAI'99

submitted by: George Paliouras

URL <http://www.iit.demokritos.gr/skel/eetn/acai99>

Advanced Course at Artificial Intelligence 1999 (ACAI '99) on Machine Learning and Applications, Chania, Greece, 5-16 July

Call for Paper for DOLAP '99

submitted by: Il-Yeol Song

URL <http://www.pages.drexel.edu/faculty/songiy/dolap.html>

ACM Second International Workshop on Data Warehousing and OLAP (DOLAP '99), Kansas City, Missouri, USA, November 6, 1999. In conjunction with ACM CIKM '99: Eight International Conference on Information and Knowledge Management. Submission deadline: July 1, 1999. For further information, please contact song@drexel.edu.

Advanced Course on Machine Learning and Applications at ACAI'99

submitted by: George Paliouras

URL <http://www.iit.demokritos.gr/skel/eetn/acai99>

Advanced Course at Artificial Intelligence 1999 (ACAI '99) on Machine Learning and Applications, Chania, Greece, 5-16 July

IEEE Intelligent Systems Special Issue

submitted by: Se June Hong

URL <http://computer.org/Intelligent/>

IEEE Intelligent Systems magazine is planning a special issue on data mining in late 1999. The special issue will feature papers on data mining techniques with emphasis on practical usefulness, scalability, and capability to handle noisy data. Intelligent Systems solicits papers on real applications based on data mining techniques: The domain of application can be scientific, business, or industry.