ANLY600

Course Summary

Course : ANLY600 **Title :** Data Mining
Length of Course : 8
Prerequisites : N/A **Credit Hours :** 3

Description

Course Description: This course covers data mining using the R programming language. It offers hands on experience approach through a learn-by-doing-it strategy. It further integrates data mining topics with applied business analytics to address real world data mining cases. It continues the examination of the role of “Data Mining in R”, and review statistics techniques in prescriptive analytics, and some predictive analytics. Additionally, some standard techniques and excel functions will be also covered.

Course Scope:

Data that has relevance for managerial decisions is accumulating at an incredible rate due to a host of technological advances. Electronic data capture has become inexpensive and ubiquitous as a by-product of innovations such as the internet, e-commerce, electronic banking, point-of-sale devices, bar-code readers, and intelligent machines. Such data is often stored in data warehouses and data marts specifically intended for management decision support. Data mining is a rapidly growing field that is concerned with developing techniques to assist managers to make intelligent use of these repositories. A number of successful applications have been reported in areas such as credit rating, fraud detection, database marketing, customer relationship management, and stock market investments. The field of data mining has evolved from the disciplines of statistics and artificial intelligence.

This course will examine methods that have emerged from both fields and proven to be of value in recognizing patterns and making predictions from an applications perspective. We will survey applications and provide an opportunity for hands-on experimentation with algorithms for data mining using easy-to- use software and cases.

Objectives

After successfully completing this course, you will be able to

- LO 1: Demonstrate advanced knowledge of data mining concepts and techniques.
- LO 2: Apply the techniques of clustering, classification, association finding, feature selection and visualization on real world data
- LO 3: Determine whether a real world problem has a data mining solution

- LO 4: Apply data mining software and toolkits in a range of applications
- LO 5: Set up a data mining process for an application, including data preparation, modelling and evaluation
- LO 6: Demonstrate knowledge of the ethical considerations involved in data mining.

Outline

Week 1:

Topic

The Big Picture

Learning Objectives

LO 1

LO 2

Readings

Data Mining – Chp 1

& Class Readings (Website)

Assignment

Week 1 Assignment

Week 2:

Topic

Data Cleaning and Preparation

Learning Objectives

LO 2

LO 5

LO 6

Readings

Data Mining – Chp 2

& Class Readings (Website)

Assignment

Week 2 Assignment

Week 3:

Topic

Logistic Regression

Learning Objectives

LO 1

LO 2

LO 3

LO 4

Readings

Data Mining – Chp 5,7

& Class Readings (Website)

Assignment

Week 3 Assignment

Week 4:

Topic

Review Week A

Learning Objectives

LO 1

LO 2

LO 3

LO 4

Assignment

Makeup work

Week 5:

Topic

Classification Trees

Learning Objectives

LO 1

LO 2

LO 3

LO 4

Readings

Data Mining – Chp 13

& Class Readings (Website)

Assignment

Week 5 Assignment

Week 6:

Topic

Clustering

Learning Objectives

LO 1

LO 2

LO 3

LO 4

Readings

Data Mining – Chp 15

& Class Readings (Website)

Assignment

Week 6 Assignment

Week 7:

Topic

Market Baskets

Learning Objectives

LO 1

LO 2

LO 3

LO 4

Readings

Data Mining – Chp 16

& Class Readings (Website)

Assignment

Week 7 Assignment

Week 8:

Topic

Review Week B

Learning Objectives

LO 1

LO 2

LO 3

LO 4

Assignment

Makeup work

Evaluation

Grading:

Name	Grade %

Materials

Book Title: Data Mining and Business Analytics with R

Author: Johannes Ledolter

Publication Info: Wiley

ISBN: 9781118447147

Required Course Textbooks:

Data Mining and Business Analytics with R

Johannes Ledolter

URL: https://www.vitalsource.com/referral?term=9781118593639

Additional Resources

Additional resources will be provided in the class.

Web Sites

In addition to the required course texts, the following public domain web sites are useful. Please abide by the university's academic honesty policy when using Internet sources as well. Note web site addresses are

subject to change.

Site Name	Web Site URL/Address
Machine Learning Videos	**http://work.caltech.edu/library/index.html**
OnePageR	**http:// togaware.com/onepager/**
Data Mining Resources	**http://www.ngdata.com/data-mining-resources/**
Top Free Data Mining Software	**http://www.predictiveanalyticstoday.com/top-free-data-mining-software/**

Course Guidelines

Citation and Reference Style

Attention Please: Students will follow the APA Format as the sole citation and reference style used in written work submitted as part of coursework to the University. Assignments completed in a narrative essay or composition format must follow the citation style cited in the APA Format. Citation and APA format is not required in forum discussion posts.

Late Assignments

Students are expected to submit classroom assignments by the posted due date (listed under each Assignment) and to complete the course according to the published class schedule. As adults, students, and working professionals, I understand you must manage competing demands on your time. Should you need additional time to complete an assignment, please contact me before the due date so we can discuss the situation and determine an acceptable resolution. Routine submission of late assignments is unacceptable and may result in points deducted from your final course grade. If arrangements are not made in advance, a late penalty of 10% will be assessed for any assignment submitted 1-7 days past the due date.

Assignments will not be accepted after the 7th day. No work will be accepted past the final day

of class.

Netiquette

Online universities promote the advancement of knowledge through positive and constructive debate – both inside and outside the classroom. Forums on the Internet, however, can occasionally degenerate into needless insults and “flaming.” Such activity and the loss of good manners are not acceptable in a university setting – basic academic rules of good behavior and proper “Netiquette” must persist.

Remember that you are in a place for the rewards and excitement of learning which does not include descent to personal attacks or student attempts to stifle the Forum of others.

- **Technology Limitations:** While you should feel free to explore the full-range of creative composition in your formal papers, keep e-mail layouts simple. The Sakai classroom may not fully support MIME or HTML encoded messages, which means that bold face, italics, underlining, and a variety of color-coding or other visual effects will not translate in your e-mail messages.

- **Humor Note:** Despite the best of intentions, jokes and especially satire can easily get lost or taken seriously. If you feel the need for humor, you may wish to add "emoticons" to help alert your readers: ;-), :), J

Disclaimer Statement

Course content may vary from the outline to meet the needs of this particular group.

Academic Services

The Online Library is available to enrolled students and faculty from inside the electronic campus. This is your starting point for access to online books, subscription periodicals, and Web resources that are designed to support your classes and generally not available through search engines on the open Web. In addition, the Online Library provides access to special learning resources, which the University has contracted to assist with your studies. Questions can be directed to **librarian@apus.edu**.

- ***Charles Town Library and Inter Library Loan:*** The University maintains a special library with a limited number of supporting volumes, collection of our professors' publication, and services to search and borrow research books and articles from other libraries.
- ***Electronic Books:*** You can use the online library to uncover and download over 50,000 titles, which have been scanned and made available in electronic format.
- ***Electronic Journals:*** The University provides access to over 12,000 journals, which are available in electronic form and only through limited subscription services.
- ***Tutor*.com**: AMU and APU Civilian & Coast Guard students are eligible for 10 free hours of tutoring provided by APUS. Tutor.com connects you with a professional tutor online 24/7 to provide help with assignments, studying, test prep, resume writing, and more. Tutor.com is tutoring the way it was meant to be. You get expert tutoring whenever you need help, and you work one-to-one with your tutor in your online classroom on your specific problem until it is done.
- ***Disability Accommodations***: Students are encouraged email dsa@apus.edu to discuss potential academic accommodations and begin the review process.

Request a Library Guide for your course (http://apus.libguides.com/index.php)

The AMU/APU Library Guides provide access to collections of trusted sites on the Open Web and licensed resources on the Deep Web. The following are specially tailored for academic research at APUS:

- Program Portals contain topical and methodological resources to help launch general research in the degree program. To locate, search by department name, or navigate by school.
- Course Lib-Guides narrow the focus to relevant resources for the corresponding course. To locate, search by class code (e.g., SOCI111), or class name.

If a guide you need is not available yet, please email the APUS Library: librarian@apus.edu.

Turnitin.com

Faculty may require assignments be submitted to Turnitin.com. Turnitin.com will analyze a paper and report instances of potential plagiarism for the student to edit before submitting it for a grade. In some cases professors may require students to use Turnitin.com. This is automatically processed through the Assignments area of the course.

Selected Bibliography

Book References in Data Mining and Knowledge Discovery:

Usama Fayyad, Gregory Piatetsky-Shapiro, Padhraic Smyth, and Ramasamy Uthurasamy, "Advances in Knowledge Discovery and Data Mining", AAAI Press/ The MIT Press, 1996.

A. Ross Quinlan, "C4.5: Programs for Machine Learning", Morgan Kaufmann Publishers, 1993.

Michael Berry and Gordon Linoff, "Data Mining Techniques (For Marketing, Sales, and Customer Support), John Wiley & Sons, 1997.

Sholom M. Weiss and Nitin Indurkhya, "Predictive Data Mining: A Practical Guide", Morgan Kaufmann Publishers, 1998.

Alex Freitas and Simon Lavington, "Mining Very Large Databases with Parallel Processing", Kluwer Academic Publishers, 1998.

A. K. Jain and R. C. Dubes, "Algorithms for Clustering Data", Prentice Hall, 1988.

V. Cherkassky and F. Mulier, "Learning From Data", John Wiley & Sons, 1998.

General Data Mining: [10]

Usama Fayyad, "Mining Databases: Towards Algorithms for Knowledge Discovery", Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, vol. 21, no. 1, March 1998.

[http://www.research.microsoft.com/research/db/debull/98mar/issue.htm]

Christopher Matheus, Philip Chan, and Gregory Piatetsky-Shapiro, "Systems for Knowledge Discovery in Databases", IEEE Transactions on Knowledge and Data Engineering, 5(6):903-913, December 1993.

Rakesh Agrawal and Tomasz Imielinski, "Database Mining: A Performance Perspective", IEEE Transactions on Knowledge and Data Engineering, 5(6):914-925, December 1993.

Usama Fayyad, David Haussler, and Paul Stolorz, "Mining Scientific Data", Communications of the ACM, vol. 39, no. 11, pp. 51-57, November 1996.

David J. Hand, "Data Mining: Statistics and more?", The American Statistician, vol. 52, no. 2, pp 112-118, May 1998.

Tom M. Mitchell, "Does machine learning really work?", AI Magazine, vol. 18, no. 3, pp. 11-20, Fall 1997.

Clark Glymour, David Madigan, Daryl Pregibon, and Padhraic Smyth, "Statistical Inference and Data Mining", Communications of the ACM, vol. 39, no. 11, pp. 35-41, November 1996.

Hillol Kargupta, Ilker Hamzaoglu, and Brian Stafford, "Scalable, Distributed Data Mining using An Agent Based Architecture", Proceedings of the Third International Conference on Knowledge Discovery and Data

Mining (KDD-97), August 1997, Newport Beach, California, USA.

M-S. Chen, Jiawei Han, and Philip S. Yu, "Data Mining: An Overview from a Database Perspective", IEEE Transactions on Knowledge and Data Engineering, 8(6): 866-883, 1996. [http://db.cs.sfu.ca/sections/publication/kdd/kdd.html]

Surajit Chaudhuri, "Data Mining and Database Systems: Where is the Intersection?", Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, vol. 21, no. 1, March 1998.

[http://www.research.microsoft.com/research/db/debull/98mar/issue.htm]

Classification: [20]

M. Mehta, R. Agrawal and J. Rissanen, "SLIQ: A Fast Scalable Classifier for Data Mining", Proc. of the Fifth Int'l Conference on Extending Database Technology, Avignon, France, March 1996.

[http://www.almaden.ibm.com/cs/quest/publications.html]

J.C. Shafer, R. Agrawal, M. Mehta, "SPRINT: A Scalable Parallel Classifier for Data Mining", Proc. of the 22th Int'l Conference on Very Large Databases, Mumbai (Bombay), India, Sept. 1996.

[http://www.almaden.ibm.com/cs/quest/publications.html]

Bing Liu and Wynne Hsu, "Post Analysis of Learned Rules", Proceedings of the Thirteenth National Conference on Artificial Intelligence (AAAI-96), Aug 4-8, 1996, Portland, Oregon, USA, pp. 828-834.

[http://www.comp.nus.edu.sg/~whsu/biblio.html]

Bing Liu, Wynne Hsu and Shu Chen, "Discovering Conforming and Unexpected Classification Rules," IJCAI-97 Workshop on Intelligent Data Analysis in Medicine and Pharmacology (IDAMAP-

97), August 23-29, 1997, Nagoya, Japan

[http://www.comp.nus.edu.sg/~whsu/biblio.html]

Bing Liu, Wynne Hsu and Shu Chen, "Using General Impressions to Analyze Discovered Classification Rules", Proceedings of the Third International Conference on Knowledge Discovery and Data Mining (KDD-97), pp. 31-36, August 14-17, 1997, Newport Beach, California, USA.

[http://www.comp.nus.edu.sg/~whsu/biblio.html]

A. Wang and B. Liu, "Concurrent Discretization of Multiple Attributes", PRICAI 98, August 1998,

Singapore

[http://www.comp.nus.edu.sg/~wangk/publication.html]

Khaled Alsabti, Sanjay Ranka, and Vineet Singh, "CLOUDS: A Decision Tree Classifier for Large Datasets", Proc. of International Conference on Knowledge Discovery and Data Mining (KDD- 98), August 1998, New York City.

[http://www.cis.ufl.edu/~ranka]

Bing Liu, Wynne Hsu, Yiming Ma, "Integrating Classification and Association Rule Mining." Proc. of the Fourth International Conference on Knowledge Discovery and Data Mining (KDD-98), New York, USA, 1998.

[http://www.comp.nus.edu.sg/~whsu/biblio.html]

Mahesh V. Joshi, George Karypis and Vipin Kumar, "ScalParC: A New Scalable and Efficient Parallel Classification Algorithm for Mining Large Datasets", Proc. of 12th International Parallel Processing Symposium (IPPS/SPDP), April 1998, Orlando, USA.

[ftp://ftp.cs.umn.edu/dept/users/kumar/WEB/papers.html]

Anurag Srivastava, Eui-Hong (Sam) Han, Vipin Kumar, and Vineet Singh, "Parallel Formulations of Decision-Tree Classification Algorithms", Proc. of the 1998 International Conference on Parallel Processing.

[ftp://ftp.cs.umn.edu/dept/users/kumar/WEB/papers.html]

Frank E. and Witten I.H., "Generating Accurate Rule Sets Without Global Optimization", Proc International Conference on Machine Learning, 1998.

[http://www.cs.waikato.ac.nz/~ml/publications.html]

Frank E., Wang Y., Inglis S., Holmes G., and Witten I.H., "Using Model Trees for Classification", Machine Learning 32(1), pp 63-76, 1998.

[http://www.cs.waikato.ac.nz/~ml/publications.html]

Frank E. and Witten I.H., "Using a permutation test for attribute selection in decision trees",

Proc International Conference on Machine

[http://www.cs.waikato.ac.nz/~ml/publications.html]

M. J. Zaki, C-T. Ho, R. Agrawal, "Scalable Parallel Classification for Data Mining on Shared- Memory Multiprocessors", IEEE International Conference on Data Engineering, March 1999. [http://www.cs.rpi.edu/~zaki/papers.htm]

W. Cohen and H. Hirsh, "Joins thta Generalize: Text Classification Using WHIRL", Proc. International Conference on Knowledge Discovery and Data Mining (KDD-98), August 1998, New York City.

[http://www.cs.rutgers.edu/~hirsh]

W. Cohen, "Fast Effective Rule Induction", Proceedings of the Twelfth International Conference on Machine Learning, Lake Tahoe, California, 1995.

[http://portal.research.bell-labs.com/orgs/ssr/people/wcohen/pubs.html]

J. Rachlin, S. Kasif, S. Salzberg, and D. Aha, "Towards a Better Understanding of Memory-Based and Bayesian Classifiers", Proc. 1994 Internatl. Conf. on Machine Learning (pp. 242--250). New Brunswick, NJ,

[http://www.cs.jhu.edu/~salzberg]

S.K. Murthy, S. Kasif, and S. Salzberg, "A System for Induction of Oblique Decision Trees", Journal of Artificial Intelligence Research 2:1 (1994),

[http://www.cs.jhu.edu/~salzberg]

M. Mehta, J. Rissanen, and R. Agrawal, "MDL-based Decision Tree Pruning", Proc. of the 1st Int'l Conference on Knowledge Discovery in Databases and Data Mining, Montreal, Canada, August, 1995.

[http://www.almaden.ibm.com/cs/quest/publications.html]

Jan C. Bioch, Onno van der Meer, Rob Potharst, "ivariate Decision Trees", Principles of Data Mining and Knowledge Discovery, First European Symposium, PKDD'97, Komorowski and Zytkow (eds.), Springer LNAI 1263, pp. 232-242, Trondheim, Norway, 1997.

Rajeev Rastogi and Kyuseok Shim, "PUBLIC: A Decision Tree Classifier that Integrates Pruning and

Building", In Proceedings of the Twenty-fourth International Conference on Very Large Data Bases, New York, New York, 1998.

[http://www.bell-labs.com/project/serendip]

J. E. Gehrke, Raghu Ramakrishnan, and Venkatesh Ganti, "RAINFOREST - A Framework for Fast Decision Tree Construction of Large Datasets", In Proceedings of the Twenty-fourth International Conference on Very Large Data Bases, New York, New York, 1998.

[http://www.cs.wisc.edu/~johannes/publications.html]

N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian Network Classifiers", Machine Learning 29:131-163, 1997.

[http://http.cs.berkeley.edu/~nir]

D. Wettschereck, D.W. Aha, and T. Mohri "A Review and Empirical Evaluation of Feature Weighting Methods for a Class of Lazy Learning Algorithms", Special AI Review Issue on Lazy Learning, Volume 11, issues 1-5, February, 1997

[http://www.aic.nrl.navy.mil/~aha/ai-review/]

Association Rules: [24]

R. Agrawal, T. Imielinski, A. Swami, "Mining Associations between Sets of Items in Massive Databases", Proc. of the ACM SIGMOD Int'l Conference on Management of Data, Washington D.C., May 1993, 207-216.

[http://www.almaden.ibm.com/cs/quest/publications.html]

R. Agrawal, H. Mannila, R. Srikant, H. Toivonen and A. I. Verkamo, "Fast Discovery of Association Rules", Advances in Knowledge Discovery and Data Mining, Chapter 12, AAAI/MIT Press, 1995.

R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules", Proc. of the 20th Int'l Conference on Very Large Databases, Santiago,

[http://www.almaden.ibm.com/cs/quest/publications.html]

R. Srikant, Q. Vu, R. Agrawal, "Mining Association Rules with Item Constraints", Proc. of the 3rd Int'l Conference on Knowledge Discovery in Databases and Data Mining, Newport Beach, California, August 1997.

[http://www.almaden.ibm.com/cs/quest/publications.html]

R. Srikant, R. Agrawal: "Mining Generalized Association Rules", Proc. of the 21st Int'l Conference on Very Large Databases, Zurich, Switzerland,

[http://www.almaden.ibm.com/cs/quest/publications.html]

Jong Soo Park, Ming-Syan Chen, Philip S. Yu, "Using a Hash-Based Method with Transaction Trimming for Mining Association Rules", IEEE Transactions on Knowledge and Data Engineering, vol. 9, no. 5, pp. 813-825, Sept/Oct 1997.

Mohammed Zaki, Srinivasan Parthasarathy, Mitsunori Ogihara, Wei Li, "New Algorithms for Fast Discovery of Association Rules", 3rd International Conference on Knowledge Discovery and Data Mining (KDD'97), pp 283-286, Newport Beach, California, August, 1997.

[http://www.cs.rpi.edu/~zaki/papers.htm]

Mohammed Zaki, Mitsunori Ogihara, Srinivasan Parthasarathy, and Wei Li, "Parallel Data Mining for Association Rules on Shared-memory Multi- processors", Supercomputing'96, Pittsburg, PA, Nov 17-22, 1996.

[http://www.cs.rpi.edu/~zaki/papers.htm]

Mohammed Zaki, Srinivasan Parthasarathy, Wei Li, "A Localized Algorithm for Parallel Association Mining", 9th Annual ACM Symposium on Parallel Algorithms and Architectures

(SPAA), pp 321-330, Newport, Rhode Island, [http://www.cs.rpi.edu/~zaki/papers.htm]

A. Amir, R. Feldman, and R. Kashi, "A New and Versatile Method for Association Generation", Principles of Data Mining and Knowledge Discovery, First European Symposium, PKDD'97, Komorowski and Zytkow (eds.), Springer LNAI 1263, pp. 221-231, Trondheim, Norway, 1997.

Charu Aggarwal and Philip Yu, "Mining Large Itemsets for Association Rules", Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, vol. 21, no. 1, March 1998. [http://www.research.microsoft.com/research/db/debull/98mar/issue.htm]

R. Ng, L. V. S. Lakshmanan, J. Han and A. Pang, "Exploratory Mining and Pruning Optimizations of Constrained Associations Rules", Proc. of 1998 ACM-SIGMOD Conf. on Management of Data, Seattle, Washington, June 1998.

[http://db.cs.sfu.ca/sections/publication/kdd/kdd.html]

Sergey Brin, Rajeev Motwani, Craig Silverstein, "Beyond Market Baskets: Generalizing Association Rules to Correlations", Proceedings of 1997 ACM SIGMOD, Montreal, Canada, June 1997.

[http://www-db.stanford.edu/midas/midas.html]

Sergey Brin, Rajeev Motwani, Dick Tsur, Jeffrey Ullman, "Dynamic Itemset Counting and Implication Rules for Market Basket Data", Proceedings of 1997 ACM SIGMOD, Montreal, Canada, June 1997.

[http://www-db.stanford.edu/midas/midas.html]

Dick Tsur, Jeffrey Ullman, Chris Clifton, Serge Abiteboul, Rajeev Motwani, Svetlozar Nestorov, Arnie Rosenthal, "Query Flocks: a Generalization of Association-Rule Mining", Proceedings of 1998 ACM SIGMOD, Seattle,

[http://www-db.stanford.edu/midas/midas.html]

K. Yoda, T. Fukuda, Y. Morimoto, S. Morishita, and T. Tokuyama, "Computing Optimized Rectilinear Regions for Association Rules", Proceedings of the Third Conference on Knowledge Discovery and Data Mining (KDD'97), pages 96-103, Los Angels, August 1997

[http://platinum.ims.u-tokyo.ac.jp/~moris/paper-list.htm]

K. Wang, W. Tay, B. Liu, "Interestingness-based Interval Merger for Numeric Association Rules" , Proc. International Conference on Knowledge Discovery and Data Mining (KDD-98), August 1998, New York City.

[http://www.comp.nus.edu.sg/~wangk/publication.html]

K. Alsabti, S. Ranka and V. Singh. "A One-Pass Algorithm for Accurately Estimating Quantiles for Disk-Resident Data", In Proc. of VLDB'97

[http://www.cis.ufl.edu/~ranka]

R. Agrawal and J. C. Shafer, "Parallel Mining of Association Rules: Design, Implementation and Experience", IEEE Transactions on Knowledge Data and Engg., 8(6):962-969, December 1996. [http://www.almaden.ibm.com/cs/quest/publications.html]

Eui-Hong (Sam) Han, George Karypis and Vipin Kumar, "Scalable Parallel Data Mining for Association Rules", Proc. of 1997 ACM-SIGMOD International Conference on Management of Data, May 1997.

[ftp://ftp.cs.umn.edu/dept/users/kumar/WEB/papers.html]

A. Shintani and M. Kitsuregawa, "Hash Based Parallel Algorithm for Mining Association Rules", Proceedings of IEEE Fourth International Conference on Parallel and Distributed Information Systems, pp.19-30, 1996.

[http://www.tkl.iis.u-tokyo.ac.jp/Kilab/Research/Paper/paper1996.html]

H. Mannila and H. Toivonen, "Multiple uses of frequent sets and condensed representations", Proc. Second International Conference on Knowledge Discovery and Data Mining (KDD'96), 189- 194, Portland, Oregon, August 1996.

[http://www.cs.Helsinki.FI/research/pmdm/datamining/datamine.html]

H. Toivonen, M. Klemettinen, P. Ronkainen, K. Hdtvnen, and H. Mannila, "Pruning and grouping discovered association rules", In MLnet Workshop on Statistics, Machine Learning, and Discovery in Databases, pp. 47-52, Heraklion, Crete, Greece, April 1995.

[http://www.cs.Helsinki.FI/research/pmdm/datamining/datamine.html]

Jose Borges and Mark Levene, "Mining Association Rules in Hypertext Databases", Proc. International Conference on Knowledge Discovery and Data Mining (KDD-98), August 1998, New York City.

[http://www.bell-labs.com/project/serendip]

Mohammed J. Zaki and Mitsunori Ogihara, "Theoretical Foundations of Association Rules", 3rd SIGMOD'98 Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD), Seattle, WA, June 1998.

[http://www.cs.rpi.edu/~zaki/papers.htm]

Clustering: [17]

R. Ng and J. Han, "Efficient and Effective Clustering Method for Spatial Data Mining", Proc. of 1994 Int'l Conf. on Very Large Data Bases (VLDB'94), Santiago, Chile, September 1994, pp. 144- 155.

[http://db.cs.sfu.ca/sections/publication/smmdb/smmdb.html]

Sander J., Ester M., Kriegel H.-P., Xu X., "Density-Based Clustering in Spatial Databases: The Algorithm GDBSCAN and its Applications", Data Mining and Knowledge Discovery, An International Journal, Kluwer Academic Publishers, Vol. 2, No. 2, 1998, pp. 169-194.

[http://www.dbs.informatik.uni-muenchen.de/index_e.html]

Tian Zhang, Raghu Ramakrishnan, Miron Livny, "BIRCH: An Efficient Data Clustering Method for Very Large Databases", Proc. of ACM SIGMOD Int'l Conf. on Data Management, June 1996, Canada.

[http://www.cs.wisc.edu/~zhang/zhang.html]

Peter Cheeseman and John Stutz, "Bayesian Classification (AutoClass): Theory and Results", in

A. M. Fayyad, G. Piatetsky-Shapiro, P. Smith, and R. Uthurusamy (eds.), "Advances in Knowledge Discovery and Data Mining", pp 153-180, AAAI/MIT Press, 1996.

Rakesh Agrawal, Johannes Gehrke, Dimitrios Gunopulos, Prabhakar Raghavan, "Automatic Subspace Clustering of High Dimensional Data for Data Mining Applications", Proc. of the ACM SIGMOD Int'l

Conference on Management of Data, Seattle, Washington, June 1998.

[http://www.almaden.ibm.com/cs/quest/publications.html]

Ester M., Kriegel H.-P., Sander J., Wimmer M., Xu X., "Incremental Clustering for Mining in a Data Warehousing Environment, Proc. 24th Int. Conf. on Very Large Data Bases, New York, 1998, 323-333.

[http://www.dbs.informatik.uni-muenchen.de/index_e.html]

Eui-Hong (Sam) Han, George Karypis, Vipin Kumar and Bamshad Mobasher, "Clustering In A High-Dimensional Space Using Hypergraph Models", Technical Report 97-063, Department of Computer Science, University of Minnesota, 1997.

[ftp://ftp.cs.umn.edu/dept/users/kumar/WEB/papers.html]

P.S.Bradley, U. Fayyad, C. Reina, "Scaling Clustering Algorithms to Large Datasets", Proc. International Conference on Knowledge Discovery and Data Mining (KDD-98), August 1998, New York City.

A. Fayyad, C. Reina, and P.S.Bradley, "Initialization of Iterative Refinement Clustering Algorithms", Proc. International Conference on Knowledge Discovery and Data Mining (KDD- 98), August 1998, New York City.

O. Zamir, O. Etzioni, O. Madani, and R. M. Karp, "Fast and Intuitive Clustering of Web Documents", Proc. of the Third International Conference on Knowledge Discovery and Data Mining (KDD-97), August 1997, Newport Beach, California, USA.

G. D. Ramkumar and A. Swami, "Clustering Data Without Distrance Functions", Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, vol. 21, no. 1, March 1998. [http://www.research.microsoft.com/research/db/debull/98mar/issue.htm]

Sudipto Guha, Rajeev Rastogi, Kyuseok Shim, "CURE: An Efficient Clustering Algorithm for Large Databases", ACM SIGMOD Conference 1998, pp. 73-84

[http://www.bell-labs.com/project/serendip]

Sudipto Guha, Rajeev Rastogi, Kyuseok Shim, "ROCK: A Robust Clustering Algorithm for Categorical Attributes", Proceedings of IEEE Conference on Data Engineering, Australia, 1999. [http://www.bell-labs.com/project/serendip]

Venkatesh Ganti, Raghu Ramakrishnan, and Johannes Gehrke, "Clustering Large Datasets in Arbitrary Metric Spaces", Proceedings of IEEE Conference on Data Engineerings, Austratlia, 1999.

[http://www.cs.wisc.edu/~raghu/raghu.html]

F. Chen, S. Figlewski, J. Heisler, and A. S. Weigend, "Uncovering Hidden Structure in Bond Futures Trading", Proc. of IEEE?IAFE/INFORMS Conference on Computational Intelligence for Financial Engineering (CIFEr'98), New York, March 1998.

[http://www.stern.nyu.edu/~aweigend/Research/Research.html]

Padhraic Smyth, "Clustering Sequences with Hidden Markov Models", in "Advances in Neural Information Processing 9", M. C. Mozer, M. I. Jordan, and T. Petsche (eds.), MIT Press, 1997. [http://www.ics.uci.edu/~mlearn/MLPapers.html]

A. Ketterlin, "Clustering Sequences of Complex Objects", Proceedings of the Third International Conference on Knowledge Discovery and Data Mining (KDD-97), August 1997, Newport Beach, California, USA.

[http://dpt-info.u-strasbg.fr/~alain]

Mining Sequential (or Temporal) Data: [25]

Philip Laird, "Identifying and Using Patterns in Sequential Data", in Algorithmic Learning Theory (ALT'93), Fourth International Workshop, pp. 1-18, Springer-Verlag.

J. T.-L. Wang, G.-W. Chirn, T. G. Marr, B. Shapiro, D. Shasha, and K. Zhang, "Combinatorial Pattern Discovery for Scientific Data: Some Preliminary Results", Proc. of ACM SIGMOD Conf. on Management of Data (SIGMOD'94), pp. 115-125, June 1994.

[http://www.cs.nyu.edu/cs/faculty/shasha/papers/papers.html]

H. Mannila, H. Toivonen, and A. Inkeri Verkamo, "Discovery of Frequent Episodes in Event Sequences", Technical Report C-1997-15, Department of Computer Science, University of Helsinki, Finland, 1997.

[http://www.cs.Helsinki.FI/research/pmdm/datamining/datamine.html]

R. Agrawal, R. Srikant, "Mining Sequential Patterns", Proc. of the Int'l Conference on Data Engineering (ICDE), Taipei, Taiwan, March 1995.

[http://www.almaden.ibm.com/cs/quest/publications.html]

R. Srikant and R. Agrawal, "Mining Sequential Patterns: Generalizations and Improvements", Proc. of the Fifth Int'l Conference on Extending Database Technology (EDBT), Avignon, France,

March 1996.

[http://www.almaden.ibm.com/cs/quest/publications.html]

C. Bettini, X. Sean Wang, S. Jajodia, "Testing Complex Temporal Relationships Involving Multiple Granularities and Its Application to Data Mining", Proc. of 15th ACM PODS Symp., Montreal, Canada, June 1996, pages 68-78.

[http://www.isse.gmu.edu/~jlin/tdb]

G. Berger and A. Tuzhilin, "Discovering Unexpected Patterns in Temporal Data Using Temporal Logic", in "Temporal Databases: Research and Practice", O. Etzion, S. Jajodia, and S. Sripada (eds.), Springer-Verlag, 1998.

B. Padmanabhan and A. Tuzhilin, "A Belief-Driven Method for Discovering Unexpected Patterns", In Proceedings of the 4rd International Conference on Knowledge Discovery and Data Mining (KDD-98), August 1998.

T. Shintani and M. Kitsuregawa, "Mining Algorithms for Sequential Patterns in Parallel: Hash Based Approach", in Research and Development in Knowledge Discovery and Data Mining: Second Pacific-Asia Conference (PAKDD'98), pp. 283-294, Melbourne, Australia, April 1998. [http://www.tkl.iis.u-tokyo.ac.jp/Kilab/Research/Paper/paper1998.html]

M. Joshi, G. Karypis, and V. Kumar, "Parallel Algorithms for Mining Sequential Associations", Presented at the Minisymposium on High Performance Data Mining at Ninth SIAM Conference on Parallel Processing for Scientific Computing, San Antonio, Texas, March 1999.

Mohammed Zaki, "Efficient Enumeration of Frequent Sequences", 7th International Conference on Information and Knowledge Management, Washington DC, November 1998.

[http://www.cs.rpi.edu/~zaki/papers.htm]

Ke Wang, "Discovering Patterns from Large and Dynamic Sequential Data", Special Issues on Data Mining and Knowledge Discovery, Journal of Intelligent Information Systems, 9(1), 8-33, 1997, Kluwer Academic Publishers.

[http://www.comp.nus.edu.sg/~wangk/publication.html]

I. Guralnik, D. Wijesekera, and J. Srivastava, "Pattern Directed Mining for Frequent Episodes", Proc. of the Fourth International Conference on Knowledge Discovery and Data Mining (KDD- 98), New York, USA, 1998.

K. Hatonen, M. Klemettinen, H. Mannila, P. Ronkainen, and H. Toivonen, "TASA: Telecommunications Alarm Sequence Analyzer, or How to enjoy faults in your network", In IEEE/IFIP 1996 Network Operations and Management Symposium (NOMS'96), 520-529, Kyoto, Japan, April 1996.

[http://www.cs.Helsinki.FI/research/pmdm/datamining/datamine.html]

E. Arjas, H. Mannila, M. Salmenkivi, R. Suramo, and H. Toivonen, "BASS: Bayesian analyzer of event sequences", In Proceedings in Computational Statistics (COMPSTAT'96) 199-204, Barcelona, Spain, August 1996, Physica-Verlag.

[http://www.cs.Helsinki.FI/research/pmdm/datamining/datamine.html]

Gary M. Weiss and Haym Hirsh, "Learning to Predict Rare Events in Event Sequences", Proc. of the Fourth International Conference on Knowledge Discovery and Data Mining (KDD-98), New York, USA, 1998.

B. Lent, R. Agrawal, R. Srikant, "Discovering Trends in Text Databases", Proc. of the 3rd Int'l Conference on Knowledge Discovery in Databases and Data Mining (KDD'97), Newport Beach, California, August 1997.

[http://www.almaden.ibm.com/cs/quest/publications.html]

J. Han, W. Gong, and Y. Yin, "Mining Segment-Wise Periodic Patterns in Time-Related Databases", Proc. of 1998 Int'l Conf. on Knowledge Discovery and Data Mining (KDD'98) , New York City, NY, Aug. 1998.

[http://db.cs.sfu.ca/sections/publication/kdd/kdd.html]

T. H. Hinke, J. Rushing, H. Ranganath, and S. J. Graves, "Target-Independent Mining For Scientific Data: Capturing Transients and Trends for Phenomena Mining", Proc. of the 3rd Int'l Conference on Knowledge Discovery in Databases and Data Mining (KDD'97), Newport Beach, California, August 1997.

Honghua Dai, "Trend Directed Learning: A Case Study", in Research and Development in Knowledge Discovery and Data Mining: Second Pacific-Asia Conference (PAKDD'98), pp. 283- 294, Melbourne, Australia, April 1998.

K.-R. Muller, A. J. Smola, G. Ratsch, B. Scholkopf, J. Kohlmorgen, and V. Vapnik, "Predicting Time Series with Support Vector Machines", Proceedings ICANN'97, Springer Lecture Notes in Computer Science, 1997.

[http://svm.first.gmd.de/publications.html]

R. Bharat Rao, S. Rickard, and F. Coetzee, "Time Series Forecasting from High-Dimensional Data with Multiple Adaptive Layers", Proc. of 1998 Int'l Conf. on Knowledge Discovery and Data Mining (KDD'98) , New York City, NY, Aug. 1998.

E. J. Keogh and M. J. Pazzani, "An enhanced representation of time series which allows fast and accurate classification, clustering, and relevance feedback", AAAI Workshop on Predicting the Future: AI Approaches to Time Series Analysis. Madison, Wisc., 1998.

[http://www.ics.uci.edu/~mlearn/MLPapers.html]

S.-P. Shieh and V. D. Gligor, "On a Pattern-Oriented Model for Intrusion Detection", IEEE Trans. on Knowledge and Data Engineering, vol. 9, no. 4, July/August 1997, pp. 661-667.

S. Chakrabarti, S. Sarawagi, and B. Dom, "Mining surprising patterns using temporal description length, Proc. of the 24th Int'l Conference on Very Large Databases (VLDB), 1998.

[http://www.almaden.ibm.com/cs/quest/publications.html]

University Policies

Student Handbook

- Drop/Withdrawal policy
- Extension Requests
- Academic Probation
- Appeals
- Disability Accommodations

The mission of American Public University System is to provide high quality higher education with emphasis on educating the nation's military and public service communities by offering respected, relevant, accessible, affordable, and student-focused online programs that prepare students for service and leadership in a diverse, global society.