Eui-Hong (Sam) Han

Contact Information

Address:	6 Whitman Terrace, Hawthorn Woods, IL 60047
Phone:	(847) 307-5929, Cell (224) 678-5691
Electronic:	han@cs.umn.edu, http://www.cs.umn.edu/∼han

Research Areas

Data Mining Information Retrieval Parallel Algorithms Artificial Intelligence

Education

1999	Ph.D.,	Computer Science,	University of Minnesota.
1990	M.S.,	Computer Science,	University of Texas, Austin.
1987	B.S.,	Computer Science,	University of Iowa.

Professional Experience

Sears Holdings Corporation	Director, Machine Learning Group	Mar. 2010 – current date
nXn Tech. LLC	Senior Scientist	Nov. 2008 – Feb. 2010
iXmatch Inc.	Senior Scientist	Mar. 2001 – Oct. 2008
University of Minnesota	Research Associate	Nov. 1999 – Jan. 2003
University of Minnesota	Research Fellow	Jan. 1999 – Oct. 1999
SGI/Cray Research Inc.	Summer Intern	June – Sept., 1996,1997,1998
IBM, Rochester, MN	Associate Programmer	Nov. 1993 – July 1995
CogniSeis Development, Houston	UNIX Systems Engineer	July 1991 – Oct. 1993

Professional and Scholarly Associations

Member, Association for Computing Machinery
Member, Phi Beta Kappa

Research Grants

P.I., Leadership Agent for Multi-source Information Fusion in Counter-Terrorism,
SBIR Phase I, the Office of the Secretary of Defense (OSD), 2/2004 - 6/2004.

P.I., Concept-based Information Retrieval and Fusion Engines,
SBIR Phase II, USAF/AFMC, 1/2007 - 6/2008.

Professional Activities

Program Committee, ACM Int. Conf. on Research and Development in Information Retrieval (SIGIR 2010), July 2010.
Program Committee, ACM Int. Conf. on Research and Development in Information Retrieval (SIGIR 2009), July 2009.
Program Committee, Int. Conf. on Data Warehousing and Knowledge Discovery (DaWaK 2008), September 2008.
Program Committee, IEEE International Conference on Data Mining (ICDM'07), October 2007.
Program Committee, Int. Conf. on Data Warehousing and Knowledge Discovery (DaWaK 2007), September 2007.
Program Committee, IEEE International Conference on Data Mining (ICDM'06), December 2006.
Program Committee, ECML/PKDD 2006, September 2006.
Program Committee, Int. Conf. on Data Warehousing and Knowledge Discovery (DaWaK 2006), September 2006.
Organizing Committee, PAKDD workshop on Knowledge Discovery in Life Science Literature, April 2006.
Program Committee, IEEE International Conference on Data Mining (ICDM'05), November 2005.
Organizing Committee, PAKDD workshop on Text Mining, May 2002.
Tutorial Chair, The First SIAM International Conference on Data Mining, April, 2001

Journal Publications

1. *Scalable Parallel Data Mining for Association Rules*, E.H. Han, G. Karypis and V. Kumar. IEEE Transactions on Knowledge and Data Engineering, Vol. 12, No. 3, May/June 2000.
2. *Chameleon: A Hierarchical Clustering Algorithm Using Dynamic Modeling*, G. Karypis, E.H. Han, and V. Kumar. IEEE Computer, Vol 32, No 8, pp 68-75, August 1999.
3. *Document Categorization and Query Generation on the World Wide Web Using WebACE*, D. Boley, M. Gini, R. Gross, E.H. Han, K. Hastings, G. Karypis, V. Kumar, B. Mobasher, J. Moore. AI Review, Vol. 13, No. 5-6, 1999.
4. *Parallel Formulations of Decision-Tree Classification Algorithms*, A. Srivastava, E.H. Han, V. Kumar, and V. Singh. Data Mining and Knowledge Discovery: An International Journal, Vol 3, No 3, pp 237-261, 1999.
5. *Partitioning-Based Clustering for Web Document Categorization*, D. Boley, M. Gini, R. Gross, E.H. Han, K. Hastings, G. Karypis, V. Kumar, B. Mobasher, J. Moore. Decision Support Systems Journal, Vol 27, No. 3, pp 329-341, 1999.
6. *Hypergraph Based Clustering in High-Dimensional Data Sets: A Summary of Results*, E.H. Han, G. Karypis, V. Kumar and B. Mobasher. Bulletin of the Technical Committee on Data Engineering, Vol. 21, No. 1, March 1998

Book Chapters

1. *Parallel Algorithms for Data Mining*, M. Joshi, E.H. Han, G. Karypis and V. Kumar. CRPC Parallel Computing Handbook, Morgan Kaufmann, 2000.
2. *Parallel Algorithm Scalability Issues in Petaflops Architectures*, A. Grama, A. Gupta, E.H. Han, and V. Kumar. Ultrascale Computing, 2000.
3. *Efficient Parallel Algorithms for Mining Associations*, M. Joshi, E.H. Han, G. Karypis and V. Kumar. Large-scale Parallel and Distributed Data Mining, Lecture Notes in Computer Science/Lecture Notes in Artificial Intelligence (LNCS/LNAI), vol. 1759, Springer-Verlag, 2000.

Conference/Workshop Publications

1. *Content-Based Methods for Predicting Web-Site Demographic Attributes*, S. Kabbur, E.H. Han, G. Karypis. Tenth IEEE International Conference on Data Mining, 2010.
2. *Feature-Based Recommendation System*, E.H. Han, G. Karypis. Fourteenth International Conference on Information and Knowledge Management, 2005.
3. *Intelligent Metasearch Engine for Knowledge Management*, E.H. Han, G. Karypis, D. Mewhort, and K. Hatchard. Twelfth International Conference on Information and Knowledge Management, 2003.
4. *High Performance Data Mining*, V. Kumar, M.V. Joshi, E.H. Han, P.N. Tan, M. Steinbach. VECPAR 2002, 5th International Conference, 2002.
5. *Text Categorization Using Weight Adjusted k-Nearest Neighbor Classification*, E.H. Han, G. Karypis, and V. Kumar. PAKDD 5th Pacific-Asia Conference, 2001.
6. *Personalized Profile Based Search Interface with Ranked and Clustered Display*, S. Kumar, B. Uygar Oztekin, L. Ertoez, S. Singhal, E-H. Han and V. Kumar. International Conference on Intelligent Agents Web Technologies and Internet Commerce, 2001.
7. *Centroid-Based Document Classification: Analysis & Experimental Results*, E.H. Han and G. Karypis. The Fourth European Conference on Principles and Practice of Knowledge Discovery in Databases, 2000.
8. *Fast Supervised Dimensionality Reduction Algorithm with Applications to Document Categorization & Retrieval*, G. Karypis and E.H. Han. Ninth International Conference on Information and Knowledge Management, 2000.
9. *Efficient Parallel Algorithms for Mining Associations*, M.V. Joshi, E.H. Han, G. Karypis and V. Kumar. Workshop on Large-Scale Parallel KDD Systems, SIGKDD, 1999.
10. *Automated Morphological Classification of Galaxies and the Morphology-Density Relation*, R. Kriessler, E.H. Han, S.C. Odewahn, and T.C. Beers. Abstract in the 193rd Meeting of the American Astronomical Society, 1999.
11. *WebACE: A Web Agent for Document Categorization and Exploration*, E.H. Han, D. Boley, M. Gini, R. Gross, K. Hastings, G. Karypis, V. Kumar, B. Mobasher, J. Moore. 2nd International Conference on Autonomous Agents (Agents'98), 1998.
12. *Parallel Formulations of Decision-Tree Classification Algorithms*, A. Srivastava, E.H. Han, V. Kumar, and V. Singh. International Conference on Parallel Processing , 1998.
13. *Dynamic Load Balancing of Unstructured Computations in Decision Tree Classifiers*, A. Srivastava, E.H. Han, V. Kumar and V. Singh. IPPS'98 Workshop on High Performance Data Mining, 1998
14. *Scalable Parallel Data Mining for Association Rules*, E.H. Han, G. Karypis and V. Kumar. ACM SIGMOD Conference on Management of Data, 1997.
15. *Clustering Based On Association Rule Hypergraphs*, E.H. Han, G. Karypis, V. Kumar and B. Mobasher. Workshop on Research Issues on Data Mining and Knowledge Discovery, 1997.
16. *Web Page Categorization and Feature Selection Using Association Rule and Principal Component Clustering*, J. Moore, E.H. Han, D. Boley, M. Gini, R. Gross, K. Hastings, G. Karypis, V. Kumar, and B. Mobasher. Workshop on Information Technologies and Systems, 1997.

Projects

- **Projects at Sears Online Business Unit**

  Led a team of researchers and engineers for the following projects:

  - *Recommendation*
    - Scalable Product Recommendation
    - Personalized Product Recommendation
  - *Data Quality*
    - Product Category Classification
    - Product Image Classification
    - Customer Review and User Generated Content Classification
    - Sentiment Analysis of Customer Reviews and Discussions
  - *Search/Browse*
    - Auto-complete with keywords and product images
    - Search Evaluation Platform
    - Browse Ranking Optimization
  - *Web Intelligence*
    - Analysis of competitors' and vendors' product offerings
    - Emerging Trend Discovery

- **R&D Projects at nXn Tech**

  - *Patent Tools*
    Developed algorithms and tools for Intellectual Property research and analysis (http://www.patenttools.com).
  - *Content-Based Demographics Predictions for Websites*
    Developed algorithms for predicting websites' audience demographics based on contents of web pages without violating user privacy.

- **R&D Projects at iXmatch**

  - *Advanced Matching/Search Engine*
    Developed iXfind engine APIs that seamlessly integrates several cutting-edge text mining and information-retrieval techniques. Features of iXfind include unified query for both unstructured text and field data, patented scoring schemes, bidirectional matching, fusion, dynamic clustering, categorization, and dynamic differentiations.
  - *Feature-Based Product Recommendations*
    Developed feature-based recommendation algorithms that provide recommendations in the domains where no sufficient historical data exist for measuring similarity between products or users.
  - *Advanced Information A.I.D.E. System*
    Developed Advanced Information A.I.D.E. System (Acquisition, Interrogation/Inquiry, Discovery and Exploration) for DoD/Air Force SBIR project utilizing Concept Indexing, Query Expansion/Augmentation, and Pattern Discovery.

- *Predictive Analytics for Customer Relationship Management*
  Developed predictive modeling algorithms to identify subset of customers and prospects for sales, marketing, and customer retention related actions. These algorithms utilize potentially rich and accurate data for existing customers, and common demographic data (available from Axiom, etc) for both existing customers and prospects.
- *Sentiment Analysis*
  Developed sentiment analysis techniques for identifying customer moods on particular brands, products, and services from text documents.
- *Transportation Optimization*
  Developed trailer load and route optimization techniques.

- **Data Mining Projects at University of Minnesota**
  - *Personalized Web Search Interface Using Categorization and Clustering*
    Developed web search interface (Scout) that filters/reranks search results based on the user profiles and that also clusters/indexes search results.
  - *Text Categorization*
    Developed text categorization algorithms based on document centroids and weight adjustment.
  - *Clustering In A High-Dimensional Space Using Hypergraph Models*
    Developed a clustering algorithm for finding clusters of data in a high-dimensional space by transforming the original data into a hypergraph and partitioning the hypergraph.
  - *WebACE: A Web Agent for Document Categorization and Exploration*
    Developed WebACE system for exploring and categorizing documents on the World Wide Web based on a user profile.
  - *Parallel Formulations of Decision-Tree Classification Algorithms*
    Implemented scalable parallel algorithms for finding classification decision trees.
  - *Scalable Parallel Data Mining for Association Rules*
    Implemented scalable parallel algorithms for finding association rules for large transaction data on Cray T3D and IBM SP2 using MPI.

- **Summer Intern Projects at Cray Research**
  - *Automated Morphological Classification of Galaxies and the Morphology-Density Relation*
    Analyzed the feasibility and effectiveness of various data mining tools on the morphological classification of galaxies based on the image data obtained from the APS scans of the POSS I.
  - *Scalability analysis of parallel classification algorithms using pthread*
    Analyzed classification algorithms in SGI's MLC++ library that is written in C++, and proposed and evaluated several parallel algorithms in C and C++ using pthread on SGI Origin 2000.
  - *Scalable Parallel Data Mining for Association Rules*