

Arun Qamra

Department of Computer Science Phone: 1.805.252.0410
Harold Frank Hall arun@cs.ucsb.edu
University of California http://www.cs.ucsb.edu/~arun/
Santa Barbara, CA 93106, USA

Education

Ph.D. in Computer Science, Expected Summer 2007
University of California Santa Barbara, CA, USA. (GPA - 3.83)
Advisor: Prof. Edward Y. Chang
Thesis: Techniques for Scalability in Multimedia Retrieval

M.S. in Computer Science, June 2003
University of California Santa Barbara, CA, USA. (GPA - 4.00)

B.E. in Chemical Engineering, June 1999,
Panjab University, India

Research Interests

Data mining, applied machine learning, and retrieval for large datasets

Professional Experience

Research Intern, Ricoh Innovations, Menlo Park, Jun 06-Sep 06
Supervisor: Michael Gormish. Developed novel techniques for document image retrieval and mangement.

Research Intern, NEC Labs America, Cupertino, Jun 05-Sep 05
Supervisor: Belle Tseng. Researched and implemented system for blog mining.

Graduate Student Researcher, UC Santa Barbara, Oct 03-Present
Supervisor: Prof. Edward Y. Chang. Conducted research in the areas of machine learning, data mining, and scalable retrieval.

Engineering Intern, Eternal Systems, Santa Barbara, Jun 02-Sep 02
Worked on reliable ordered group communication protocols at startup building solutions for fault-tolerant distributed systems.

Programmer, Dept of Economics, UC Santa Barbara, Jan 02-Sep 03
Supervisor: Prof. Gary Charness. Developed software for experimental economics and game theory research.

Software Engineer, Infosys Technologies, India, Jul 99-Aug 01
Worked on development, maintenance of a supply chain management product.

Selected Research Projects

Robust Similarity Measurement: A pair of objects may be considered semantically similar even if only a subset of their features match closely, but traditional distance measures fail to capture this. In this project, a distance measure that can identify such *partial matches* was developed. Unlike existing partial-match methods, the proposed measure is metric, does not require training, allows scalable retrieval, and can be used for clustering and classification.

Embedding for Complex Distance Measures: In various applications, distance measures are often computationally expensive, and may even be non-metric. In this project, methods were developed to embed objects into a Eu-

clidean space, allowing approximate distances to be quickly computed, and large datasets to be efficiently indexed. A key contribution was a method to create high-quality embeddings at low embedding cost.

Indexing for Support Vector Machine Queries: Support Vector Machines (SVMs) are often used to learn query concepts. Objects that match this concept (as defined by the trained SVM) can then be retrieved from a database. The traditional approach to answer such a query is to retrieve, and evaluate against the SVM, each database object. The goal of this project was to reduce the number of objects that need to be retrieved in this scenario.

Near-Replica Detection for Images and Text: Near-replica detection is an important problem with applications such as detection of copyright violations for images, and duplicate elimination for text. In this project, we developed a scalable system for detecting near-replicas with high accuracy.

Labeling Personal Photographs Using Content and Context: With the increasing use of personal digital cameras, there is a need for automating organization of personal photograph collections. This project investigated using content analysis along with contextual information, such as time-stamp, location, and social networks, to assign semantic labels, and to detect and recognize faces and landmarks in the images. A prototype system is being built (at fotofiti.com) to provide such a service.

Bayesian Model to Mine Communities and Stories from Blogs: Blogs are an important new medium of online content, with distinct characteristics such as their personal nature, associated time-stamps, and the formation of online communities. A Bayesian model was developed to mine coherent stories seen in communities in the blogosphere, based on content, entry timestamps, and blogger community structure.

Publications

A. Qamra and E. Y. Chang, **Scalable Indexing for Perceptual Data**, *International Workshop on Multimedia Content Analysis and Mining*, WeiHai, China, June 2007.

A. Qamra and E. Y. Chang, **RCMap: Efficiently Creating High-Quality Euclidean Embeddings**, *SIAM International Conference on Data Mining*, Minneapolis, MN, April 2007.

A. Qamra and E. Y. Chang, **Scalable Landmark Recognition Using EX-TENT**, *Multimedia Tools and Applications Journal*, 2007.

A. Qamra, B. Tseng, and E. Y. Chang, **Mining Blog Stories Using Community based and Temporal Clustering**, *ACM Conference on Information and Knowledge Management*, Arlington, VA, November 2006.

N. Panda, E. Y. Chang, and A. Qamra, **Hypersphere Indexer**, *International Conference on Database and Expert Systems Applications*, Krakow, Poland, September 2006.

C. Tsai, A. Qamra, and E. Y. Chang, **Inferring Image Metadata from Context and Content**, *IEEE International Conference on Multimedia*, Amsterdam, Netherlands July 2005.

A. Qamra and E. Y. Chang, **Using Pivots to Index for Support Vector Machine Queries**, *International Workshop on Computer Vision meets Databases*, in cooperation with ACM SIGMOD, Baltimore MD, June 2005.

A. Qamra, Y. Meng, and E. Y. Chang, **Enhanced Perceptual Distance Functions and Indexing for Image Replica Recognition**, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume 27, Number 3, March 2005.

A. Qamra and E. Y. Chang, **Using Feature Patterns to Assist Automatic Image Categorization**, *IEEE International Conference on Multimedia*, Taipei, Taiwan, June 2004.

A. Qamra, K. Goh, and E. Y. Chang, **Using MEGA to Predict Molecular Bio-Activity** (poster), *ACM International Conference on Research in Computational Molecular Biology*, San Diego CA, March 2004.

Manuscripts:

A. Qamra and E. Y. Chang, **Partial-Match Based Similarity Search**, March 2007 (submitted).

A. Qamra, Y. Wu, and E. Y. Chang, **Embedding Non-Metric Perceptual Distances in a Metric Space**, March 2004.

Graduate Coursework

Data Mining, Advanced Machine Intelligence, Multimedia Systems, Game Theory, Computational Geometry, Advanced Algorithms and Applications, Advanced Networking, Distributed Systems, Modern Programming Languages, Network Security, Grid Computing, Media Networks, Mixed Reality Systems

Awards

UCSB Mericos Scholarship, 2003, 2005
President's Work Study Award, 2003-2006

Programming Skills

C, Java, Matlab, Unix shell scripting

Other Activities

Reviewer for TKDE 2005, 2006, MMSJ 2007
Volunteer, ACM Multimedia Conference, Santa Barbara, Oct 2006
Radio Jockey, KCSB 91.9 FM, Santa Barbara