

HYUK CHO

3370-B Lake Austin Blvd.

Austin, Texas 78703-5518

H : (512) 708-0661 O : (512) 232-7421

hyukcho@cs.utexas.edu

<http://www.cs.utexas.edu/users/hyukcho/>

OBJECTIVE

To get AN INTERNSHIP or A FULL-TIME JOB POSITION on challenging and creative Data Mining Research topics. Especially for (1) text data mining tasks such as Query Retrieval, Classification, Clustering, and Co-clustering and (2) bioinformatics tasks on gene expression matrices using clustering/Co-clustering algorithms.

EDUCATION

<i>August 1999 ~ Current</i>	Ph.D Program Computer Science The University of Texas, Austin, Texas
	Advisor Dr. Inderjit S. Dhillon
<i>March 1996 ~ February 1998</i>	M.S. Computer Science Korea University, Seoul, Korea(ROK)
	Thesis Synthesis of Neural Associative Memories Using Qualitative Analysis and LMI Techniques
	GPA 4.44/4.5
<i>March 1989 ~ August 1994</i>	B.E. Computer Engineering Chonbuk National University, Chonju, Korea(ROK)
	Completed with seven semesters, ranked 1st
	GPA 4.10/4.5

PROJECT EXPERIENCE

<i>May 2005 ~ Current</i>	SPHERICAL CO-CLUSTERING ALGORITHM: THEORY AND APPLICATION (Dr. Inderjit S. Dhillon)
	<ul style="list-style-type: none">• Implement a new co-clustering algorithm that utilizes a special normalization and an objective function.• Use both text datasets and bioinformatics datasets.
<i>January 2005 ~ Current</i>	CENTERING AND SCALING OF MICROARRAYS: A COMPARATIVE STUDY FOR CLUSTERING ALGORITHMS (Dr. Inderjit S. Dhillon)
	<ul style="list-style-type: none">• Update three co-clustering algorithms (Information-theoretic, Euclidean, and Residue) in C++.• Use Human cancer microarrays: COLON, LEUKEMIA, and MLL data.
<i>January 2004 ~ December 2004</i>	CO-CLUSTERING OF HUMAN CANCER MICROARRAYS USING MINIMUM SQUARED RESIDUE CO-CLUSTERING ALGORITHM (Dr. Inderjit S. Dhillon)
	<ul style="list-style-type: none">• Implement three co-clustering algorithms (Information-theoretic, Euclidean, and Residue) in C++.• Use Human cancer microarrays: COLON, LEUKEMIA, LUNG, and MLL data.
<i>June 2003 ~ December 2003</i>	MINIMUM SQUARED RESIDUE CO-CLUSTERING OF GENE EXPRESSION DATA (Dr. Inderjit S. Dhillon, Yuqiang Guan, and Suvrit Sra)
	<ul style="list-style-type: none">• Propose two new co-clustering algorithms (Euclidean and Residue), having spectral initialization step, batch update step. and local search step.• Use YEAST and HUMAN LYMPHOMA data.

- January 2003 ~ May 2003* **FEATURE CLUSTERING ON CLUSTERING AND CLASSIFICATION OF .GOV TRECWeb DATA**
EE380L - Practicum in Data Mining (Dr. Joydeep Ghosh)
- Evaluate how feature clustering affects on both clustering and classification of huge text data set.
 - Apply different distance (or similarity) measures of document clustering algorithms and compare feature selection vs. feature clustering.
 - Use **.GOV TRECWeb** data and some other artificial data for experiments.
- September 2002 ~ May 2003* **MDL-Based FORMULATION OF DISTRIBUTIONAL CLUSTERING**
(Dr. Inderjit S. Dhillon, Subramanyam Mallela, and Dr. Byron Dom)
- Apply **MDL**(Minimum Description Length) formulation in order to predict optimal number of feature clusters from Distributional Clustering for better classification of text documents.
 - Use **NEWS20** data and some other artificial data for experiments.
- May 2002 ~ August 2002* **SEMISUPERVISED LEARNING FOR CLASSIFICATION OF LARGE TEXT DATA USING FEATURE CLUSTERING**
CS395T - Conference Course (Dr. Inderjit S. Dhillon)
- Do feature clustering for better classification of text documents for which only small training data's labels are available, and make use of label information to enhance classification accuracy.
 - Apply different metrics (distance, similarity, and divergence).
 - Use **NEWS20** data for experiments.
- January 2002 ~ May 2002* **SCHEME INTERPRETER USING JAVA LANGUAGE**
CS386L - Programming Languages (Dr. Greg Lavender)
- August 2001 ~ December 2001* **CLASSIFICATION ALGORITHMS ON GENE EXPRESSION DATA**
CH391L - Bioinformatics (Dr. Edward M. Marcotte)
http://www.cs.utexas.edu/users/hyukcho/ch391l_project.html
- Compare the performance of supervised machine learning algorithms for classification based on gene expression data.
 - Evaluate the feasibility and performance of traditional classification algorithms.
 - Use **YEAST** Gene Expression data for experiments.
- January 2001 ~ December 2001* **COMPARISONS ON CLASSIFICATION ALGORITHMS**
TEXT MINING: CLUSTERING AND QUERYING
Research Assistant (Dr. Inderjit S. Dhillon)
http://www.cs.utexas.edu/users/hyukcho/classification_algorithms.html
- Implement in **MATLAB** well-known classification algorithms: Naive Bayesian(NB), K-Nearest Neighbor(KNN), Centroid-based(CB), and Support-Vector Machine(SVM).
 - Propose variations of algorithms and compare their classification performance in accuracy, precision and recall classification measure.
 - Use **CLASSIC3** and **NEWS20** for classification experiments.
 - Apply different normalizations and query expansions to Keyword Matching(KM) and Generalized VSM model for Query Retrieval.
 - Use **FBIS** and **LATIMES** (in TREC) for Query Retrieval experiments.
- August 2000 ~ December 2000* **EVALUATION OF ALGORITHMS ON DOCUMENT RETRIEVAL**
EE380L - Data Mining (Dr. Joydeep Ghosh)
- Use **SVDPACKC** for SVD(Singular Value Decomposition) of VSM(Vector Space Model) for Query Retrieval.
 - Modifying **SVDPACKC**(in C Language) is essential to get sparse matrix storage format of CCS.
 - Use **classic3** data(CISI, CRAN, MED) for experiments.
- January 2000 ~ May 2000* **COMPARISONS ON PARTITIONING ALGORITHMS**
CS395T - Conference Course (Dr. Inderjit S. Dhillon)
- Use **METIS** and **hMETIS** for text document clustering (here we do partitioning).
 - Compare with other graph partitioning algorithms such as **CHACO** and **FM**
 - Modifying **METIS**(in C Language) is essential.

- Convert VSM into graph model.

<i>August 1999 ~ May 2000</i>	SPECTRAL GRAPH PARTITIONING CS383C - Numerical Analysis: Linear Algebra (Dr. Inderjit S. Dhillon)
	<ul style="list-style-type: none"> • Use existing <i>Lanczos</i>-based software for computing eigenvectors of both adjacency and Laplacian matrices. • Some experience with <i>Lanczos</i> algorithm (in FORTRAN and C Language) is essential. • Apply spectral algorithm to special graphs such as Clique and Roach graphs.
<i>July 1997 ~ July 1998</i>	DESIGN AND APPLICATION OF INTELLIGENT SYSTEM USING CLUSTERING TECHNIQUE AND EVOLUTION PROGRAM Korea University Research Foundation
<i>April 1997 ~ March 1998</i>	EFFICIENT CLUSTERING ALGORITHM Research Assistant Korea University Research Foundation
<i>August 1996 ~ July 1997</i>	AN OPTIMAL DESIGN PROCEDURE FOR BSB(Brain-State-in-a-Box) NEURAL NETWORKS Research Assistant Korea Research Foundation

PROFESSIONAL EXPERIENCE

<i>June 2004 ~ Current</i>	RESEARCH ASSISTANT Research on Co-clustering Algorithms (Dr. Inderjit S. Dhillon) The University of Texas, Austin, Texas
<i>January 2002 ~ May 2004</i>	TEACHING ASSISTANT Elements of Software Design (Java Programming) The University of Texas, Austin, Texas
<i>January 2001 ~ December 2001</i>	RESEARCH ASSISTANT Research on Classification Algorithms (Dr. Inderjit S. Dhillon) The University of Texas, Austin, Texas
<i>August 1999 ~ May 2000</i>	TEACHING ASSISTANT Computer Programming: C++ The University of Texas, Austin, Texas
<i>March 1999 ~ July 1999</i>	PART-TIME INSTRUCTOR C and Assembly Language Korea University, Seoul, Korea
<i>March 1996 ~ February 1998</i>	RESEARCH AND TEACHING ASSISTANT Database and Fuzzy System Laboratory Korea University, Seoul, Korea

PUBLICATIONS

- H. Cho and I. S. Dhillon, "SPHERICAL CO-CLUSTERING ALGORITHM: THEORY AND APPLICATION," *In preparation*.
- H. Cho and I. S. Dhillon, "CENTERING AND SCALING OF MICROARRAYS: A COMPARATIVE STUDY FOR CLUSTERING ALGORITHMS," *In preparation*.
- H. Cho and I. S. Dhillon, "CO-CLUSTERING OF HUMAN CANCER MICROARRAYS USING MINIMUM SUM-SQUARED RESIDUE CO-CLUSTERING (MSSRCC) ALGORITHM," *In preparation*.
- H. Cho, I. S. Dhillon, Y. Guan, and S. Sra, "MINIMUM SUM-SQUARED RESIDUE CO-CLUSTERING OF GENE EXPRESSION DATA," *Proceedings of the fourth SIAM International Conference on Data Mining*, pp. 114-125, 2004.

- H. Cho, Y. Im, J. Park, J. Moon, and D. Park, "DESIGN OF GBSB NEURAL ASSOCIATIVE MEMORIES USING SOLUTION SPACE PARAMETERIZATION AND OPTIMIZATION APPROACH," *International Journal of Fuzzy Logic and Intelligent System*, vol, I, no. 1, pp. 35-43, 2001.
- J. Park, H. Cho, and D. Park, "ON THE DESIGN OF BSB NEURAL ASSOCIATIVE MEMORIES USING SEMIDEFINITE PROGRAMMING," *Neural Computation*, vol, 11, pp. 1985-1994, 1999.
- J. Park, H. Cho, and D. Park, "DESIGN OF GBSB NEURAL ASSOCIATIVE MEMORIES USING SEMIDEFINITE PROGRAMMING," *IEEE Transactions on Neural Networks*, vol, 10, no. 4, pp. 946-950, July 1999.
- H. Cho, "SYNTHESIS OF NEURAL ASSOCIATIVE MEMORIES USING QUALITATIVE ANALYSIS AND LMI TECHNIQUES," *M.S. Thesis*, Korea University, Feb. 1998.
- Four other papers in Korean are available upon request.

SKILLS

<i>Language</i>	HIGH LEVEL: BASIC, FORTRAN, COBOL, PASCAL, C, C++, JAVA LOW LEVEL (Assembly Language): Z80, I8051, I80286, M68000 OTHERS: HTML, UNIX Shell, Perl
<i>ToolBoxes</i>	FUZZY LOGIC TOOLBOX IN MATLAB LMI TOOLBOX IN MATLAB NEURAL NETWORK TOOLBOX IN MATLAB
<i>Others</i>	FUZZY SET THEORY, ROUGH SET THEORY, OPTIMIZATION THEORY, GENETIC ALGORITHMS, NEURAL NETWORKS, ORACLE, SVDPACKC, CHACO, (h)METIS, RAINBOW

REFERENCES

Dr. INDERJIT S. DHILLON

Department of Computer Sciences
The University of Texas at Austin
Austin, TX 78712

office : (512) 471-9725 ext. 253
Email : inderjit@cs.utexas.edu

Dr. JOYDEEP GHOSH

Department of ECE
The University of Texas at Austin
Austin, TX 78712

office : (512) 471-8980
Email : ghosh@ece.utexas.edu

Dr. DAIHEE PARK

Department of Computer Science
Korea University
Seoul, Korea(ROK)

office : (82) 2-3290-3987
Email : dhpark@korea.ac.kr