

Arvind Balasubramanian

2200 Waterview Parkway Apt. 2134
Richardson, TX 75080

Phone: 972-342-0091
Email: arvindb1986@gmail.com

OBJECTIVE

Seeking a challenging internship position where I can apply my skills in research and software development.

EDUCATION

PhD in Computer Science

Aug 2008 – Present

Advisor: Dr. Balakrishnan Prabhakaran

MS in Computer Science (GPA: 3.94/4.0)

Aug 2008 – Dec 2010

The University of Texas at Dallas, Richardson, TX

B.Tech. in Information Technology (GPA: 8.67/10.00)

Aug 2004 – May 2008

Bengal Institute of Technology, Kolkata, India

TECHNICAL SKILLS

Research Areas: Machine Learning, Data Mining, Statistical Analysis, Multimedia Applications.

Languages and Skills: C, C++, Java, Python, HTML, CSS, JavaScript, JSP.

Operating Systems: Windows, UNIX.

Databases: MySQL, Microsoft SQLServer, Microsoft Access.

Additional Skills: MATLAB, Weka, OpenCV.

RESEARCH EXPERIENCE

Research Assistant

May 2009 – Present

Multimedia and Networking Lab, Department of Computer Science

The University of Texas at Dallas, Richardson, TX

Worked with data mining techniques for knowledge discovery and pattern recognition. Emphasis on analyzing motion tracking data of the human body and tongue, with highly significant applications in medicine and tele-immersive computing. Involved in optimizing index and retrieval structures for human motion datasets.

Working on feature extraction techniques for tongue motion data, using machine learning algorithms to extract factors useful in automatic classification of speech pronunciations. This would greatly assist in diagnosis and therapy of speech disorders.

Also worked on analysis of patterns in scores from communication disorder tests for children (speech, language and hearing disorders). Actively collaborating with the Callier Center for Communication Disorders.

ACADEMIC PROJECTS

Feature Class Discovery in Domain-Specific Information Extraction

Developed and demonstrated an information extraction framework to identify and classify relevant details from text data pertaining to a specific domain. Developed a novel feature class discovery technique using a word clustering approach. The words are clustered using the K-means clustering technique on the basis of their contextual similarities provided by Normalized Google Distance. The obtained word clustering is expected to contribute new label(s) to the existing known class set. Points of focus were the architecture and the functional details of the system, and performance evaluation. The system was developed in Java and the open source machine learning toolkit Weka.

Query-Based Document Retrieval System Using Vector Space Model

Designed and implemented a query based document retrieval system, which returns a list of documents from a database based on relevance to a user-specified query string. The system incorporates the Vector Space Model (where each word is a feature, and each document a feature vector) uses Term frequency and inverse document frequency (TF-IDF) for term weighting and cosine similarity as similarity measure between the query and documents. The system was developed primarily in Python.

Part-of-Speech Tagger

Developed a highly accurate Part of Speech Tagger for English Text using the Hidden Markov Model technique. The model is trained from a number of training files, and Viterbi Decoding algorithm is used to predict the best POS tag for a word, taking into consideration the sequence of words preceding this word. Smoothing and Katz Backoff were used to enhance accuracy of the system. The tagger was developed in Python.

Automatic License Plate Recognition System

Developed an automatic vehicle license plate recognition system, which involves modules for detection and extraction of the license plate region from images of vehicles and for optical character recognition to extract the text information on the license plate. The system was developed using C++, the open source computer vision toolkit OpenCV, and the Google Tesseract OCR engine.

Text Categorization

Evaluated several supervised learning classifiers on a text categorization task for a dataset of movie-review documents labeled with respect to their overall sentiment polarity (positive or negative). Classifiers used include support vector machines, adaptive boosting, decision trees, K-nearest-neighbor and Naive Bayes classifiers, and were evaluated using 10-fold cross validation. Weka and Python were the primary tool and language of implementation.

Ensuring Consistency of Replicated Files in a Fault-Tolerant Distributed System

Simulated a distributed system in which a single file is replicated at a number of sites. The Suzuki-Kazami token-based mutual exclusion algorithm is employed to handle requests from processes to modify the respective local copies of the file. The system is also made tolerant against partitioning due to site failures by incorporating the Majority-based Dynamic Voting Protocol. An update to the file occurs only on acquiring a majority of votes from sites with the latest version. The updates are reflected across all the sites that participated in the voting, ensuring consistency. The project was done using multithreading and socket programming in Java.

Nine Men's Morris Game

Implemented the strategy game Nine Men's Morris using both the Minimax and Alpha-Beta game tree techniques. The user plays against the computer, which is programmed to compute and evaluate all possible moves for the next seven turns based on an optimization function, and play the most optimal move. The game and the accompanying interactive GUI were developed in Java.

City Traffic Regulation System using Optimized DVRP

Proposed a City Traffic Regulation System based on Distance Vector Routing Protocol (DVRP). The project involved an in-depth analysis of the nature and effects of vehicular traffic congestion, presented a method to overcome the adverse effects of traffic congestion, and elaborated on the elements and implementation details of the system proposed. An optimized version of DVRP was designed and implemented, and a simulation of the system working on the optimized algorithm was developed to illustrate the functioning of the system when implemented in a real-world scenario. The implementation was developed using Java and the RouterSim Simulator.

PROFESSIONAL EXPERIENCE**Teaching Assistant****Sept 2008 – May 2009**

Department of Computer Science

The University of Texas at Dallas, Richardson, TX

Assisted in teaching and grading undergraduate courses in C++, Java, and Object Oriented Analysis and Design, and also tutored students in the Undergraduate Programming Lab.

RELEVANT COURSEWORK

Design and Analysis of Computer Algorithms

Machine Learning

Advanced Machine Learning

Advanced Data Mining

Natural Language Processing

Artificial Intelligence

Multimedia Systems

Quantitative Analysis of Risk and Uncertainty in Business

Advanced Operating Systems

Database Design

INDUSTRIAL/ACADEMIC PROGRAMS***Infosys Campus Connect Program***

An academic training program with focus on applied industry-related topics in Relational Database Management Systems, System Development Methodology, Client Server Concepts, Web Technology, User Interface Development, and other topics in Computer Science, conducted by *Infosys Technologies Limited*.

IBM Corporation's Database Management Concepts and PL/SQL.**PRESENTATIONS**

Arvind Balasubramanian, Christine Dollaghan, Thomas Campbell and Balakrishnan Prabhakaran, '*Predicting Poor Language Scores at Age 6 From Combination of Scores at Ages 3 and 4*', Symposium on Research in Child Language Disorders 2010, Madison WI.

AVAILABILITY: Summer 2011