I. FACULTY MEMBER INFORMATION

Name: Rick Chow
Title: Associate Professor of Computer Science
Email: rchow@uscupstate.edu
Phone: 864-503-5206
Department: Division of Math and CS

II. STUDENT INFORMATION

Name: Frank Darwin
Major: Mathematics
Email: fsdarwin@uscupstate.edu
Phone: N/A
Department: Division of Math and CS

III. CHAIR/DEAN INFORMATION

Department Chair: Dr. York Bradshaw
Department Email: ybradshaw@uscupstate.edu

IV. PROJECT SUMMARY

Semester student will work: summer

Project Title:
An Evolutionary Algorithm-Based Feature Selection Engine for a Clinical Costs Prediction System

Project Summary:
- A brief overview of the project, current status, importance of the work, etc.

Healthcare costs are escalating in the U.S., partly due to the rapid rate of increase of people diagnosed with chronic diseases such as diabetes. Sound policies for funding at governmental agencies and cost management at the hospitals depend on an effective system that can predict the costs of hospital stays.

Such a prediction system will first process large
amount of data from national databases with hundreds of variables and millions of records to establish a prediction model. Given a patient’s demographic information, insurance coverage, diagnoses, and/or treatments, the system will then try to predict the hospital charges and length of stay as accurate as possible.

In a joint effort to develop such a prediction system, Drs. Rick Chow and Wei Zhong in Computer Science teamed up with Dr. Marsha Dowell in Nursing and Dr. Richard Stolz in Business. Currently, a prototype of the prediction system is developed with some promising results. However, the dataset, HCUP-3, has a large number of variables (high-dimensions) and substantial amount of erroneous data caused by human errors and missing data (noise), prediction results vary significantly across different clusters of the data.

Evolutionary Algorithm (EA) is a class of algorithms that mimic Nature’s evolutionary process for solving optimization problems. To address the high-dimensional and noisy problems of the HCUP-3 dataset, an EA based engine will be added to the prediction prototype to reduce the dimensions by selecting only the core variables/features that are optimal for effective predictions and to reduce noise by eliminating irrelevant variables. Such a feature selection engine should improve the prediction capability significantly.

This project focuses on the development of an EA based feature selection engine that will work seamlessly with the prediction system prototype. Experimental results will be gathered for further enhancements of the prediction system. A summary report will be generated to present experimental results to the research group. A paper based on the experimental results will be submitted to an international conference.

- Student objectives – what will be accomplished by the end of the semester.

If this research assistant position is funded, the student will be involved in the development of the EA based feature selection engine for the prediction system. A number of objectives or milestones are listed below:
1. Learning the basics of an Evolutionary Algorithm. (1 week)
2. Learning the prediction prototype. (1 week)
3. Developing the EA based feature selection engine (5 weeks)
   a. Developing the core engine (3 weeks)
   b. Developing the interface with the prediction systems (2 weeks)
4. Testing and debugging (2 week)
5. Writing a summary report (1 week)

**Student Qualifications:**
- The student’s academic performance to date (include GPA), relevant course work, background knowledge of the project, etc.
- Exactly what the student’s responsibilities will be.
- Why the student would be successful in this project.

Frank Darwin has a USC System GPA of XXX. He has prior experience in undergraduate research and presented at the third annual USC Upstate Research Symposium. His research commitment and capability is highly praised by his advisor, Dr. Gamal Elnagar.

This project involves techniques and procedures that require a solid background in Math and Statistics. Being a Math major, Frank has the necessary background to perform effectively in this research project.

Frank also completed a number of software development courses such as SCSC 321, and is performing very well in Dr. Zhong’s Data Mining class, SCSC 525, this semester. In SCSC 525, he has learned the basic data mining techniques that are essential for our research project. More importantly, Frank is a conscientious and reliable student. He has shown great enthusiasm in exploring new ideas. He is expected to make significant contributions to this research effort.

As for responsibility, first, the student will be responsible for reading literatures relevant to the topics and attending presentations and discussions sessions conducted by Drs. Zhong and Chow this semester to have a head start. Second, the student will learn the existing programs in great details. Third, the student will be involved in the development of the EA based feature selection engine. Next, he will test and conduct experiments, and write a lab reports to present the experimental results to the research group.

**Resulting External Submission:**
- Include detailed information about the external publication, presentation, performance, or exhibit that will most likely result from this work.

Results generated in this project will be included in a paper to be submitted to one of the two major conferences in Evolutionary Computing: Genetic and Evolutionary Computation Conference (GECCO) 2008 and
IEEE Congress on Evolutionary Computation (CEC) 2008. The results will also support the broader research effort of an NIH AREA Grant (R15).

V. VITA

Employment:
Rick Chow, Associate Professor, USC Upstate 1999 - 2007

Education:
Ph.D. in Computer Science, University of Louisiana Lafayette, 1994.

Research Interests:
Evolutionary Computation, Bioinformatics, Middleware and Databases

Recent Publications:


Please report any website problems to curs @ uscupstate.edu