

## A Recursive Application of a Support Vector Machine for Protein Spot Detection in 2-Dimensional Gel Electrophoresis

Gary D. Boetticher, Hisham Al-Mubaid, Karen Frasier-Scott

THE ANALYSIS OF LARGE COLLECTIONS of proteins has historically utilized two-dimensional polyacrylamide gel electrophoresis. Proteomic technology, however, needs less labor-intensive protein spot detection methodology. UHCL researchers are studying techniques for recursively applying a Support Vector Machine (SVM) in identifying protein. An SVM can be programmed to optimize differences between classes, which correspond to the presence or absence of a protein.



Dr. Gary D. Boetticher

### Publication List

Boetticher, G., H. Al-Mubaid, and K. Frasier-Scott. "Automated Hybridization of Machine Learners for Recursive Spot Identification, Optimization, and Gel Matching of 2-Dimensional Gel Electrophoresis," *J. Comp. Sci.* (2005). (Accepted.)

Dasika, M. "A User Driven Protein Spot Detection Web Service On 2-Dimensional Gel Electrophoresis," Master's thesis, U. of Houston-Clear Lake, Houston, TX. Chair: Dr. Boetticher, 2005.

### Funding and Proposals

"Mucosal Biomarkers of Viral Induced CAP," SEPSIS and CAP: Partnerships for Diagnostics Development, National Institute of Allergy and Infectious Diseases, National Institutes of Health, involving the University of Texas Medical Branch (UTMB), 5 years, \$251,362. (Not funded.)

"Using Machine Learners To Predict Infant RSV Infections," National Heart Lung Blood Institute, Clinical Proteomics Programs, RFA-HL-04-019, National Institutes of Health, involving the University of Texas Medical Branch (UTMB), 4 years, \$199,282. (Not funded.)

## The Impact of Chromosome Lineage upon Genetic Program Modeling

Gary D. Boetticher, Kim Kaminsky

ONE OF THE CHALLENGES IN DATA MINING IS PROVIDING SUFFICIENT coverage of the search space in order to produce an acceptable model. Traditionally, Genetic Programs consider all chromosomes within a population for breeding purposes. Considering the expanse of the search space, it is imperative to focus breeding efforts in Genetic Programs in order to attain a better solution in less time. This research examines the lineage of Genetic Programs in order to identify any breeding patterns. Five separate experiments are conducted where chromosomes are grouped into five classes. Lineage patterns are assessed for the best, middle, and worst-class parental chromosomes. Based upon the results, a new Genetic Programming (GP) process is proposed.

The technique utilized in this study solves problems by genetically breeding a population of individuals, or chromosomes, over a series of generations.

### Results

The results are very promising from these experiments. We anticipate that the proposed approach will lead to significant improvements in genetic programming performance.

### Publications

Boetticher, G. and K. Kaminsky. "Assessing the Impact of Chromosome Lineage in Genetic Program Modeling," *Data Mining and Knowledge Discovery J.* (Submitted.)

Month	Lat	Long	Speed	Wind	
330	September	14.0	58.0	18.0	35.0
331	September	13.3	32.7	11.0	35.0
332	September	13.2	33.8	11.0	40.0
333	September	13.0	34.8	10.0	45.0
334	September	9.1	47.0	16.0	38.0
335	September	8.3	48.7	18.0	38.0
336	September	9.5	50.0	13.0	30.0
337	October	8.5	43.6	17.0	35.0
338	October	10.1	45.9	17.0	40.0
339	October	10.7	46.4	17.0	45.0
340	November	15.8	75.5	12.0	35.0
341	November	15.8	76.5	10.0	35.0
342					