INDUSTRIAL MATHEMATICS MODELING WORKSHOP FOR GRADUATE STUDENTS 2000 ABSTRACTS

Problem 1. *Text Classification Using Bayesian Nets and Support Vector Machines*

Laura Mather
La Jolla Research Lab
Britannica.com, Inc.
3252 Holiday Court, Suite 208
La Jolla, CA 92037

Team Members: Marguerite Etheridge, Vahan Grigoryan, Steven Lord, Scott Pope, Charlotte Schulze-Hewett and Ling Wu

Faculty Consultant: Naihuan Jing

Working Sessions held in **Room 227** Harrelson

Classification of text documents into topical categories is one of the current areas of interest in the information retrieval community. The definition of text classification is as follows. Given a document or set of documents pertaining to a particular topic, find other documents that are similar to the given set. The applications of text classification are broad, including email filtering, improved text retrieval, and improved document clustering.

This problem will explore the use of several mathematical classification techniques and their applicability to the text classification problem.
Classification techniques to be considered include Bayesian Nets and
Support Vector Machines. In most cases the words of a document are used as the features for the classification algorithm. Since there are usually many more words in a document set than most classification techniques can handle, it is likely that some feature reduction will be necessary. In addition, mapping the feature set of text documents to the various classification techniques may be a challenge. Finally, a comparison of the performance of the various algorithms will be conducted.

Problem 2. *Modeling Cash Flows in Bond Structures*

Christian Wypasek
Structured Transactions and Analytical Research Group
First Union Securities
NC0166
One First Union Center, TW6
Charlotte, NC 28288-0166

Team Members: Brian Adams, Jean Cadet, Lingu Du, Xiaohua Du, Chotiros Surapholchai, Xuelei Wang

Faculty Consultant: Jean-Pierre Fouque

Working Sessions held in **Room 221** Harrelson

In the financial arena, future cash flows from income-producing assets are pledged to pay bonds that are sold in the capital markets. Often the income-producing assets are loans. A bond structure facilitates the investor purchase of slices (or tranches) of the aggregate cash flow in a manner designed to mitigate certain forms of risk to the investor inherent in the underlying assets.

For a given transaction, bonds of varying degrees of risk are issued. The bond issuer may retain the residual tranche, the riskiest interest of the cash flows, as it is often not salable. This retained interest is valued on the Company's books as the present value of its projected cash flows. The cash flows are highly sensitive to assumptions (models for loss and prepayment) surrounding the underlying assets. Incorrect estimates of the future cash flows to the retained interest will cause the value of that interest to be incorrectly stated on the Company's books. The team will be asked to investigate the effect on evaluation of a bond structure caused by risk interactions in the underlying assets.

The team will have an opportunity to analyze real world data and build or apply models for the asset cash flow from origination through delinquency, prepayment or end of observation. Distributions for prepayment and default will be simulated at the loan level and aggregated into a simple bond structure. For the residual tranche, the variability of the realized economic benefit is sought; finding a characterization in terms of prepayment/default/interest rate interaction are of particular interest. If time allows, an additional issue of interest is an investigation of the effect model mis-specification on the value of the residual.

Problem 3. *A Competition Model for Viral Inhibition of Host Cell Proliferation*

Sarah Holte
Fred Hutchinson Cancer Research Center
1100 Fairview Ave. North MW-500
Seattle, WA 98109

Team Members: David Bortz, Bob Guy, Jeff Hood, Kay Kirkpatrick, Vinh Nguyen, Victoria Shimanovich

Faculty Consultant: H.T. Banks

Working Sessions held in **Room 233** Harrelson

Some viruses encode proteins that promote cell proliferation while others, such as the human immunodeficiency virus (HIV), encode proteins that prevent cell division. It has been hypothesized that the selective advantage determining which strategy evolves depends on the ability of the virus to induce a cellular environment which maximizes both virus production and cell lifespan. In HIV, the protein that causes cell cycle arrest is Vpr.

Recently we developed a mathematical model based on discrete difference equations (Holte S, Emerman M. to appear, Mathematical Biosciences) to study the competition between two genotypes of HIV---one that encodes this protein (Vpr+) and one that does not (Vpr-). We would like to extend this model to a continuous PDE model that can account for the dependence of cellular production of virus on the age (i.e. time since infection) of the infected cells. The model will be used to assess differences between laboratory ({\it in vitro}) and human infection ({\it in vivo}) conditions.
In particular, we are interested in parameters that could be different between the {\it in vitro} condition, where the Vpr- genotype dominates, and the {\it in vivo} condition, where the Vpr+ genotype dominates. The model will also be used to estimate a variety of parameters from data produced in laboratory experiments.

Problem 4: *Carrier Phase Synchronization for Advanced Digital Communications Systems*

Tien M. Nguyen
The Aerospace Corporation
P.O. Box 92957
Los Angeles, CA 90009-2957

Team Members: Hafid Chrifi Alaoui, Xiuxia Du, Tung Nguyen, Jean Steiner, Chao-Jen Wong, Yi Yang

Faculty Consultant: H.T. Tran

Working Sessions held in **Room 266** Harrelson

Synchronization is the heart of digital communications systems. The functions of a synchronizer are to generate a set of reference signals, such as carrier phase, carrier frequency and timing references, for demodulation and detection processes. In this workshop we will concentrate only on the carrier phase synchronization problem.

Theory for a simple carrier phase tracking loop will be developed, and it will be used as a guideline to analyze tracking loops employed by advanced digital communications systems. Using the theory developed here, we will analyze the carrier phase tracking performance of a Gaussian Minimum Shift-Keying (GMSK) communication system.

Problem 5: *Diffusion Models of Photobleaching Experiments for Articular Cartilage*

Farshid Guilak
Orthopaedic Research Laboratories
Department of Surgery
Duke University Medical Center
Box 3093, Med Sci Res Bldg Rm 375
Durham, NC 27710

Team Members: Jeff Anderson, Feng-Nan Hwang, Sung Ha Kang, Bahareh Momken, Richard Schugart and Caroline Torcaso

Faculty Consultant: Mansoor Haider

Working Sessions held in **Room 353** Harrelson

Articular cartilage is a hydrated biological soft tissue that acts as a load-bearing surface in joints. The structural matrix in cartilage is comprised of collagen fibers and charged macromolecules (proteoglycans) with embedded cells (chondrocytes) that are responsible for matrix synthesis and repair. The maintenance of healthy cartilage and its progressive degradation (osteoarthritis) are known to be highly dependent on the mechanical and electrochemical properties of the extracellular matrix. The diffusion properties of the extracellular matrix can be quantified via photobleaching experiments in which fluorescent tracer particles are introduced into a region of tissue and their diffusion is monitored using confocal microscopy. The goal is to model the photobleaching experiment for cartilage and characterize the dependence of effective diffusion coefficients on anisotropy and fixed charge density in the extracellular matrix.

Problem 6: *Frictional Sliding on Surface of Variable Curvature*

Tony Royal
Jenike & Johanson, Inc.
One Technology Park Drive
Westford, MA 01886-3189

Team Members: Gerard Awanou, Gregory Baramidze, Jeffrey Housman, Christos Kavouklis, Julia Sulakova and Michael Zager

Faculty Consultant: Pierre Gremaud

Working Sessions held in **Room 215** Harrelson

Streams of powders are important in studying the behavior of filling of bins, belt-to-belt transfers and pneumatic conveying. Applications are relevant to a wide range of problems in industries such as pharmaceutical and food industries, large scale power plants, ship-loading and mining operations. The problem posed will be the formulation and solution of equations of motions of a particle sliding on a frictional surface of arbitrary curvature. Equations for a sloping plane, cylinder and cone have been derived. The hope is that a general algorithm can be developed.