Summer 2007 Program on Challenges in Dynamic Treatment Regimes and Multistage Decision-Making
June 18-29, 2007

SPEAKER ABSTRACTS

Victoria Chen
University of Texas at Arlington
Department of Industrial and Manufacturing Systems Engineering
vchen@uta.edu

"An Adaptive Dynamic Programming Decision Making Framework"

Dynamic programming (DP) is a method for optimizing a system changing over time that has been successfully applied in manufacturing systems, environmental engineering, business, and many other fields. Exact solutions are only possible for small problems or under very limiting restrictions, but recent advances in computational power have given rise to many approximate DP methods. These advances now provide the potential for application of DP to complex dynamic decisions, such as adaptive interventions or dynamic treatment regimes. The key advantage of DP is its ability to account for future decisions when optimizing a current decision. This presentation will describe a DP based decision-making framework in the context of an air quality problem and pain management.

Damien Ernst
Supélec
damien.ernst@supelec.fr

"Clinical Data Based Optimal STI Strategies for HIV: A Reinforcement Learning Approach"

Nowadays, the design of drug-scheduling strategies plays a critical role in the cure of several diseases such as AIDS or cancers. Since for such diseases, precise scheduling and combination of drugs are determinant in the global efficiency of the therapy, physicians face the difficult problem of deciding when and which drugs to administer to a patient. Moreover, for such diseases, the dynamics of their interaction with the immune system are so complex that the design of some close-to-optimal treatment strategies requires the use of some specific techniques able to capture the subtle aspects of these dynamics. In this talk, we address the particular problem of computing optimal Structured Treatment Interruptions (STI) strategies for HIV infected patients. Such STI strategies consist of cycling a patient on and off anti-retroviral drugs in order to bring him into a state in which he can maintain immune control over the virus in the absence of treatment. We argue that reinforcement learning techniques may be useful to extract such strategies

directly from clinical data, without the need of an accurate mathematical model of the HIV infection dynamics. To support our claims, we present simulation results obtained by running a recently proposed batch-mode reinforcement learning algorithm, known as fitted Q iteration, on numerically generated clinical data.

Adam Gaweda
University of Louisville
Department of Medicine
aegawe01@gwise.louisville.edu

"From Population to Individual Drug Dosing in Chronic Illness - Intelligent Control for Management of Renal Anemia"

Variability in drug response between individuals makes effective treatment of chronic illnesses challenging. This presentation demonstrates how we use control theory and machine learning to transition the process of drug delivery from population to the individual, patient-specific standpoint. Our work covers development of dose-response models from clinical data, as well as the use model-based and model-free control methods in chronic drug dosing. Our application, used in the presentation, is the pharmacological management of anemia due to kidney failure using recombinant human erythropoietin as the drug of interest. Throughout the talk we will outline and emphasize open questions and challenges that have arisen in the course of the presented research effort.

Miguel Hernan
Harvard School of Public Health
Department of Epidemiology
mhernan@hsph.harvard.edu

"Introduction to Causal Inference"

Erica Moodie
McGill University
Department of Epidemiology and Biostatistics
erica.moodie@mcgill.ca

"Asymptotic Bias Correction for Estimates of Optimal Dynamic Treatment Regimes"

One approach to inference about optimal dynamic regimes in a multi-interval trial is Robins' (2004) g-estimation, which always yields consistent estimates. However, the estimates may be asymptotically biased under certain longitudinal distributions of the treatments and covariates, termed exceptional laws. In this talk, I will introduce exceptional laws and describe a new approach to recursive g-estimation which we call Zeroing Instead of Plugging in (ZIPI). ZIPI shares all of the nice asymptotic properties of recursive g-estimates at non-exceptional laws while providing a reduction in the asymptotic bias at exceptional laws when decision rule parameters are not shared across intervals.

Susan Murphy
University of Michigan
Department of Statistics
samurphy@umich.edu

"Introduction on Nonstandard Statistical Inference"

In this tutorial I will illustrate why methods/algorithms for constructing decision rules lead to non-regular estimators. This fact is problematic when measures of confidence (e.g. confidence intervals, hypothesis tests) are desired due to the small size of the training set. Measures of confidence would aid in reducing the number of variables in the decision rules and in ascertaining when there is no evidence in the data that two actions lead to different results.

Ron Parr
Duke University
Department of Computer Science
parr@cs.duke.edu

"RL with Additional Discussion of Connections to Classification"

Joelle Pineau
McGill University
Department of Computer Science
jpineau@cs.mcgill.ca

(Tuesday, June 19, 2007)
"Computational Challenges with High Dimensional Data"

This tutorial will present reinforcement learning techniques for handling realistic dynamic treatment design problems. We will discuss the challenge of learning a value function in large state and action spaces, and present possible solutions using both linear function approximation and kernel-regression techniques. A case study will be presented pertaining to treatment design for major depressive disorder. The talk will also outline challenges pertaining to efficient action selection during learning though an examination of concepts such as exploration and bayesian reinforcement learning. Finally we will discuss the selection and design of an appropriate reward function through preference elicitation methods.

(Thursday, June 21, 2007)
"Adaptive Stimulation Design for the Treatment of Epilepsy"

Brain stimulation has recently emerged as a promising therapy for patients with medically-intractable epilepsy. However little is known about the best stimulation patterns to use when applying electrical stimulation, such that we get maximal seizure reduction, while also minimizing long-term damage to the brain. The overall goal of this

project is to automatically optimize a closed-loop strategy for the control of deep brain stimulation using reinforcement learning methods.

In this talk, I will outline recent progress on this project, including: (1) use of ensemble methods to automatically detect seizures, (2) design of a computational model of epilepsy to provide synthetic training data for the reinforcement learning agent, (3) initial results of applying SARSA-based reinforcement learning within the computational model, (4) formal results pertaining to transfer of control strategies between models (e.g. from the computational model to the biology).
This is joint work with Robert Vincent (McGill University), Aaron Courville (Universite de Montreal) and Massimo Avoli (Montreal Neurological Institute).

Daniel Rivera
Arizona State University
Department of Chemical Engineering
daniel.rivera@asu.edu

"Introduction to Mechanistic Models and Control Theory"

The dynamic behavior of engineering systems can be described by mechanistic models based on the conservation and accounting of physical properties such as mass, momentum, and energy. The set of differential and algebraic equations arising from mechanistic modeling can be used in control system design methodologies whose ultimate purpose is to alter dynamical system behavior from undesirable conditions to desirable ones. A brief overview of mechanistic modeling and control system design will be presented, illustrated with examples. Some thoughts on how these ideas can inform the development of dynamic treatment regimes will be described. For instance, what does control theory tell us about model accuracy requirements when the intended purpose of the model is to design a dynamic treatment regime? The tutorial will conclude by providing a glimpse of some emerging paradigms in control systems engineering that appear to have particular relevance to dynamic treatment regimes and multi-stage decision-making.

Andrea Rotnitzky
Di Tella University and Harvard University
Department of Statistics
arotnitzky@utdt.edu

"Estimation of Dynamic Treatment Regimes Effects under Flexible Dynamic Visit Regimes"

Often in the management of chronic diseases, doctors indicate the patient when the next clinic visit should be according to medical guidelines. Of course, in spite of their doctors' indication, patients are free to return to the clinic earlier if they need to do so. Naturally, at every clinic visit, whether planned or not, treatment decisions are made, e.g. whether to start, switch, continue, discontinue or, alter the dose of a, treatment. It is thus of public health interest to estimate the effect of dynamic treatment regimes that are to be implemented in settings in which: i) medical guidelines are used by doctors to indicate their patients when the next clinic visit should be and these indications may depend on

the patient health status, ii) patients may come to the clinic earlier than the indicated return date and iii) doctors have the opportunity to intervene and alter the treatment each time the patient comes to the clinic. Marginal structural mean (MSM) models (Robins, 1999, 2000, Murphy et. al., 2003) for the effects of dynamic treatment regimes assume the frequency of clinic visits is the same for all patients. In this talk we present a method, based on an extension of the MSM model, which allows estimation from observational data of the effects of dynamic treatment regimes that are to be implemented in settings in which i) ii) and iii) hold.
This is joint work with Liliana Orellana.

Peng Sun
Duke University
Department of The Fuqua School of Business
psun@duke.edu

"Bias and Variance in Value Function Estimates"

We consider a finite-state, finite-action, infinite-horizon, discounted reward Markov decision process and study the bias and variance in the value function estimates that result from empirical estimates of the model parameters. We provide closed-form approximations for the bias and variance, which can then be used to derive confidence intervals around the value function estimates. We illustrate and validate our findings using a large database describing the transaction and mailing histories for customers of a mail-order catalog firm.
Joint work with Shie Mannor, Duncan Simester and John Tsitsiklis.

Ambuj Tewari
University of California, Berkeley
Department of Computer Science
ambuj@cs.berkeley.edu

"Sample Complexity of Policy Search with Known Dynamics"

Policy search methods are one of the several approaches to solving large scale Markov decision processes (MDPs). Assuming access to a simulator of the system, Ng and Jordan (2000) showed that sharing random bits between different policy evaluations allows us to put the problem of choosing a near optimal policy from a given class in the standard statistical learning theory framework. For this approach, we can prove uniform bounds on the difference between empirical and true value functions of policies under an assumption on the policy class, reward functions and state transition dynamics. These bounds then easily lead to sample complexity estimates. I will try to show why the policy search setting is more complex than the usual statistical learning setting.

Peter Thall
University of Texas MD Anderson Cancer Center
Department of Biostatistics Department

rex@mdanderson.org

“Comparing Two-Stage Treatment Strategies Based on Sequential Failure Times Subject to Interval Censoring”

For many diseases, therapy involves multiple stages, with the treatment in each stage chosen adaptively based on the patient's current disease status and history of previous treatments and clinical outcomes. Physicians routinely use such multi-stage treatment strategies, also called dynamic treatment regimes or treatment policies. This talk presents a Bayesian framework for a clinical trial comparing two-stage strategies based on the time to overall failure, defined as either second disease worsening or discontinuation of therapy. Each patient is randomized among a set of treatments at enrollment, and if disease worsening occurs the patient is then re-randomized among a set of treatments excluding the treatment received initially. The goal is to select the best two-stage strategy, defined as that having either the largest mean or median overall failure time. A parametric model is formulated to account for non-constant failure time hazards, regression of the second failure time on the patient's first disease worsening time, and the complications that the failure time in either stage may be interval censored and there may be a delay between first worsening and the start of the second stage of therapy. The method is applied to a trial of six two-stage strategies involving treatments for metastatic kidney cancer. A simulation study in the context of this trial is presented.

Anastasios Tsiatis
North Carolina State University
Department of Statistics
tsiatis@stat.ncsu.edu

“Introduction to Dynamic Treatment Regimes”