Peter Krafft: Previous Research

I have a strong background in computer science and Bayesian statistics, which I have applied to several projects during my undergraduate and master's programs.[1] My work has appeared in four venues: at a local undergraduate research conference, at a workshop associated with an international conference, in an international conference, and in a book chapter. I have also given two talks at my school and one at Cambridge University.

Graduate Research:

My first graduate project was joint work with Evan Ray and Prof. John Staudenmayer. One of the problems with developing new statistical methods is that practitioners such as clinical researchers, especially those who are not familiar with programming, are often hesitant to implement the new methods themselves. Even when statisticians develop packages for their code, these packages depend on users being familiar with the programming language for which the packages are written. These difficulties make the rate of adoption of new methods much slower. **In this project, I helped develop a software package which allows statisticians to quickly deploy a web-based interface for their algorithms.** We used the package ourselves to create a website for some of Prof. Staudenmayer's recent research, and this work was recognized as a runner-up for best student presented poster at an international conference [4]. We will be deploying the website with a forthcoming journal paper [2] and releasing the general package with a separate paper.

I am doing my master's project with Prof. Hanna Wallach (a computer scientist) and Prof. Bruce Desmarais (a political scientist). In this project we are using the email inboxes and outboxes from the county department managers of New Hanover County, North Carolina. We want to use this dataset to infer attributes of managers who are particularly effective at getting their own agendas on their county's legislative docket as measured by the number of times each manager's department is mentioned in the county's legislative agendas. **This project could help bureaucrats design more effective communication strategies.** However, this research poses a major modeling challenge. Email data contain both text attributes (the subject and body of the each email) and network attributes (the author and recipients of each email). Furthermore, an initial analysis of these data suggested that a joint approach to modeling the text and network attributes would be more effective than treating these features as independent, but there has been very little work in joint text and network modeling. As a result, **I had to brainstorm many possible instances of this new class of models**. After extensive discussions with Prof. Wallach and Prof. Desmarais about the properties of each of these models, I selected one to investigate further. By extending a model from the network literature to a framework that is appropriate for email data and combining it with a standard probabilistic text model, the model I developed can identify who each managers communicates with about particular topics. This model will allow us to discover whether managers with broader communication strategies are more effective than managers with more targeted communication strategies. I am currently writing up our preliminary results in the form of a tech report which we plan to submit in February to the International Conference on Machine Learning. [1] I will release the data and my code from this project after we have published our results.

[1] For full papers from these projects, see my website: `http://www.cs.umass.edu/~pkrafft/papers/`.

Undergraduate Research:

One relevant research project from my undergraduate career that helped inspire my current interest in Bayesian statistics was a project with Arthur Baines and Prof. Michael Lavine. In this project we studied a popular model that is used to incorporate spatial effects into statistical analyses. Previous researchers had shown that this model is sensitive to its specification. In their example, Reich, Hodges, and Zadnik showed how slight changes to the model specification determine whether or not to draw the conclusion that the data provide evidence for a relationship between socioeconomic status and stomach cancer rates in Slovenia. Our project investigated a different issue with this same model. Through simulation, we observed that the precision parameters of the model could not always be estimated reliably, an important result to researchers interested in knowing the precise amount of variation explained by each component of the model. **From this research I learned about how to analyze statistical models through simulation, and I learned that it is important to be aware of the unexpected properties that complex models introduce.**

Supervisory Roles:

I have also mentored two undergraduate students. The first of these students was Jessie Hamelin who was an NSF REU student working with Prof. Sridhar Mahadevan last summer. I helped Jessie formulate and implement her research plan for the summer, which involved working with another undergraduate, Nathaniel (Nate) May, on a project largely of Nate's conception: developing artificial intelligence for the game of Othello. Although Jessie went back to her school in the fall, I am now mentoring Nate who is an undergraduate in my department. Nate has been helping with the representation and analysis of the email data that I am using in my master's project. He is currently looking at whether subsets of the communication network relating to particular topics display different patterns than the network as a whole. **Nate has expressed interest in continuing his work on this project next semester.** Both of these experiences have been rewarding and have helped me understand the perspective of my own mentors and advisors.

Publications and Presentations (accepted and in preparation):

[1] P. Krafft, H. Wallach, and B. Desmarais. Mixed-membership modeling of government email networks. UMass Computer Science Department Tech Report: UM-CS-2011-045, In preparation.

[2] J. Staudenmayer, K. Lyden, S. Keadle, E. Ray, P. Krafft, and P. Freedson. A novel method to estimate energy expenditure from an accelerometer. *Biometrics*, In preparation.

[3] C. Wang, P. Krafft, and S. Mahadevan. Manifold alignment. In *Manifold Learning: Theory and Applications*. Taylor and Francis. CRC Press, In press.

[4] E. Ray, P. Krafft, P. Freedson, and J. Staudenmayer. Novel analytic methods to estimate physical activity from accelerometer data: an opensource web-based tool. *International Conference on Ambulatory Monitoring of Physical Activity and Movement*, 2011. **Runner-up, Best Student Presented Poster**.

[5] P. Krafft and S. Mahadevan. Feature-preserving embeddings for topic transfer. *NIPS Workshop on Transfer Learning by Learning Deep Generative Models*, 2010.

[6] P. Krafft. Applying deep belief networks to the game of Go. Undergraduate Thesis. University of Massachusetts Amherst, 2010.

[7] P. Krafft and M. Lavine. Modeling light in Harvard Forest. *The 15th Annual Massachusetts Statewide Undergraduate Research Conference*, 2009.