

## **CptS 595: AI Seminar**

**Fall 2015**

**Tuesdays 9:10-10:25**

**EME 130**

**INSTRUCTOR:** Diane J. Cook  
**OFFICE:** EME 121  
**OFFICE PHONE:** 335-4985  
**CLASS SITE:** <http://eecs.wsu.edu/~cook/aiseminar>

### **COURSE OVERVIEW:**

The purpose of this class is to serve as a journal club. During our weekly classes we will meet to critically evaluate recent articles in the Artificial Intelligence academic literature. We will not focus on the research areas you are each pursuing for your dissertation. Instead, we will look at papers introducing emerging topics, papers with demonstrated excellence, and papers describing topics of interest to the AI Group at Washington State University.

Each class we will focus on one paper or topic. Each student will lead one of these classes. Unlike paper critiques, the main focus of the presentation will be to educate the class on the topic and techniques found in the paper(s). You are encouraged to go beyond the selected paper as appropriate to either 1) survey other papers on the topic, 2) create examples to illustrate the ideas, 3) implement and evaluate an aspect of the idea, and/or 4) create discussion questions for the class.

### **EVALUATION:**

- (1) 1 Presentation, 60%. You will need to propose a topic or paper one week before your presentation, before the beginning of the class period. Send your proposal (2-3 sentences) to [cook@eecs.wsu.edu](mailto:cook@eecs.wsu.edu). Include a link to a paper of focus or email the paper itself. I will then make this available to the class to read before the presentation date.
- (2) Discussion, 40%. We want this group to be as engaged as possible, this is the best way to learn the material. You will have access to the paper ahead of time which gives you a chance to think of at least one question for discussion each class.

One key component of the discussion this semester is the advance discussion points we will ask you to identify each week. In particular, we will ask you to email two questions and/or discussion points to me at [cook@eecs.wsu.edu](mailto:cook@eecs.wsu.edu) and to the speaker no later than Monday noon, one day before the presentation. This will allow the speaker to organize the discussion and will make it easier for you to raise these questions and discussion points in class.

<u>Date</u>	<u>Presenter / Topic</u>
Aug 25	Yinghui Wu
Sep 1	
Sep 8	
Sep 15	
Sep 22	
Sep 29	
Oct 6	
Oct 13	Guest speaker: Andrew Sixsmith, 1:30pm – 2:30pm
Oct 20	Guest speaker: Jesse Hoey, 1:30pm – 2:30pm
Oct 27	
Nov 3	
Nov 10	
Nov 17	
Nov 24	Thanksgiving vacation
Dec 1	Guest speaker: Sherry Willis, 1:30pm – 2:30pm
Dec 8	

**Safety Information:** <http://safetyplan.wsu.edu> ; <http://oem.wsu.edu/emergencies>

Washington State University has developed a resource in support of commitment to the safety of students, faculty, staff and visitors. The Campus Safety Plan provides a comprehensive listing of university policies, procedures, statistics, and information relating to campus safety, emergency management, and the health and welfare of the campus community.

## Suggested Topic List

- Time series analysis
- GPU computing for machine learning (e.g., CUDA)
- Crowd sourcing for machine learning
- ML for game design and adaptation
- Integration of neuroscience into AI / machine learning
- HPC approaches to Big Data ML
- AI inspired by multidisciplinary research (e.g., neuroscience, cognitive psychology, information theory, biological computing)

## Papers to Consider

### (Best / Most Cited / Suggested Papers Close to Topic List)

- E. Papalexakis, A. Fyshe, N. Sidiropoulos, P. Talukdar, T. Mitchell, C. Faloutsos. Good-enough brain model: Challenges, algorithms and discoveries in multi-subject experiments, KDD 2014.
  - Given a simple noun such as apple, and a question such as "is it edible?", what processes take place in the human brain? More specifically, given the stimulus, what are the interactions between (groups of) neurons (also known as functional connectivity) and how can we automatically infer those interactions, given measurements of the brain activity? Furthermore, how does this connectivity differ across different human subjects? In this work we present a simple, novel good-enough brain model, or GeBM in short, and a novel algorithm Sparse-SysId, which are able to effectively model the dynamics of the neuron interactions and infer the functional connectivity. Moreover, GeBM is able to simulate basic psychological phenomena such as habituation and priming (whose definition we provide in the main text). We evaluate GeBM by using both synthetic and real brain data. Using the real data, GeBM produces brain activity patterns that are strikingly similar to the real ones, and the inferred functional connectivity is able to provide neuroscientific insights towards a better understanding of the way that neurons interact with each other, as well as detect regularities and outliers in multi-subject brain activity measurements.
- C.-L. Liu, T.-H. Tsai, and C.-H. Lee. Online Chinese restaurant process, KDD 2014.
  - Processing large volumes of streaming data in near-real-time is becoming increasingly important as the Internet, sensor networks and network traffic grow. Online machine learning is a typical means of dealing with streaming data, since it allows the classification model to learn one instance of data at a time. Although many online learning methods have been developed since the development of the Perceptron algorithm, existing online methods assume that the number of classes is available in advance of classification process. However, this assumption is unrealistic for large scale or streaming data sets. This work proposes an online Chinese restaurant process (CRP) algorithm, which is an online and nonparametric algorithm, to tackle this problem. This work proposes a relaxing function as part of the prior and updates the parameters with the likelihood function in terms of the consistency between the true label information and predicted result. This work presents two Gibbs sampling algorithms to perform posterior inference. In the experiments, the online CRP is applied to three massive data sets, and compared with several online learning and batch learning algorithms. One of the data sets is obtained from Wikipedia, which comprises approximately two million documents. The experimental results reveal that the proposed online CRP performs well and efficiently on massive data sets. Finally, this work proposes two methods to update the hyperparameter  $\alpha$  of the online CRP. The first method is based on the posterior distribution of  $\alpha$ , and the second exploits the property of online learning, namely adapting to change, to adjust  $\alpha$  dynamically.
- Janardhan Rao Doppa, Alan Fern, and Prasad Tadepalli. HC-Search: Learning Heuristics and Cost Functions for Structured Prediction, AAAI 2013 (\* covered in Jana's class)
  - Structured prediction is the problem of learning a function from structured inputs to structured outputs. Inspired by the recent successes of search-based structured prediction, we introduce a new framework for structured prediction called HC-Search. Given a structured input, the framework uses a search procedure guided by a learned heuristic H to uncover high quality candidate outputs and then uses a separate learned cost function C to select a final prediction among those outputs. We can decompose the regret of the overall approach into the loss due to H not leading to high quality outputs, and the loss due to C not selecting the best among the generated outputs. Guided by this decomposition, we minimize the overall regret in a greedy stage-wise manner by first training H to quickly uncover high quality outputs via imitation learning, and then training C to correctly rank the outputs generated via H according to their true losses. Experiments on several benchmark domains show that our approach significantly out-performs the state-of-the-art methods.

- Elias Bareinboim, Jin Tian, Judea Pearl. Recovering from Selection Bias in Causal and Statistical Inference, AAI 2014.
  - Selection bias is caused by preferential exclusion of units from the samples and represents a major obstacle to valid causal and statistical inferences; it cannot be removed by randomized experiments and can rarely be detected in either experimental or observational studies. In this paper, we provide complete graphical and algorithmic conditions for recovering conditional probabilities from selection biased data. We also provide graphical conditions for recoverability when unbiased data is available over a subset of the variables. Finally, we provide a graphical condition that generalizes the backdoor criterion and serves to recover causal effects when the data is collected under preferential selection.
- Gary Doran and Soumya Ray. SMILe: Shuffled multiple-instance learning, AAI 2013.
  - Resampling techniques such as bagging are often used in supervised learning to produce more accurate classifiers. In this work, we show that multiple-instance learning admits a different form of resampling, which we call “shuffling.” In shuffling, we resample instances in such a way that the resulting bags are likely to be correctly labeled. We show that resampling results in both a reduction of bag label noise and a propagation of additional informative constraints to a multiple-instance classifier. We empirically evaluate shuffling in the context of multiple-instance classification and multiple-instance active learning and show that the approach leads to significant improvements in accuracy.
- Xiang Chen, Tovi Grossman, Daniel Wigdor, and Goerge Fitzmaurice. Duet: Exploring joint interactions on a smart phone and a smart watch, CHI 2014.
  - The emergence of smart devices (e.g., smart watches and smart eyewear) is redefining mobile interaction from the solo performance of a smart phone, to a symphony of multiple devices. In this paper, we present Duet -- an interactive system that explores a design space of interactions between a smart phone and a smart watch. Based on the devices' spatial configurations, Duet coordinates their motion and touch input, and extends their visual and tactile output to one another. This transforms the watch into an active element that enhances a wide range of phone-based interactive tasks, and enables a new class of multi-device gestures and sensing techniques. A technical evaluation shows the accuracy of these gestures and sensing techniques, and a subjective study on Duet provides insights, observations, and guidance for future work.
- Todd Kuesza, Saleema Amershi, Rich Caruana, Danyel Fisher, and Denis Charles. Structured labeling for facilitating concept evolution in machine learning, CHI 2014.
  - Labeling data is a seemingly simple task required for training many machine learning systems, but is actually fraught with problems. This paper introduces the notion of *concept evolution*, the changing nature of a person's underlying *concept* (the abstract notion of the target class a person is labeling for, e.g., *spam* email, *travel related* web pages) which can result in inconsistent labels and thus be detrimental to machine learning. We introduce two *structured labeling* solutions, a novel technique we propose for helping people define and refine their concept in a consistent manner as they label. Through a series of five experiments, including a controlled lab study, we illustrate the impact and dynamics of concept evolution in practice and show that structured labeling helps people label more consistently in the presence of concept evolution than traditional labeling.
- Jian Tang, Zhaoshi Meng, Xuanlong Nguyen, Ziaozhu Mei, and Ming Zhang. Understanding the limiting factors of topic modeling via posterior contraction analysis, ICML 2014.
  - Topic models such as the latent Dirichlet allocation (LDA) have become a standard staple in the modeling toolbox of machine learning. They have been applied to a vast variety of data sets, contexts, and tasks to varying degrees of success. However, to date there is almost no formal theory explicating the LDA's behavior, and despite its familiarity there is very little systematic analysis of and guidance on the properties of the data that affect the inferential performance of the model. This paper seeks to address this gap, by providing a systematic analysis of factors which characterize the LDA's performance. We present theorems elucidating the posterior contraction rates of the topics as the amount of data increases, and a thorough supporting empirical study using synthetic and real data sets, including news and web-based articles and tweet messages. Based on these results we provide practical guidance on how to identify suitable data sets for topic models, and how to specify particular model parameters.
- Thanawin Rakthanmanon, Bilson Campana, Abdullah Mueen, Gustavo Batista, Brandon Westover, Qiang Zhu, Jesin Zakaria, and Eamonn Keogh. Searching and mining trillions of time series subsequences under dynamic time warping, KDD 2014.
  - Most time series data mining algorithms use similarity search as a core subroutine, and thus the time taken for similarity search is the bottleneck for virtually all time series data mining algorithms. The difficulty of scaling search to large datasets largely explains why most academic work on time series data mining has plateaued at considering a few millions of time series objects, while much of industry and science sits on billions of time series objects waiting to be explored. In this work we show that by using a combination of four novel ideas we can search and mine truly massive time series for the first time. We demonstrate the following extremely unintuitive fact; in large datasets we can exactly search under DTW much more quickly than the current state-of-the-art *Euclidean distance* search algorithms. We demonstrate our work on the largest set of time series experiments ever attempted. In particular, the largest dataset we consider is larger than the combined size of all of the time series datasets considered in all data mining papers ever published. We show

that our ideas allow us to solve higher-level time series data mining problem such as motif discovery and clustering at scales that would otherwise be untenable. In addition to mining massive datasets, we will show that our ideas also have implications for real-time monitoring of data streams, allowing us to handle much faster arrival rates and/or use cheaper and lower powered devices than are currently possible.

- Vias C. Raykar, Shipeng Yu, Linda H. Zhao, Gerardo Hermosillo Valadez, Charles Florin, Luca Bogoni, and Linda Moy. Learning From Crowds. JMLR 11:1297-1322, 2010.
  - For many supervised learning tasks it may be infeasible (or very expensive) to obtain objective and reliable labels. Instead, we can collect subjective (possibly noisy) labels from multiple experts or annotators. In practice, there is a substantial amount of disagreement among the annotators, and hence it is of great practical interest to address conventional supervised learning problems in this scenario. In this paper we describe a probabilistic approach for supervised learning when we have multiple annotators providing (possibly noisy) labels but no absolute gold standard. The proposed algorithm evaluates the different experts and also gives an estimate of the actual hidden labels. Experimental results indicate that the proposed method is superior to the commonly used majority voting baseline.