Course Form for PKU Summer School International 2018

Course Title	**Probabilistic Models for Structured Data**
	结构化数据的概率模型
Teacher	SUN Yizhou
First day of classes	July 2, 2018
Last day of classes	July 13, 2018
Course Credit	2 credits

Course Description

Objective

The course aims to introduce probabilistic models for structured data, where data points are no longer independent with each other, such as sequential data and graph/network data. The course will cover modeling, inference, and learning of state-of-the-art probabilistic models, including Hidden Markov Model, Markov Random Field, Conditional Random Field, and Factor Graph. Applications across different domains, such as text mining, medical domain, and social network analysis. At the end of the course, the students are expected to be able to do the following: (1) understanding the mathematical formulation of different probabilistic models that work for structured data, including intuition and mathematical derivations and proof; (2) apply these models to real-world applications; (3) potential of developing novel models for structured data for publications.

Pre-requisites /Target audience

Pre-requisites: basic knowledge in statistics and probability, linear algebra, optimization, programming.
Target audience: Senior undergraduate students and graduate students in various disciplines (computer science, statistics, economics, finance, electronic engineering, biology, physics)

Proceeding of the Course

No

Assignments (essay or other forms)

Readings, In-class and online discussions, and take-home exercises

Evaluation Details

Attendance and Discussions: 25%
Assignments: 45%
Exam: 30%

Text Books and Reading Materials

- Daphne Koller and Nir Friedman (2009). Probabilistic Graphical Models. The MIT Press;
- Charles Sutton and Andrew McCallum (2014). An Introduction to Conditional Random Fields. Now Publishers.
- http://deepdive.stanford.edu/inference
- Additional readings are given in the Class Schedule.

Academic Integrity (If necessary)

Students are allowed to discuss readings and assignments among classmates in and outside the class, but are discouraged to seek help from any living person outside the class. However, individual-based writing assignments must be independently completed (i.e., without any plagiarism).

CLASS SCHEDULE
(Subject to adjustment)

Session 1: Introduction to Probabilistic Models and Structured Data	Date: 7/2/2018

【Description of the Session】(purpose, requirements, class and presentations scheduling, etc.)
Review of basics of probability theory and statistics; introduction to probabilistic models and structured data; MLE and MAP principles; applications.

【Questions】
What are probabilistic models? What are structured data? What will structure bring in to probabilistic models? What would be the standard procedure involved in probabilistic modeling? What will be the useful applications of such models? What are the principles of inference and learning on such models?

【Readings, Websites or Video Clips】

【Assignments for this session (if any)】
Problems on basic stats.

Session 2: Probabilistic Models for Unstructured Data	Date: 7/3/2018

【Description of the Session】(purpose, requirements, class and presentations scheduling, etc.)
Introduction of two well-known probabilistic models: Naïve Bayes and Logistic Regression, and discuss of their limitations.

【Questions】
What is Naïve Bayes? What is logistic regression? What are their limitations? What are generative models and discriminative models? How to extend Naïve Beyes to semi-supervised setting?

【Readings, Websites or Video Clips】

【Assignments for this session (if any)】
Theoretical questions;
Implementation of two algorithms and apply them on a text classification task.

Session 3: Warm up: Hidden Markov Models	Date: 7/4/2018
【Description of the Session】(purpose, requirements, class and presentations scheduling, etc.) Introduce HMM, which is a well-known probabilistic model for sequential data. Introduce the concepts of modeling, inference, and learning via HMM.	
【Questions】 What is sequential data? What are the applications? What is HMM? What are the standard modeling, inference and learning procedure of HMM?	
【Readings, Websites or Video Clips】	
【Assignments for this session (if any)】	

Session 4: Markov Random Fields	Date: 7/5/2018
【Description of the Session】(purpose, requirements, class and presentations scheduling, etc.) Introduce MRF, a more general undirected graphical model. Introduce a simple pairwise MRF.	
【Questions】 What are MRFs? What is Markov property? What is collective inference? What is Gibbs distribution? How to construct a simple pairwise MRF?	
【Readings, Websites or Video Clips】	
【Assignments for this session (if any)】 Theoretical questions; Implementation of simple pairwise MRF, with the application of text classification.	
Session 5: Gaussian Markov Random Fields	Date: 7/6/2018
【Description of the Session】(purpose, requirements, class and presentations scheduling, etc.) Introduce another special case of MRF, where random variables can take numerical values.	
【Questions】 What is Gaussian MRF? What is the modeling, inference and learning procedure involved?	
【Readings, Websites or Video Clips】	
【Assignments for this session (if any)】 Theoretical questions; Implementation.	
Session 6: Hinge Loss Markov Random Fields	Date: 7/9/2018
【Description of the Session】(purpose, requirements, class and presentations scheduling, etc.) Introduce another special case of MRF, where reasoning can be performed. An application on medical inference will be shown.	
【Questions】	

How to model human knowledge encoded as logic rules and apply these rules in the modeling of MRF?	
【Readings, Websites or Video Clips】	
【Assignments for this session (if any)】 Brain storming on possible applications.	
Session 7: Conditional Random Fields	Date: 7/10/2018
【Description of the Session】(purpose, requirements, class and presentations scheduling, etc.) Introduce a directed probabilistic graphical model, which is CRF; introduce a special case of CRF, linear-chain CRF, and its application on named entity recognition (NER).	
【Questions】 What are CRFs? What is the difference between CRFs and MRFs? What are the pros and cons.	
【Readings, Websites or Video Clips】	
【Assignments for this session (if any)】 Theoretical questions; Implementation and apply it on text classification.	
Session 8: Skip-Chain Conditional Random Field	Date: 7/11/2018
【Description of the Session】(purpose, requirements, class and presentations scheduling, etc.) Introduce another CRF, and its application to relation extraction.	
【Questions】 What is the limitation of linear-chain CRF? What is general CRF? How does skip-chain overcome the limitation of linear-chain CRF? How can skip-chain be applied to relation extraction?	
【Readings, Websites or Video Clips】	
【Assignments for this session (if any)】	
Session 9: Factor Graph	Date: 7/12/2018
【Description of the Session】(purpose, requirements, class and presentations scheduling, etc.) Introduce a general form of probabilistic model, and its inference algorithm, sum-product algorithm.	
【Questions】 What is factor graph? How to do inference on it? What is sum-product algorithm? What are the relationship between factor graph and the previous two models.	
【Readings, Websites or Video Clips】	
【Assignments for this session (if any)】	

Session 10: Student Presentation / Exam	Date: 7/13/2018
【Description of the Session】 (purpose, requirements, class and presentations scheduling, etc.) Student presentation of previous homeworks and exam.	
【Questions】	
【Readings, Websites or Video Clips】	
【Assignments for this session (if any)】	