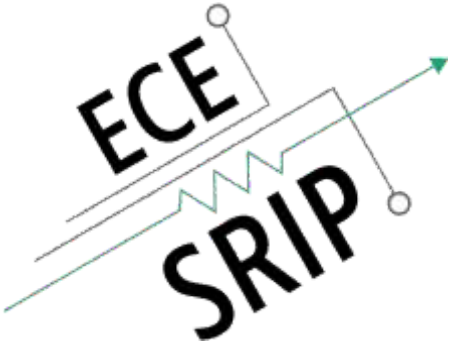



## FACULTY MENTOR
Xie, Pengtao

## PROJECT TITLE
Interpretable Generation of Diagnosis Reports from Medical Images

## PROJECT DESCRIPTION
Medical imaging is widely used in clinical practice for diagnosis and treatment.
For each medical image, a diagnosis report needs to be written to narrate the medical findings in the image. Writing diagnosis reports can be error-prone for inexperienced physicians, and time-consuming and tedious for experienced physicians. To address these issues, there have been researching works on automatically generating medical imaging reports. In these works, the report generators are mostly black-box deep learning models that make the generated reports hard to interpret. In this project, we will study the interpretable generation of diagnosis reports, which aims to not only generate high-quality reports but also make the generated reports explainable. We will adopt a retrieve-and-synthesize
a strategy which searches for the diagnosis guidelines relevant to the input image
from medical textbooks, and synthesizes the retrieved guidelines for interpretation.

Will accommodate either in person or remotely.

## INTERNS NEEDED
4 Students

## PREREQUISITES
1. Have a machine learning background
2. Proficient in PyTorch or TensorFlow

## FACULTY MENTOR
Xie, Pengtao

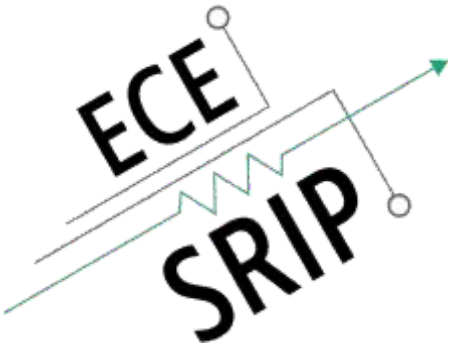
## PROJECT TITLE
Multi-task Learning with Graph Neural Networks

## PROJECT DESCRIPTION
Multi-task learning (MTL) is a widely studied machine learning problem, where a set of correlated tasks are learned jointly. The existing approaches for MTL are inadequate in capturing the nonlinear relationship among different tasks. In this project, we will propose a framework for nonlinear MTL based on graph neural networks (GNN). The relationship between tasks is represented using a graph and it is assumed that the models of correlated tasks have similar weight parameters. We generate the weight parameters of individual tasks by transforming meta weights. The transformation is conducted using a graph neural network built on top of the task relationship graph. The developed methods will be applied for individualized medical diagnosis.

Will accommodate either in person or remotely.

## INTERNS NEEDED
4 Students

## PREREQUISITES
1. Have a machine learning background
2. Proficient in PyTorch or TensorFlow




## FACULTY MENTOR
Xie, Pengtao

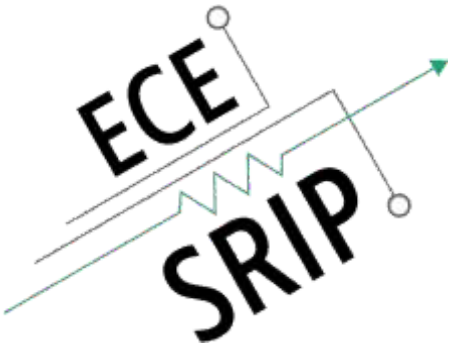
## PROJECT TITLE
Discriminative Learning of Graph Structure

## PROJECT DESCRIPTION
Given a set of covariates, learning a graph structure among them is a widely studied machine learning problem that has many downstream applications, such as estimating brain networks from FMRI data, inferring gene regulatory networks, the understanding of the stock market, etc. In existing approaches, such a structure is typically learned in an unsupervised way, mostly by maximizing the data likelihood of the covariates. The structure learned in an unsupervised way may not be optimal for discriminative tasks, is not amenable for interpretation, and is not response-specific. In this project, we will study supervised graph structure learning, where the structure is learned to optimally predict the responses. We will develop a general framework that uses graph neural networks to characterize the predictive power of a graph structure and use reinforcement learning to search for the optimal structure. To improve computational efficiency, we will also study a differentiable structure learning approach. The developed methods will be applied for estimating brain networks from FMRI data and inferring gene regulatory networks.

Will accommodate either in person or remotely.

## INTERNS NEEDED
4 Students

## PREREQUISITES
1. Have a machine learning background
2. Proficient in PyTorch or TensorFlow




## FACULTY MENTOR
Xie, Pengtao

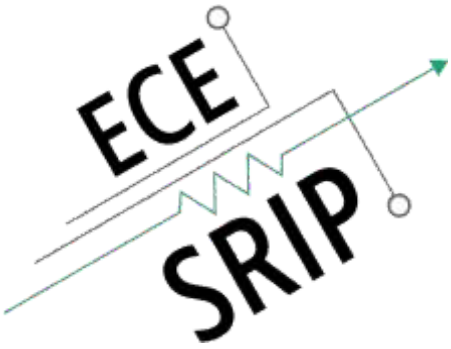
## PROJECT TITLE
Few-shot Multi-instance Learning

## PROJECT DESCRIPTION
Multi-instance learning (MIL) is a widely studied problem in machine learning, which aims to classify a bag of instances to a single class label. In many real-world applications, both the number of bags for each class and the number of instances in each bag are small, which results in a doubly few-shot MIL problem. For example, in a medical diagnosis task where the diseases are classes, patients are bags, and different hospital admissions of the same patient are instances in the bag. For rare diseases, the number of patient cases is small. And for most young patients, the number of admissions is small. In this project, we will solve the few-shot MIL problem. We assume each bag has a bag-specific distribution from which the instances of this bag are generated. We measure the similarity of two bags directly on their distributions, instead of on the individual instances. In this way, the global properties of bags can be better captured, which helps to improve classification accuracy. We will study two ways to measure the similarity of two distributions, based on the maximum mean discrepancy and optimal transport. Then a nonparametric classification approach will be developed based on the bag similarity. The developed methods will be applied for clinical diagnosis based on multiple episodes of patient visits.

Will accommodate either in person or remotely.

## INTERNS NEEDED
4 Students

## PREREQUISITES
1. Have a machine learning background
2. Proficient in PyTorch or TensorFlow




## FACULTY MENTOR
Xie, Pengtao

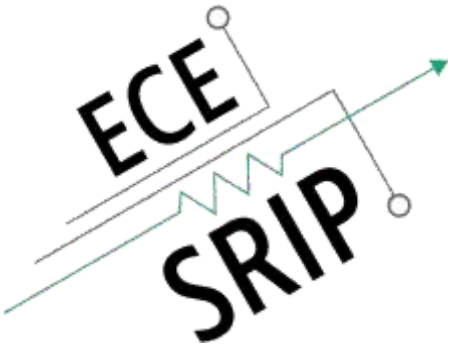
## PROJECT TITLE
Automatic Medical Knowledge Graph Construction Using Deep Reinforcement Learning

## PROJECT DESCRIPTION
We aim at developing an ML system that automatically builds a medical knowledge graph (KG) consisting of 1) entities such as symptoms, diseases, medications, and 2) the relationships between entities such as “disease A can be treated by medication B”, “disease A causes symptom C”. This system continuously and iteratively performs the following steps: 1) collecting medical documents from a broad range of sources such as clinical notes from electronic healthcare systems, medical literature from PubMed, clinical guidelines from textbooks; 2) extracting candidate entities and relations from these documents by performing syntactic and semantic analysis; 3) selecting the authentic entities and relations from the candidates by inspecting their compactness with the existing knowledge graph. At the core of this system is deep reinforcement learning (RL)
method. We build a virtual agent that takes actions to construct a KG by interacting with the environment, with the goal of improving the preciseness and coverage of the KG.

This project will accommodate either in person or remotely.

## INTERNS NEEDED
4 Students

## PREREQUISITES
1. Have a machine learning background
2. Proficient in PyTorch or TensorFlow




## FACULTY MENTOR
Xie, Pengtao

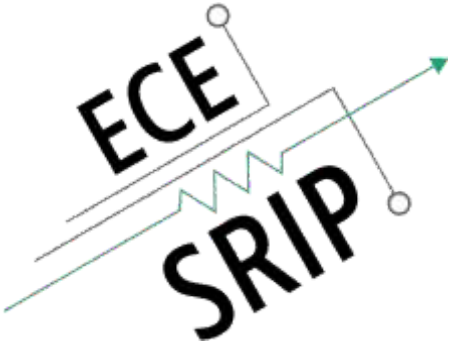
## PROJECT TITLE
Diversity-promoting Learning of Hierarchical Latent Space Models

## PROJECT DESCRIPTION
Hierarchical latent space models (HLSMs) are widely utilized to discover latent patterns that inherently have a hierarchical structure. The patterns discovered by existing HLSMs are biased towards frequent signals, have a lot of redundancy, and are not amenable for interpretation. To address this issue, we will propose to diversify the components in HLSMs. We will develop two hierarchical diversity-promoting priors including the hierarchical mutual angular process and hierarchical determinantal point process, and apply them to two HLSMs including hierarchical topic models and hierarchical sparse coding. We will develop posterior inference algorithms based on variational inference and MCMC sampling. The developed methods will be applied for extracting diverse topic hierarchy from documents.

This project will accommodate either in person or remotely.

## INTERNS NEEDED
4 Students

## PREREQUISITES
1. Have a machine learning background
2. Proficient in PyTorch or TensorFlow




## FACULTY MENTOR
Xie, Pengtao

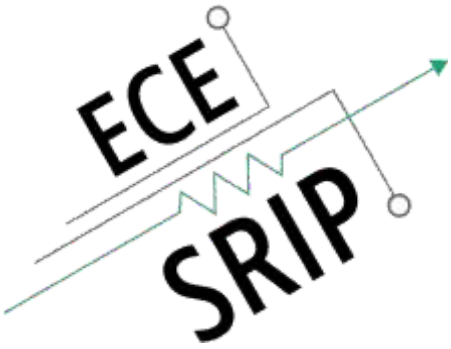
## PROJECT TITLE
Curriculum Determinantal Point Process for Diverse Subset Selection

## PROJECT DESCRIPTION
Diverse subset selection has wide applications in machine learning, such as document summarization, product recommendation, etc. In many real-world applications, it is oftentimes difficult to select the subset at one stroke and the selection is conducted in multiple rounds. For example, in hiring, an employer typically conducts multiple rounds of interviews and selects a subset of candidates in each round. In this project, we propose curriculum diverse subset selection by developing a curriculum determinantal point process that sequentially selects subsets with decreasing sizes. Our method will be applied for two applications: extractive document summarization and product recommendation.

This project will accommodate either in person or remotely.

## INTERNS NEEDED
4 Students

## PREREQUISITES
1. Have a machine learning background
2. Proficient in PyTorch or TensorFlow




## FACULTY MENTOR
Xie, Pengtao

## PROJECT TITLE
When Gaussian Process Meets Determinantal Point Process: Learning Diverse Inducing Points

## PROJECT DESCRIPTION
Gaussian process (GP) is widely used for regression and classification, but is computationally prohibitive. The inducing point method, which utilizes an auxiliary set of data points to form a low-rank approximation of the kernel matrix in GP, is a widely utilized approach to scale GP up. It is desirable to use the smallest number of inducing points to achieve the best approximation. In this project, we will study how to achieve this goal. We will propose to diversify the inducing points in a way that a compact set of inducing points are expressive enough to approximate the kernel matrix with high fidelity. Specifically, the determinantal point process will be leveraged to achieve diversity. The developed methods will be applied for large-scale patient risk stratification.

This project will accommodate either in person or remotely.

## INTERNS NEEDED
4 Students

## PREREQUISITES
1. Have a machine learning background
2. Proficient in PyTorch or TensorFlow