Multi-view Relationships for Analytics and Inference

Abstract:
An interesting area of machine learning is methods for multi-view data, relational data whose features have been partitioned. Multi-view learning exploits relationships between views, giving it certain advantages over traditional single-view techniques, which may struggle to find these relationships or only learn them implicitly. These relationships are often especially salient in understanding the data or performing prediction. This work explores an underutilized approach in multi-view learning: to focus on multi-view relationships—the variables that govern relations between views—themselves as units of analysis. We investigate how this approach impacts analytics and inference in ways that standard multi-view and single-view learning cannot. We hypothesize that by ignoring relations between views or factoring them in only indirectly, standard approaches risk overlooking key structure. Accordingly, our goal is to investigate the extent multi-view relationships can be characterized and employed as units of analysis in descriptive analytics and inference. We present novel methods to do so, either using domain knowledge or by learning from data, which reveal structure that alternative methods do not or have competitive performance with the state of the art. Empirical results are presented in several application domains. First, we use domain knowledge to assume a known form for multi-view relationships in the task of gamma source detection. We aggregate the views by filtering their inferences collectively to perform classification. Second, we assume multi-view relationships are linear and learn them from data in a different approach toward gamma source detection. Our method detects anomalies when these relationships are disrupted. Third, we relax the assumption of linearity and propose a novel clustering method that finds cluster-wise linear relationships. This method discovers explanatory structure in a medical problem. Fourth, we extend this method to classification and demonstrate its superior performance on a load monitoring problem.

Speaker:
Eric Lei

Thesis Committee:
Artur Dubrawski (Chair)
Barnabas Poczos
Mario Berges
Simon Labov (Lawrence Livermore National Laboratory)

July 30, 2019
3:00pm
GHC 8102