
Discriminative Embeddings of Latent Variable Models for Structured Data*

Le Song

Abstract: Kernel classifiers and regressors designed for structured data, such as sequences, trees and graphs, have significantly advanced a number of interdisciplinary areas such as computational biology and drug design. Typically, kernels are designed beforehand for a data type which either exploit statistics of the structures or make use of probabilistic generative models, and then a discriminative classifier is learned based on the kernels via convex optimization. However, such an elegant two-stage approach also limited kernel methods from scaling up to millions of data points, and exploiting discriminative information to learn feature representations.

In this talk, I will present `structure2vec`, an effective and scalable approach for structured data representation based on the idea of embedding latent variable models into feature spaces, and learning such feature spaces using discriminative information. Interestingly, `structure2vec` extracts features by performing a sequence of function mappings in a way similar to graphical model inference procedures, such as mean field and belief propagation. In applications involving millions of data points, we showed that `structure2vec` runs 2 times faster, produces models which are 10,000 times smaller, while at the same time achieving the state-of-the-art predictive performance.

Bio: Le Song is an assistant professor in the Department of Computational Science and Engineering, College of Computing, Georgia Institute of Technology. He received his Ph.D. in Machine Learning from University of Sydney and NICTA in 2008, and then conducted his post-doctoral research in the Department of Machine Learning, Carnegie Mellon University, between 2008 and 2011. Before he joined Georgia Institute of Technology, he was a research scientist at Google. His principal research direction is machine learning, especially nonlinear methods and probabilistic graphical models for large scale and complex problems, arising from artificial intelligence, social network analysis, healthcare analytics, and other interdisciplinary domains. He is the recipient of the NSF CAREER Award¹⁴, AISTATS'16 Best Student Paper Award, IPDPS'15 Best Paper Award, NIPS13 Outstanding Paper Award, and ICML10 Best Paper Award. He has also served as the area chair for leading machine learning conferences such as ICML, NIPS and AISTATS, and action editor for JMLR.

*Internal Talk, Gatsby Unit, July 21, 2016.