



Estimating Conversion Rates of Rare Events through a Multidimensional Dynamic Hierarchical Bayesian Framework

Hongia Yang

Yahoo, Inc., Sunnyvale, CA, USA – hongxia@yahoo-inc.com

We consider the problem of estimating occurrence rates of rare events for extremely sparse data using pre-existing hierarchies and selected features to perform inference along multiple dimensions. In particular, we focus on the problem of estimating click rates for {Advertiser, Publisher, User} tuples where both the Advertisers and the Publishers are organized as hierarchies that capture broad contextual information at different levels of granularity. Typically, the click rates are low and the coverage of the hierarchies and dimensions is sparse. To overcome these difficulties, we decompose the joint prior of the three-dimensional Click-Through-Rate (CTR) using tensor decomposition and propose a Multidimensional dynamic Hierarchical Bayesian framework (abbreviated as MadHab). We set up a specific framework of each dimension to model dimension-specific characteristics. More specifically, a hierarchical beta prior for the Advertiser dimension, a hierarchical dynamic Bayesian prior for the Publisher dimension and a feature-dependent mixture prior for the User dimension. Besides the centralized implementation, we propose two distributed algorithms for inference which make the model highly scalable and suited for large scale data mining applications. We demonstrate that on a real world ads campaign platform our framework can effectively discriminate extremely rare events in terms of their click propensity.

Keywords: Display Advertising, Hierarchical Beta Prior, Dynamic Hierarchical Bayesian Model, Dependent Mixture Modeling, MapReduce, Spark.