Stochastic Blockmodeling for Online Advertising

Li Chen
Department of Applied Mathematics and Statistics
Johns Hopkins University
Baltimore, MD 21210
lichen.jhu1@gmail.com [*]

Matthew Patton
AOL Advertising.com
1020 Hull St
Baltimore, MD 21230
matthew.patton@teamaol.com

Keywords: Online Advertising, Graph Inference, Clustering.

Introduction

Online advertising is market communication over the internet. This form of advertising has proven its importance in the golden digital age. In this work, we approach the business problem of improving the effectiveness of ad campaigns and targeting a wider range of online audience via discovering the intrinsic structures of the websites. We introduce a stochastic blockmodeling framework for discovering the website structures, propose two vertex clustering algorithms based on the Bayesian information criterion, and compare the performance with a goodness-of-fit method and a deterministic graph partitioning method. We demonstrate the effectiveness of our proposed algorithms on simulation and the AOL website dataset.

Background

A graph $G = (V, E)$ represents a collection of interacting objects, where the objects are vertices in the vertex set $V := \{1, 2, \ldots, \}$ containing n vertices, and the interactions are edges in the edge set E. The adjacency matrix A of the graph G is an $n \times n$ matrix. In this work, we assume that the graph G is undirected, unweighted, and not loopy. That implies the adjacency matrix A is symmetric, binary and hollow. A random graph is a graph-valued random variable: $\mathcal{G} : \Omega \to \mathcal{G}_n$, where $\mathcal{G}_n$ represents the collection of all $2^{\binom{n}{2}}$ possible graphs on the vertex set V. Associated with the adjacency matrix A, there exists a communication probability matrix $P \in [0, 1]^{n \times n}$, where each entry P_{ij} denotes the probability of edge existence between vertex i and vertex j.

One very general random graph is the latent position graph (Hoff, Raftery, and Handcock 2002). Each vertex v_i is associated with a latent position X_i drawn independently from some distribution F on $\mathbb{R}^D$. The edge probability between vertices i and j is $P_{ij} = \text{Bernoulli}(l(X_i, X_j))$, where $l : \mathbb{R}^D \times \mathbb{R}^D \to [0, 1]$ is the link function. The stochastic blockmodel (Holland, Laskey, and Leinhardt 1983) is a

special case of the latent position graph, where its latent positions are point masses. The vertices are partitioned in to K blocks, and the edges are conditional independent Bernoulli trials.

Definition 1. *Stochastic blockmodel (SBM) Let K be the number of block memberships. Let π be the block membership function of length K vector such that $\sum_{k=1}^{K} \pi_k = 1$. Let $\{Y_i\}_{i=1}^{n} \sim Categorical([K], \pi)$ be the block memberships. Let $B \in [0, 1]^{K \times K}$ be a symmetric probability matrix denoting the block probabilities. Then $A \sim SBM([n], B, \pi)$ if*

$$
\begin{aligned}
\mathbf{P}_{ij} &= Prob(A_{ij} = 1 | X_i, X_j) \\
&= Prob(A_{ij} = 1 | Y_i, Y_j) = B_{Y_i, Y_j}.
\end{aligned}
$$

The BIC-Based Vertex Clustering Algorithms

In the classical setting for unsupervised learning, we observe independently identically distributed feature vectors $X_1, X_2, \ldots, X_n$, where each $X_i : \Omega \to \mathbb{R}^D$ is a random vector for some probability space Ω. Here we consider the case when the feature vectors $X_1, X_2, \ldots, X_n$ are unobserved. Instead we observe a latent position random graph $\mathcal{G}(X_1, X_2, \ldots, X_n)$ on n vertices. We intend to cluster the vertices using the observed graph.

For stochastic blockmodels with K blocks and a known model dimension D, (Sussman et al. 2012) and (Rohe et al. 2011) respectively have shown that adjacency spectral embedding and Laplacian spectral embedding are consistent estimates of the latent positions. The resulted embedding is a K-mixture and D-variate Gaussian distributions asymptotically. Such results motivate us to propose a model-based clustering approach on the embedded space of the stochastic blockmodel. The optimal number of clusters and covariance structure correspond to the model selection criterion by the Bayesian information criterion (BIC). Our approach is presented in Algorithm 1. We denote the algorithms by ASE and LAP respectively, if M is either the adjacency matrix or the Laplacian matrix.

We compare ASE and LAP with the integrated classification likelihood (ICL) method (Daudin, Picard, and Robin 2008), which is a likelihood maximization method for stochastic blockmodels, and the Louvain algorithm (Blondel et al. 2008), which optimizes graph modularity and performs efficiently for large graphs.

Algorithm 1 The BIC-based Vertex Clustering Approach

Input: An input square matrix M of order n, an integer $K \geq 1$, and an embedding dimension D.

Step 1 : Compute the first D orthonormal eigenpairs of M, denoted by $(U_M, S_M) \in \mathbb{R}^{n \times D} \times \mathbb{R}^D$.

Step 2: Define the D-dimensional embedding of M to be $\hat{M} := U_M S_M^{1/2}$.

Step 3:

for k in $1 : K$ **do**

    Fit Gaussian mixture models with different covariance types and k clusters to $\hat{M}$, and compute the BIC.

end for

Step 4: Cluster the vertices using the optimal model selected via the maximum BIC.

Experiments

Clustering Validation

For simulation, we measure clustering performance using the adjusted rand index (ARI) (Hubert and Arabie 1985). The higher ARI indicates a better clustering performance. For real data experiment, one challenge for clustering validation is the lack of ground truth. We examine the significance of the clusters using external datasets including website topics, revenue per website, number of clicks on ads per website, and impressions (the volume of ad display) per website.

Simulation

In this experiment, only ASE and LAP correctly select the number of blocks with an average ARI above 90%. As the number of clusters increases, the ARIs of ASE and LAP decrease due to the phenomenon of bias-variance tradeoff.

AOL Website Graph

We use the relational events of online users who visit the websites during the day July 1, 2014 under one campaign, and build a graph containing websites as vertices. An edge exists between websites v_i and v_j, if they share at least one common user. The resulted adjacency matrix is symmetrized, binarized, hollow, and of size 1569×1569. In practice, the true model dimension D of the stochastic blockmodel is unknown. We estimate D using a profile likelihood maximization method (Zhu and Ghodsi 2006).

The ARIs for comparing the 6 pairs of algorithms indicate the partitions by ASE and LAP are most similar, while ASE detects most block signal. The partitions by ICL and Louvain are least similar. In terms of website topics, Cluster 1 discovered by ASE mainly contains references and popular sites. Cluster 5 discovered by ASE mainly contains websites on politics, baby, teen, gallary. The clusters discovered by the other clustering algorithms do not correspond to any significant topic clusters.

In addition, we evaluate the clusters using revenue, clicks and impressions. For each pair of clusters in each clustering algorithm, we apply two-sided Wilcoxon rank sum test with the null hypothesis that the revenue/clicks/impressions are the same. At a significance level of 5%, ASE discovers more statistically significant clusters based on the business metrics than other algorithms.

See https://sites.google.com/site/lichenjhuresearch/home for details.

Discussion

In this work, we introduce a stochastic blockmodeling framework for online advertising, and propose two BIC-based vertex clustering algorithms: ASE and LAP. We demonstrate in simulation that our proposed algorithms are able to detect the correct number of blocks. In the AOL website graph experiment, our proposed algorithm ASE is able to detect significant website clusters validated via impressions, revenue and number of clicks. The applications of our approach not only extend to further cluster-based inference, but also can serve as a basic guidance for website inventory acquisition. While our proposed approach is presented for undirected and unweighted graphs, it adapts to directed and weighted graphs. We are optimistic that random graph framework is valuable for online advertising research.

References

Blondel, V. D.; Guillaume, J.-L.; Lambiotte, R.; and Lefebvre, E. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* 2008(10):P10008.

Daudin, J.-J.; Picard, F.; and Robin, S. 2008. A mixture model for random graphs. *Statistics and computing* 18(2):173–183.

Hoff, P. D.; Raftery, A. E.; and Handcock, M. S. 2002. Latent space approaches to social network analysis. *Journal of the american Statistical association* 97(460):1090–1098.

Holland, P. W.; Laskey, K. B.; and Leinhardt, S. 1983. Stochastic blockmodels: First steps. *Social networks* 5(2):109–137.

Hubert, L., and Arabie, P. 1985. Comparing partitions. *Journal of classification* 2(1):193–218.

Rohe, K.; Chatterjee, S.; Yu, B.; et al. 2011. Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics* 39(4):1878–1915.

Sussman, D. L.; Tang, M.; Fishkind, D. E.; and Priebe, C. E. 2012. A consistent adjacency spectral embedding for stochastic blockmodel graphs. *Journal of the American Statistical Association* 107(499):1119–1128.

Zhu, M., and Ghodsi, A. 2006. Automatic dimensionality selection from the scree plot via the use of profile likelihood. *Computational Statistics & Data Analysis* 51(2):918–930.