Implementation of Bayesian Hierachical Clustering

Lina Yang, Xiaodi Qin

April 29, 2017

1 Abstract

Hierarchical clustering is commonly used in unsupervised learning. There are several limitations to the traditional hierarchical clustering, including no guide to pre-specify number of clusters, difficult to choose suitable distance metric, and lack of probability based criterion for model evaluation. To overcome these limitations, Heller and Ghahramani (2005) developed a Bayesian hierarchical clustering based on Dirichlet process mixtures. It uses marginal likelihoods and Bayesian hypothesis testing as criteria for merging cluster.

2 Background

3 Algorithm

4 Optimization

5 Application

5.1 Simulated data sets

5.2 Real data sets

6 Comparative analysis

7 Discussion

8 Code

The repository can be found at `https://github.com/qxxxd/Bayesian-Hierarchical-Clustering`.

Algorithm 1: Bayesian Hierachical Clustering

Input: Data $X = (X_0, X_1, ..., X_N)$, $family \in \{niw\}$, hyperparameter α, scaling factor on the prior precision of the mean r.

Output: A linkage matrix Z

Set $\mathbb{S} = \emptyset$
For $l \in \{0, 1, \ldots, N-1\}$**:**
    $n_l^0 = 1$, $d_l^0 = \alpha$, $X_l^0 = X_l$, $ml_l = \textsc{family}(X_l^0)$
    $\mathbb{S} = \mathbb{S} \cup \{l\}$
Set $t = 0$, $\mathbb{P} = \emptyset$
For $i \in \{0, 1, \ldots, N-2\}$**:**
    For $j \in \{i+1, \ldots, N-1\}$**:**
        $c_{1,t} = i$, $c_{2,t} = j$, $n_t = n_i^0 + n_j^0$
        $X_t = (X_i^0, X_j^0)^T$, $d_t = \alpha\Gamma(n_t) + d_i^0 d_j^0$
        $P_{1,t} = \textsc{family}(X_t)\alpha\Gamma(n_t)/d_t$, $P_{2,t} = ml_i ml_j (d_i^0 d_j^0 / d_t)$
        $logodds_t = \log P_{1,t} - \log P_{2,t}$
        $\mathbb{P} = \mathbb{P} \cup \{t\}$, $t = t+1$
Set $p = 0$, $Z = []$
While *1***:**
    $idx = \arg\max_{idx \in \mathbb{P}} logodds$
    $Z.\textsc{append}([c_{1,idx}, c_{2,idx}, logodds_{idx}, n_{idx}])$
    $n_{N+p}^0 = n_{idx}$, $X_{N+p}^0 = X_{idx}$, $d_{N+p}^0 = d_{idx}$, $ml_{N+p} = P_{1,idx} + P_{2,idx}$
    $rm = \{c_{1,idx}, c_{2,idx}\}$, $\mathbb{S} = \mathbb{S} \setminus rm$
    If $\mathbb{S} = \emptyset$**:**
        break
    For $q \in \mathbb{S}$**:**
        $c_{1,t} = N + p$, $c_{2,t} = q$, $n_t = n_{N+p}^0 + n_q^0$
        $X_t = (X_{N+p}^0, X_q^0)^T$, $d_t = \alpha\Gamma(n_t) + d_{N+p}^0 d_q^0$
        $P_{1,t} = \textsc{family}(X_t)\alpha\Gamma(n_t)/d_t$, $P_{2,t} = ml_{N+p} ml_q (d_{N+p}^0 d_q^0 / d_t)$
        $logodds_t = \log P_{1,t} - \log P_{2,t}$
        $\mathbb{P} = \mathbb{P} \cup \{t\}$, $t = t+1$
    $\mathbb{P} = \mathbb{P} \setminus \{r : c_{1,r} \in rm \vee c_{2,r} \in rm\}$
    $\mathbb{S} = \mathbb{S} \cup \{N+p\}$, $p = p+1$
return Z

9 References

Heller, K. A. and Z. Ghahramani (2005). Bayesian Hierarchical Clustering. In *Proceedings of the 22Nd International Conference on Machine Learning*, ICML '05, New York, NY, USA, pp. 297–304. ACM.