Predicting Average Medical Payment using Physician Referral Network at the Hospital Service Area Level

Song Wang, Ruosi Guo, Daniel Ricci

Nov 10th, 2015

Background

- standard SC = eigen-decomposition + k-means.
- In many cases, the people doesn't belong to a single community. Instead they belong to multiple ones and may have the characteristics of multiple communities.
- SC is a good algorithm, can we adjust it to tackle this issue?

Basic Idea

- Levnina in Michigan has some work on this. I am going to use a model similar to their model called overlapping continuous community assignment model (OCCAM)

$$\mathcal{A} := \mathbb{E}A,\ \mathcal{A} = \alpha_n \Theta Z B Z^T \Theta^T$$

- Its eigenvectors: $\mathcal{X} = \Theta Z (Z^T \Theta^2 Z)^{-1/2} U$, U is a orthonormal matrix. i.e. $\mathcal{X}_i = \theta_i Z_i V$ where $V = (Z^T \Theta^2 Z)^{-1/2} U$. $\mathcal{X}_i^* = \|Z_i V\|^{-1} Z_i V$. Especially, if node i is a pure node, $\mathcal{X}_i = V_{Z_i}$ while the those mixed nodes are in the cone spanned by these pure nodes.

Algorithm:(Given adjacency matrix A, number of clusters K, parameters γ)

- S1 Calculate eigenvectors X of A or $L = D^{-1/2}AD^{-1/2}$, $X \in \mathbb{R}^{n \times K}$;
- S2 Normalize each row in X. $X_i^* = \frac{1}{\|X_i\| + \tau_n} X_i$.
- S3 Use random project to select pure nodes; (Need to determine η_n
  - for $i = 1, 2, .., N = 10,000$:
  - generate $w \in \mathbb{R}^{K \times 1}$, $y = Xw$, calculate the maximum denoted as Max, similar Min, select the indices producing values $\geq (1 - \eta_n)Max$ or $\leq (1 + \eta_n)Min$ if $Min > 0$ otherwise $\leq (1 - \eta_n)Min$
  - combined all the indices, select the indices which appears $\geq N * \gamma$, and their corresponding row vectors in X, denoted as X_p
- S4 run k-kmeans on X_p, get centers $X_c \in \mathbb{R}^{K \times K}$, membership matrix $\hat{Z}_p$
- S5 calculate $\hat{Z}_m = (X \backslash X_p)(X_c)^{-1}$
- S6 return $\hat{Z} = [\hat{Z}_p, \hat{Z}_m]$.

Method 2

Theorem (Bounds on individual rows)

If $\mathcal{A}$ satisfies the following assumptions:
Assumption 1 (A1) : $\frac{m_0}{n} \leq \|\mathcal{X}_i\|_2^2 \leq \frac{M}{n}$, $1 \leq i \leq K$.
Assumption 2 (A2) : $\frac{2\sqrt{2K}\|A-\mathcal{A}\|}{\lambda(\mathcal{A})} \to 0, n \to 0.$
We have that X and $\mathcal{X}$ can be well-bounded. i.e. for any $\epsilon > 0$, except a degenerate index set T,

$$\|X_i^* - \mathcal{X}_i^*\| \leq \frac{\epsilon}{1-\epsilon}, \; i \notin T$$

.

Proof:

Let

$$T := \{i \in \{1, 2, .., n\} : \|X_i - \mathcal{X}_i\| \geq \epsilon \|\mathcal{X}_i\|_2\}$$

$$\begin{aligned} \left(\frac{2\sqrt{2K}\|A - \mathcal{A}\|}{\lambda_K(\mathcal{A})}\right)^2 \geq & \quad \|X - \mathcal{X}\|_F^2 \\ = & \quad \sum_{i=1}^{n} \|X_i - \mathcal{X}_i\|_2^2 \\ \geq & \quad \sum_{i \in T} \|X_i - \mathcal{X}_i\|_2^2 \\ \geq & \quad \sum_{i \in T} \epsilon^2 \|\mathcal{X}_i\|_2^2 \\ \geq & \quad |T| \epsilon^2 m_0^2 / n \end{aligned} \tag{1}$$

Therefore, under (A1 -A2),

$$\delta_n = \frac{|T|}{n} \leq \frac{8K\|A - \mathcal{A}\|^2}{\lambda_K(\mathcal{A})^2 \cdot \epsilon^2 \cdot m_0^2} \to 0$$

For $i \in T^c$,

$$\begin{aligned} \|X_i^* - \mathcal{X}_i^*\| \leq & \quad \tfrac{\|X_i - \mathcal{X}_i O\|}{\min\{\|X_i\|, \|\mathcal{X}_i\|\}} \\ \leq & \quad \tfrac{\epsilon \|\mathcal{X}_i\|}{(1-\epsilon)\|\mathcal{X}_i\|} = \tfrac{\epsilon}{1-\epsilon} \end{aligned} \tag{2}$$

This inequality indicates, those sample rows are located within a ball, with center at common population row and a radius $\frac{\epsilon}{1-\epsilon}$.

Theorem (Random projection on the $\mathcal{X}$)

For $\mathcal{X}^*$ defined as above, adopt the following projection rule: For every projection, we pick the the indices producing the **biggest and smallest** *values under each projection.*
When the number of random projections ,N, is big enough, we can distinguish the endpoint and points inside the cone. When N is large enough, we have that :

$$\frac{\textit{times selecting node } i}{N} \rightarrow \begin{cases} 1/(1+n/K), & \textit{if } i \textit{ is a mixed node} \\ n/K > 0, & \textit{if } i \textit{ is a pure node} \end{cases}$$

Proof.

Every projection will lands on some endpoints, which contains a a lot of index. The frequency will evenly distributed among mixed nodes, while concentrating on the pure nodes. □

Theorem

For sample eigenvectors X^* defined as above, adopt the following projection rule: For every projection, look at its maximum Max and minimum Min, and keep those indices producing values $\geq (1-\eta_n)Max$ or $\leq (1+\eta_n)Min$, for $\eta_n > 0$.($(1-\eta_n)Min$ if $Min < 0$). When the number of random projections is big enough, we can distinguish the endpoint and points inside the cone using the above algorithm. When $N > f(\eta_n, \epsilon, etc.)$, we have that:

$$\frac{\textit{number of mixed points selected}}{\textit{number selected points}} = O(1/n).$$

Remark 1: We can consider this as a weighted sampling idea. through increasing random projections, we put more and more weight on the pure nodes. **Remark 2:**

K-means on selected pure points

How far are $\|\mathcal{X}_i^*\|$ and $\|\mathcal{X}_j^*\|$ away from each other?

Lemma

Assuming $Z_i \sim Dirchilet(\alpha)$, and

$$EZ_iZ_i^T = \frac{1}{\alpha_0(\alpha_0+1)} diag(\alpha_1, ..., \alpha_K) + uu^T$$

We can simplify a little bit :

$$\begin{aligned} \mathcal{V}^*(\mathcal{V}^*)^T &= \quad normalized\ Of\ (Z^T\Theta^2 Z)^{-1} \\ &= \quad Diag^{-1} - \frac{Diag^{-1}uu^TDiag^{-1}}{1+u^TDiag^{-1}u} \end{aligned}$$

New Ideas on detecting pure nodes

- Karl suggested that for X^*, by looking at the how variables are the points in the neighborhood of each points, we have a sense of which is potentially a pure points – the one with smaller variance.
- Some analogy between the two ideas, both assuming in the sample, there are more points centering around the pure nodes. Random project has the potentials to fewer computation when the geometric structure is clear.