

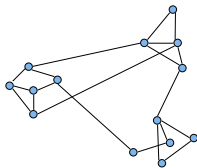
スペクトラルクラスタリング

茂木 哲矢

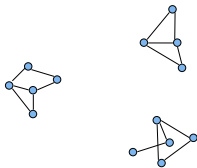
2008年6月24日

スペクトラルクラスタリングとは

- ▶ クラスタリングをグラフの分割問題にする .
- ▶ グラフの表現方法



- ▶ 評価関数を用いて切り取るエッジを決める .



カットのための評価関数

評価関数の最適解は固有値問題の解に対応する。

扱う評価関数，

- ▶ Mcut
- ▶ Ncut

サブグラフ A と B の類似度 $cut(A, B)$

$$cut(A, B) = W(A, B) = \sum_{a \in A, b \in B} sim(a, b)$$

また，

$$W(A) = W(A, A)$$

Mcut

評価関数

$$Mcut = \frac{cut(A, B)}{W(A)} + \frac{cut(A, B)}{W(B)}$$

が小さいほどよいクラスタリング。

求め方

- ▶ $I - D^{-1/2}WD^{-1/2}$ の固有値問題を解く。
- ▶ 2番目に小さい固有値に対する固有ベクトル¹ u_2 を得る。
- ▶ $\hat{q} = D^{-1/2}u_2$ を求める。
- ▶ \hat{q} をソートしある値以上を A , それより小さい値をを B にインデックスでクラスタリングする。

W はデータ間の類似度 , $D = diag(We)$ ($diag$ は対角要素の行列 , $e = (1, 1, \dots, 1)^t$)

¹Fielder ベクトルって言うらしいよ。

Ncut

評価関数

$$Ncut = \frac{cut(A, B)}{d_A} + \frac{cut(A, B)}{d_B}$$

$$d_X = \sum_{i \in X} d_i$$

求め方

- ▶ $I - D^{-1/2}WD^{-1/2}$ の固有値問題を解く .
- ▶ Fielder ベクトルを求める .

→ 求め方は Mcut と同じ .

→ 結果も Mcut と同じになる .

数値例 1/2

W, D から Fielder ベクトルを求める .

$$W = \begin{pmatrix} 10.0000 & 0.7071 & 0.3333 & 0.2774 & 0.3714 \\ 0.7071 & 10.0000 & 0.4472 & 0.2774 & 0.2857 \\ 0.3333 & 0.4472 & 10.0000 & 0.5000 & 0.3124 \\ 0.2774 & 0.2774 & 0.5000 & 10.0000 & 0.4851 \\ 0.3714 & 0.2857 & 0.3124 & 0.4851 & 10.0000 \end{pmatrix}$$

$$D = \begin{pmatrix} 11.6892 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 11.7174 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 11.5929 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 11.5399 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 11.4546 \end{pmatrix}$$

数値例 2/2

fileder ベクトルは ,

$$z_2 = (0.5028, 0.5252, -0.0748, -0.5202, -0.4417)^t$$

これより \hat{q} を求めソートして ,,

$$\text{sort}(\hat{q}) = (0.1534, 0.1471, -0.0220, -0.1305, -0.1531)$$

クラスタ A に 2,1 , クラスタ B に 3,5,4 と分けた場合が Mcut · Ncut とともに評価関数が最小になる (Mcut=0.1427, Ncut=0.1329) .

→ クラスタ A に 2,1 , クラスタ B に 3,5,4 と分けた場合が最良のクラスタリング結果である .