

クラスタリング手法の評価

AF16009 池辺 颯一

SIT

2018 年 12 月 12 日

- 情報化社会の発展によりデータ量が膨大になりつつある
- それらのデータを自動的に分類するクラスタリングに着目

目的

目的



目標

- 各クラスタリング手法について比較及び評価を行う

実験方法

実験評価方法

クラスタリング手法

eFCMA

$$\underset{u, v, \pi}{\text{minimize}} \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \|x_k - v_i\|_2^2 + \lambda^{-1} \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \log\left(\frac{u_{i,k}}{\pi_i}\right)$$

$$d_{i,k} = \|x_k - v_i\|_2^2,$$

$$u_{i,k} = \frac{\pi_i \exp(-\lambda \|x_k - v_i\|_2^2)}{\sum_{j=1}^C \pi_j \exp(-\lambda \|x_k - v_j\|_2^2)},$$

$$v_i = \frac{\sum_{k=1}^N u_{i,k} x_k}{\sum_{k=1}^N u_{i,k}},$$

クラスタリング手法

qFCMA

$$\underset{u, v, \alpha}{\text{minimize}} \sum_{i=1}^C \sum_{k=1}^N (\alpha_i)^{1-m} (u_{i,k})^m \|x_k - v_i\|_2^2 + \frac{\lambda^{-1}}{m-1} \sum_{i=1}^C \sum_{k=1}^N (\alpha_i)^{1-m} (u_{i,k})^m$$

$$d_{i,k} = \|x_k - v_i\|_2^2,$$

$$u_{i,k} = \frac{\alpha_i (1 + \lambda(1-m) \|x_i - v_k\|_2^2)^{\frac{1}{1-m}}}{\sum_{j=1}^C \alpha_j (1 + \lambda(1-m) \|x_j - v_k\|_2^2)^{\frac{1}{1-m}}},$$

$$v_i = \frac{\sum_{k=1}^N (u_{i,k})^m x_k}{\sum_{k=1}^N (u_{i,k})^m},$$

クラスタリング手法

sfcma

$$\begin{aligned} & \underset{u, v, \alpha}{\text{minimize}} \quad \sum_{i=1}^c \sum_{k=1}^n (\alpha_i)^{1-m} (u_{i,k})^m \|x_k - v_i\|_2^2 \\ & \text{subject to} \quad \sum_{i=1}^c u_{i,k} = 1, \sum_{i=1}^c \alpha_i = 1 \text{ and } u_{i,k} \in [0, 1] \quad m > 1 \end{aligned}$$

$$d_{i,k} = \|x_k - v_i\|_2^2 = \left(\sqrt{\sum_{\ell=1}^m (x_{k,\ell} - v_{i,\ell})^2} \right)^2,$$
$$u_{i,k} = \frac{1}{\sum_{j=1}^c \frac{\alpha_j}{\alpha_i} \left(\frac{d_{j,k}}{d_{i,k}} \right)^{\frac{1}{1-m}}}, \quad v_i = \frac{\sum_{k=1}^n (u_{i,k})^m x_k}{\sum_{k=1}^n (u_{i,k})^m},$$

Yeast Data Set

実験結果

考察



まとめ

