Stat 210B Homework1

Weitong Liang

1 BIRCH

, Wordle , , , , , $R = (r_1, \ldots, r_7)$, r_i i (i=7). R W , R W . $BIRCH$. BIRCH , CF-tree(Clustering Feature Tree), . clustering feature:

$$CF = (N, \overrightarrow{LS}, SS)$$

N , $\overrightarrow{LS}$, $R_i, i \in [N]$,

$$\overrightarrow{LS} = \sum_{i=1}^{N} R_i = (\sum_{i=1}^{N} r_{i1}, \ldots, \sum_{i=1}^{N} r_{i7}).$$

SS , :

$$SS = \sum_{i=1}^{N} R_i^T R_i.$$

C_1, C_2, clustering feature CF_1, CF_2. , clustering feature :

$$CF_1 + CF_2 = (N_1 + N_2, \overrightarrow{LS_1} + \overrightarrow{LS_2}, SS_1 + SS_2)$$

, $d,$:

$$d = \sqrt{\frac{\sum_{m=1}^{N_1} \sum_{n=1}^{N_2} (r_{1m} - r_{2n})^2}{N_1 N_2}}.$$

, D

$$\sqrt{\frac{2NSS - 2LS^T LS}{N(N-1)}}$$

,CF-tree , , CF-tree , . $c1$ $c2$ CF , $c3$ CF . CF-tree . CF-tree . R_1 . R_2, R_1 R_2 $c3$. , R_1, R_2 CF, R_2 CF CF. , . , $c1$. R_i , R_j , . c_2 . , CF-tree .