3 multiple alignment

Generel

We want to find similar parts of multiple strings. ~~These m~~ Similar strings might be evolutionary or functional related even though much of them differ.

Sum of Pairs

We need to score our alignment using some score.

$$M = \begin{bmatrix} \text{---} m_1 \text{---} \\ \text{---} m_2 \text{---} \\ \vdots \\ \text{---} m_k \text{---} \end{bmatrix}$$

$$SP(M) = \sum_{0 < i < j \leq k} PairScore(m_i, m_j)$$

Can we use ideas from global alignment - Ie is SP-score column-based?

$$SP(M) = \sum_{0 < i < j \leq k} PairScore(M_i, M_j)$$

$$= \sum_{0 < i < j \leq k} \sum_{0 < c \leq l} d(M_i[c], M_j[c])$$

$$= \sum_{0 < c \leq l} \sum_{0 < i < j \leq k} d(M_i[c], M_j[c])$$

$$= \sum_{0 < c < l} SP\begin{bmatrix} M_1[c] \\ \vdots \\ M_k[c] \end{bmatrix}$$

Yes it is.

Exact

Definer $D(i_1, \dots, i_k) \equiv$ Cost of $OPT(S_1[1,\dots,i_1], \dots, S_K[1,\dots,i_k])$

Compute D by maximizing over all lastcolumns.

$\#$last columns $= 2^k - 1$

$$D(i_1, \dots, i_k) = \min_{\text{last column}} \left[D(i_1', \dots, i_k') + \text{cost(last column)} \right] \quad \begin{bmatrix} S_1[i_1] \text{ or } - \\ \vdots \\ S_K[i_k] \text{ or } - \end{bmatrix}$$

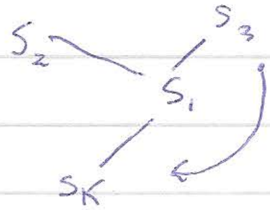
Time $= O(n^k \cdot 2^k \cdot$ "time to compute score"$)$

Space $= O(n^k)$

Star algorithm

1) Compute $S_1 = \min \sum_{S \in F - S_1} D(S_1, S)$

$O(K^2 n^2)$



2) Optalign S_1, S_i = Extend M

$O(kn^2)$ Show how to extend

Total $\underline{O(K^2 n^2)}$

Showing $\frac{SP(M)}{SP(M^*)} < 2$

$$SP(M) = \frac{1}{2} \sum_{i=1}^{k} \sum_{j=1 \neq i}^{k} d(M_i, M_j)$$

metric triangle inequality

$$\leq \frac{1}{2} \sum_{i=1}^{k} \sum_{j=1 \neq i}^{k} d(M_i, M_1) + d(M_1, M_j)$$

metric symetry

$$= \frac{1}{2} \sum_{i=1}^{k} \sum_{j=1 \neq i}^{k} d(M_1, M_i) + d(M_1, M_j)$$

rewrite

$$= \frac{1}{2} \sum_{l=2}^{k} 2(k-1)\, d(M_1, M_l)$$

$$= (k-1) \sum_{l=2}^{k} d(M_1, M_l)$$

$$= (k-1) \sum_{l=2}^{k} D(S_1, S_l)$$

$$SP(M^*) = \frac{1}{2} \sum_{i=1}^{k} \sum_{j=1 \neq i}^{k} d(M^*_i, M^*_j)$$

nothing is better than the optimal scores

$$\geq \frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{k} D(S_i, S_j)$$

by choice of S_1

$$\geq \frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{k} D(S_1, S_j)$$

$$= \frac{1}{2} k \sum_{j=1}^{k} D(S_1, S_j)$$

rewrite and $D(S_1, S_1) = 0$

$$= \frac{1}{2} k \sum_{l=2}^{k} D(S_1, S_l)$$

$$\frac{SP(M)}{SP(M^*)} \leq \frac{(k-1) \sum_{l=2}^{k} D(S_1, S_l)}{\frac{1}{2} k \sum_{l=2}^{k} D(S_1, S_l)}$$

$$= \frac{2(k-1)}{k}$$

$$< 2$$