Multiple String Comparison

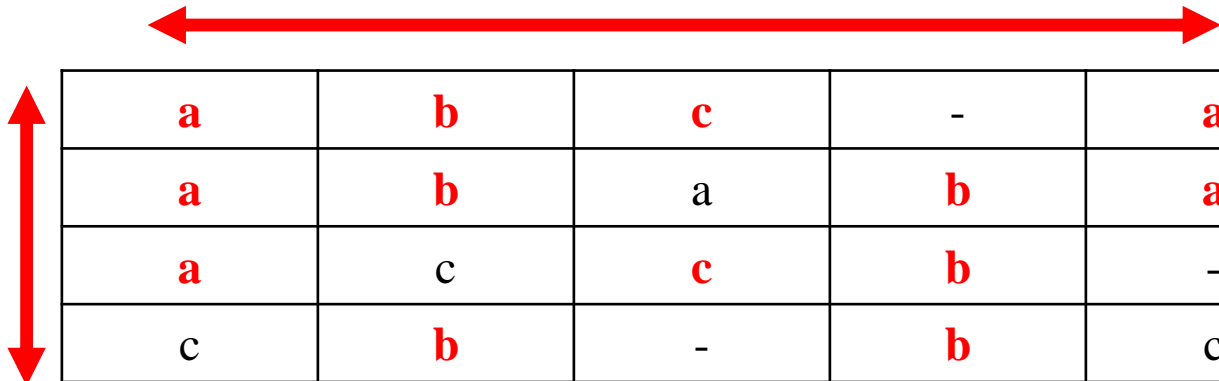이강배

Contents

Multiple String Comparision

- A **multiple sequence alignment** (**MSA**)
    - a sequence alignment of three or more biological sequences, generally protein, DNA, or RNA.

- From the resulting MSA,
    - **sequence homology** can be inferred
    - **phylogenetic analysis** can be conducted to assess the sequences' shared evolutionary origins.

Why multiple string comparison

- **Central technique for multiple string comparison involves _multiple alignment_**

  - Ex) $S_1 =$ abca, $S_2 =$ ababa, $S_3 =$ accb, $S_4 =$ cbbc, $k = 4$, $l = 5$

5 columns

4 rows

a	b	c	-	a
a	b	a	b	a
a	c	c	b	-
c	b	-	b	c

Family representations and alignments with profiles

- **Three common representations**

  1) ***profile* representations**

  2) *consensus* sequence representations

  3) s*ignature* representations

Family representations and alignments with profiles

- **Aligning a string to a profile**
  - How well new string S, or some substring of S fits the Profile $\mathcal{P}$?
    - It is naturally formalized as an easy generalization of **pure string alignment**

  - Consider a string C of profile column positions,
    - Align S to C by inserting spaces into S and C etc.

  - Ex) S = aabbc, profile $\mathcal{P}$ = abcba

a	b	c	-	a
a	b	a	b	a
a	c	c	b	-
c	b	-	b	c

a	a	b	b	c

a	-	b	c	-	a
a	-	b	a	b	a
a	-	c	c	b	-
c	-	b	-	b	c
a	a	b	-	b	c

Family representations and alignments with profiles

- **Aligning a string to a profile**
  - How well new string S, or some substring of S fits the Profile $\mathcal{P}$?
    - It is naturally formalized as an easy generalization of **pure string alignment**

  - Consider a string C of profile column positions,
    - Align S to C by inserting spaces into S and C etc.

  - Ex) S = aabbc, profile $\mathcal{P}$ = abcba

	C1	C2	C3	C4	C5
a	**0.75**	0	0.25	0	**0.50**
b	0	**0.75**	0	**0.75**	0
c	0.25	0.25	**0.50**	0	0.25
-	0	0	0.25	0.25	0.25

a	a	b		b	c
1		**2**	**3**	**4**	**5**

Family representations and alignments with profiles

- **How to score a string/profile alignment**
  - Ex) S = aabbc, profile $\mathcal{P}$,

the alphabet-weight scheme

	a	b	c	-
a	**2**	-1	-3	-1

Align string to profile

a	a	b		b	c
1		2	3	4	5

profile

	C1	C2	C3	C4	C5
a	**0.75**	0	0.25	0	**0.50**
b	0	**0.75**	0	**0.75**	0
c	0.25	0.25	**0.50**	0	0.25
-	0	0	0.25	0.25	0.25

MSA

a	-	b	c	-	a
a	-	b	a	b	a
a	-	c	c	b	-
c	-	b	-	b	c
a	a	b	-	b	c

- **C1 score** : $0.75 \times 2 - 0.25 \times 3 = 0.75$

Family representations and alignments with profiles

- **How to score a string/profile alignment**
  - Ex) S = aabbc, profile $\mathcal{P}$,

the alphabet-weight scheme

	a	b	c	**-**
a	2	-1	-3	**-1**

Align string to profile

a	**a**	b		b	c
1		2	3	4	5

profile

	C1	*C2*	*C3*	*C4*	*C5*
a	**0.75**	0	0.25	0	**0.50**
b	0	**0.75**	0	**0.75**	0
c	0.25	0.25	**0.50**	0	0.25
-	0	0	0.25	0.25	0.25

MSA

a	-	b	c	-	a
a	-	b	a	b	a
a	-	c	c	b	-
c	-	b	-	b	c
a	a	b	-	b	c

- **$C2$ score** : -1

Family representations and alignments with profiles

- **How to optimally align a string to a profile**
  - The recurrences for computing optimal string to profile alignment

  - Base condition
    - $V(0, j) = \sum_{k \leq j} S(\_, k)$
    - $V(i, 0) = \sum_{k \leq i} s(S_1(k), \_)$

  - $V(i,j) = \max[V(i\text{-}1,j\text{-}1)+S(S_1(i),j),\ V(i\text{-}1,j)+s(S_1(i),\_),\ V(i,j\text{-}1)+S(\_,j)]$

Family representations and alignments with profiles

- **How to optimally align a string to a profile**
  - $V(i, 0) = \sum_{k \leq i} s(S_1(k), \_)$

V(i,j)		**0**	**1**	**2**	**3**	**4**	**5**
			a	a	b	b	c
0		0	-1	-2	-3	-4	-6
1	C1	-1.25					
2	C2	-2.5					
3	C3	-3.75					
4	C4	-4.5					
5	C5	-5.5					

Profile matrix

	C1	C2	C3	C4	C5
a	**0.75**	0	0.25	0	**0.50**
b	0	**0.75**	0	**0.75**	0
c	0.25	0.25	**0.50**	0	0.25
-	0	0	0.25	0.25	0.25

Scoring matrix

	a	b	c	-
a	2	-1	-3	-1
b		4	0	-1
c			1	-2
-				0

Family representations and alignments with profiles

- **How to optimally align a string to a profile**
  - $V(i,j) = \max[V(i-1,j-1)+S(S_1(i),j), V(i-1,j)+s(S_1(i),\_), V(i,j-1)+S(\_,j)]$

$V(i,j)$		**0**	**1**	**2**	**3**	**4**	**5**
			a	a	b	b	c
0		0	-1	-2	-3	-4	-6
1	C1	-1.25					
2	C2	-2.5					
3	C3	-3.75					
4	C4	-4.5					
5	C5	-5.5					

Profile matrix

	C1	C2	C3	C4	C5
a	**0.75**	0	0.25	0	**0.50**
b	0	**0.75**	0	**0.75**	0
c	0.25	0.25	**0.50**	0	0.25
-	0	0	0.25	0.25	0.25

Scoring matrix

	a	b	c	-
a	2	-1	-3	-1
b		4	0	-1
c			1	-2
-				0

- $\max[V(i-1,j-1)+S(S_1(i),j), V(i-1,j)+s(S_1(i),\_), V(i,j-1)+S(\_,j)]$
- $= 0 + (s(\text{a,a}) \times p(\text{a},1)) + (s(\text{a,c}) \times p(\text{c},1))$
- $= 0 + (2 \times 0.75) + (-3 \times 0.25) \quad = \textbf{0.75}$

Family representations and alignments with profiles

- **How to optimally align a string to a profile**
  - $V(i,j) = \max[V(i-1,j-1)+S(S_1(i),j), V(i-1,j)+s(S_1(i),\_), V(i,j-1)+S(\_,j)]$

$V(i,j)$		**0**	**1**	**2**	**3**	**4**	**5**
			a	a	b	b	c
0		0	-1	-2	-3	-4	-6
1	C1	-1.25 →					
2	C2	-2.5					
3	C3	-3.75					
4	C4	-4.5					
5	C5	-5.5					

Profile matrix

	C1	C2	C3	C4	C5
a	**0.75**	0	0.25	0	**0.50**
b	0	**0.75**	0	**0.75**	0
c	0.25	0.25	**0.50**	0	0.25
-	0	0	0.25	0.25	0.25

Scoring matrix

	a	b	c	-
a	2	-1	-3	-1
b		4	0	-1
c			1	-2
-				0

- $\max[V(i-1,j-1)+S(S_1(i),j), \boldsymbol{V(i-1,j)+s(S_1(i),\_)}, V(i,j-1)+S(\_,j)]$

$= -1.25 + s(a,\_)$

$= -1.25 + (-1)$ $\qquad = \textbf{-2.25}$

Family representations and alignments with profiles

- **How to optimally align a string to a profile**

  - $V(i,j) = \max[V(i-1,j-1)+S(S_1(i),j),\ V(i-1,j)+s(S_1(i),\_),\ V(i,j-1)+S(\_,j)]$

$V(i,j)$		**0**	**1**	**2**	**3**	**4**	**5**
			a	a	b	b	c
0		0	-1	-2	-3	-4	-6
1	C1	-1.25					
2	C2	-2.5					
3	C3	-3.75					
4	C4	-4.5					
5	C5	-5.5					

Profile matrix

	C1	C2	C3	C4	C5
a	**0.75**	0	0.25	0	**0.50**
b	0	**0.75**	0	**0.75**	0
c	0.25	0.25	**0.50**	0	0.25
-	0	0	0.25	0.25	0.25

Scoring matrix

	a	b	c	-
a	2	-1	-3	-1
b		4	0	-1
c			1	-2
-				0

- $\text{Max}[V(i-1,j-1)+S(S_1(i),j),\ V(i-1,j)+s(S_1(i),\_),\ \boldsymbol{V(i,j-1)+S(\_,j)}]$

= -1 + ($s(\_,a) \times p(a,1)$) + ($s(\_,c) \times p(c,1)$)

= -1 + (-1 × 0.75) + (-2 × 0.25)   = **-2.25**

Family representations and alignments with profiles

- **How to optimally align a string to a profile**

  - $V(i,j) = \max[V(i-1,j-1)+S(S_1(i),j), V(i-1,j)+s(S_1(i),\_), V(i,j-1)+S(\_,j)]$

$V(i,j)$		**0**	**1**	**2**	**3**	**4**	**5**
			a	a	b	b	c
0		0	-1	-2	-3	-4	-6
1	C1	-1.25					
2	C2	-2.5					
3	C3	-3.75					
4	C4	-4.5					
5	C5	-5.5					

↘ **1.5**
→ **-2.25**
↓ **-2.25**

Profile matrix

	$C1$	$C2$	$C3$	$C4$	$C5$
a	**0.75**	0	0.25	0	**0.50**
b	0	**0.75**	0	**0.75**	0
c	0.25	0.25	**0.50**	0	0.25
-	0	0	0.25	0.25	0.25

Scoring matrix

	a	b	c	-
a	2	-1	-3	-1
b		4	0	-1
c			1	-2
-				0

- $\max[V(i-1,j-1)+S(S_1(i),j), V(i-1,j)+s(S_1(i),\_), V(i,j-1)+S(\_,j)]$

= $\max[\textbf{1.5}, -2.25, -2.25]$

Family representations and alignments with profiles

- **How to optimally align a string to a profile**

  - $V(i,j) = \max[V(i-1,j-1)+S(S_1(i),j), V(i-1,j)+s(S_1(i),\_), V(i,j-1)+S(\_,j)]$

$V(i,j)$		**0**	**1**	**2**	**3**	**4**	**5**
			a	a	b	b	c
0		0	-1	-2	-3	-4	-6
1	$C1$	-1.25	**1.5**				
2	$C2$	-2.5					
3	$C3$	-3.75					
4	$C4$	-4.5					
5	$C5$	-5.5					

Profile matrix

	$C1$	$C2$	$C3$	$C4$	$C5$
a	**0.75**	0	0.25	0	**0.50**
b	0	**0.75**	0	**0.75**	0
c	0.25	0.25	**0.50**	0	0.25
-	0	0	0.25	0.25	0.25

Scoring matrix

	a	b	c	-
a	2	-1	-3	-1
b		4	0	-1
c			1	-2
-				0

- $\max[V(i-1,j-1)+S(S_1(i),j), V(i-1,j)+s(S_1(i),\_), V(i,j-1)+S(\_,j)]$

= $\max[\textbf{1.5}, -2.25, -2.25]$

Family representations and alignments with profiles

- **How to optimally align a string to a profile**
  - $V(i,j) = \max[V(i-1,j-1)+S(S_1(i),j), V(i-1,j)+s(S_1(i),\_), V(i,j-1)+S(\_,j)]$

$V(i,j)$		**0**	**1**	**2**	**3**	**4**	**5**
			a	a	b	b	c
0		0	-1	-2	-3	-4	-6
1	C1	-1.25	0.75	-0.25	-1.25	-2.25	-4.25
2	C2	-2.5	-0.5	-0.75	2.75	1.75	-0.25
3	C3	-3.75	-1.75	-1.75	1.5	2.25	1
4	C4	-4.5	-2.5	-2.5	1	4.25	2.25
5	C5	-5.5	-3.5	-2.5	0	3.25	**2.5**
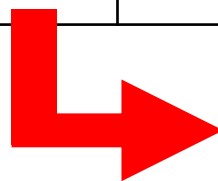
a	**a**	**b**		**b**	**c**
1		**2**	**3**	**4**	**5**

Family representations and alignments with profiles

- **How to optimally align a profile to a profile**

Profile 1 MSA

a	b	c	-	a
a	b	a	b	a
a	c	c	b	-
c	b	-	b	c

Profile 1 matrix

	C1	C2	C3	C4	C5
a	**0.75**	0	0.25	0	**0.50**
b	0	**0.75**	0	**0.75**	0
c	0.25	0.25	**0.50**	0	0.25
-	0	0	0.25	0.25	0.25

Profile 2 MSA

a	b	a	-	a
a	b	a	b	a
a	-	a	b	-
b	b	-	b	c

Profile 2 matrix

	C1	C2	C3	C4	C5
a	**0.75**	0	**0.75**	0	**0.50**
b	0.25	**0.75**	0	**0.75**	0
c	0	0	0	0	0.25
-	0	0.25	0.25	0.25	0.25

Family representations and alignments with profiles

- **How to optimally align a profile to a profile**

  - Base condition
    - $V(0, j) = \sum_{k \leq j} S(\_, k)$
    - $\cancel{V(i, 0) = \sum_{k \leq i} s(S_1(k), \_)}$
    - → $V(i, 0) = \sum_{k \leq i} \textcolor{red}{S(\boldsymbol{k}, \_)}$

  - $V(i,j) = \cancel{\max[V(i\text{-}1,j\text{-}1)+S(S_1(i),j), V(i\text{-}1,j)+s(S_1(i),\_), V(i,j\text{-}1)+S(\_,j)]}$

    → $\max[V(i\text{-}1,j\text{-}1)+\textcolor{red}{S(i,j)}, V(i\text{-}1,j)+\textcolor{red}{S(i,\_)}, V(i,j\text{-}1)+S(\_,j)]$

Family representations and alignments with profiles

- **How to optimally align a profile to a profile**

  - $V(0, j) = \sum_{k \le j} S(\_, k)$

V(i,j)		**0**	**1**	**2**	**3**	**4**	**5**
			$C1_{p2}$	$C2_{p2}$	$C3_{p2}$	$C4_{p2}$	$C5_{p2}$
0		0					
1	$C1_{p1}$	-1.25					
2	$C2_{p1}$	-2.5					
3	$C3_{p1}$	-3.75					
4	$C4_{p1}$	-4.5					
5	$C5_{p1}$	-5.5					

Profile 1 matrix

	C1	*C2*	*C3*	*C4*	*C5*
a	**0.75**	0	0.25	0	**0.50**
b	0	**0.75**	0	**0.75**	0
c	0.25	0.25	**0.50**	0	0.25
-	0	0	0.25	0.25	0.25

Profile 2 matrix

	C1	*C2*	*C3*	*C4*	*C5*
a	**0.75**	0	**0.75**	0	**0.50**
b	0.25	**0.75**	0	**0.75**	0
c	0	0	0	0	0.25
-	0	0.25	0.25	0.25	0.25

Scoring matrix

	a	b	c	-
a	2	-1	-3	-1
b		4	0	-1
c			1	-2
-				0

Family representations and alignments with profiles

- **How to optimally align a profile to a profile**

  - $V(i, 0) = \sum_{k \leq i} S(k, \_)$

V(i,j)		**0**	**1**	**2**	**3**	**4**	**5**
			$C1_{p2}$	$C2_{p2}$	$C3_{p2}$	$C4_{p2}$	$C5_{p2}$
0		0	-1				
1	$C1_{p1}$	-1.25					
2	$C2_{p1}$	-2.5					
3	$C3_{p1}$	-3.75					
4	$C4_{p1}$	-4.5					
5	$C5_{p1}$	-5.5					

- $\sum_{k \leq i} [s(y, \_) \times p(y, k)]$

$= s(\_,a) \times p(a,1) + s(\_,b) \times p(b,1)$

$= (-1 \times 0.75) + (-1 \times 0.25)$

Profile 1 matrix

	C1	**C2**	**C3**	**C4**	**C5**
a	**0.75**	0	0.25	0	**0.50**
b	0	**0.75**	0	**0.75**	0
c	0.25	0.25	**0.50**	0	0.25
-	0	0	0.25	0.25	0.25

Profile 2 matrix

	C1	**C2**	**C3**	**C4**	**C5**
a	**0.75**	0	**0.75**	0	**0.50**
b	0.25	**0.75**	0	**0.75**	0
c	0	0	0	0	0.25
-	0	0.25	0.25	0.25	0.25

Scoring matrix

	a	b	c	-
a	2	-1	-3	-1
b		4	0	-1
c			1	-2
-				0

Family representations and alignments with profiles

- **How to optimally align a profile to a profile**
  - $V(i, 0) = \sum_{k \leq i} S(k, \_)$

V(i,j)		**0**	**1**	**2**	**3**	**4**	**5**
			$C1_{p2}$	$C2_{p2}$	$C3_{p2}$	$C4_{p2}$	$C5_{p2}$
0		0	-1	-1.75			
1	$C1_{p1}$	-1.25					
2	$C2_{p1}$	-2.5					
3	$C3_{p1}$	-3.75					
4	$C4_{p1}$	-4.5					
5	$C5_{p1}$	-5.5					

Profile 1 matrix

	C1	*C2*	*C3*	*C4*	*C5*
a	**0.75**	0	0.25	0	**0.50**
b	0	**0.75**	0	**0.75**	0
c	0.25	0.25	**0.50**	0	0.25
-	0	0	0.25	0.25	0.25

Profile 2 matrix

	C1	*C2*	*C3*	*C4*	*C5*
a	**0.75**	0	**0.75**	0	**0.50**
b	0.25	**0.75**	0	**0.75**	0
c	0	0	0	0	0.25
-	0	0.25	0.25	0.25	0.25

Scoring matrix

	a	b	c	-
a	2	-1	-3	-1
b		4	0	-1
c			1	-2
-				0

- $\sum_{k \leq i}[s(y, \_) \times p(y, k)]$
$= [s(\_,a) \times p(a,1) + s(\_,b) \times p(b,1)] + s(\_,b) \times p(b,2) + s(\_,\_) \times p(\_,2)$
$= -1 + (-1 \times 0.75) + (0 \times 0.25)$

Family representations and alignments with profiles

- **How to optimally align a profile to a profile**

  - $V(i, 0) = \sum_{k \leq i} S(k, \_)$

V(i,j)		**0**	**1**	**2**	**3**	**4**	**5**
			$C1_{p2}$	$C2_{p2}$	$C3_{p2}$	$C4_{p2}$	$C5_{p2}$
0		0	-1	-1.75	-2.5	-3.25	-4.25
1	$C1_{p1}$	-1.25					
2	$C2_{p1}$	-2.5					
3	$C3_{p1}$	-3.75					
4	$C4_{p1}$	-4.5					
5	$C5_{p1}$	-5.5					

Profile 1 matrix

	C1	*C2*	*C3*	*C4*	*C5*
a	**0.75**	0	0.25	0	**0.50**
b	0	**0.75**	0	**0.75**	0
c	0.25	0.25	**0.50**	0	0.25
-	0	0	0.25	0.25	0.25

Profile 2 matrix

	C1	*C2*	*C3*	*C4*	*C5*
a	**0.75**	0	**0.75**	0	**0.50**
b	0.25	**0.75**	0	**0.75**	0
c	0	0	0	0	0.25
-	0	0.25	0.25	0.25	0.25

Scoring matrix

	a	b	c	-
a	2	-1	-3	-1
b		4	0	-1
c			1	-2
-				0

Family representations and alignments with profiles

- **How to optimally align a profile to a profile**
  - $V(i,j) = \max[V(i-1,j-1)+S(i,j),\ V(i-1,j)+S(i,\_),\ V(i,j-1)+S(\_,j)]$

$V(i,j)$		**0**	**1**	**2**	**3**	**4**	**5**
			$C1_{p2}$	$C2_{p2}$	$C3_{p2}$	$C4_{p2}$	$C5_{p2}$
0		0	-1	-1.75	-2.5	-3.25	-4.25
1	$C1_{p1}$	-1.25					
2	$C2_{p1}$	-2.5					
3	$C3_{p1}$	-3.75					
4	$C4_{p1}$	-4.5					
5	$C5_{p1}$	-5.5					

Profile 1 matrix

	C1	*C2*	*C3*	*C4*	*C5*
a	**0.75**	0	0.25	0	**0.50**
b	0	**0.75**	0	**0.75**	0
c	0.25	0.25	**0.50**	0	0.25
-	0	0	0.25	0.25	0.25

Profile 2 matrix

	C1	*C2*	*C3*	*C4*	*C5*
a	**0.75**	0	**0.75**	0	**0.50**
b	0.25	**0.75**	0	**0.75**	0
c	0	0	0	0	0.25
-	0	0.25	0.25	0.25	0.25

Scoring matrix

	a	b	c	-
a	2	-1	-3	-1
b		4	0	-1
c			1	-2
-				0

- $\max[\mathbf{V(i\text{-}1,j\text{-}1)+S(i,j)},\ V(i-1,j)+S(i,\_),\ V(i,j-1)+S(\_,j)]$

$= 0 + (s(a,a) \times p(a,1) \times p(a,1)) + (s(a,c) \times p(a,1) \times p(c,1))$ → $1.125 + (-0.562) = 0.563$

$+ (s(b,a) \times p(c,1) \times p(c,1)) + (s(a,c) \times p(c,1) \times p(c,1))$

Family representations and alignments with profiles

- **How to optimally align a profile to a profile**
  - $V(i,j) = \max[V(i-1,j-1)+S(i,j),\ V(i-1,j)+S(i,\_),\ V(i,j-1)+S(\_,j)]$

$V(i,j)$		**0**	**1**	**2**	**3**	**4**	**5**
			$C1_{p2}$	$C2_{p2}$	$C3_{p2}$	$C4_{p2}$	$C5_{p2}$
0		0	-1	-1.75	-2.5	-3.25	-4.25
1	$C1_{p1}$	-1.25					
2	$C2_{p1}$	-2.5					
3	$C3_{p1}$	-3.75					
4	$C4_{p1}$	-4.5					
5	$C5_{p1}$	-5.5					

Profile 1 matrix

	C1	*C2*	*C3*	*C4*	*C5*
a	**0.75**	0	0.25	0	**0.50**
b	0	**0.75**	0	**0.75**	0
c	0.25	0.25	**0.50**	0	0.25
-	0	0	0.25	0.25	0.25

Profile 2 matrix

	C1	*C2*	*C3*	*C4*	*C5*
a	0.75	0	**0.75**	0	**0.50**
b	0.25	**0.75**	0	**0.75**	0
c	0	0	0	0	0.25
-	0	0.25	0.25	0.25	0.25

Scoring matrix

	a	b	c	-
a	2	-1	-3	-1
b		4	0	-1
c			1	-2
-				0

- $\text{Max}[\mathbf{V(i-1,j-1)+S(i,j)},\ V(i-1,j)+S(i,\_),\ V(i,j-1)+S(\_,j)]$
- $= 0 + (s(a,a) \times p(a,1) \times p(a,1)) + (s(a,c) \times p(a,1) \times p(c,1))$ ➜ $1.125 + (-0.562) = 0.563$
  - $+ (s(b,a) \times p(c,1) \times p(c,1)) + (s(a,c) \times p(c,1) \times p(c,1))$

Family representations and alignments with profiles

- **How to optimally align a profile to a profile**
  - $V(i,j) = \max[V(i-1,j-1)+S(i,j),\ V(i-1,j)+S(i,\_),\ V(i,j-1)+S(\_,j)]$

Profile 1 matrix

	C1	C2	C3	C4	C5
a	0.75	0	0.25	0	0.50
b	0	0.75	0	0.75	0
c	0.25	0.25	0.50	0	0.25
-	0	0	0.25	0.25	0.25

Profile 2 matrix

	C1	C2	C3	C4	C5
a	0.75	0	0.75	0	0.50
b	0.25	0.75	0	0.75	0
c	0	0	0	0	0.25
-	0	0.25	0.25	0.25	0.25

Scoring matrix

	a	b	c	-
a	2	-1	-3	-1
b		4	0	-1
c			1	-2
-				0

$V(i,j)$		**0**	**1**	**2**	**3**	**4**	**5**
			$C1_{p2}$	$C2_{p2}$	$C3_{p2}$	$C4_{p2}$	$C5_{p2}$
0		0	-1	-1.75	-2.5	-3.25	-4.25
1	$C1_{p1}$	-1.25					
2	$C2_{p1}$	-2.5					
3	$C3_{p1}$	-3.75					
4	$C4_{p1}$	-4.5					
5	$C5_{p1}$	-5.5					

- $\max[\mathbf{V(i-1,j-1)+S(i,j)},\ V(i-1,j)+S(i,\_),\ V(i,j-1)+S(\_,j)]$

$= 0 + (s(a,a) \times p(a,1) \times p(a,1)) + (s(a,c) \times p(c,1) \times p(c,1))$ ➔ $1.125 + (-0.562) = 0.563$

$+ (s(b,a) \times p(a,1) \times p(b,1)) + (s(b,c) \times p(b,1) \times p(c,1))$ ➔ $-0.187 + 0 = -0.187$

Family representations and alignments with profiles

- **How to optimally align a profile to a profile**
  - $V(i,j) = \max[V(i-1,j-1)+S(i,j),\ V(i-1,j)+S(i,\_),\ V(i,j-1)+S(\_,j)]$

$V(i,j)$		**0**	**1**	**2**	**3**	**4**	**5**
			$C1_{p2}$	$C2_{p2}$	$C3_{p2}$	$C4_{p2}$	$C5_{p2}$
0		0	-1	-1.75	-2.5	-3.25	-4.25
1	$C1_{p1}$	-1.25					
2	$C2_{p1}$	-2.5					
3	$C3_{p1}$	-3.75					
4	$C4_{p1}$	-4.5					
5	$C5_{p1}$	-5.5					

Profile 1 matrix

	C1	*C2*	*C3*	*C4*	*C5*
a	**0.75**	0	0.25	0	**0.50**
b	0	**0.75**	0	**0.75**	0
c	0.25	0.25	**0.50**	0	0.25
-	0	0	0.25	0.25	0.25

Profile 2 matrix

	C1	*C2*	*C3*	*C4*	*C5*
a	**0.75**	0	**0.75**	0	**0.50**
b	0.25	**0.75**	0	**0.75**	0
c	0	0	0	0	0.25
-	0	0.25	0.25	0.25	0.25

Scoring matrix

	a	b	c	-
a	2	-1	-3	-1
b		4	0	-1
c			1	-2
-				0

- $\max[\mathbf{V(i-1,j-1)+S(i,j)},\ V(i-1,j)+S(i,\_),\ V(i,j-1)+S(\_,j)]$

$= 0 + (s(a,a) \times p(a,1) \times p(a,1)) + (s(a,c) \times p(c,1) \times p(c,1))$ ➜ $1.125 + (-0.562) = 0.563$

$+ (s(b,a) \times p(a,1) \times p(b,1)) + (s(b,c) \times p(b,1) \times p(c,1))$ ➜ $-0.187 + 0 = -0.187$

$= 0 + 0.563 + (-0.187) = \mathbf{0.376}$

Family representations and alignments with profiles

- **How to optimally align a profile to a profile**

  - $V(i,j) = \max[V(i-1,j-1)+S(i,j),\ V(i-1,j)+S(i,\_),\ V(i,j-1)+S(\_,j)]$

$V(i,j)$		**0**	**1**	**2**	**3**	**4**	**5**
			$C1_{p2}$	$C2_{p2}$	$C3_{p2}$	$C4_{p2}$	$C5_{p2}$
0		0	-1	-1.75	-2.5	-3.25	-4.25
1	$C1_{p1}$	-1.25 →					
2	$C2_{p1}$	-2.5					
3	$C3_{p1}$	-3.75					
4	$C4_{p1}$	-4.5					
5	$C5_{p1}$	-5.5					

Profile 1 matrix

	C1	C2	C3	C4	C5
a	**0.75**	0	0.25	0	**0.50**
b	0	**0.75**	0	**0.75**	0
c	0.25	0.25	**0.50**	0	0.25
-	0	0	0.25	0.25	0.25

Profile 2 matrix

	C1	C2	C3	C4	C5
a	0.75	0	**0.75**	0	**0.50**
b	0.25	**0.75**	0	**0.75**	0
c	0	0	0	0	0.25
-	0	0.25	0.25	0.25	0.25

Scoring matrix

	a	b	c	-
a	2	-1	-3	-1
b		4	0	-1
c			1	-2
-				0

- $\max[V(i-1,j-1)+S(i,j),\ \mathbf{V(i-1,j)+S(i,\_)},\ V(i,j-1)+S(\_,j)]$

$= \text{-1.25} + (s(a,\_) \times p(a,1)) + (s(b,\_) \times p(b,1))$

$= \text{-1.25} + (-1) + (-0.25) \qquad = \mathbf{-2.50}$

Family representations and alignments with profiles

- **How to optimally align a profile to a profile**
  - $V(i,j) = \max[V(i-1,j-1)+S(i,j),\ V(i-1,j)+S(i,\_),\ V(i,j-1)+S(\_,j)]$

$V(i,j)$		**0**	**1**	**2**	**3**	**4**	**5**
			$C1_{p2}$	$C2_{p2}$	$C3_{p2}$	$C4_{p2}$	$C5_{p2}$
0		0	-1	-1.75	-2.5	-3.25	-4.25
1	$C1_{p1}$	-1.25					
2	$C2_{p1}$	-2.5					
3	$C3_{p1}$	-3.75					
4	$C4_{p1}$	-4.5					
5	$C5_{p1}$	-5.5					

Profile 1 matrix

	C1	**C2**	**C3**	**C4**	**C5**
a	**0.75**	0	0.25	0	**0.50**
b	0	**0.75**	0	**0.75**	0
c	0.25	0.25	**0.50**	0	0.25
-	0	0	0.25	0.25	0.25

Profile 2 matrix

	C1	**C2**	**C3**	**C4**	**C5**
a	**0.75**	0	**0.75**	0	**0.50**
b	0.25	**0.75**	0	**0.75**	0
c	0	0	0	0	0.25
-	0	0.25	0.25	0.25	0.25

Scoring matrix

	a	b	c	-
a	2	-1	-3	-1
b		4	0	-1
c			1	-2
-				0

- $\max[V(i-1,j-1)+S(i,j),\ V(i-1,j)+S(i,\_),\ \mathbf{V(i,j-1)+S(\_,j)}]$

$= -1 + (s(\_,a) \times p(a,1)) + (s(\_,c) \times p(c,1))$

$= -1 + (-1 \times 0.75) + (-2 \times 0.25) \quad = \mathbf{-2.25}$

Family representations and alignments with profiles

- **How to optimally align a profile to a profile**
  - $V(i,j) = \max[V(i-1,j-1)+S(i,j),\ V(i-1,j)+S(i,\_),\ V(i,j-1)+S(\_,j)]$

V(i,j)		**0**	**1**	**2**	**3**	**4**	**5**
			$C1_{p2}$	$C2_{p2}$	$C3_{p2}$	$C4_{p2}$	$C5_{p2}$
0		0	-1	-1.75	-2.5	-3.25	-4.25
1	$C1_{p1}$	-1.25					
2	$C2_{p1}$	-2.5					
3	$C3_{p1}$	-3.75					
4	$C4_{p1}$	-4.5					
5	$C5_{p1}$	-5.5					

↘ **0.376**
→ **-2.50**
↓ **-2.25**

- $\max[V(i-1,j-1)+S(i,j),\ V(i-1,j)+S(i,\_),\ V(i,j-1)+S(\_,j)]$
- = Max[**0.376,** -2.50, -2.25]

Profile 1 matrix

	C1	*C2*	*C3*	*C4*	*C5*
a	**0.75**	0	0.25	0	**0.50**
b	0	**0.75**	0	**0.75**	0
c	0.25	0.25	**0.50**	0	0.25
-	0	0	0.25	0.25	0.25

Profile 2 matrix

	C1	*C2*	*C3*	*C4*	*C5*
a	**0.75**	0	**0.75**	0	**0.50**
b	0.25	**0.75**	0	**0.75**	0
c	0	0	0	0	0.25
-	0	0.25	0.25	0.25	0.25

Scoring matrix

	a	b	c	-
a	2	-1	-3	-1
b		4	0	-1
c			1	-2
-				0

Family representations and alignments with profiles

- **How to optimally align a profile to a profile**
  - $V(i,j) = \max[V(i-1,j-1)+S(i,j),\ V(i-1,j)+S(i,\_),\ V(i,j-1)+S(\_,j)]$

Profile 1 matrix

	C1	C2	C3	C4	C5
a	0.75	0	0.25	0	0.50
b	0	0.75	0	0.75	0
c	0.25	0.25	0.50	0	0.25
-	0	0	0.25	0.25	0.25

$V(i,j)$		0	1	2	3	4	5
			$C1_{p2}$	$C2_{p2}$	$C3_{p2}$	$C4_{p2}$	$C5_{p2}$
0		0	-1	-1.75	-2.5	-3.25	-4.25
1	$C1_{p1}$	-1.25	**0.376**				
2	$C2_{p1}$	-2.5					
3	$C3_{p1}$	-3.75					
4	$C4_{p1}$	-4.5					
5	$C5_{p1}$	-5.5					

Profile 2 matrix

	C1	C2	C3	C4	C5
a	0.75	0	0.75	0	0.50
b	0.25	0.75	0	0.75	0
c	0	0	0	0	0.25
-	0	0.25	0.25	0.25	0.25

Scoring matrix

	a	b	c	-
a	2	-1	-3	-1
b		4	0	-1
c			1	-2
-				0

- $\max[V(i-1,j-1)+S(i,j),\ V(i-1,j)+S(i,\_),\ V(i,j-1)+S(\_,j)]$

= $\max[\textbf{0.376},\ -2.50,\ -2.25]$

Family representations and alignments with profiles

- **How to optimally align a profile to a profile**
  - $V(i,j) = \max[V(i-1,j-1)+S(i,j), \, V(i-1,j)+S(i,\_), \, V(i,j-1)+S(\_,j)]$

$V(i,j)$		**0**	**1**	**2**	**3**	**4**	**5**
			$C1_{p2}$	$C2_{p2}$	$C3_{p2}$	$C4_{p2}$	$C5_{p2}$
0		0	-1	-1.75	-2.5	-3.25	-4.25
1	$C1_{p1}$	-1.25	0.376	-0.624	-1.374	-2.124	-3.124
2	$C2_{p1}$	-2.5	-0.125	2.312	1.562	0.812	-0.188
3	$C3_{p1}$	-3.75	-1.375	1.062	1.062	0.876	-0.124
4	$C4_{p1}$	-4.5	-2.125	0.501	0.312	2.938	1.938
5	$C5_{p1}$	-5.5	-3.125	-0.499	0.252	1.938	**2.25**
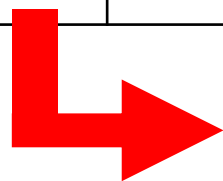
1_{p2}	2_{p2}	3_{p2}	4_{p2}	5_{p2}
1_{p1}	2_{p1}	3_{p1}	4_{p1}	5_{p1}

a	b	a	b	a
a	b	c	b	a