

אלגוריתמים בביולוגיה חישובית – תרגיל 1

(1) נראה כי ישנם 3^n אופציות לעימוד שרשראות T, S באורך n .
 נגדיר $T = t_1 t_2 \dots t_n$, $S = s_1 s_2 \dots s_n$, אזי, לכדי לבחור את העימוד בעמדה הראשונה יש לנו את האופציות הבאות:

s_1			
—			

—			
t_1			

s_1			
t_1			

כאשר אופציה של לעמד גאפ מול גאפ היא טריוויאלית ולא נתחשב בה.
 כעת, ראינו שיש לנו 3 אופציות לעימוד הראשון, נמשיך את ההוכחה באינדוקציה, שאת הצעד הראשון הצגנו כרגע.
 הנחת האינדוקציה: נניח נכונות עבור רצפים באורך $n - 1$, שלהם לפחות 3^{n-1} עימודים, ונראה עבור רצפים באורך n .
 צעד האינדוקציה:

עבור S', T' שני רצפים באורך n ש- $n-1$ הנקלאוטידים הראשונים מועמדים בהם, נגיע לשתי האותיות האחרונות, כאשר הן יכולות להיות כמו בצעד הבסיס, או שתיהן אות, או שאחד הרצפים ביניהם נגמר ויש להציב גאפ מול האות הנותרת, כלומר יש 3 אפשרויות, כאשר שלושתן תקיפות לכל עימוד באורך $n - 1$ לכן ישנן $3^n = 3^{n-1} * 3$ אופציות לעימוד.

(2) כאשר אנו משתמשים בשיטת affine gap penalty על-מנת להפיק את העימוד האופטימאלי בין שני רצפים, כאשר אנו נמצאים במיקום (i, j) עלינו "לזכור" את התוצאה הטובה ביותר אם אנו לא משתמשים ב-gap וכן את התוצאה הטובה ביותר אילו השתמשנו ב-gap.

על-מנת לעשות זאת נצטרך להתבונן בכמה מונים שונים עבור העמדת x_1, \dots, x_i אל מול y_1, \dots, y_j :

$F(i, j)$ – תוצאה אם x_i עומד מול y_j

$G(i, j)$ – תוצאה אם x_i עומד אל מול gap אחרי הופעת y_j

$H(i, j)$ – תוצאה אם y_j עומד אל מול gap אחרי x_i

נגדיר את $V(i, j)$ להיות התוצאה הטובה ביותר עבור העמדת x_1, \dots, x_i אל מול y_1, \dots, y_j .

נגדיר את $Ptr(i, j)$ להיות העימוד האופטימאלי עבור העמדת x_1, \dots, x_i אל מול y_1, \dots, y_j .

פסאודו-קוד עבור אלגוריתם Needleman-Wunsch בשיטת affine gap penalty (תחת ההנחה כי $d > e$):
אתחול:

$$V(0, 0) = 0$$

$$V(i, 0) = d + (i-1)*e$$

$$V(0, j) = d + (j-1)*e$$

איטרציה:

For each $i = 1, \dots, M$:

For each $j = 1, \dots, N$:

$$V(i, j) = \max\{F(i, j), G(i, j), H(i, j)\}$$

$$F(i, j) = V(i-1, j-1) + s(x_i, y_j)$$

$$G(i, j) = \max \begin{cases} V(i-1, j) + d \\ G(i-1, j) + e \end{cases}$$

$$H(i, j) = \max \begin{cases} V(i, j-1) + d \\ H(i, j-1) + e \end{cases}$$

$$Ptr(i, j) = \begin{cases} (x_i, y_j), & F(i, j) > G(i, j) \text{ AND } H(i, j) \\ (x_i, gap), & G(i, j) > F(i, j) \text{ AND } H(i, j) \\ (gap, y_j), & H(i, j) > F(i, j) \text{ AND } G(i, j) \end{cases}$$

סיום ואופן החזרת העימוד האופטימאלי:

$V(M, N)$ הינה התוצאה הטובה ביותר של העימוד האופטימאלי.

$Ptr(i, j)$ הינו העימוד האופטימאלי.