

9/03/12

אוניברסיטת תל אביב – ביה"ס למדעי המחשב

גנומיקה חישובית 0382.3102.01 -

תשע"ב

מרצים: פרופ' רון שמיר, פרופ' רודד שרן

מתרגל: ירון אורנשטיין

משך הבחינה: שלוש שעות ללא אפשרות הארכה

חומר עזר: שני דפים כתובים בכתב יד (משני הצדדים)

יש לענות על שלוש מתוך ארבע השאלות. לכל השאלות ניקוד שווה.

יש לענות בצורה קצרה, ברורה ומדויקת ולנמק כל טיעון. תשובה לא מנומקת לא תזכה בנקודות.

בשאלות בהן נדרש לתאר אלגוריתם יש לציין במפורש את סיבוכיות הזמן.

בשאלות בהן השתמשת באלגוריתם שתואר בשיעור או בתרגיל אין צורך לחזור עליו לפרטיו אלא רק לפרט במדויק את השינויים לעומת מה שנלמד.

שאלה 1

א. (20 נק') תת-מחרוזת לא ריקה של מחרוזת s נקראת תת-מחרוזת "יחודית" אם היא מופיעה פעם אחת בדיוק ב- s . המחרוזת s עצמה נחשבת תת-מחרוזת ייחודית. תארי' אלגוריתם יעיל ככל האפשר המחשב את מספר תת-המחרוזות הייחודיות במחרוזת קלט s .

ב. (13 נק') נקרא לתת-מחרוזת α של מחרוזת s מחרוזת "יחודית מינימלית" של s , אם α מופיעה רק פעם אחת ב- s וכל רישא ממש של α מופיעה יותר מפעם אחת ב- s . תארי' אלגוריתם יעיל ככל האפשר שמוצא את כל המחרוזות הייחודיות מינימליות במחרוזת נתונה s .

שאלה 2

ברצוננו לעמד ברמת ה-דנא שני רצפי גנים $S(1,..,n)$ ו- $T(1,..,m)$ ממינים שונים. אנו יודעים שאיננו יכולים להתייחס באותו אופן לכל עמדה בקודון (למשל, כי העמדה השלישית בכל קודון היא מנונת יחסית ופחות מבדילה בין חומצות אמינו). נניח שהגדרנו ציון דמיון בין בסיסים x, y באותה עמדה i בקודון כ- $\delta(x, y)$ וציון דמיון בין בסיסים שאינם באותה עמדה כ- $\sigma(x, y)$. ציון indel לבסיס בודד יהא δ תמיד בלי תלות בעמדה וציון לביטול או הוספה של קודון שלם - שיכולים לקרות רק בין קודונים - יהא Δ ($3\delta < \Delta$). תארי' במלואו אלגוריתם יעיל ככל האפשר לחישוב העימוד האופטימלי.

www-112

שאלה 3

ברצונך לחשב העמדה של משפחת רצפים נתונה בעזרת מודל Profile HMM שלא מאפשר ביטולים (deletions).

א. (13 נק') תאר/י את המודל בפירוט כולל דיאגרמת המצבים והמעברים ביניהם.

ב. (20 נק') רשום/מי את אלגוריתם ויטרבי לחיפוש רצפים נוספים במשפחה בעזרת מודל זה (בהנחה שלמדת כבר את פרמטרי המודל). על האלגוריתם לחשב ישירות לכל רצף את לוג יחס הנראות שהרצף מהמשפחה לעומת שהרצף מקרי (היינו שהנוקליאוטידים בו שווי התפלגות ובלתי תלויים).

שאלה 4

יהי $x = x_1 x_2 \dots x_n$ מחרוזת רנא (RNA) מעל א"ב ACGU. קיפול דו-ממדי של המחרוזת הוא אוסף של זוגות זרים של אינדקסים בין 1 ל-n שהוא מקון, כלומר אם מקומות a, b מזווגים, ומקומות c, d מזווגים, וכן $a < b$, $c < d$ וגם $a < c$ אזי לא ייתכן $c < b < d$. בסיס יכול להיות מזווג עם בסיס אחר ברצף, ואם אינו מזווג הוא נקרא חופשי. זיווג ייתכן בין הבסיסים A ל-U ובין C ל-G. עבור זוג (i, j) נגדיר את הזוג $(i+1, j-1)$ בתור הצמוד לו.

נגדיר מודל אנרגטי פשוט של קיפול בצורה הבאה. אם בסיס חופשי, אין לו תרומה אנרגטית. אם בסיס מזווג, יש לו תרומה (שלילית) אך ורק אם הזוג הצמוד לו גם מזווג. יש לתאר אלגוריתם תכנון דינמי יעיל ככל האפשר המוצא קיפול הממקסם את מספר הזוגות המזווגים שהזוג הצמוד להם גם הוא מזווג (היינו קיפול בעל אנרגיה מינימלית).

בהצלחה!

www.611