22.12.16

בס"ד.

אלגוריתמים - הרצאה

תזכורת: קוד זו פונק' מא"ב Σ שמוגדרת $C:\Sigma\to\{0,1\}^*$

לדוג' -

a	b	c
01	10	11

קוד רישות: קוד שבו $a,b\in\Sigma$ $c(a)$ אינו רישא של $c(b)$.

בעיה:

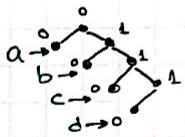
קלט: $\Sigma=<a_1,...,a_n>$ ותדירויות $w_1,...,w_n$

פלט: קוד רישות אופטימלי c אשר $\sum w_i l_i$ מינימלי (כאשר l_i = מס' הביטים ב- $c(a_i)$)

ייצוג של קוד ע"י עץ בינארי:

כל קוד ניתן להציג כעץ בינארי לדוג' -

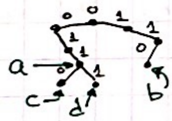
a	b	c	d
0	10	110	1110



ניתן לראות שזה קוד רישות ולכן כל אות מופיעה כעלה.

נסתכל על עץ של קוד שאינו קוד רישות -

a	b	c	d
01	110	010	011



הצעת אלגוריתמים:

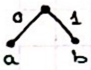
1) אלגוריתם של Huffman (1952) ← האלגוריתם החמדני הראשון שנראה בקורס.

(האלגוריתם קוני לבחירה שנבצע את מה שנראה לנו הכי טוב כרגע ונתקדם משם.

תהליך האלגוריתם:

(1) נמיין את המשקולות $w_1\geq w_2\geq\dots\geq w_n$

(2) תנאי העצירה אם $n=2$, נבנה עץ



אחרת, ניקח את שתי המשקלות המינימליות (w_{n-1},w_n), נוציא אותן מהרשימה (יחד עם a_n ו- a_{n-1}) ומחזיר לרשימה משקל חדש $w_{n-1}+w_n$ ומו חדש b.

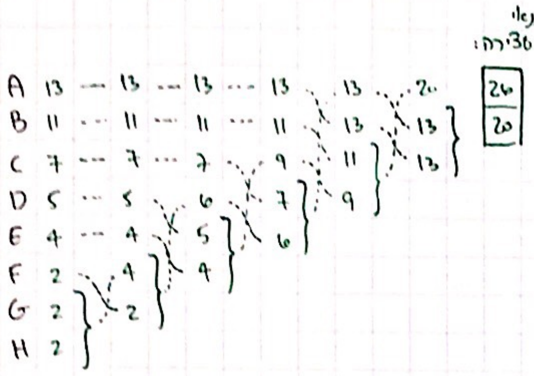
(הרשימה: $a_1,...,a_{n-2},b$ עם המשקלות $w_1,...,w_{n-2},w_{n-1}+w_n$) וקוראים רקורסיבית לאלגוריתם עם $w_1,...,w_{n-2},w_{n-1}+w_n$.

(3) את העץ שנקבל כפלט מהקריאה הרקורסיבית נעדכן ע"י הוספה שני בנים ל- "b" ונצמיד אחד ל- a_{n-1} ואחד ל- a_n.

(*) נשים ♡!

הנחנו שכאשר סידרנו את המשקלות גם סידרנו לפיהם את $a_1,...,a_n$
וכנוסף נשים לב שבכל קריאה דיצמנו מיון מחדש על מערך כמעט ממוין.

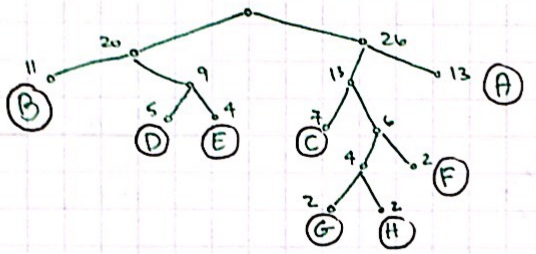
דוגמה:



נראה את ריצת האלגו

A	13
B	11
C	7
D	5
E	4
F	2
G	2
H	2

וכעת במעלה הרקורסיה האלגו בונה את העץ כך:



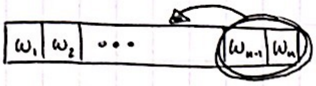
ראינו איך האלגוריתם עובד, כעת צריך להראות נכונות וזמן ריצה.

זמן ריצה:

(1) מיון של n איברים: $O(n\log(n))$

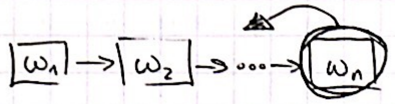
נסתכל על איזה מבני נתונים כדאי לנו...

<u>מערך:</u>



לכל סיבוב ייקח $O(n)$ להכניס את שני המינימלים (למצוא אותם)

<u>רשימה מקושרת:</u>



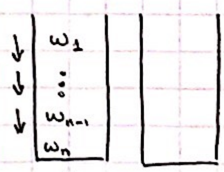
שוב, לכל סיבוב $O(n)$

⇐ יש לנו n סיבובים אזי מערך/רשימה מקושרת ייקחו $O(n^2)$

<u>ערימה:</u> לכל סיבוב יש 2 הוצאות והכנסה אחת ← $O(\log(n))$ כפול n סיבובים $O(n\log n)$

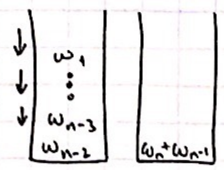
האם יש יותר טוב?!

<u>פתרון ע"י שני תורים:</u>



ריק. (ממוינים)

⇐ ניקח את השניים הקטנים ביותר, נחבר אותם ונכניס לתור השני.



וכמובן כל פעם האיבר החדש יהיה יותר גדול מהאיבר שנמצא בראש התור החדש לכן פשוט נדחוף אותו מלמעלה לתור.

איך נמצא את השניים הקטנים ביותר?

22.12.16

בס"ד.

בפתרון עם שני התורים אין קושי למצוא את שני האיברים המינימליים.

מכיוון שהתור הראשון האיברים כבר ממוינים (כי לא הכנסנו אליו איבר חדש)

והתור השני האיברים גם כן ממוינים כי בהכרח איבר שנכנס מאוחר יותר גדול יותר

לכן יש 3 אפשרויות: 1. שני האיברים שנמצאים בתחתית התור הראשון.

2. שני האיברים שנמצאים בתחתית התור השני.

3. האיבר שנמצא בתחתית התור הראשון והאיבר שנמצא בתחתית התור השני.

לכן צריך $O(1)$ למצוא אותם.

לסיכום: זמן הריצה של האלגוריתם הוא $O(\text{זמן מיון}_{\omega_1,\dots,\omega_n}) + O(n)$

לאחר שחישבנו זמן ריצה נוכיח נכונות:

ברור שקיבלנו קוד רישות כי כל האותיות בעלים, נותר להוכיח שהוא אופטימלי, כלומר מקיים $\sum\omega_i l_i$ מינימלי.

סימון: עבור עץ בינארי T שמתאים לקוד רישות c עבור $\omega_1,\dots,\omega_n$ נסמן $\omega(T)=\sum\omega_i l_i$

משפט: האלגוריתם של הופמן מחזיר עץ T אופטימלי (כלומר, $\omega(T)$ מינימלי)

למה 1: בעץ אופטימלי לכל קודקוד פנימי יש שני בנים.

הוכחה - נניח בשלילה שקיים עץ אופטימלי T כך שקיים קודקוד u כך שיש לו בן יחיד v, נסמן ב-x את האב של u, $\omega(T)=\sum\omega_i l_i$

ניצור מ-T את T_2 ע"י החלפת הבן u (של x) ב-v ונוציא אותו מהעץ. ב-T_2 נקבל: $\omega(T_2)=\sum\omega_i l_i'$ וקיבלנו:

א. $\forall i: l_i' \le l_i$

ב. לכל y עלה בתת עץ של v $l_y' < l_y$

$\Rightarrow \omega(T_2)<\omega(T)$ בסתירה למינימליות של T ∎

מסקנה: בעץ אופטימלי, ברמה התחתונה יש לפחות 2 עלים

הוכחה - אחרת, יש קודקוד בודד ברמה התחתונה ואביו קודקוד פנימי עם בן יחיד בסתירה.

למה 2: בעץ אופטימלי העלים עם המשקל ω_{n-1} ו-ω_n ברמה התחתונה.

הוכחה - נניח בה"כ ש-ω_n אינו ברמה התחתונה של T העץ האופטימלי. לפי המסקנה

יש לפחות 2 עלים ברמה התחתונה $\Leftarrow$ יש ברמה התחתונה עלה עם משקל $\omega_n<\omega_j$ (חוץ מ-ω_{n-1})

ניצור T_2 בדיוק כמו T עם החלפה ω_j ו-ω_n.

נקבל ש- $\omega(T)=\sum\omega_i l_i$ ו-

$\omega(T_2)=\sum\omega_i l_i-\omega_j(l_n-l_j)+\omega_n(l_n-l_j)=\omega(T)+(\omega_n-\omega_j)(l_j-l_n)<\omega(T)$ $\Rightarrow$ סתירה לאופטימליות

בס"ד

<u>למה 3</u>: קיים עץ אופטימלי כך ש- ω_n ו- ω_{n-1} אחים. (ברמה התחתונה)

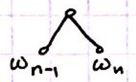
← הוכחה - יהי T עץ אופטימלי, ω_n ו- ω_{n-1} ברמה התחתונה לפי למה 2.

אם הם אחים, סיימנו. אחרת, נובע מלמה 1 שחייב להיות ל- ω_n אח ω_j.

נגדיר T_2 ע"י החלפה ω_{n-1} ב- ω_j ונקבל $\omega(T)=\omega(T_2)$ כאשר T_2 מקיים את הדרוש.

הוכחת המשפט -

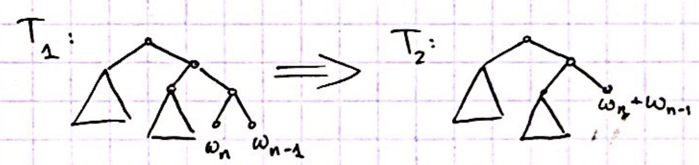
נוכיח באינדוקציה על n (מס' העלים)

בסיס האינדוקציה: $n=2$, במקרה זה הופמן מחזיר  עץ אופטימלי.

הנחת האינדוקציה: נניח נכונות על $n-1$ ונוכיח ל- n, נסתכל רגע על סתם עץ אופטימלי (ללא קשר להופמן)

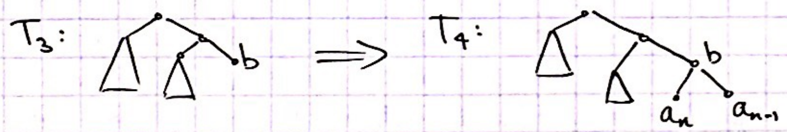
לפי למה 3, קיים עץ אופטימלי T_1 כך ש- ω_{n-1} ו- ω_n אחים.

נגדיר T_2 מ- T_1 ע"י מחיקת העלים ω_n ו- ω_{n-1} וסימון האב המשותף b ב- $\omega_n+\omega_{n-1}$ כך:



טענה: T_2 אופטימלית עבור המשקלות $\omega_1,\dots,\omega_{n-2},\omega_n+\omega_{n-1}$ ועבור תווים $a_1,\dots,a_{n-2},b$

← הוכחה - נניח בשלילה שקיים עץ T_3 כך ש- $\omega(T_3)<\omega(T_2)$, נבנה T_4 מ- T_3 ע"י הוספה שני בנים ל- b, אחד נסמן, בתור $a_n|\omega_n$ והשני בתור $a_{n-1}|\omega_{n-1}$



$$\omega(T_4)= \omega(T_3)+\omega_n+\omega_{n-1}$$

ולכן,

$$\omega(T_1)=\omega(T_2)+\omega_n+\omega_{n-1} > \omega(T_3)+\omega_n+\omega_{n-1}=\omega(T_4)$$

אך T_4 עץ עם משקלות $\omega_1,\dots,\omega_n$ בסתירה לאופטימליות של T_1.

$\Leftarrow$ T_2 אופטימלי.

נחזור לאלגוריתם של הופמן:

באלגוריתם של הופמן בצעד 3 בוחרים את ω_n ו- ω_{n-1}, מאחדים את המשקל ל- $\omega_{n-1}+\omega_n$

ומקבלים את המשקלות $\omega_1,\dots,\omega_{n-2},\omega_n+\omega_{n-1}$, לפי הנחת האינדוקציה העץ שהופמן מחזיר H_{n-1}

ל- $\omega_1,\dots,\omega_{n-2},\omega_n+\omega_{n-1}$ הוא אופטימלי. לכן $\omega(H_{n-1})=\omega(T_2)$.

האלגוריתם של הופמן מחזיר H_n ע"י הוספה שני בנים ל- b שנמצא ב- H_{n-1} ולכן

(המשך בעמוד הבא) ←

21.12.16

בס״ד.

נקבל ש- $\omega(H_n) = \omega(H_{n-1}) + \omega_n + \omega_{n-1} = \omega(T_2) + \omega_n + \omega_{n-1} = W(T_1)$

אך T_1 אופטימלי, לכן H_n אופטימלי. $\square$

Scanned by CamScanner