הוכחת סיבוכיות משוערכת $O(\log^* n)$ למימוש Union-Find באמצעות עצים הפוכים עם איחוד לפי גודל וכיווץ מסלולים

מבוסס על הוכחה של Claire Mathieu, שבתורה התבססה על הוכחה מהמהדורה הראשונה של ספר הלימוד

ניתוח זמנים

<u>משפט</u>: במימוש באמצעות אוסף עצים של n איברים עם איחוד לפי גודל וכיווץ מסלולים, סיבוכיות זמן בצוע m פעולות Union/Find חסומה ע"י $O(m\log^* n)$.
לפיכך הזמן המשוערך (amortized time) לכל פעולה חסום ע"י $O(\log^* n)$.

כאשר הפונקציה $\log^* n$ מוגדרת כדלהלן:

$$\log^{(i)}(n) = \overbrace{\log_2\log_2 \dots \log_2}^{i} n$$

$$\log^*(n) = \min\{i \ge 0 \mid \log^{(i)}(n) \le 1\}$$

כאמור, בספר הלימוד ניתן למצוא הוכחה לחסם הדוק עוד יותר. בשקפים הבאים נראה הוכחה לחסם הנ"ל, תוך שימוש בשיטת הצבירה לניתוח משוערך.

הוכחה

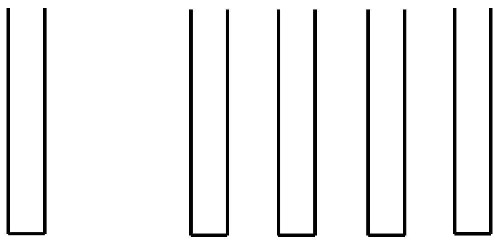
<u>הגדרה:</u> יהיה v צומת ביער הUnion-Find.
יהא T_v העץ הגדול ביותר ש-v היה שורשו עד כה. אזי נסמן:
$r(v) \triangleq \lfloor \log|T_v| \rfloor$.
נקרא ל- $r(v)$ <u>הדרגה</u> של v.

<u>הבחנות:</u>

1. הדרגה המקסימלית היא לכל היותר $\log_2 n$.
2. ברגע שצומת v מפסיק להיות שורש, הוא לעולם לא יהפוך שוב לשורש, ודרגתו לא תשתנה יותר לעולם (כי v לא יהיה שורש של עץ גדול יותר).
3. לכל צומת יש דרגה גדולה ממש מכל צאצאיו.
<u>הוכחה:</u> באינדוקציה על מספר הפעולות Union ו-Find.
4. בעץ מגודל k יש לכל היותר $k/2^r$ צמתים מדרגה r.
<u>הוכחה:</u> לפי הגדרת הדרגה, מתקיים כי $|T_w| \ge 2^r$ לכל צומת w מדרגה r.
יהיו u ו-v שני צמתים מדרגה r באותו עץ. אזי T_u ו-T_v זרים (תרגיל).
מכאן שלא ייתכנו למעלה מ- $k/2^r$ צמתים מדרגה r בעץ מגודל k.

הוכחה – תאים

כדי לנתח את הסיבוכיות נאגד צמתים שאינם שורשים בעלי דרגה דומה בתאים (בדומה לדרגה, תאים אלו לא נשמרים בזיכרון ולמעשה אינם חלק מהאלגוריתם, אלא רק משמשים אותנו בהוכחה).



<u>צמתים עם דרגות בתחום:</u> 0 1 $2,3$ $4,\dots,15$ … $B,\dots,2^B$-1

נבחין כי הדרגות הקטנות ביותר בכל תא נתונות ע"י הסדרה הבאה:

$$0,\quad 2^0=1,\quad 2^1=2,\quad 2^2=4,\quad 2^{2^2}=16,\quad 2^{2^{2^2}}=65536,\dots$$

לפני שנראה שימוש לתאים בניתוח שלנו, נתעניין במספר התאים ובגודלם.

תכונות התאים (1)

טענה 1: מספר התאים הוא לכל היותר $\log^* n + 1$.

הוכחה:

- לפי הבחנה 1, הדרגה המקסימלית היא לכל היותר $\log_2 n$.
- כפי שהבחנו, הדרגות הקטנות ביותר בכל תא נתונות ע"י הסדרה:

$$0, \quad 2^0 = 1, \quad 2^1 = 2, \quad 2^2 = 4, \quad 2^{2^2} = 16, \quad 2^{2^{2^2}} = 65536, \ldots$$

- מכאן שמספר התאים הוא לכל היותר

$$1 + \min\left\{i: 2^{2^{\cdots^{2^0}}} \geq \log_2 n\right\} =$$

- ע"י העלאה בחזקת שתיים של שני האגפים בסוגריים נקבל:

$$1 + \min\left\{i: 2^{2^{\cdots^{2^2}}} \geq n\right\} =$$

- ע"י הוצאת $\log$ לפי בסיס שתיים i פעמים בשני האגפים בסוגריים נקבל:

$$1 + \min\{i: 1 \geq \log^{(i)} n\} = 1 + \log^* n$$

תכונות התאים (2)

טענה 2: בעץ בן k צמתים, מספר הצמתים בתא המכיל צמתים עם דרגות בתחום $[B, 2^B - 1]$ הוא לכל היותר $2k/2^B$.

הוכחה:

- לפי הבחנה 4 בנוגע לדרגות, מספר הצמתים מדרגה r בעץ בן k צמתים הוא לכל היותר $k/2^r$.
- מכאן שסה"כ הצמתים בתא $[B, 2^B - 1]$ בעץ בן k צמתים הוא לכל היותר

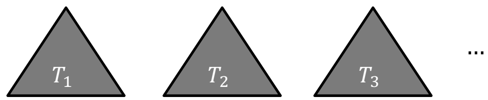
$$\sum_{r=B}^{2^B-1} k/2^r$$

- סכום זה הוא סכום של סדרה הנדסית עם מנה חצי ומכאן שהוא לכל היותר פעמיים האיבר הראשון בסדרה, כלומר $2k/2^B$.

הוכחת המשפט

משפט: במימוש באמצעות אוסף עצים של n איברים עם איחוד לפי גודל וכיווץ מסלולים, סיבוכיות זמן בצוע m פעולות Union/Find חסומה ע"י $O(m \log^* n)$.

הוכחה: לאחר m פעולות Union ו-Find ניוותר עם מספר עצים, $T_1, T_2, \ldots$ מגודל $n_1, n_2, \ldots$ בהתאמה אשר "בהם" בוצעו עד כה $m_1, m_2, \ldots$ פעולות. כלומר m_i פעולות Union/Find בוצעו על צמתים\תתי קבוצות של T_i.



כל צומת מתחיל כקבוצה נפרדת ולכן כדי שהעץ T_i יהיה מגודל n_i חייבים היו להתבצע בעץ T_i לפחות $n_i - 1$ פעולות Union, ובפרט $m_i \geq n_i - 1$.

בשקפים הבאים נראה כי לכל עץ T_i עלות m_i הפעולות שבוצעו בו חסומה על ידי

$$O\big((m_i + n_i)\log^* n\big) = O(m_i \log^* n)$$

נסכום על פני כל העצים ונקבל כי עלות $m = \sum_i m_i$ הפעולות היא $O(m \log^* n)$.

הוכחת החסם לכל עץ

טענה: יהא T_i עץ מגודל $n_i \leq n$ שהתקבל ע"י m_i פעולות Union/Find, כאשר האיחוד נעשה לפי גודל ובחיפוש מתבצע כיווץ מסלולים. אזי עלות m_i הפעולות היא $O(m_i \log^* n)$.

הוכחה:
עלות כל פעולת Union היא $O(1)$ ומכאן שפעולות אלו עולות לכל היותר $O(m_i)$.

עלות פעולות Find לינארית במספר הקשתות בהם נעלה.
נחלק קשתות אלו ל-3 סוגים:

1) קשתות בין בן של שורש לשורש.
2) קשתות בין צומת ואביו, הנמצאים בתאים שונים.
3) קשתות בין צומת ואביו, הנמצאים באותו תא.

נעלה במעלה קשת אחת מסוג (1) לכל היותר בכל חיפוש, ומכאן שקשתות מסוג זה תורמות $O(m_i)$ על פני m_i הפעולות.

קשתות בין צמתים בתאים שונים

<u>הוכחה (המשך):</u>
עלות פעולות Find לינארית במספר הקשתות בהם נעלה.
נחלק קשתות אלו ל-3 סוגים:

1) קשתות בין בן של שורש לשורש. **עלות:** $O(m_i)$ על פני m_i הפעולות.
2) קשתות בין צומת ואביו, הנמצאים בתאים שונים.
3) קשתות בין צומת ואביו, הנמצאים באותו תא.

נבחין כי לפי הבחנה 3 על דרגות, בעת עליה במסלול חיפוש של צומת אנו אך ורק עולים בדרגה, לכן קשת מסוג (2) מתאימה למעבר לתא עם דרגות גדולות יותר ואין מעברים לתא עם דרגות קטנות יותר.
כיוון שמספר התאים הוא לכל היותר $\log^* n + 1$ (לפי טענה 1), הרי שבכל פעולת חיפוש יש לכל היותר $\log^* n$ קשתות מסוג (2).

לכן על פני כל הפעולות קשתות אלו עולות לכל היותר $O(m_i \log^* n)$.

קשתות בין צמתים באותו תא

<u>הוכחה (המשך):</u>
נותר לנו לחסום את מספר הקשתות בין צומת ואביו הנמצאים באותו תא. את מספר זה ניתן לרשום כך:

$$\sum_{cell\ c} \sum_{u \in c} |\{(u,v) traversed, v \in c\}|$$

נתמקד בצומת $u \in c$, בתא c המכיל צמתים עם דרגות בתחום $[B, 2^B - 1]$:

כיוון שאנו נבצעים כיווץ מסלולים וכיוון שלכל צומת דרגה גדולה ממש מצאצאיו (הבחנה 3), כל פעם שעולים בקשת (u,v) כנ"ל דרגת האב (החדש) של u גדלה. למעשה, דרגת האב של u לא יכולה לקטון.
לכן מספר הפעמים שדרגת האב של u יכולה לגדול עד שהאב של u יהיה בתא אחר הוא לכל היותר 2^B.
מכאן שהסכום הנ"ל קטן\שווה מ:

$$\sum_{\text{תא } c} \sum_{u \in c} 2^B$$

קשתות בין צמתים באותו תא

<u>הוכחה (סיום):</u>
לפי טענה 2, מספר הצמתים בתא $[B, 2^B - 1]$ הוא לכל היותר $2n_i/2^B$, ומכאן שהסכום הנ"ל קטן\שווה מ-

$$\sum_{\text{תא } c} \frac{2n_i}{2^B} \cdot 2^B = \sum_{\text{תא } c} 2n_i$$

לפי טענה 1, מספר התאים הוא לכל היותר $\log^* n + 1$ ולכן הסכום הנ"ל הוא לכל היותר

$$2n_i \cdot (\log^* n + 1) = O(n_i \log^* n)$$

וזו העלות הנגרמת מקשתות בין צמתים באותו תא.

סיכום – הוכחת חסם לכל עץ

<u>עבור עץ מגודל n_i בו בוצעו m_i פעולות, עלויות הפעולות השונות הוא:</u>

כל פעולות ה-Union עולות $O(m_i)$ זמן.

העלות הנגרמת מפעולות Union, עם הפרדה לסוגי קשתות בהן עולים:

1) קשתות בין בן של שורש לשורש.
עלות: $O(m_i)$ על פני m_i הפעולות.

2) קשתות בין צומת ואביו, הנמצאים בתאים שונים.
עלות: $O(m_i \log^* n)$ על פני m_i הפעולות.

3) קשתות בין צומת ואביו, הנמצאים באותו תא.
עלות: $O(n_i \log^* n) = O(m_i \log^* n)$על פני m_i הפעולות.

כשהמעבר האחרון נובע מכך ש $m_i \geq n_i - 1$.

סה"כ עלות הפעולות בעץ הנ"ל היא $O(m_i \log^* n)$, ובזאת סיימנו את ההוכחה.