מ.מ.למדמ"ח ∼ עמית ווינשטין ∼ קודים לתיקון שגיאות

שחר פרץ

19 ליוני 2024

1 חזרה

הודעה מקודית באורך k.
נשלח הודעה באורך m.
פונ' קידוד $E\colon \{0,1\}^k \to \{0,1\}^m$.
יתקיים $|\mathrm{Im}(E)| = 2^k$, ונסמן $C = \mathrm{Im}(E)$. מרחק האמינג: $\Delta(x,y) = |\{i \mid x_i \neq y_i\}|$.
מרחק של קוד: $d = \Delta(C) = \min_{x \neq y \in \{0,1\}\}} \Delta(E(x), E(y))$.
מטרה: d גדול ו־n קטן. החסם יהיה $d \leq n - k + 1$.
כמה שגיאות ניתן לזהות? $d-1$; עבור כדור האמינג סביב $E(x)$ (כל האפשרויות במרחק האמינג $d-1$ בהכרח $E(y)$ לא יכול להיות שם ולכן בוודאות נוכל לזהות זאת)
כמה שגיאות ניתן לתקן? $\lfloor \frac{d-1}{2} \rfloor$ כי כך כדורי ה־humming לא ידעו זה בזה ($\lfloor \frac{d-1}{2} \rfloor + \lfloor \frac{d-1}{2} \rfloor < d$) כלומר אין חיתוך בין הכדורים, ובהכרח יהיה אפשר לזהות בצורה כזו או אחרת לתקן את הטעויות.

2 אלגו'ים

אלגו' 1: לעבור על כל 2^k המילים ולמצוא את המרחק האמיניג המינימלי של מילה נתונה x, שידוע שהוא מתחת לחסם הדרוש כדי לתקן שגיאות, כלומר באיזה כדור האמינג היא נמצאת. סיבוכיות $2^k \cdot n$. תחת ההנחה שלהפעיל את הקוד לוקח $O(1)$.

אלגו' 2: נשנה את המילים עד שנקבל משהו מינימלי. סיבוכיות: $n^{\lfloor \frac{d-1}{2} \rfloor}$.

אלו אלגוריתמים שמתבססים על ההנחה שהכי סביר להחזיר את המילה הקרובה ביותר במרחק humming.

3 דוג'ים

לדוגמה עבורים Rap_3 (לחזור על כל ביט 3 פעמים) יתקיים $k=1, n=3, d=3$, או באופן כללי $n=3k, d=3$ (עבור שינוי מינמלי של אות אחת, שיגרור מרחק האמינג של 3). עוד כמה דוגמאות:

name	k	n	d	תיאור
Rep_3	k	$3k$	3	חזרה 3 פעמים
Rep_t	k	tk	t	חזרה t פעמים
Par	2	3	2	הוספת ביט זוגיות
Par	k	$k+1$	2	ראה לעיל
משחק קלפים	5^2	6^2	4	משהו מהשבוע שעבר

הערה: הוספת ביט זוגיות הוא ה־xor של כל הערכים, שמסומן ב־$\oplus$ - ההפוך לאמ"מ, או אך ורק אחד משתי האפשרויות (כמו לשאול ילד, אתה רוצה גליה או עוגת שוקולד). בפייתון נשתמש ב־^ בשביל xor. תכונות:

$$A \oplus B = 0 \iff A = B,\ (A \oplus B) \oplus C = A \oplus (B \oplus C),\ A \oplus B = B \oplus A$$

xor של רצף ערכים בינאי ייצג את הזוגיות של העמודה.

וכפועל: קיסרנו, נקסר, הקסרה.

4 Index Code

נבחר $k = 2^\ell - 1$ עבור ℓ כלשהו.

נקודד הודעה $x_1, \ldots, x_k$ באופן הבא:

נגדיר:

$$EC(x) = \bigoplus \{\, \mathrm{bin}(i) \mid \imath i.x_i = 1 \}$$

וגם (עיגול קטן מסמל שירשור)

$$E(x) = x^\circ EC(X)$$

דוג':

```
  1234567
  0110110
2 010
3 011
5 101
6 110
---------
  010
```

כלומר $EC(x) = 010, E(x) = 0110110010$. יתקיים $n = k + \ell = k + \log(k)$.

ועבור d? טענה: $d \geq 2$. הסבר: אם $\Delta(x, y) \geq 2$, סיימנו. אחרת, אם $\Delta(x, y) = 1$, אזי בהכרח קיים אינדקס יחיד i כך ש־$x_i \neq y_i$, ולכן $EC(x)$ ו־$EC9y)$ שונים, כלומר $\Delta(EC(x), EC(y)) \geq 2$.

הוכחה ש־$d \leq 2$: נראה דוג'

$$EC(x) = 000, \qquad x\colon 0000000 \tag{1}$$
$$EC(y) = 001, \qquad y\colon 1000000 \tag{2}$$

סה"כ $d = 2$ כי הוכחנו שני חסמים.

לא מדהים.

4.1 לא מדהים, גרסה 2

רעיון: נשרשר את $EC(x)$ פעמיים: $E(x) = x \circ EC(x) \circ EC(x)$ עתה, יתקיים $= 2^\ell - 1$, וגם $n = k + \ell = k + 2\log k$, ואחרונה $d = 3$: ניתן חסם תחתון ועליון.

$d \geq 3$: אם $\Delta(x, y) \geq 3$ סיימנו. אם $\Delta(x, y) = 1$ אזי $\Delta(E(x), E(y)) \geq 3 \implies EC(x) \neq EC(y)$. אם $\Delta(x, y) = 2$ אזי $\exists i, i'.x_i \neq y_i \wedge x'_i \neq y'_i$. כלומר, $EC(x) \oplus EC(y) = i \oplus i'$ כלומר $\Delta(E(x), E(y)) \geq 4$.

כיוון שני: $d \leq 3$. נשתמש באותה הדוגמה. $x = 0000000$, $y = 1000000$, נקבל $E(x) = 000, EC(y) = 001$ וסה"כ נקבל מרחק 3.

סה"כ $d = 3$.

4.1.1 זיהוי שגיאות

תיאורטית, ניתן לתקן שגיאה אחת. איך נעשה זאת, באופן יעיל? קלט: $y = x \circ EC_1, EC_2$, שזה שיבוש לכל היותר אחד על $x' \circ EC(x') \circ EC(x')$. נפלג למקרים:

- $EC_1 \neq EC_2 \implies$ יש שגיאה ב־EC_1 או EC_2, כלומר x תקין ונחזיר אותו.
- $EC_1 = EC_2 \implies$
  - אם $EC(x) = EC_1$, אז אין שגיאה, ונחזיר את x.
  - אחרת, האינדקס ב־x בו יש טעות הוא $EC(x) \oplus EC_1(x)$, ונחזיר את x אחרי שנהפוך את הקלט הזה.

4.2 יותר מדהים

רעיון: נשרשר ביט זוגיות (נקסר את כל x):

$$E(x) = x \circ EC(x) \circ EC(x) \circ Par(x)$$

טענה: $d = 4$.

טענה: עבור קוד עם מרחק אי־זוגי d, הוספת ביט זוגיות מעלה את המחק ל־$d + 1$.

הוכחה. נפצל למקרים.

- $\Delta(E(x), E(y)) \geq d + 1$: סיימנו
- $\Delta(E(x), E(y)) = d$: ביט הזוגיות יהיה כונה ולכן נקבל מרחק $d + 1$.

■

5 קוד Hamming

שמם נגזר (למיטב זכורנו של המורה) מהעובדה שכדורי ההאמינג ברדיוס 1 מהווים ריצוף מרחק של המושלם או משהו כזה. משפחת קודים של $(n,k,d) = (2^t, 2^t-1, 3)$. הדוגמה הכי פופולארית: בפרט, קוד $(7,4,3)$. נניח שהקלט הוא x_3, x_5, x_6, x_7. נוסיף את הביטים החסרים. נוסיף את הביטים:

$$
\begin{aligned}
x_1 &= x_3 \oplus x_5 \oplus x_7 \\
x_2 &= x_3 \oplus x_6 \oplus x_7 \\
x_4 &= x_5 \oplus x_6 \oplus x_7
\end{aligned}
$$

זהו קוד עבורו $d = 3$.

הוכחה. תעברו על המילים ותבדקו ■