בדיקה הסתברותית של הוכחות | 67790

הרצאות | פרופ׳ גיא קינדלר

כתיבה | נמרוד רק

תשפ״ג סמסטר ב׳

השבוע הראשון הושלם באדיבותו (הרבה) של דויד קיסר-שמידט וסיכומו.

תוכן העניינים

שבוע $\mathbb{I}$ | מבוא

הגדרה מכונת טיורינג היא אוטומט עם סרט זיכרון שהיא יכולה לנוע עליו. מ"ט M מקבלת שפה $L \subseteq \Sigma^*$ אם היא מסיימת במצב מקבל על x אם"ם $x \in L$.

הגדרה מ"ט חישוב זו מ"ט שיש לה מצב עוצר שכשהיא מגיעה אליו הערך שרשום על הסרט הוא הפלט שלה.

הגדרה $\mathsf{P} = \{L :$ קיימת מ"ט המכריעה את L בזמן פולי' $\}$.

הגדרה נאמר כי $L \in \mathsf{NP}$ אם קיימת שפה L^π כך ש:

1. $L^\pi \in \mathsf{P}$.
2. המילים ב-L^π הן מהצורה (x, w) כאשר $x \in L$ ו-$|w| \le \mathsf{poly}(x)$.
3. לכל $x \in L$ קיים w כך ש-$(x, w) \in L^\pi$.

הערה בעיות הכרעה של שפה L הן למעשה חלוקה של Σ^* ל-$(\mathcal{Y}, \mathcal{N})$.

הגדרה בעיית הבטחה (promise problem) היא חלוקה $L = (\mathcal{Y}, \Sigma^* \setminus (\mathcal{Y} \cup \mathcal{N}), \mathcal{N})$ של Σ^*. מ"ט שמזהה את L מקבלת ודוחה נכונה מילים ב-$\mathcal{Y}, \mathcal{N}$ בהתאמה (מבטיחה את התשובה עליהם) ומילים ב-$\Sigma^* \setminus (\mathcal{Y} \cup \mathcal{N})$ יכולות להתקבל, להדחות או שהמ"ט לא תעצור (אין ערובה לתוצאת הריצה).

הערה בעיית הכרעה של שפה L היא בעיית הבטחה מהצורה $(L, \varnothing, \Sigma^* \setminus L)$.

הערה רדוקציה חשיבה לבעיות הבטחה מוגדרת בדומה לרדוקציה בבעיות הכרעה.

הגדרה נאמר כי $L \in \mathsf{NPH}$ אם לכל $L' \in \mathsf{NP}$ קיימת רדוקציה פולי' מ-L' ל-L (כאשר L בעיית הבטחה).

הגדרה נאמר כי $L \in \mathsf{NPC}$ אם $L \in \mathsf{NPH}$ וגם $L \in \mathsf{NP}$.

הגדרה בעיית $\mathsf{MAX} - \mathsf{3SAT}$ מקבלת קלט חוקי $I \in \mathsf{3CNF}$ (נוסחה המורכבת מהסגרים) והמטרה היא לתת השמה שתספק כמה שיותר הסגרים.

עבור קלט חוקי I נגדיר $\operatorname{opt}(I)$ (מסומן לעתים $\operatorname{val}(I)$) האחוז המקסימלי של הסגרים שניתן לספק ב-I.

הערה $\mathsf{3SAT} \in \mathsf{NP}$ היא שפת כל הקלטים החוקיים שהערך שלהם הוא 1 (נוסחה הניתנת לסיפוק במלואה).

הערה $\mathsf{MAX} - \mathsf{3SAT}$ אינה בעיית הכרעה או הבטחה ולכן לעת עתה ההוכחה (העד) אינה מוגדרת היטב.

הערה ל-$\mathsf{3SAT}$ יש כמה מאפיינים מיוחדים מבחינת בדיקת הוכחות. ראשית ניתן לבדוק הוכחה במקביל על כל ההסגרים אם נתון לנו כוח חישוב מקבילי מספיק. ניתן לנצל מנגנון זה לצורכי בדיקה הסת' של השמה: אם $I \in \mathsf{3SAT}$ אז בהגרלת הסגרת, ההסת' שיסופק היא $P(\mathcal{Y}) = 1$ אבל אם $I \notin \mathsf{3SAT}$ אז $P(\mathcal{Y}) \le 1 - \frac{1}{m}$ (לפחות הסגר אחד לא מסופק). כלומר ניתן להגדיר מוודא הסת' לבעיה.

הגדרה מוודא הסת' לבעיית הבטחה הוא מ"ט שמקיים את התנאים הבאים:

- (לוקליות) המ"ט מבצעת מספר גישות קבוע לעד (3 ביטים בלבד מתוך העד).
- (רנדומיות) המ"ט מגרילה $\mathcal{O}(\log n)$ ביטים.
- (שלמות) המ"ט מקבלת קלט בשפה בהסת' 1 (המוודא מושלם).
- (תקפות, Soundness) קיים לו חסם מלעל להסת' לקבלת קלט שאינו בשפה (במקרה שלנו $1-\Theta\left(\frac{1}{n}\right)$).

טענה לכל $L \in \mathsf{NP}$ קיים מוודא הסת' עם פרמטרים כמו שכתבנו למעלה.

הוכחה: ממשפט קוק-לוין, יש רדוקציה מ-L ל-3SAT ולכן מספיק לבדוק הסת' את הקלט המתקבל ל-3SAT. ■

משפט (PCP בניסוח 3SAT) לכל $L \in \mathsf{NP}$ קיים מוודא הסת' עם פרמטרים כנ"ל ו-$1 >$ soundness $<$ const (ישנו חסם מלעל קבוע קטן ממש מאחד לתקפות).

הערה כדי לקיים את הדרישה על התקפות צריך שהרדוקציה מהשפה לנוסחה ב-3CNF תיתן נוסחה שהיא בהסת' נמוכה ספיקה.

הגדרה $\mathsf{gap}-\mathsf{MAX}-\mathsf{3SAT}\,[c,s]$ היא בעיית ההבטחה עם

$$\mathcal{Y}=\{I : \text{קלט חוקי לבעיית 3SAT } I \ \wedge\ \mathrm{val}(I)\geq c\}$$

$$\mathcal{N}=\{I : \text{קלט חוקי לבעיית 3SAT } I \ \wedge\ \mathrm{val}(I)\leq s\}$$

הערה אינטואטיבית, c הוא המשלים (אחד פחות-) אחוז ה-false negative שאנחנו מוכנים לסבול ו-s הוא אחוז ה-false postiive שאנחנו מוכנים לסבול.

משפט (ניסוח מחדש של PCP עם $\mathsf{gap}-\mathsf{MAX}-\mathsf{3SAT}$) קיים $s<1$ כך ש-$\mathsf{gap}-\mathsf{MAX}-\mathsf{3SAT}\,[1,s]\in\mathsf{NPH}$.

הערה הניסוח החדש מספיק כי ל-$\mathsf{gap}-\mathsf{MAX}-\mathsf{3SAT}\,[1,s]$ יש מוודא הסת' שעונה על הקריטריונים האמורים לעיל ולכן עם רדוקציה מכל $L \in \mathsf{NP}$ נקבל את משפט ה-PCP המקורי.

המוודא מקבל $I=c_1\wedge\ldots\wedge c_m$ נוסחה חוקית ו-f השמה (העד), מגריל $i\in[m]$ ובודק האם c_i מסופקת ע"י f (צריך לבדוק את שלושת הביטים ב-f המתאימים לליטרלים ב-c_i). אם הפסוקית מסופקת יענה $\mathcal{Y}$ ואחרת $\mathcal{N}$.

- אם $I\in\mathcal{Y}$ אז יש השמה מספקת ולכן המוודא יענה $\mathcal{Y}$ על איזשהו עד (לכן תמיד נסווג נכון $I\in\mathcal{Y}$).
- אם $I\in\mathcal{N}$ אז $s\cdot m$ פסוקיות לכל היותר מסופקות ע"י כל השמה ולכן ההסת' שניפול על אחת מסופקת (שתגרום לנו לחשוב ש-I כן ספיקה) היא s, כלומר s הוא קבוע התקפות במקרה הזה.

הגדרה אלג' α-מקרב ל-$\mathsf{MAX}-\mathsf{3SAT}$ (עבור $\alpha\in[0,1]$) הוא אלג' שמקבל כקלט נוסחת 3CNF חוקית I ומחזיר מספר b שמקיים $\alpha\,\mathrm{val}(I)\leq b\leq\mathrm{val}(I)$.

מסקנה (ממשפט ה-PCP) אם $\mathsf{P}\neq\mathsf{NP}$ אז לא קיים אלג' α-מקרב פולינומי ל-$\mathsf{MAX}-\mathsf{3SAT}$ עבור $\alpha>s$ (כאשר s הקבוע ממשפט ה-PCP).

הוכחה: נניח בשלילה שקיים אלג' כזה. תהי $L \in \mathsf{NP}$, לכן קיימת רדוקציה f מ-L ל-$\mathsf{gap}-\mathsf{MAX}-\mathsf{3SAT}\,[1,s]$. יהי קלט w לבעיית ההכרעה L. נריץ את אלג' הקירוב על $f(w)$ ונקבל

$$\alpha\ \mathrm{val}\left(f\left(w\right)\right) \leq b \leq \mathrm{val}\left(f\left(w\right)\right)$$

- אם $w \in L$ אז $b \geq \alpha\underbrace{\mathrm{val}\left(f\left(w\right)\right)}_{\geq 1} \geq \alpha > s$.
- אם $w \notin L$ אז $b \leq \mathrm{val}\left(f\left(w\right)\right) \leq s$.

כלומר השוואה של b ל-s תכריע האם $w \in L$ ולכן מ"ט דטר' פולי' בזמן יכולה להכריע את L כלומר $L \in \mathsf{P}$, ולכן $\mathsf{P} = \mathsf{NP}$ סתירה. ■

מסקנה אם $\mathsf{gap}-\mathsf{MAX}-\mathsf{3SAT}\,[c,s] \in \mathsf{NPH}$ וגם $\mathsf{P} \neq \mathsf{NP}$ אז אין אלג' קירוב עם פרמטר גדול מ-$\frac{s}{c}$.

הוכחה: כנ"ל. ■

דוגמאות לאלג' קירוב לבעיות קשות ב-NP

- ראינו באלגו' אלג' $\frac{7}{8}$-מקרב ל-3SAT (מגרילים הרבה השמות עד שאחת מספקת לפחות $\frac{7}{8}$ מהפסוקיות).
- בעיית $\mathsf{MAX}-\mathsf{Exact3}-\mathsf{LIN2}$ היא בעיית האופטימיזציה מעל מערכת n משוואות, בכל אחת שלושה משתנים (שניתן לשים בהם $0,1$) שערכה הוא המספר המקס' של משוואות שניתן לספק במערכת.

  אלג' $\frac{1}{2}$-מקרב לבעיה (שראינו באלגו') בודק לכל משתנה איזו השמה עדיפה (לפי תוחלת סיפוק המשוואה) ובוחר באופן חמדני את ההשמה העדיפה.

  ידוע כי $\mathsf{gap}-\mathsf{MAX}-\mathsf{E3}-\mathsf{LIN2}\left[1-\epsilon,\frac{1}{2}+\epsilon\right]$ היא בעיה קשה ב-NP לכל $\epsilon > 0$ (כלומר אינטואיטיבית ממש קשה להבדיל בין מערכות משוואות שניתן לספק כמעט את כל המשוואות בהן לבין מערכות שניתן לספק קצת יותר מחצי ממשוואותיהן).
- בעיית $\mathsf{MAX}-\mathsf{IS}$ לכל גרף מחזירה את גודל קבוצת הקודקודים הבת"ל (אף שני קודקודים בקבוצה אינם מחוברים בצלע) המקסימלית.

  ידוע כי $\mathsf{gap}-\mathsf{MAX}-\mathsf{IS}\left[1-\frac{1}{\sqrt{2}}-\epsilon,\epsilon\right]$ קשה ב-NP לכל $\epsilon > 0$ (ראו הסבר אינטואיטיבי לעיל).

קווים לדמותו של PCP - משחק עם שני שחקנים חזקים ומוודא חלש

נתונים שני שחקנים (חזקים חישובית) שמשחקים משחק : בהינתן נוסחה, הם מתאמים עמדות (בוחרים השמה) ואז מופרדים.

שחקן אחד מקבל פסוקית ושחקן נוסף משתנה בפסוקית. הראשון מחזיר השמה למשתנים בפסוקית והאחרון השמה למשתנה.

הם מנצחים אם ההשמה של הראשון מספקת את הפסוקית ואם שני השחקנים מסכימים על הערך המושם במשתנה שניתן לאחרון מתוך הפסוקית.

הערה הרעיון מאחורי המשחק הזה הוא שקילות ה-PCP למצב בו שני שחקנים חזקים חישובית מנסים להראות הסת' למוודא חלש מאוד שניתן לספק את נוסחה מסוימת.

טענה בהינתן $\text{val}\left(I\right)\leq\alpha$ (שיעור הפסוקיות שניתנות לסיפוק בו זמנית המקסימלי), ההסת' שינצחו היא $P\left(\text{success}\right)\leq1-\frac{1-\alpha}{3}$.

הוכחה: נניח שהשחקנים משחקים באסטרטגיה עם שיעור הצלחה β. לכן

$$\begin{aligned}
E_{c\in I}\left[\mathbb{1}_{\{c\text{ מפסידים על }\}}\right] & \overset{(*)}{\leq}E_{c\in I}\left[\mathbb{1}_{s_{1}(c)\neq s_{2}(c)}\right]\\
 & \overset{(**)}{\leq}3\cdot E_{c\in I}\left[\frac{\sum_{i=1}^{3}\mathbb{1}_{s_{1}(c_{i})\neq s_{2}(c_{i})}}{3}\right]\\
 & \overset{(***)}{=}3\cdot\left(1-\beta\right)
\end{aligned}$$

$(*)$ מונוטוניות ההסת' : השחקנים אידאליים ולכן אם הפסוקית ניתנת להשמה תחת ההשמה (אסטרטגיה) שהוסכמה בהתחלה, שניהם ייתנו אותה. אם היא לא מסופקת תחת ההשמה שחקן 1 ידע את זה וישנה את ההשמה (שתספק ובתקווה תהיה זהה להשמת שחקן 2 למשתנה). לכן אם הם מפסידים הם בהכרח לא מסכימים על ההשמה לפסוקית (של שחקן 1 זו החדשה שהמציא עכשיו ממנה הוא חושף 3 ערכים למוודא ושל 2 היא המוסכמת במקור ממנה הוא חושף ערך אחד למוודא). $s_{1}\left(c\right),s_{2}\left(c\right)$ הן וקטורים ב-$\left\{ 0,1\right\} ^{3}$.

$(**)$ הכפלה וחלוקה ב-3 וגם חסם האיחוד על אי ההסכמה על ההסגר (לפחות אחד מהליטרלים לא מוסכם).

$(***)$ הצלחה היא לשכנע את המוודא שניתן לספק את הפסוקית (במרמה או לאו), ואי הסכמה יש רק כשההשמה המקורית לא מספקת את הפסוקית (כלומר הנוסחה לא ספיקה). לכן ההסת' לכישלון $1-\beta$ היא ההסת' לאי הסכמה בין השחקנים, שזה בדיוק תוחלת ממוצע אי ההסכמה במשוואה למעלה.

ולכן

$$P\left(\text{success}\right)=\beta\leq1-\frac{E_{c\in I}\left[\mathbb{1}_{\{c\text{ לא מסופק }\}}\right]}{3}\leq1-\frac{1-\alpha}{3}$$

$\blacksquare$

הגדרה משחק בין שני שחקנים עם סיבוב אחד (2 Player 1 Round Game) הוא שלשה $G=\left\langle V,P_{1},P_{2}\right\rangle$ כאשר:

- $P_{1}=\left\langle Q_{1},\Sigma_{1}\right\rangle ,P_{2}=\left\langle Q_{2},\Sigma_{2}\right\rangle$ הם השחקנים כאשר Q_{1},Q_{2} אוסף שאלות ו-Σ_{1},Σ_{2} אוסף תשובות.
- $V=\left\langle D,P\right\rangle$ הוא מוודא כאשר D התפלגות מעל $Q_{1}\times Q_{2}$ (לא בהכרח ב"ת) ו-P "פרדיקט" שהוא פ' $Q_{1}\times Q_{2}\times\Sigma_{1}\times\Sigma_{2}\mapsto\left\{ 0,1\right\}$.

ערך ההצלחה של המשחק הוא $\text{val}\left(G\right)=\sup_{\text{strategies}}P\left(\text{success}\right)$.

טענה נניח שאנחנו משחקים את המשחק למעלה עם שני השחקנים והנוסחה I שעבורה מתקיים $\text{val}\left(I\right)=\alpha$. אז ניתן לחשב את $\text{val}\left(G\right)$ בזמן סופי.

הוכחה: תוחלת ההצלחה במשחק היא α (ההסת' שניפול על פסוקית שסופקה ע"י ההשמה המקסימלית שלנו) כלומר

$$\alpha = E\left[\mathbb{1}_{\text{success}}\right] = E_{r_1,r_2}\left[E_{\textsf{strategies}}\left[\mathbb{1}_{\text{success}}\right]\right]$$

כאשר r_1, r_2 סרטי ביטים אקראיים (ככה ממודלת גישה לערכים אקראיים), והאסטרטגיות בתוחלת הפנימית למעשה עוברות דטרמיניזציה כי בהינתן סרט עם הערכים האקראיים שלו, האסטרטגיה נהפכת לדטר'. מתכונות התוחלת, יש לפחות אסטרטגיה אחת (א"ד שנהיית דטר' תחת סרט מקרי כלשהו) עם לפחות ערך α, שזה הכי הרבה שאנחנו יכולים להשיג. לכן מספיק שנעבור על כל האסטרטגיות הדטר' ונקבל $\mathrm{val}\left(G\right) = \max_{\text{det}' \text{ strategies}} P\left(\text{success}\right)$ כלומר שהאסטרטגיה שמשיגה sup היא מתוך קבוצה סופית. ■