מקליד: עידן אריה
מרצה: פרופ' עידו דגן
תאריך: 2015-11-29

מודלים הסתברותיים יישומיים במדעי המחשב
89-919-01

החלקת Held Out

r שכיחות באימון

$$t_r = \sum_{x:C^T(x)=r} C^H(x)$$

$$p_{HO}\left(x \middle| C^T(x) = r\right) = \frac{t_r}{N_r \cdot |S^H|}$$

N_0 מספר ערכי X שלא נצפו בS^T

נשים ♡ שיש כאן הנחה מובלעת שגודל X ידוע:

הנחה: $|X|$ ידוע.

פרקטית במידע ולא ידוע, צריך להניח את $|X|$, וזה הופך להיות פרמטר של השיטה

נשים ♡: הנוסחה לא מחייבת $\left|S^H\right| = \left|S^T\right|$.

$\left|S^T\right|$ משמש לאומדן הסתברות משמש לאומדן הסתברות של ערכי X ספציפיים

$\left|S^H\right|$ משמש לאומדן הסתברות של מחלקות שכיחות(ערכי r אפשריים)

במקרים שבהם יש יותר ערכי X מאשר מחלקות שכיחות נעדיף S^T גדול יותר

נראה שסכום האומדנים מסתכם ל1:

$$\sum_{x \in X} p_{HO}(x) = \sum_{r=0}^{\infty} \left(\frac{t_r}{\cancel{N_r} |S^H|} \cdot \cancel{N_r} \right) = \frac{\sum_{r=0}^{\infty} t_r}{|S^H|} = \frac{|S^H|}{|S^H|} = 1$$

ביצוע הפחתה(discount)רק לערכי r נמוכים יחסית(בכל שיטת הפחתה)

$(n_r \equiv N_r)$

מוטיבציה:
- ההפחתה חשובה לr נמוכים
- בr גבוהים ייתכנו אנומליות בהחלקה

נבצע החלקה רק ל$r \leq R$ עבור R גדול כלשהו. כדי שסכום האומדנים יסתכם ל1, נעשה זאת על ידי כך שנחלק לערכים עם $r = 0$ את אותה מסת הסתברות שהתפנתה מההחלקה. נרצה שיתקיים:

$$1 = n_0 \cdot p\left(x \middle| C^T(x) = 0\right) + \sum_{r=1}^{R} n_r \cdot p_d\left(x \middle| C^T(x) = r\right) + \sum_{r=R+1}^{\infty} n_r \cdot p_{MLE}\left(x | C(x) = r\right)$$

$$\Downarrow$$

$$p_d\left(x \middle| C^T(x) = 0\right) = \frac{1}{n_0}\left[1 - \sum_{r=1}^{R} n_r \cdot p_d(r) - \sum_{r=R+1}^{\infty} n_r \cdot p_{MLE}(r)\right]$$

נשים ♡: שיטת תיקון זו מניחה ש$\sum_{r=1}^{R} n_r p_d(r) < \sum_{r=1}^{R} n_r p_{MLE}(r)$. כלומר - ש$P_d$ אכן מבצע בפועל הפחתה של האומדנים ל$r = 1, \ldots R$. (אחרת נקבל אומדן שלילי ל$P_d(r=0)$).

שיטת החלקת Back-Off

מוטיבציה: אומדן בהסתברות מותנית לn-gram. למשל: ביגרם: $p(w'|w)$.

ניזכר - זו התפלגות מותנית נפרדת לכל w מותנה: $\sum_{w'} p(w'|w) = 1$.

האומדן דורש החלקה, בפרט לא נרצה להשתמש ב$p_{MLE}(w'|w) = 0$ למילים w' שלא נצפו אחרי w.

אם נשתמש בהחלקה רגילה(כגון Lid, H.O.), כל המילים w' כך ש$\text{freq}(w,w') = 0$ ($p_{MLE}(w'|w) = 0$) יקבלו את אותה הסתברות מותנית בw.

אבל: נסתכל על w', w'' כך ש

$$p_{MLE}(w'|w) = p_{MLE}(w''|w) = 0$$

$$p(w') > p(w'')$$

נרצה שייתקיים:

$$p_d(w'|w) > p(w''|w)$$

זו מטרת שיטת הB.O. - נחלק את המסה עבור מילים שלא נצפו אחרי w באימון באופן פרופורציונלי לשכיחות(היוניגרם) שלהם:

$$p_B(w'|w) = \begin{cases} p_d(w'|w) & C(w,w') > 0 \\ \alpha(w) \cdot p_d(w') & \text{otherwise} \end{cases}$$

כאשר:

P_d	שיטת ההחלקה הבסיסית שבה משתמשים
C	המונה במדגם האימון
$\alpha(w)$	מקדם נרמול שיבטיח סכום 1 להסתברויות(יקצה לw' במקרה השני($C(w,w') = 0$) את המסה שהתפנתה במקרה הראשון($C(w,w') > 0$))

חישוב $\alpha(w)$: נסמן $\beta(w)$ - מסת ההתסברות שהתפנתה ב(1) ע"י ההחלקה:

$$\beta(w) = 1 - \sum_{w':C(w'|w)>0} p_d(w'|w)$$

צריך שיתקיים:β שווה למסה שנקצה במקרה השני:

$$\beta(w) = \sum_{w':C(w,w')=0} \alpha(w) \cdot p_d(w')$$

$$\implies \alpha(w) = \frac{\beta(w)}{\sum_{w':C(w,w')=0} p_d(w')} = \frac{1 - \sum_{w':C(w,w')>0} p_d(w'|w)}{1 - \sum_{w':C(w,w')>0} p_d(w')}$$

(נשים לב שעוברים מ$\sum_{w':C(w,w')=0} p_d(w')$ ל$1 - \sum_{w':C(w,w')>0} p_d(w')$ בגלל שחישובית הרבה יותר קל לעבור על המילים שכן הופיעו **אחרי** w מאשר על המילים שלא הופיעו אחרי w)

אם $C(w) = 0$ אזי $\alpha(w) = 1$, וההסתברות $p(w'|w)$ מנוונת ל$p_d(w')$.