דילקמ: הירא ודיע
הצרמ: וגד ודיע 'פורפ
דיראת: 2015-12-06

שיערוך Back off

הרעיון: אם לא מוצאים בn-גרם, נחפש ב$n-1$-גרם

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

נסמן n-gram: $w_1,\dots w_n$. נשערך: $p\left(w_n\middle|\overbrace{w_1,\dots w_{n-1}}^{w_1^{n-1}}\right)$

למשל: הרעיון ל3-gram:

- אם אין סטטיסטיקה(שכיחות 0) לw_1^3, נעשה backoff לw_2^3
- אם אין סטטיסטיקה לw_2^3, נעשה backoff לw_3

נוסחה כללית לP_B לn-gram:

$$p_B\left(w_n\middle|w_1^{n-1}\right)=\begin{cases}p_d\left(w_n\middle|w_1^{n-1}\right) & C\left(w_1^n\right)>0\\ \alpha\left(w_1^{n-1}\right)\cdot p_B\left(w_n\middle|w_2^{n-1}\right) & \text{otherwise}\end{cases}$$

עבור יוניגרם: $p_B\left(w\right)=p_d\left(w\right)$, כי אין למה לעשות backoff...

מה זה α?

עבור מילה מתנה בודדת:

$$\alpha\left(w\right)=\frac{1-\sum\limits_{w':C(w,w')>0}p_d\left(w'|w\right)}{1-\sum\limits_{w':C(w,w')>0}p_d\left(w'\right)}$$

ועבור סדרת מתנה של $n-1$ מילים:

$$\alpha\left(w_1^{n-1}\right)=\frac{1-\sum\limits_{w_n:C\left(w_1^n\right)>0}p_d\left(w_n\middle|w_1^{n-1}\right)}{1-\sum\limits_{w_n:C\left(w_1^n\right)>0}p_d\left(w_n\middle|w_2^{n-1}\right)}$$

במילים: הסיכוי למילה לא ידועה אחרי w_1^{n-1} חלקי הסיכוי למילה לא ידועה אחרי w_2^{n-1}.

הרעיון הוא שרק אם אין לנו מידע אנחנו עושים backoff.

דילקמ: הירא ודיע
הצרמ: וגד ודיע 'פורפ
דיראת: 2015-12-06

גישות אינטרפולציה לאומדן

יש לנו כמה דרכים שונות לשערך התסברויות, ואנחנו רוצים לעשות אינטרפולציה בין התוצאות שלהן.

הרעיון: נשלב בצורה ממושקלת מספר אומדנים.

דוגמאות:

- שילוב בין מודל ביגרם למודל נאיבי($0 < \lambda < 1$):

$$p_{\text{int}}\left(w'|w\right) = \lambda \cdot p_{ML}\left(w'|w\right) + \left(1-\lambda\right) \cdot P\left(w'\right)$$

- מודל יוניגרם לטקסטים ברפואה: נשלב מודל שנלמד מטקסטים ברפואה למודל מטקסטים כלליים:

$$p\left(w\right) = \lambda \cdot P_{\text{med}} + \left(1-\lambda\right) \cdot p_{\text{gen}}\left(w\right)$$

- באופן דומה לשקלול מספר מודלים:

$$p\left(w_3|w_1,w_2\right) = \lambda_1 \cdot p\left(w_3|w_1,w_2\right) + \lambda_2 \cdot p\left(w_3|w_2\right) + \lambda_3 \cdot p\left(w_3\right)$$

כך ש $\sum_{i=1}^{3} \lambda_i = 1$

איך מכיילים?

- במקרה פשוט, של מעט פרמטרים λ_i, ניתן לבצע כיול אמפירי ע"י ניסוי מספר ערכים.

נשים ♡: במודל n-gram ניתן לקבוע צירוף מקדמים שונה לכל התפלגות, כלומר עבור כל מאורע מתנה.
בפרט: נצפה לערכי λ אופטימליים גבוהים יותר מאורעות מתנים שכיחים יותר.
ניתן לקבץ מאורעות מתנים לפי שכיחויות ולכייל מקדמים ביחד לכל קבוצה.

- יש מודלים של אינטרפולציה שעבורם קיימים אלגוריתמים יעילים לכיול המקדמים.

בפרט: לn-gram יש אלגוריתם ME

דילקמ: הירא ודיע
הצרמ: וגד ודיע 'פורפ
דיראת: 2015-12-06

משתנים חבויים - Hidden Variables

לפעמים נרצה לחשב אספקטים נוספים של התופעה, שאנחנו לא צופים בהם, כדי למדל יותר טוב את התופעה.

סכימה כללית

נתונה תופעה נצפית Y, בד"כ מורכבת $y = (\langle y_1, \dots y_n \rangle)$.

דוגמה: סקירת מילים/ידיעה/מסמך

מניחים שקיימת התפלגות משותפת של Y עם משתנה חבוי נוסף X

דוגמה: הדסק במערכת שבו נכתבה הידיעה

בדר"כ נתעניין בX דיסקרטי שמקבל מספר סופי של ערכים "אטומיים".

בתהליך הגנרטיבי שנניח: כדי לייצר תצפית $y \in Y$ קודם נבחר $x \in X$ לפי $p(x)$, ואח"כ נייצר y לפי $p(y|X = x)$.

שימושים במודלים חבויים

כללית: גם כשאין שימוש לערכי X, מידול בעזרת X יכול לשפר את מידול Y.

במקרים אחרים יש עניין יישומי בערכי X:

- בסיטואציות אופייניות של למידה/סיווג:

1. למידה/סיווג מבוקר/מפוקח(Supervised):
   ערכי X ידועים, ויש מדגם אימון שבו לכל $y \in Y$ שנצפה ידוע(מתוייג) מה ערך $x \in X$ המתאים לאותו y.
   - ממדגם כזה נשערך את ההתפלגות המשותפת.
   - יישום: בהינתן y, לשערך את $p(x|y)$. ובפרט: לסווג את הy לx המתאים.
2. למידה לא מפוקחת(Unsupervised)
   לא ידועה מראש זהות ערכי X, ואין מדגם מתוייג.
   המטרה: לקבל(קלסטרינג) את התצפיות של הy לפי ערכי $x \in X$ שכנראה ייצרו אותם.
   מטרת הלמידה: שערוך θ, אוסף הפרמטרים של $p(x, y; \theta)$

מודל חבוי - עירוב היסטוגרמות Mixture of Histograms

(זה בעצם עירוב של מולטינומים)

נמחיש: כמודל ליצירת מסמכים(Y) לפי נושאים(X).

נתייחס למסמך כהיסטוגרמת שכיחויות של המילים שמופיעות בו(זה המודל המולטינומי שלמדנו ליוניגרם).

נסמן: המילים נלקחות ממילון בגודל v: $w_1, \dots w_v$

מדגם מסמכים: $y = y_1, \dots y_N$(N מסמכים במדגם)

נניח אי תלות בין המסמכים.

ייצוג מסמך y_t: כהיסטוגרמה של v שכיחויות $y_t = (n_{t1}, n_{t2}, \dots n_{tv})$ כאשר n_{tk} היא שכיחות המילה הw_k במסמך y_t

לפשטות: נניח אורך מסמכים קבוע n: $\forall_t \sum_{k=1}^{v} n_{tk} = n$

דילקמ: הירא ודיע
הצרמ: וגד ודיע 'פורפ
דיראת: 2015-12-06

מודלים הסתברותיים יישומיים במדעי המחשב
89-919-01

מודל עם משתנה חבוי

נניח קיום:

- מ"מ נוסף X עם ערכים $x_1, \ldots x_{|X|}$
  (בדוגמה המסמכים:מייצג קטגוריה, או קלסטר)
- התפלגות משותפת $p(x, y; \theta)$[1]

תהליך גנרטיבי ליצירת מסמך

1. נבחר $x \in X$(קטגוריה) לפי התפלגות $p(x) = P(X = x)$
2. נקבע את המילים בy לפי התפלגות $p(y_t|X = x)$

בפרט: נניח עבור $p(y_t|x)$ מודל מולטינומי נפרד לכל $x \in X$

בהתאם - הפרמטרים θ של המודל

$$\forall_{x_i \in X} p(x_i)$$

$$\forall_{x_i \in X} \forall_{w \in w_1, \ldots w_v} p(w_k|x_i)$$

לפי המודל המולטינומי:

$$p(y_t|x_i) = \prod_{k=1}^{v} p(w_k|x_i)^{n_{tk}}$$

נתעניין ב3 שאלות:

1. $p(y_t; \theta)$ - הסתברות יצירת תצפית
2. $p(x_i|y_t; \theta)$ - הסתברות הסיווג/שיוך לקלסטר
3. אומדן $\hat{\theta}_{ML}$ - נסיון ערכי ML לפרמטרים

[1] מימין לנקודה-פסיק(;) רושמים את הפרמטרים של ההסתברות