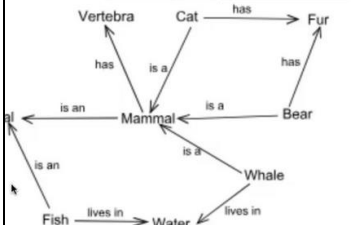

$$\theta^{t+1} = \theta^t - \mu \cdot (\hat{y} - y) \cdot f$$

רשתות סמנטיות (Semantic Nets)



האם אוסף השינונים באמת מייצג יחס

דמיון לשוני אמין?

- ניתן לשאול אנשים עד כמה המילים דומות. זה סובייקטיבי.
- נשתמש בדמיון קוסינוס

$$\cosine(\mathbf{v}, \mathbf{w}) = \frac{\mathbf{v} \cdot \mathbf{w}}{\|\mathbf{v}\| \|\mathbf{w}\|} = \frac{\sum_{i=1}^N v_i w_i}{\sqrt{\sum_{i=1}^N v_i^2} \sqrt{\sum_{i=1}^N w_i^2}}$$

מטריצת מילים/מסמכים: ניצג כל מילה

לפי אוסף המסמכים שהיא מופיעה בו

וכמה בכל מסמך.

חסומות של המטריצה הנ"ל:

- וקטורים ארוכים לכל מילה.
- מילה כמו תהיה דומה לשרא המילים functionablen ויהיה קשה

ללמוד עליהן משהו.

מטריצת שכנויות- נספור הפעוּת של זוגות מילים

באותו מסמך w המילה שמעניינת אותנו, k ההקשר

	computer	data	result	pie	sugar	count(w)
cherry	2	8	9	442	25	486
strawberry	0	0	1	60	19	80
digital	1670	1683	85	5	4	3447
information	3325	3982	378	5	13	7703

count(context) 4997 5673 473 512 61 11716

PMI - זהו: נוסחא שעוזרת לנו להתגבר על כך שחלק מהמילים שיכוחות יתור וחלק מהמילים פחות בכך שאנחנו מנרמלים את ההסתברויות

$$PMI(w, c) = \log_2 \frac{P(w, c)}{p(w)p(c)}$$

$P(w, c)$  - ההסתברות ש w תופיע במסמך עם ההקשר c

$P(w)$  - ההסתברות שש תופיע במסמך

$P(c)$  - ההסתברות של ההקשר

כשהתוצאה תהיה שווה לאפס זה אומר שיש אי תלות בין המילים

PMI - ביצוע של RELU על pmi (רק חיובי)

שיכון ע"י חיזוי הקשרים מקומי - Skip

Gram. נפועל אלגוריתם חיזוי, אבל נשמר

רק את המשקולות הנלמדים.

האלגוריתם: הבינתן מילה w, ננסה

לחזות את השכנות שלה c בהסתברות

גבוהה ככל הניתן, ונעבור כך על כל

המילים בקורפוס. בעצם נחשב מה

ההסתברות ש w, c מופיעות ביחד

$$P(+|w, c)$$

לכל מילה יהיה וקטור מטרה בתפקידה

כחזרה, ווקטור הקשר בתפקידה כנחזית.

הפר פרמטרים- רוחב החלון (בכל פעם בדגום רק "חלון" בסביבת המילה וכך נגדיר מה יהיו מילות ההקשר, כמות הדוגמאות השליליות לכל דוגמה חיובית, מימד הוקטור, בדגום את המילים השליליות מתוך התפלגות שמוסחת לטובת מילים נדירות, אלפא = 0.75 זה ערך טוב.

$$P_{\alpha}(c) = \frac{count(c)^{\alpha}}{\sum_c count(c)^{\alpha}}$$

יעד החיזוי: נשתמש בגרסיה לוגיסטית

על מנת לחשב את ההסתברות של

p(c|w) על כל דוג' חיובית ניקח כמה

שליליות:

$$P(+|w, c_{pos}) \prod_i P(-|w, c_{neg_i})$$

סקיפגראם- הפסד:

$$L_{CE} = -[\log \sigma(c_{pos} * w) + \sum_{i=1}^K \log \sigma(-c_{neg_i} * w)]$$

סקיפגראם- גרדיאנט:

$$\frac{\partial L_{CE}}{\partial c_{pos}} = [\sigma(c_{pos} * w) - 1]w$$

$$\frac{\partial L_{CE}}{\partial c_{neg}} = [\sigma(c_{neg} * w)]w$$

$$\frac{\partial L_{CE}}{\partial w} = [\sigma(c_{pos} * w) - 1]c_{pos} + \sum_{i=1}^K [\sigma(c_{neg_i} * w)]c_{neg_i}$$

סקיפגראם- עדכון:

$$w^{t+1} = w^t - \eta [\sigma(c_{pos} * w^t) - 1]c_{pos} + \sum_{i=1}^K [\sigma(c_{neg_i} * w^t)]c_{neg_i}$$

הצגה של וקטורים במימד נמוך ע"י PCA

או ע"י שיכון שכנים סטוכסטי לפי

התפלגות t (t-SNE)

מטריקות לשערור מודלי דימיון:

Pearson correlation:

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$$

את התחום. רגיש לפיזור הערכים השונים

בטווח של X.

Spearman correlation:

$$r_s = \frac{cov(R(X), R(Y))}{\sigma(R(X))\sigma(R(Y))}$$

R - דירוג, ממיינים את הציונים של כל אחת מהרשימות

סמנטיקה גבוהה- ייצוג משמעות של

ביטוי באמצעות הרכבת חלקיו.

(קשור R(AGENT,THEME) וכו').

3 גישות לסימון תפקידים תמטיים:

1. VerbNet: כתובים שמות התפקידים

התמטיים.

2. PropBank: רק מספור של

הארגומנטים, שמירה על הסדר.

3. FrameNet: כתובים שמות התפקידים

הספציפיים.

	gave	Boyang	a book
VerbNet	AGENT	RECIPIENT	THEME
PropBank	ARG0: giver	ARG2: entity given to	ARG1: thing given
FrameNet	DONOR	RECIPIENT	THEME
	taught	Boyang	algebra
VerbNet	AGENT	RECIPIENT	TOPIC
PropBank	ARG0: teacher	ARG2: student	ARG1: subject
FrameNet	TEACHER	STUDENT	SUBJECT

רובוטריקים

השלים בself-attention:

1. השוואה בין הווקטורים בקלט ומתן ציון

לדמיון ביניהם.

2. נרמול הציונים שקיבלנו ע"י שימוש

באמצעות softmax (הסתברויות).

3. חישוב הפלט הנוכחי.

$$y_i = \sum_{j=1} \alpha_{ij} x_j$$

כל מילה שואלת כמה היא צריכה לקחת מכל מילה

ברמת הייצוג הבאה

Query - המילה המתשאלת,

Key - המילה כגורם השוואה

Value - המילה כמרכיב של שכבה הבאה

$$Self\ Attn(Q, K, V) = softmax\left(\frac{Q \cdot K^T}{\sqrt{d_k}}\right) \cdot V$$

בהינתן 3 מילים  $x_1, x_2, x_3$  ו-  $v_1, v_2, v_3$  בהתאמה, נקבל לדוגמה עבור  $y_1$ :

$$y_1 = \sum_{i=1}^3 q(x_1) \cdot k(x_i)^T \cdot v(x_i)$$

$$Q = X \cdot W_q$$

$$K = X \cdot W_k$$

$$V = X \cdot W_v$$

$$q(\vec{x}_i) = \vec{x}_i \cdot W_q$$

$$k(\vec{x}_i) = \vec{x}_i \cdot W_k$$

$$v(\vec{x}_i) = \vec{x}_i \cdot W_v$$

$$P(+|w, c)$$

מבנה הרובוטריקי:



הקשרים השיווריים נותנים לשכבות

העליונות גישה ישירה למידע המגיע

מהשכבות הנמוכות.

$$x = \text{LayerNorm}(x + \text{SelfAttn}(x))$$

$$y = \text{LayerNorm}(x + \text{FFNN}(x))$$

בנומליזציה נחשב:

$$\mu = \frac{1}{d_k} \sum_{i=1}^d x_i$$

$$\sigma = \sqrt{\frac{1}{d_k} \sum_{i=1}^d (x_i - \mu)^2}$$

$$\hat{x} = \frac{(x - \mu)}{\sigma}$$

ואז ננרמל:

צומי מרובה ראשים

Attention: כמה שכבות של self-

attention שנקראות ראשים, לכל ראש יש

סט מטריצות משלו. כל ראש יכול למצוא

קשורות שונות בין המילים במשפט.

\*נוסיף לשכבת הצומי מיקום עם

הpositional embeddings.

למידה בהעברה: למידת מערכת למטרה

מסוימת ושימוש באלמנטים הנלמדים

שלה למשימה אחרת. (skip-gram)

Fine-tuning: עדכון הפרמטרים שנלמדו

למשימה המקורית על גבי המשימה

החדשה.

חיולול טקסט (seq to seq)

קלט: רצף מילים או משהו שכולל רצף

מילים כמו תמונות או קולות.

פלט: רצף מילים כלשהו.

קשיים במעבר בין שפות:

1. סדר המילים שונה.

2. אי תאימות במשמעות.

3. ביטוי בהטיה מורפולוגית (זמן וגוף)

לעומת מילות עזר.

4. אלמנטים חסרים (כינוי וגוף)

משולש וקוואל: >syntax- >semantics->interlingua

מודל ה"ערנץ הרועש": בהינתן המשפט

ה"מרועש" מהו ה"מקור" שאותו הרעישו

בשפת היעד.

$$w^* = \arg \max_w P(w|a) = \arg \max_w P(a|w) \frac{P(w)}{P(a)} = \arg \max_w P(a|w) \frac{P(w)}{P(a)}$$

$$p(w^{(1)}, \dots, w^{(T)} | a) = \prod_{t=1}^T p(w_t^{(1)}, \dots, w_t^{(T)} | m, M^{(1)}, M^{(T)})$$

$$= \prod_{t=1}^T \prod_{i=1}^n p(w_t^{(i)} | m_t, M_t^{(1)}, M_t^{(T)})$$

מודל תרגום ברמת המילה:

תרגום צירופים (many2many):

	Asha	gave	Boyang	a book
VerbNet	AGENT	RECIPIENT	THEME	
PropBank	ARG0: giver	ARG2: entity given to	ARG1: thing given	
FrameNet	DONOR	RECIPIENT	THEME	
	Asha	taught	Boyang	algebra
VerbNet	AGENT	RECIPIENT	TOPIC	
PropBank	ARG0: teacher	ARG2: student	ARG1: subject	
FrameNet	TEACHER	STUDENT	SUBJECT	

מטריקות לבדיקת חיולול טקסט:

Backtranslation- נאמן מודל תרגום

משפט היעד לשפת המקור, ואז נשתמש

במה שיצרו בתור דאטה לאימון לתרגום

מהמקור ליעד. דוגמה ללמידה חצי

מפוקחת

1. גישת החפיפה - נחליט שיש טקסט מושלם אחד

(reference) ונשווה בין הרצפים שלו ושל תרגום

המערכת (אפשר עם רצפי תווים או אותיות

CHRF2,2) (17 unigrams)

$$\text{unigram P: } 17/17 = 1 \quad \text{unigram R: } 17/18 = .944$$

$$\text{bigram P: } 13/16 = .813 \quad \text{bigram R: } 13/17 = .765$$

$$\text{chrP} = (17/17 + 13/16)/2 = .906$$

$$\text{chrR} = (17/18 + 13/17)/2 = .855$$

$$\text{chrF2,2} = 5 \cdot \frac{\text{chrP} + \text{chrR}}{4} = .86$$

REF: witness for the past, HYPI: witness of the past

19

2. מרחב- נלמד מודל שפה גדול ונשווה בין המשפטים ע"י

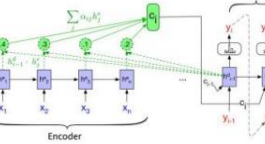
ציוני מודל השפה

הגדול (משפט למשפט או מילה למילה לפי יישור)

מקודד (encoder)- לוקח טקסט (או אובייקט אחר)

ומחזיר וקטור

מפוענח (decoder)- לוקח וקטור (או כמה) ומחזיר טקסט



מקודד רובוטריקי- כמו קודם אבל יכול לראות את כל

הקלט

מפוענח רובוטריקי- יכול להסתכל רק אחורה בפלט, אבל

יכול להסתכל על כל הקלט בעזרת cross-attention

המקודד והמפוענח:

1. מודל שפה- מסתכל רק אחורה autoregressive

ללמידה מעוברת

2. מודל חיזוי או מקודד לבעיות תרגום או מוסקר (מסתכל

קדימה)

מודל שפה דו כיווני ELMo: חיבור של מודל קדימה

ומודל אחורה

BERT- רובוטריקי כמודל שפה דו כיווני. נסתיר לחלק

את המילים את עצמן. יכול למשמש כמקודד תרגום.

חיזוי קורפרנס- אזכורים של אותה ישות יוצרים cluster

ואם מניין לפי סדר הופעה נקבל שרשרת

Mention pairs - לכל אזכור חדש נחשב התבחרות

קישור עבור כל אזכור קודם (האם הם קשורים לאותה

יישות) - אז לאזכור החדש עם כל אזכור קודם

Mention rank - נבדוק עבור האזכור את כל האזכורים

לפניו ובחר את זה עם ההסתברות המקסימלית, האזכור

עם הציון הגבוה הוא זה שיש לו הכי הרבה משמעות

עבור הבנת האזכור הקיים (barak Obama, president)

Entity based - בהינתן אזכור נבחר את השרשראות

(הקלאסטריות)=ישויות) שהוא יכול להצטרף אליהן

בגדרת שימוש בפצ'רים תלוי ישות ותלוי שרשרת

דוגמאות: אזכורים שכבר ראינו: vc.her,mb

האזכור החדש: she

Mention pairs - (she,her), (she,vc), (she,mb)

Mention rank - (she,vc,her,mb)

Entity based - (she,vc,her,mb)

לניחות טקסט- קוהרנטי- טקסט הוא קוהרנטי אם ניתן

למצוא קשר בין שני משפטים עוקבים

קוהסיטי- טקסט הוא קוהסיטי אם יש נושא משותף

דני הלך לים, הוא אוהב לשחות- קוהסיטי וקוהרנטי,

המשפט הראשון הוא הגרעין והשני תלוי

דני הלך לים, הוא אוהב תות - קוהסיטי ולא קוהרנטי

קורפוס reference - ציטוטים

שומעים על אותו יישות.

סוגי אזכורים: (ושמות)

1. צירופים שמייים בלתי מידעיים: קראתי

אתמול כתבה מעניינת.

2. צירופים שמייים מידעיים: לחצתי על

קישור לכתבה במקרה.

3. כינוי: גוף: לקחתי אותן מהגן.

4. אזכור ריק: יובל קרא את הכתבה ואז

הלך הביתה.

מטריקות:

בעיה קשה. האם מעניינים אותנו האזכורים,

הישויות, או הקשרים שבין הישויות?

נמדוד קשרים: לכל אזכור, כמה מבין הקשרים

לשאר אזכורי הישויות נמצאו ע"י המערכת?

נבנה מערכת שמקשרת כל אזכור לזה שקדם לו

(שרשרת יחידה) ונקבל תוצאה מאוד גבוהה

נמדוד אזכורים: לכל אזכור, כמה מתוך

השרשרת שהענקנו נמצאת גם בשרשרת

האמיתית?

נבנה מערכת ששירתת רק יחידים (singletons) ונקבל

recall=1, precision=1 בלי לפגוע בלי

נמדוד יישויות (צבירים): האם נמצאו צבירים

דומים