

תאריך: 15.2.2017

**מבחן מועד א' בקורס עיבוד שפה טבעית (67658)**

שנת לימודים התשע"ז 2016/2017

משך המבחן: שעותיים

מרצה הקורס: ד"ר עמרי אבנד

השימוש בכל חומר עזר אסור, נא כתבו על הכריכה על אילו שאלות עניתם.

**חלק א' (80 נקודות): ענו על בדיוק שתי שאלות מבין שאלות 1-3**

**שאלה 1 (40 נקודות):**

- א. תארו את שיטת ההחלקה Add- $\delta$  (Laplace) Smoothing עבור מודל שפה יוני-גרם (Unigram Language Model) המחושב ע"פ אומד נראות מרבית (Maximum Likelihood Estimator, MLE). בהינתן שאנו משערכים את המודל על פי קורפוס בו יש 100 מילים שונות, ובו השכיחות של המילה  $w$  היא 5% (כלומר 5% מה- word tokens הם  $w$ ), האם שיטת החלקה זו מעלה את ההסתברות שלה  $\Pr(w)$  או מקטינה אותה? הסבירו את תשובתכם.
- ב. נתון מודל שפה בי-גרם (Bigram Language Model) בו הסתברויות המעבר  $\Pr(w_i|w_{i-1})$  משוערכות בעזרת אינטרפולציה ליניארית עם מודל שפה יוני-גרם. הסבירו כיצד אינטרפולציה עם Unigram Model שמחושב ע"פ אומד נראות מרבית עשויה להקל את בעיית ה-sparsity (דלילות).
- ג. הציגו את שיטת Kneser-Ney במקרה של מודל בי-גרם. הסבירו איזו בעיה בשיטת האינטרפולציה עם מודל יוני-גרם המשווער ע"י אומד נראות מרבית (אותו תיארתי בסעיף הקודם) מודל זה פותר. לוו את ההסבר בדוגמא.
- הערה:** ניתן להניח שה-discounted term, כלומר הגורם המוחסר, מקיים  $0 < d < 1$ .
- ד. נתבונן בשיטת החלקה אחרת למודל שפה בי-גרם. נסמן ב-  $\text{Count}(w, u)$  את מספר הפעמים בהם הופיע הבי-גרם  $(w, u)$  בקורפוס האימון. נסמן ב-  $\text{Count}(u)$  את סך כל מספר הפעמים ש-  $u$  הופיע באיזושהו בי-גרם. נסמן ב-  $N$  את סך מספר הבי-גרמים בקורפוס. יהי  $0 < C < 1$  קבוע. נגדיר מודל שפה כך שהסתברויות המעבר שלו הן:

$$q_{BO}(w_i|w_{i-1}) = \begin{cases} \frac{\text{Count}(w_{i-1}, w_i) - C}{\text{Count}(w_{i-1})} & \text{if } \text{Count}(w_{i-1}, w_i) > 0 \\ \alpha(w_{i-1}) \cdot \frac{\text{Count}(w_i)}{N} & \text{if } \text{Count}(w_{i-1}, w_i) = 0 \end{cases}$$

הראה שלכל  $w_{i-1}$  ניתן למצוא ערך ל- $\alpha(w_{i-1})$  כך ש- $q_{BO}(w_i|w_{i-1})$  יהווה התפלגות (כלומר כל הסתברויות המעבר יהיו אי-שליליות ויסתכמו ל-1 על פני כל הערכים האפשריים ל- $w_i$ ).

**שאלה 2 (40 נקודות):**

- א. נתון מודל Hidden Markov Model (HMM) כך שה-emitted value (הערך הגלוי) הוא זוג. נסמן את הערך הגלוי ב-  $(x_i, w_i) \in X \times W$  לכל  $i$  בשרשרת. המצב החבוי (hidden state) יסומן ע"י  $y_i \in Y$ . רשמו את הנוסחא להתפלגות המשותפת של המשתנים המקריים בשרשרת:

$$(x_1, w_1, y_1), (x_2, w_2, y_2), \dots, (x_n, w_n, y_n)$$

ב. מהם הפרמטרים של המודל בסעיף א? כמה פרמטרים בסה"כ יש למודל (הניחו שגדלי הקבוצות  $X, Y, W$  הן  $n_x, n_y, n_w$  בהתאמה)?

ג. הניחו כעת שלכל  $i$ , המשתנים המקריים  $x_i$  ו- $w_i$  הם בלתי תלויים בהינתן  $y_i$ . רשמו מחדש את הנוסחא להתפלגות המשותפת של המשתנים המקריים במודל. כמה פרמטרים יש במודל זה?  
ד. הסבירו מהי בעיית ה-sparsity (דלילות) בשערוך פרמטרים. כיצד הנחת אי-התלות בסעיף הקודם עשויה להקל עליה? ניתן להניח ששערוך הפרמטרים נעשה בשיטת אמידת נראות מרבית (Maximum Likelihood Estimation).

שאלה 3 (40 נקודות):

א. הגדירו את מודל הערוץ הרועש לתרגום (Noisy Channel Model).  
ב. הגדירו את מודל התרגום IBM Translation Model 1 בו השתמשנו כמודל לבעיית ה-word alignment. אילו פרמטרים יש למודל?  
ג. בהינתן ערכים לפרמטרים של IBM Model 1, ובהינתן זוג משפטים  $e, f$  כך ש- $e$  הוא תרגום של  $f$ , הראו כיצד ניתן לחשב את ה-word alignment הסביר ביותר בין המילים של שני המשפטים.  
ד. הראו ש-IBM Model 1 איננו רגיש לחילוף בסדר המילים. כלומר הראו שאם  $f_1$  ו- $f_2$  משפטים בשפת המקור שמתקבלים זה מזה ע"י החלפת שתי המילים הראשונות (כלומר, המילה הראשונה ב- $f_1$  היא המילה השנייה ב- $f_2$ , המילה השנייה ב- $f_1$  היא המילה הראשונה ב- $f_2$  וכל שאר המילים ב- $f_1$  ו- $f_2$  זהות), אז  $\Pr(f_1|e) = \Pr(f_2|e)$ , בהינתן פרמטרים למודל ומשפט בשפת היעד  $e$ .

**חלק ב' (20 נקודות): ענו על בדיוק שאלה אחת מבין שאלות 4-5**

שאלה 4 (20 נקודות):

א. עבור משימת תיוג חלקי דיבר לא מבוקרת (Unsupervised Part of Speech Tagging), הסבירו מדוע לא ניתן להשתמש בממד Accuracy המשמש במשימת תיוג חלקי דיבר מבוקרת (Supervised Part of Speech Tagging).  
ב. הגדירו את ממד Mapping-based Accuracy עבור Unsupervised Part of Speech Tagging עבור גרסת many-to-one mapping. הסבירו כיצד ניתן לחשב את הממד בהינתן קורפוס מתויג ע"י מערכת לא מבוקרת לתיוג חלקי דיבר וגם ע"י תיוג Gold Standard.

שאלה 5 (20 נקודות):

א. בשאלה זאת נעסוק ב-MST Parser שנלמד בכיתה עבור Dependency Parsing. רשמו את פונקציית הניקוד (scoring function) עבור עצי גזירה (parse trees) בה האלגוריתם משתמש. מהם הפרמטרים של המודל?  
ב. כתבו פסאודו-קוד לאלגוריתם הפרספטרון (Perceptron) המשמש ללמידת הפרמטרים של ה-MST Parser. ניתן להניח שקיימת פונקציה מוכנה המקבלת גרף מכוון מלא ממושקל (weighted complete directed graph) ומחזירה עץ פורש מכוון מקסימלי שלו (maximum directed spanning tree), ואין צורך לרשום כיצד היא ממומשת.