

202-2-5211 עיבוד שפה טבעית — תרגיל 3

מועד הגשה: 20 בפברואר 2024, 13:59

משקל התרגיל: 8 נקודות מהציון הסופי. מספר הנקודות של כל שאלה מופיע בסוגריים. ניתן להגיש את התרגיל בצמדים. ניתן להתייעץ עם סטודנטים אחרים אך בסופו של דבר על הפתרון להיות של המגישה. יום בלבד.

1 רשתות נשנות (4)

נזכיר את ההגדרה של מידול שפה: מידול שפה הוא הבעיה של חיזוי ההסתברות עבור טקסט נתון:

$$P(w_1, w_2, \dots, w_N)$$

ואשר לרוב נהוג לפרק לחיזוי מילה-מילה:

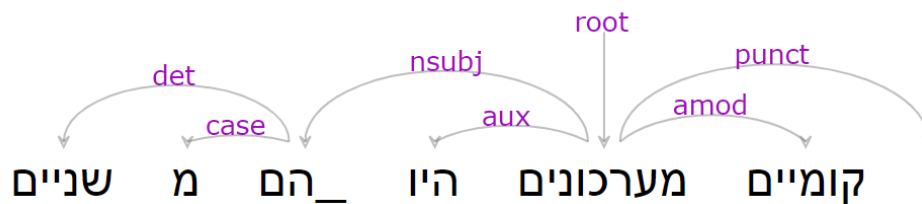
$$= P(w_1 | \text{START}) \cdot P(w_2 | w_1) \cdot \dots \cdot P(w_k | w_1, \dots, w_{k-1}) \cdot \dots \cdot P(w_N | w_1, \dots, w_{N-1}),$$

כאשר START הוא תבנית מיוחדת שלא מופיעה בטקסט ומסמנת תחילת מחרוזת. בניגוד למידול שפה באנגרמים מהתרגיל הראשון, באמצעות רשתות נשנות אפשר להתייחס לכל אחת מההסתברויות לעיל במלואן, תיאורטית ללא כל אובדן הקשר. נגדיר עבור הבעיה רשת נשנית רגילה, חד-כיוונית, שבה המילים מוזנות דרך שכבת שיכון (embedding) והמצב הסמוי משמש כקלט לשכבה לינארית שתפקידה לחזות את ההסתברות המילה הבאה.

מומלץ בשלב זה לצייר סקיצה של הרשת, אך אין צורך להגיש כזו.

1. רשמו במדויק את משוואת פונקציית ההפסד של המודל במהלך למידת משפט בן עשרים מילים. תנו שמות עבור כל המשתנים והקבועים הנחוצים ורשמו מקרא.

2. בשלב המבחן נרצה לראות האם המודל משלים לבד משפט באורך $n > 5$ לאחר שהוזנו לו חמש המילים הראשונות. איזו בעיה יכולה להיווצר בהערכה של ביצועי המודל? הסבירו. (יש הרבה תשובות אפשריות).



איור 1: עץ תלויות עבור המשפט בשאלה 2.

2 ניתוח תלויות (4)

לפניכם משפט מתוך פרויקט התלויות האוניברסליות בעברית:

שניים מהם היו מערכונים קומיים.

העץ שלו, כפי שתויג ידנית ע"י מומחה.ית, נתון באיור 1 משמאל לימין על-מנת לא לבלבל את שמות המעברים שלמדנו.

1. בתרגום האנגלי של המשפט, מה סוג (label) הקשת שתופיע בכיוון הפוך לזה שבמקור העברי?
 2. רשמו את סדר פעולות המעבר (transition) לפי הסכימה עליה עברנו בכיתה, arc-standard.
 3. רשמו את המצב המלא של המנתח, על כל חלקיו, לאחר הפעולה השביעית מהסדר שבסעיף הקודם.
 4. חשבו את פעולות המעבר הנדרשות בסכימת arc-eager, עליה תוכלו למצוא פרטים בפרק 18.2.4 בספר של סטנפורד (אין צורך לרשום אותן). כמה פעולות נוספות נדרשות לעומת התשובה מהסעיף הראשון? למספרם של איזה אלמנטים מהגרף שווה המספר הזה?
- שימו לב: לצורך שאלה זו הניחו כי פעולת rightArc מחולקת לחיבור הקשת ולהזחה בנפרד.