שאלה 1:

סעיף a:

1. משפטים עם מילים בעלות דוד משמעות שמית:

Ford makes the best cars.

המילה Ford במשפט מכוונת ליצרנית המכוניות, אף אותה מילה יכולה לשמש גם כשם של בן אדם.

I can get to JFK with a taxi.

המילה JFK מתייחסת לשדה התעופה בניו יורק, אך היא גם יכולה לשמש ככינוי שך נשיא ארה"ב לשעבר.

ii על ידי שימוש בפיצ'רים למעט המילה עצמה ניתן לתת לה הקשר בתוך המשפט, שיעזור לפתור את בעיית הדו-משמעות. בנוסף, מילים שהן named entities עלולות להיות נדירות באימון ולכן הפיצ'רים עוזרים בהכללה.

iii דוגמאות לפיצ'רים: (נניח כי המילה היא wi, במיקום i במשפט)

* Wi-1 is part of a named entity **and** wi starts with a capital letter.
* Wi-1 is a prepesition that refers to a location (e.g. at, inside, out of…)

סעיף b:

מימדים:

, ,

סיבוכיות חיזוי עבור משפט באורך T: (נניח גודל חלון w)

חישוב : מכפילים פעמים את וקטור המילה (one hot) עם המטריצה E המכילה וקטורי קידוד באורך D – סה"כ .

חישוב : .

חישוב **:**

סה"כ עבור מילה בודדת במשפט:

ולכן עבור משפט באורך T:

סעיף d:

התוצאות הטובות ביותר עם דיוק של 83% F1:

2019-06-08 11:25:13,782:DEBUG: Token-level confusion matrix:

go\gu PER ORG LOC MISC O

PER 2967 49 55 12 66

ORG 143 1638 120 53 138

LOC 60 94 1873 20 47

MISC 40 54 52 1009 113

O 58 54 15 26 42606

2019-06-08 11:25:13,782:DEBUG: Token-level scores:

label acc prec rec f1

PER 0.99 0.91 0.94 0.92

ORG 0.99 0.87 0.78 0.82

LOC 0.99 0.89 0.89 0.89

MISC 0.99 0.90 0.80 0.85

O 0.99 0.99 1.00 0.99

micro 0.99 0.98 0.98 0.98

macro 0.99 0.91 0.88 0.90

not-O 0.99 0.89 0.87 0.88

2019-06-08 11:25:13,782:INFO: Entity level P/R/F1: 0.82/0.85/0.83

ניתן להסיק מה-confusion matrix שהטעויות המשמעותיות הן:

True class: ORG, predicted class: PER,LOC,O

True class: MISC predicted class: O

באופן כללי רואים שהמודל הכי מתקשה עם זיהוי ORG לפי מספר השגיאות בשורה זאת בטבלה.

מגבלות המודל:

1. המודל לא משתמש בהמשכיות של תיוגים שביצע על מילים קודמות, כלומר הוא לא מתחשב בעובדה שרצף של מילים יכול להוות שם אחד של אותה יישות ולכן יש לתייג את כולן בצורה זהה. לדוגמה:

x : Jordan won the Samsung Tel Aviv marathon

y': LOC O O ORG ORG LOC O

האירוע "מרתון סמסונג תל אביב" מהווה ישות אחד, אך המודל תייג חלק מהמילים בצורה שונה.

1. במשפטים ארוכים המודל אינו מסוגל להתחשב בהקשר של מילה כדי לפתור דו משמעות. לדוגמה:

x : in his interview, Jordan said that he was very excited from his victory

y': O O O LOC O O O O O O O O O

ברור שהמשפט עוסק בראיון הניצחון של Jordan במרתון, אך המודל לא מסוגל להבחין בכך ולכן מתייג את השם ב LOC.

שאלה 2:

סעיף a:

1. מספר הפרמטרים של מודל ה RNN לעומת מודל החלון:

(instead of in window model)

(instead of 0 in window model)

1. *סיבוכיות החישוב לחיזוי תיוגים של משפט באורך* T*:*
   1. *חישוב : - כאורך קידוד המילה*
   2. *חישוב : –* 
      1. *כפל וקטור בגודל* H *עם מטריצה בגודל* HxH*.*
      2. *כפל וקטור בגודל* D *עם מטריצה בגודל* DxH*.*
      3. *הוספת* bias *בגודל* H*.*
   3. *חישוב :* 
      1. *כפל וקטור בגודל* H *עם מטריצה בגודל* HxC*.*
      2. *הוספת* bias *בגודל* C*.*
   4. *סה"כ:*

סעיף b:

1. ניקח לדוגמה את השם הבא, כולל התיוגים שלו:

Tel/LOC Aviv/LOC Yaffo/LOC

עם התיוגים הבאים:

1. O O LOC
2. LOC O LOC

במקרה זה ערך ה cross entropy ירד מכיוון שמ-1 ל-2 חזינו תיוג נכון אחד יותר, ובנוסף ערך ה F1 גם ירד כי ה precision

1. קשה לעשות אופטימיזציה ל F1 מכיוון שעל מנת לחשב אותו נדרשים התיוגים של המודל על כל הטקסט, ולכן אין קשה למקבל את התהליך.

סעיף d:

ללא masking ה loss וה gradient יתחשבו בתיוגים שהמודל ביצע על האפסים שהוספנו בסוף כל משפט וזה יבוא לידי ביטוי בשינוי הפרמטרים של המודל. השימוש ב masking מבטל את ההשפעות האלה ע"י איפוס הרכיבים הלא רלוונטיים ב loss וב gradient.

סעיף g:

מגבלה ראשונה של מודל ה RNN היא חוסר היכולת לתייג באותה צורה רצפים של מילים שכולן חלק מאותו שם, לדוגמה:

x : Ariel Sharon Park

y': PER LOC LOC

במשפט כתוב שם הפארק "פארק אריאל שרון" שכולו אמור להיות מתוייג LOC.

דרך לפתרון – שימוש ב Attention שייתן משקל גבוה לכל מילים הרצופות הקודמות שהן חלק משם של יישות.

מגבלה שניה היא חוסר יכולת להסתכל על המילים הבאות במשפט. לדוגמה במשפט שהצגנו למעלה אם המודל היה יודע שהמילה Park עומדת להגיע בסוף השם, סביר יותר שהיה מתייג את המילה Ariel כ LOC.

דרך לפתרון – שימוש ב bi-directional RNN.

שאלה 3:

סעיף c:

1. האנטרופיה המקסימלית האפשרית מתקבלת כאשר ההתפלגות המתקבלת אחידה – והערך הוא:
2. Sss

סעיף d:

המודל מתקשה בתיוג נכון של שמות בעלי רצף מילים ארוך, לדוגמה:

x : 6,739-yard Brown Deer Park Golf Course after the first round

y\*: O LOC LOC LOC LOC LOC O O O O

y': O PER PER LOC LOC O O O O O

p : 0.97 1.00 0.95 0.61 0.82 0.77 1.00 1.00 1.00 1.00

בנוסף, בדומה לשאלה 2, למודל אין "ראיה לעתיד" ולכן מתקשה במקרים שבהם המילה השניה-שלישית בשם עוזרת בהכרעה כי מדובר ביישות מסוג מסוים:

x : France expels African , Air France unions protest .

y\*: LOC O MISC O ORG ORG O O O

y': LOC O MISC O MISC LOC O O O

p : 1.00 1.00 1.00 1.00 0.47 0.91 1.00 1.00 1.00

המודל לא קישר בין המילים Air ו- France כחלק מאותה יישות.

לעומת זאת כאשר יש הפרדה מספיק גדולה בין שמות, והם לא ארוכים מדי (עד כ-2 מילים) המודל יותר מצליח לתייג נכון את המשפט:

x : Scores : Australia 228-9 in 50 overs , Sri Lanka 232-6 in 45.5 overs .

y\*: O O LOC O O O O O LOC LOC O O O O O

y': O O LOC O O O O O LOC LOC O O O O O

p : 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00