

NLP EX3 RESULT

Aviv ohayon 313410458

1.השוואה בין one hot vec ל W2V במודל הלוג לינארי:

LOSS - כפי שניתן לראות מהגרפים המצורפים כאשר בדקנו מהו ההפסד כל צורה ,
ההפסד על האימון בשני המקרים נתן הפסד שהוא פחות או יותר דומה, הנע בסביבות ה 0.68 עבור
ה loss של ה one hot ו 0.6 עבור ה loss של ה W2V. עם זאת כאשר בדקנו את המודלים על ה
validation set הפער היו הרבה יותר גדולים, כאשר ה one hot נע בצורה חדשה בין 0.69
ועד 0.7005 בעוד ה W2V נע בצורה פחות קיצונית ותמיד נישאר בטווח של 0.68 פחות או יותר.
ACC - הדיוק בשלב האימון בשימוש ב one hot נתן ערכים משמעותית פחות טובים ואף קבועים עם
ערכים של 0.56 בעוד שבשימוש ב W2V הדיוק עלמה משמעותית ליותר מ 0.65. גם בסט הולידציה
יש שיפורים גבוהים מאוד כאשר בשימוש ב one hot הערכים היו כמעט תמיד בעלי ערך של 0.51
ואילו ב W2V ככל שכמות ה epoch עלו הדיוק התמרכז לסביבות ה 0.57.
הסיבה לכך לדעתי היא שכאשר השתמשנו ב one hot vec לא ניכסנו משקל לתכיפות בה מילים
מגיעות יחדיו במשפט, הקשר זה תומך בחובו מידע סמנטי רב אשר בשימוש ב w2v אשר עושה
אינטרפולציה בין הייצוג של המילים במשפט מצליח להגיע לתוצאות טובות יותר הן באימון והן בשלב
הולידציה גם עבור ההפסד וגם עבור הדיוק.

2.השוואה של המודלים הקודים עם מודל ה LSTM בשימוש ב: W2V

כאשר השתמשנו במודל ה LSTM התוצאות היו כבר משמעותית טובות יותר מהשימוש במודל
הלינארי.
LOSS - ההפסד של המודל כבר לאחר אימונו ב 4 epoch לעומת epoch 20 ים במודלים הלוג
לינאריים) מתכנס במהירות לכמעט 0 בסט האימון, ובסט הולידציה הוא לא עובר את ה 0.45 ועף
ניתן לראות שהוא רק משתפר ומצמצם את ההפסד.
ACC - הדיוק בשלב האימון הוא כמעט מושלם, ניתן לראות שהמודל משתפר במהירות על סט האימון
שלו וכבר לאחר כמות קטנה של epoch מצליח לעלות מ 0.84 ל 0.9, ובסט הולידציה התוצאות גם כן
טובות כאשר המודל מצליח לזהות נכון על מידע חדש לגמרי בדיוק הנע מ 0.78 בהתחלה ומגיע עד
דיוק של 0.8 לאחר כמות איטרציות קטנה על ה epoch.
הסיבה לכך היא שהשתמשנו ב bi direction LSTM שהוא צורה חכמה יותר של מודל RNN ומנצל
בצורה חכמה יותר את ההיקשר הסמנטי בין מילים במשפט, כך שיחד עם השימוש ב W2V שכאמור
מאפשר את הדחיפות של מילים ביחד במשפט על ידי מיצוע הערכים של המילים שבמשפט ולא רק
על ידי יצוג המילים כבלתי תלויות אחת מהשנייה (כמו ב one hoe vector ה bi LSTM מנצל את
העובדה שלמילים יש הקשר משני הצדדים, כלומר, שמילה מסוימת באמצע המשפט יכולה להתקשר
למילה מתחילו או למילה מסופו ובעצם זה מה שמשפיע על הסמנטיקה של כל המשפט. כיוון
שבמודל ה bi LSTM אנחנו בודקים את ההיקשר של משפט לא רק מצד אחד (כמו ב RNN רגיל) אלא
משני צידו, ההסתברות שנזהה נכון את משמעות המילים במשפט ולכן גם את משמעות המשפט
עצמו עולה בעשרות מונים, וזו הסיבה שאנחנו וראים תוצאות הרבה יותר טובות במודל זה הן ב
train והן ב validation.

3השוואת התוצאות על ה subset השונים "rare": ו "negative"

התוצאות של כלל המודלים על סוגי ה subset השונים הינם:

One hot vector log linear model evaluation with the trained model

NEGAT_POLARITY LOSS: 0.706814465800389, NEGAT_POLARITY ACC: 0.4796875

One hot vector log linear model evaluation with the trained model

RARE LOSS: 0.7586384910345078, REARE ACC: 0.28

W2V log linear model evaluation with the trained model

NEGAT_POLARITY LOSS: 0.7882740104453183, NEGAT_POLARITY ACC:
0.47180059523809526

W2V log linear model evaluation with the trained model

RARE LOSS: 0.7416032855212689, RARE ACC: 0.3

W2V LSTM model evaluation with the trained model

NEGAT_POLARITY LOSS: 0.7441478397223342, NEGAT_POLARITY ACC:
0.6758928571428572

W2V LSTM model evaluation with the trained model

RARE LOSS: 0.3435355707537383, RARE ACC: 0.84

כפי שניתן לראות מהתוצאות הללו, כאשר בחנו את המודלים עבור קבוצות שהן תחת הגדרה של negated polarity או rare גם פה מודל ה bi LSTM היה המודל בעל הדיוק הגבוה ביותר עם דיוק בעל תוצאות כפולות משל המודלים האחרים, הסיבה לכך כפי שפירטתי בעמוד הקודם היא שהוא מנצל את ההיקשר של מילים במשפט מ 2 הצדדים ולכן מצליח לזהות את המשמעות הסמנטית של מילים שהן מהצורה של negated polarity בצורה הטובה ביותר. בנוסף, בהקשר של המילים הנדירות, מודל ה bi LSTM ידע לקתלג מילים אלו בצורה נכונה נטו מהמשמעות של שאר המילים במשפט מבלי צורך אמיתי ב"לנחש" מה המשמעות שלהם מבלי הצורך להיתקל במילה הזו בשלב האימון. דרך זו דומה לדרך שבה אנשים יכולים להבין מה מילה מסויימת אומרת רק מההקשר במשפט בה היא נאמרת מבלי הצורך בהכרח לדעת את משמעות המילה.