

עיבוד שפה טבעית תרגיל בית 2 – דו"ח

אימון: - גרפים והערות על המודלים יש בסעיף מבחן.

תיאור מודל בסיסי: המודל מבוסס על המודל המתואר במאמר שהוצג בכיתה וכנדרש בתרגיל בית. המודל מורכב מ 2 word embedding, אחד עבור התגים ואחד עבור המילים עצמם. תחילה אימנו את המודל על סט האימון ועשינו hyperparameter tuning על סט הולידציה. אחרי שמצאנו את הפרמטרים הטובים ביותר אימנו את המודל מחדש על סט האימון והולידציה גם יחד ונתנו תיוג על הקובץ comp.unlabeled. את התיוגים שמנו בקובץ comp.labeled.

תיאור ה LSTM: מבוסס על המאמר שנדרשנו לממש. בעל 2 שכבות, דו כיווני, מימד חבוי בגודל 140 ומימד הכניסה בגודל 190 (כגודל שרשרת וקטורי ה embedded של התג והמילה). אחרי זה לוקחים את כל וקטורי ה hidden של ה lstm ויוצרים וקטורים של כל הקומבינציות האפשריות. את התוצאה מכניסים לתוך multilayer perceptron, ומשנים את המימד למטריצה של גודל המשפט על גודל המשפט וזה מהווה את המטריצה ה scores עבור כל זוג head and modifier ועל זה ביצענו הסקה. בדקנו מספר hyperparameters על סט הולידציה וכך הגענו לערכים אותם הצגנו (מספר השכבות ב LSTM, המימד החבוי ב LSTM וב MLP) ופונקציית אקטיבציה ומספר השכבות השתמשנו במה שהוצג במאמר.

אימון המודל הבסיסי:

הוספנו dropout של כפי שמותאר במאמר והחלפנו מילים ב <unk>. השתמשנו ב Adam עם learning rate של $1e-2$ עם multiplicative learning rate decay של 0.8 ו Cross Entropy Loss יתר ההיפרפרמטרים לקחנו את מה שהציעו במאמר. גרפים וזמנים למטה.

תיאור מודל מתקדם: לקחנו את המודל הקודם והכנסנו לו את השיפור הבא: במקום לקחת את כל הקומבינציות האפשריות של כל זוג מילים (מה hidden של ה LSTM) החלטנו לקחת מטריצה בי אפניית, להכפיל פעמים במטריצה של ה hidden שיוצאת וכך לקבל מטריצה של מספר המילים על מספר המילים והיא תהיה המטריצה עליה נפעיל את האלגוריתם של chu li Edmonds.

המוטיבציה לשינוי היה המאמר <https://arxiv.org/pdf/1611.01734.pdf> שהציג שינויים שהוא עשה והוא הכניס את המטריצה הבי אפניית. לקחנו ממנו השראה גם בהיפר פרמטרים (אחרי שניסו עוד ערכים שונים אך הגענו למסקנה שאלו הטובים ביותר באמצעות בחינת הביצועים על סט הולידציה).

HYPERPARAMETER CONFIGURATION

Param	Value	Param	Value
Embedding size	100	Embedding dropout	33%
LSTM size	400	LSTM dropout	33%
Arc MLP size	500	Arc MLP dropout	33%
Label MLP size	100	Label MLP dropout	33%
LSTM depth	3	MLP depth	1
α	$2e^{-3}$	β_1, β_2	.9
Annealing	$.75 \frac{t}{5000}$	t_{max}	50,000

אימון המודל המתקדם: אימנו במשך 24 אפוקים עם scheduler של Cosine Annealing LR עם $t_{max}=5000$. הוספנו דרופ אאוט ב LSTM (כמו בבסיסי) וגם ב MLP לפני הפעלת המטריצה הבי אפניית. גודל באצ' לקחנו בגודל 50 בדומה לבסיסי. בנוסף השתמשנו בהצעה של המאמר המקורי למימוש loss להוספה של 1 לכל מה שהוא לא head אמיתי בזמן אימון. גרפים למטה.

הסקה:

השתמשנו באלגוריתם של Chu-Liu Edmonds כדי לבצע הסקה עם המימוש שקיבלנו מצוות הקורס. השתמשנו ספציפית בפונקציה decode mst כשהקלט הוא משקלי הקשתות והפלט הוא מה שאנו בודקים איתו את ה UAS.

מבחן:

ביצועים:

Test-UAS	Train-UAS	
0.8912	0.98	מודל בסיסי
0.905	0.98	מודל מתקדם

זמנים:

מודל בסיסי:

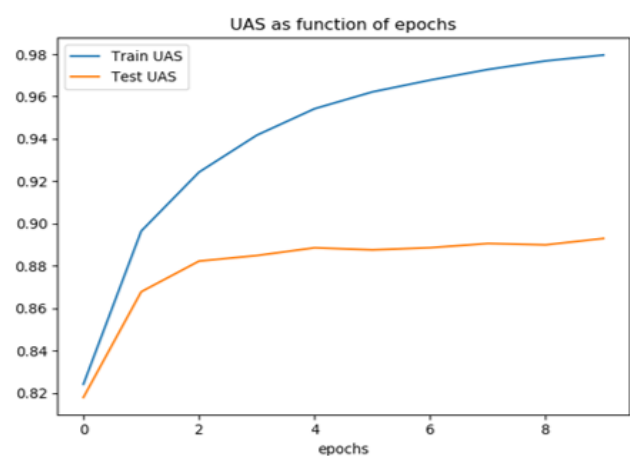
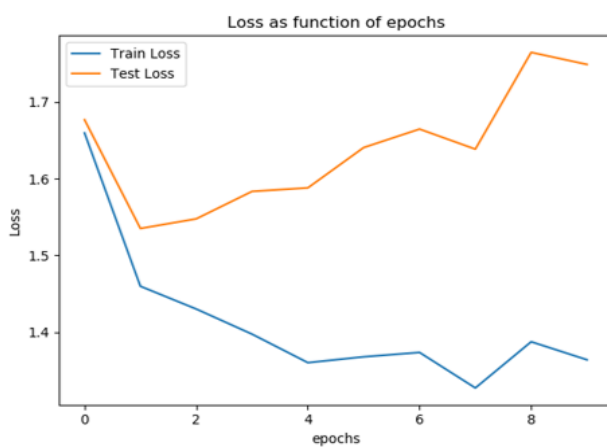
זמן אימון עבור איפוק אחד: 2 דקות

זמן חיזוי עבור סט הולידציה: 10 שניות

מודל מתקדם:

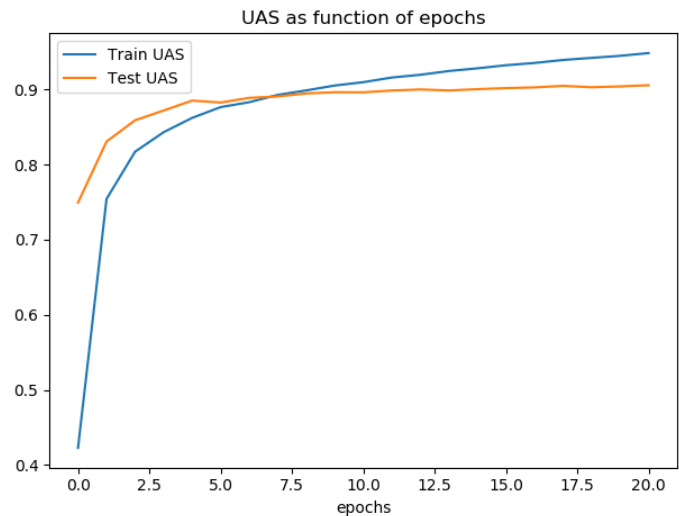
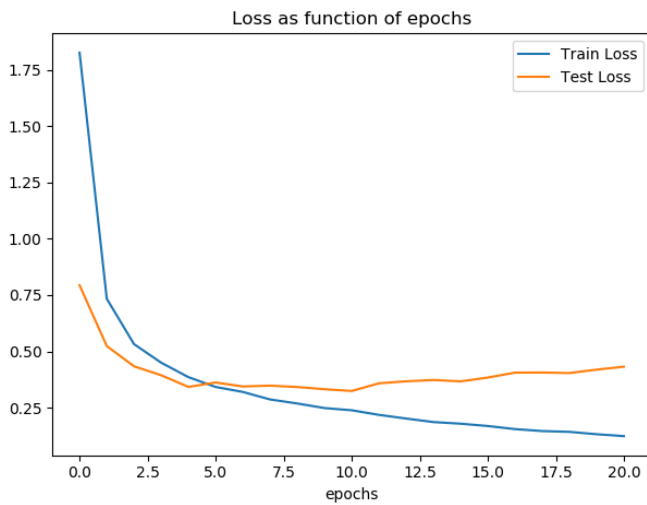
זמן אימון עבור איפוק אחד: 3 דקות

זמן חיזוי עבור סט הולידציה: 17 שניות



מודל בסיסי:

מודל מתקדם:



ההבדל בביצועים יכול להגיע ממגוון סיבות ביניהם:

- המודל המתקדם הרבה יותר גדול במימדים ולכן מסוגל להגיע לביצועים טובים יותר.
- ניסינו פחות הייפר פרמטרים מכיוון שניסינו להיצמד למאמר המקורי במודל הבסיסי.
- הוספנו שיפורים כגון המטריצה ה biaffine והרחבת הלוס כפי שמתואר במאמר המקורי במקום שימוש ב $\text{Cross Entropy Loss}$.

תחרות: ביצענו מגוון שינויים למודל הבסיסי על בסיס המאמר שצינו. ביניהם: שיפור פונקציית ה LOSS , שימוש בסוג $\text{learning rate decay}$ ושימוש במטריצה בי אפינית וגם ביצענו אופטימיזציות הייפר פרמטרים כדי לקבל את הטובים ביותר על מנת להגיע לתוצאות טובות. מעבר לזה אחרי שדיווחנו תוצאות על סט הולידציה (test) אימנו מחדש את המודל עם אותם פרמטרים על כל הדאטה סט, גם על ה itrain וגם על ה test ועם המודל המאומן החדש אימנו ונתנו תיגום לתחרות.

חלוקת עבודה:

העבודה התבצעה ביחד כאשר קטיה כתבה את הקובץ שמייצר את הדאטה סט ולדאוג שזה יכנס בפורמט המתאים למודל ואליעזר כתב את המודל ואימן כך שיהיו ביצועים טובים. על התחרות עבדנו ביחד בהבנת המאמר הנ"ל והמימוש שלו נעשה ביחד.