

תרגיל בית 3 - רשתות RNN

1. הקדמה

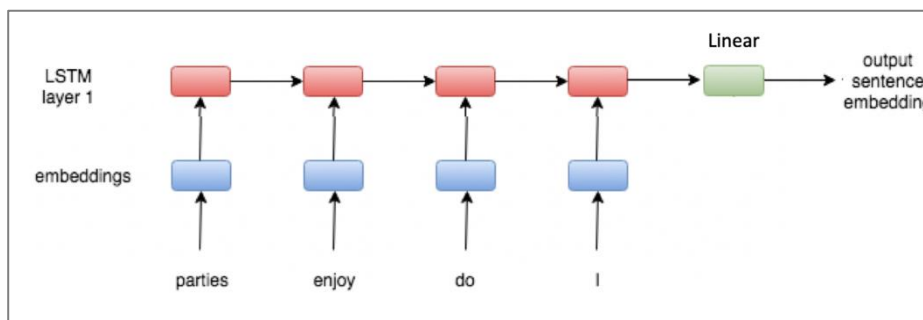
במסגרת עבודה זו מימשנו מספר רשתות היזון חוזר - RNN על מנת ללמוד ולייצר משפטים בעלי הגיון ומשמעות. תהליך האימון בוצע על מאגר Penn Treebank. נבחנו מספר רשתות, כאשר בסופו של דבר נבחרה הרשת בעלת מדד ה- Preplexity הנמוך ביותר על סט הוולידציה – 97.8.

2. תהליך הלמידה

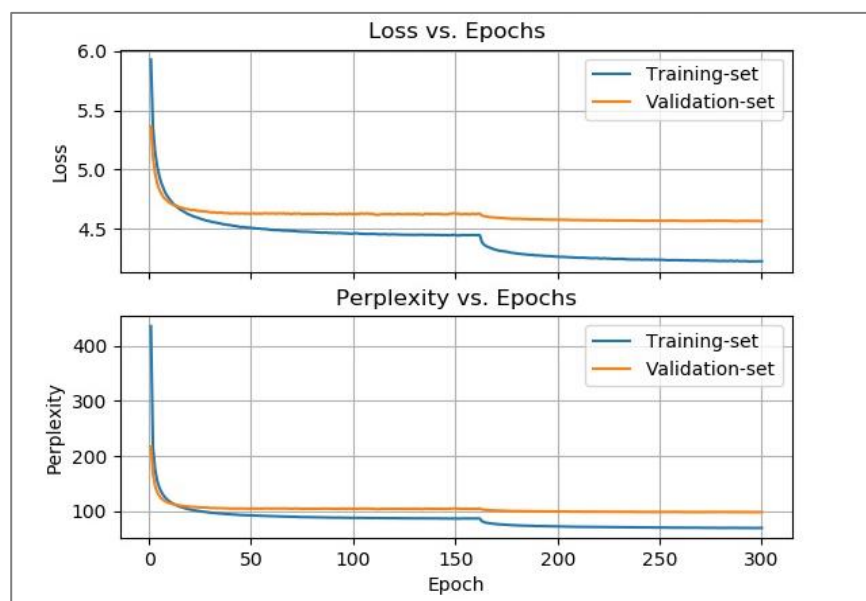
לצורך בחירת המודל המיטבי ניסינו מספר רשתות. על מנת לעמוד בקריטריון 3M פרמטרים נבחנו - רשתות LSTM עם 4 שכבות ו- 122 תאים בכל אחת, 2 ו-140, 3 ו-130, 1 ו-226. כמו גם רשת GRU עם שכבה 1 ו-250 תאים. בסופו של דבר הוחלט לנסות ולעשות שימוש ברשת LSTM עם שכבה אחת ו-249 תאים שלאחריה שכבת FC. מספר המאפיינים ששימשו ל- embedding

זהו למספר התאים ועל ידי כך ניתן לעשות שימוש באותם המשקלים ל- embedding ול- FC. כמו כן הבנו שכדאי מאוד להשתמש ב- dropouts לאחר ביצוע ה- embedding (כך שרשת ה- LSTM לומדת לחזות רצף מילים גם כאשר חלק חסרות) ו- dropouts גם ביציאה של רשת ה- LSTM, כך למעשה הרשת מצליחה ללמוד את ההקשר של המשפט גם עם מילים חסרות ולא רק כרצף מילים עוקבות (מעין fill in the blanks). סכמה של הרשת מוצגת באיור 1. לצורך האימון השתמשנו בפונקציית מחיר של אנטרופיה צולבת ואופטימיזר ADAM ומתזמן (משנה) קצב לימוד torch.optim.lr_scheduler.ReduceLROnPlateau עם קצב לימוד התחלתי 0.002 ואפסילון $1e-7$, batch size בגודל 20. Dropouts, שבוצעו גם בתאי ה- LSTM, היו בהסתברות 0.5. לרשת סה"כ 2,998,000 משקלים.

תהליך האימון לאורך 300 epochs מוצג באיור 2 - שגיאה על סט האימון וסט הוולידציה ו- Perplexity כפונקציה של epochs.



איור 1



איור 2

3. Inference

• T=2.0 :

buy low sell high is the valley cheney historically first what
sooner given blocks to review
equity
and professional exercised for fear bond assistant gregory
act then damaged

משפטים שנוצרו עם שגיאת הקלדה cell :

• T=0.5 :

buy low sell high is the <unk> of the <unk> <unk>
the <unk> <unk> <unk> <unk> <unk> <unk> <unk>
<unk> <unk> <unk> and <unk> <unk>
the <unk>

• T=1.0 :

buy low cell high is the key an rate
the work say that if <unk> officials said it 's a valuable
platform and the court would use their deduction

• T=2.0 :

buy low cell high is the deferred coach decided review in
green price vice solid good statistics and expected contras
from said gains during oil has developed orders
current

4. דיון

סביר להניח שאם היינו מבצעים עיבוד מקדים כדי להחליף את מילות ה- <unk> עם מילים בעלות משמעות (מבחינה הסתברותית) על ידי שימוש ב-ngram למשל, היינו מקבלים תוצאות טובות בהרבה כיוון שזו היא המילה שחוזרת על עצמה הכי הרבה ואין לה איזשהו המשך רצף קבוע והגיוני, אלא רנדומלי. גם הורדת ה- n/ מהמילון כנראה היתה משפרת את התוצאות במידת מה.

נשים לב קיימת חזרה על המילה "<unk>" וזאת כנראה מכיוון שהיא נמצאת הרבה בקורפוס Penn Treebank, פתרון אפשרי יכול להיות עיבוד מקדים של מאגר הנתונים שקיבלנו ולשערך או להוציא את המילה "<unk>" ואת סימן ירידת השורה "n/" כדי שלא יהיו קיימים כחלק מהמילון שלנו.

דרכים נוספות לייצור משפטים הן על ידי שימוש במודלים גנרטיבים כמו VAE (למשל זה) או מודלים הסתברותיים/תדירותיים כמו ngrams. להפתעתנו לא ניתן לייצר משפטים באופן מלאכותי על ידי GAN, כך לפי [Ian Goodfellow](#).

לאחר תהליך האימון ביצענו בדיקת ייצור משפטים בני 30 מילים כאשר המשפט מתחיל ב- "buy low sell high is the". בדקנו מספר טמפרטורות שונות והתחושה הכללית היתה שככל שהטמפרטורה נמוכה יותר המשפטים יותר הגיוניים אך ברוב ההרצות שביצענו המשפטים הכילו לרוב <unk> ו-n/ (כלומר שורת רווח) כיוון שאלו המילים הנפוצות ביותר במילון ובטמפרטורה נמוכה אין הרבה "גיוון". יחד עם זאת לפעמים התקבלו גם בטמפרטורות נמוכות משפטים שהכילו (מספר) מילים בעלות משמעות. כאשר העלנו את הטמפרטורה קיבלנו משפטים בעלי יותר מילים (ולא <unk>) אך הם נשמעו פחות טבעיים. ניתן לראות מספר דוגמאות למטה. כמו כן בשל טעות הקלדה ביצענו גם בדיקה עם המשפט "buy low cell high is the" ובאופן מפתיע התקבלו תוצאות הגיוניות לא פחות. המשפטים הטובים ביותר התקבלו בטמפרטורה של 1 ומטה. חלק מהתוצאות שהתקבלו בניסויים מוצגות מטה.

• T=0.5 :

buy low sell high is the same of the <unk> of the <unk>
the exchange said it was n't the <unk> of the <unk>
the <unk> <unk> would

• T=0.75 :

buy low sell high is the market industry <unk>
the average of these firms who are also trying to buy
assets and operating profits were improved with the
assumption

• T=0.8 :

buy low sell high is the \$ N and <unk> <unk> in the
program
analysts say he expects a substantial increase in the net
interest income and the economy

• T=1.0 :

buy low sell high is the feature of the industry 's public
<unk>
the talks on money to let orders on the names of shares
earns he says called

• T=1.5 :

buy low sell high is the important bates anger his
marketing history perspective james asarco inc
soviet people criticized latest interstate hospital arrived
approaches rather valuable above bowling hospital