🚀המאמר היומי של מייק -04.10.24: ⚡️🚀  
Were RNNs All We Needed?

המאמר הזה משך את תשומת ליבי כי יש לו ״all we needed" בכותרת. מסיבה שאינה ב-100% ברורה לי מאמרים כאלו יוצרים בי דחף חזק לסקור אותם. אז ככה הגעתי למאמר הזה שאלולא השם כנראה שלא הייתי מגיע אליו.

המאמר מציע לשפצר את ה-RNN כך שנוכל להפעיל אותו בצורה מקבילית במהלך האימון. הסיבה העיקרית ש-RNN כמעט יצא מכלל שימוש היום הוא חוסר היכולת שלו להתאמן באופן מקבילי כלומר לבצע חיזוי של כמה טוקנים ממוסכים. הטרנספורמרים לעומת זאת כן ניחנים ביכולת הזו אך יש להם מגבלה בדמות סיבוכיות ריבועית במונחי אורך הסדרה (שכואבת לנו בעיקר באינפרנס כי מאמנים אותם פעם אחת) שמקשה על השימוש (לפחות הנאיבי שלהם) לסדרות מאוד ארוכות.

מצד שני ל-RNNs יש יכולת יותר טובה לעבד סדרות מאוד ארוכות כי כל ה״זיכרון״ שלהם מקודד בכמה ווקטורים (1,2 או 3) והסיבוכיות החישובית שלהם פרופורציונלית לאורך הסדרה ולא לריבוע שלה (גם באימון וגם באינפרנס). כאמור הבעיה הגדולה של ה-RNNS שדי הרגה את הארכיטקטורה הזו היא אי יכולתה לאפשר חיזוי מקבילי באימון. זה שהופך את האימון על כמויות דאטה עצומות כמו שמקובל היום (עשרות טריליונים טוקנים) עם RNNs לארוך מדי ולא פיזיבילי.

חשוב להבין שהסיבה לחוסר יכולת לחזות בצורה מקבילי נובעת מהמעברים הלא לינאריים בין המצבים החבויים ב-RNN (גם ב-LSTM וגם ב-GRU).

לאחרונה SSMs (או State Space Models) ניסו לטפל בבעיה הזו דרך ארכיטקטורה שבה המעברים האלו כן לינאריים וארכיטקטורת ממבה (שסקרתי בהרחבה לפני כמה חודשים) ששכללה SSMs לרמת ביצועים קרובה לטרנספורמרים. בנוסף A21 Labs השתמשו בממבה כאבן בניין של הארכיטקטורה החדשה שלהם לפני כחודשיים(יחד עם הטרנספורמרים).

עכשיו אתם שואלים מה המאמר המסוקר עשה בנידון. כאמור הבעיה הגדולה ב-RNN היה מעברים לא לינאריים בין המצבים החבויים. המחברים פשוט הורידו את התלות הלא לינארית מהמשוואות של LSTM ו-GRU. מה שהתקבל כתוצאה מכך ניתן למקבול במהלך האימון (אבל דורש יותר זיכרון מהגרסאות הרגילות). יצא משהו די דומה לממבה - גם כן המצב החבוי תלוי באופן ליניארי במצב החבוי הקודם ובאופן לא לינארי בייצוג האיבר הנוכחי של סדרת הדאטה.

מה שמפתיע אותי קצת כאן זה ביצועים טובים מדי - אני קצת חשדן אבל בואו נראה מה קורה עם הארכיטקטורה הזו בעתיד.

https://arxiv.org/abs/2410.01201v1