🚀המאמר היומי של מייק 22.09.24: ⚡️🚀  
 Quiet-STaR: Language Models Can Teach Themselves to Think Before Speaking

סקירה זה ממשיכה את קו הסקירות ״בדרך ל-o1" והפעם המאמר לפחות לפי השם התקרב די מהר למה קורה לכאורה ב-o1. כלומר ״המודל חושב״ לפני שהוא מחזיר את תשובתו למשתמש. כמובן שגם המטרה כאן גם שיפור ריזונינג של המודל.

המאמר משפר את Sta שסקרתי לפני כמה ימים ועושה את דרך בניית ״שרשראות ריזונינג לוקליים״ העוזרים למודל לחזות בצורה יותר מדויקת. כל שרשרת ריזונינג כזו מורכבת ממה שנקרא במאמר ״טוקני חשיבה״ (thought tokens) שהמודל מייצר ותהליך זה ניתן לפרש בתור ״חשיבה של המודל״. מכיוון שחיזוי של רוב הטוקנים אינה משימה קשה במיוחד ולא נדרשים עבורה טוקני חשיבה המאמר מציע לשלב את הייצוג המגיע מטוקנים אלו עם ייצוג הקונטקסט המופק מהטקונים הקודמים.

לטענת המחברים (הדי הגיונית) טוקני חשיבה של טוקן נתון עוזרים לא רק לחיזוי של הטוקן הבא אלא גם לטוקנים שבאים אחריו. אז המודל מאומן למקסם את דיוק החיזוי כמה מהטוקנים הבאים. בנוסף המאמר מאמנים טוקנים מיוחדים המסמנים את ההתחלה ואת הסוף של שרשראות טוקני החשיבה: <|startofthought|> ו- <|endofthought|> שגם את הייצוגים שלהם נלמדים במהלך האימון.

האימון מתבצע בשיטת REINFORCE מאוד סטנדרטית הלקוחה מעולם למידה עם חיזוקים. בכל איטרציה עבור כל שרשרת של טוקני חשיבה אנו ממקסמים את ההפרש בין איכות החיזוי של כמה טוקנים הבאים (כלומר log-likelihood) לבין הממוצע של אותה איכות החיזוי עבור כמה שרשראות טוקני חשיבה (שנבנים כל פעם). המאמר טוען שזה מקטין את השונות שערוך ה-likelihood. ד״א מה שממקסמים זה סוג של פונקציית ה- advantage שדי נפוצה בעולם RL. כאמור השכבה (mixing head) המשלבת ייצוג טוקני חשיבה יחד עם ייצוג הקונטקסט הרגיל מאומנת גם כן.

יחד עם זה המודל עצמו (המשקלים) מאומן יחד עם טוקני חשיבה וכל השאר (ראו את האלגוריתם). המחברים שמו לב שלא צריך לבנות טוקני חשיבה לטוקנים הנחזים בקלות (אנטרופיה נמוך של וקטור ההסתברויות) ומאפשר לחסוך לא מעט כוח חישוב באינפרנס.

https://arxiv.org/pdf/2403.09629