🚀המאמר היומי של מייק 19.09.24: ⚡️🚀   
Training Chain-of-Thought via Latent-Variable Inference

ממשיכים בקו הסקירות שהובילו (לפחות לעניות דעתי) למודל החדש (יחסית, יצא כבר לפני שבוע) של openai. במאמר הקודם שסקרתי STaR דיברנו על איך ניתן לשפר יכולת ריזונינג של מודל שפה כאשר יש בידינו דאטהסט גדול יחסית של שאלות ותשובות D ודאטהסט קטן של שאלות ותשובות עם הריזונינג. בגדול הרעיון שם היא לרתום מודל שפה לייצר ריזונינג לשאלות, להוסיף שאלות שהריזונינג שלהם הוביל לתשובה נכונה לדאטהסט הקטן ולהמשיך לאמן עד ההתכנסות.

המאמר הנוכחי שיצא בערך שנה וחצי אחריו משכלל את הגישה הזו ומציע שיטה ש״ממנפת״ גם את השאלות שעבוד המודל יצר ריזונינג שלא הוביל לתשובה הנכונה. המאמר מכיל מתמטיקה די כבדה אז אנסה להעביר לכם את הרעיון הכללי יחסית בפשטות.

הרי המטרה שלנו היא לעשות פיינטיון למודל שפה כך שיכולת הריזונינג שלו תשתפר. מתמטית ניתן לתרגם את הבעיה לבעיה וריאציונית באופן הבא. אנו מעוניינים לאמן מודל שיוצר ריזונינג עבור שאלה x. מה שיש לנו זה דאטהסט של שאולות x ו-תשובות y. אז אנחנו רוצים לאמן את המודל להפיק ריזונינג z (ניתן להתייחס אלי כמו אל משתנה לטנטי) מהתפלגות בהינתן השאלה x מ-D תוך כדי ניצול של התשובה y. כלומר אנו רוצים למקסם את הנראות (likelihood) של ההתפלגות המותנית של הריזונינג z בהינתן (עבור) שאלה x ותשובה y. במילים פשוטות אנו מאפטמים את פרמטרי המודל כך שהנראות הזו תהיה מקסימלית על D.

אולם אנו לא יכולים לעשות זאת בצורה ישירה כלומר לא ניתן לדגום את הריזונינג בהינתן שאלה x ותשובה y. הסיבה לכך היא שאנו לא רוצים לאמן מודל שמייצר ריזונינג לשאלה יחד עם התשובה (כי אנו רוצים מודל שיפתור לנו שאלות בלי לדעת את התשובה). אז המאמר הקודם בחר לנצל את תשובה y על ידי פלטור החוצה של z שהובילו לתשובות לא נכונות. לעומת זאת המאמר הזה מציע שיטה שבה אנו ממנפים גם את ה- z-ים הלא נכונים לשיפור המודל.

כאמור המאמר מנצל כמה שיטות מתמטיות די כבדות לכך ואחת מהם הוא שכלול של Markov Chain Monte Carlo כאשר ה-proposal distribution (שממנו דוגמים במטרה שזו תתכנס עם הזמן להתפלגות היעד כלומר זו של ריזונינג z בהינתן שאלה x ותשובה y) משתנה עם האיטרציה להאצת התכנסות (Markovian score climbing שהוא שכלול של Robbins-Monro לחישוב ״גודל העדכון״).

מה הקשר ל-MCMC אתם שואלים? אנו כל פעם דוגמים מהמודל עם המשקלים מהאיטרציה הקודמת (באץ') ומקווים שזה יתכנס להתפלגות הרצויה. המחברים מציעים לעדכן את משקלי המודל גם עבור התשובות הלא נכונות וגם הנכונות (בכיוונים שונים כמובן). ככל שהאיטרציות עוברות השיטה מעדכנת את המודל יותר עבור דוגמאות עם ריזונינג לא נכון (מוביל לתשובה לא נכונה) כי רוב השאלות כבר מקבלות ריזונינג נכון ו״פחות שווה״ להתחשב הזה.

בנוסף המאמר משכלל את עדכון משקלי המודל המדובר על ידי כך שהוא שומר את הריזונינג האחרון z לכל דוגמא ומחשב את גודל (כמו קצב למידה) של עדכון משקלי המודל בהתאם. למשל העדכון עבור הריזונינג של דוגמא שהוביל לתשובה נכונה באיטרציה הנוכחית ולתשובה שגויה באיטרציה הקודמת גורמת לעדכון גדול יותר עבור המודל. הגישה מקורה במה שנקרא memoized wake-sleep שמציע שיטת אימון למודלים נוירו-סימבוליים גנרטיביים בכלל דרך ניצול הזכרון המצטבר של העדכונים.

וכל זה כדי לשפר את הריזונינג של המודל - מקווה שהצלחתם להבין את העיקר🙂

https://arxiv.org/pdf/2312.02179