המאמר היומי של מייק: 04.05.25

Do NOT Think That Much for 2+3=? On the Overthinking of o1-Like LLMs

המאמר מציג מחקר ראשון מסוגו המתמקד בתופעה חדשה שאותרה ב-LLMs מתקדמים, המכונים "מודלים דמויי o1" (כגון OpenAI o1 ודומיו). החידוש המרכזי של המאמר טמון בזיהוי, אפיון והצעת פתרונות לבעיית "חשיבת היתר" (Overthinking) במודלים אלו.

למיטב זכרוני זהו המאמר(פורסם בסוף דצמבר 2024) הראשון שמגדיר ומנתח באופן מקיף את תופעת "חשיבת היתר" במודלים דמויי o1. התופעה מתבטאת בכך שמודלים אלו נוטים להקצות משאבי חישוב רבים (המתבטא ביצירת טוקנים מיותרים לפעמים) גם עבור בעיות פשוטות מאוד (כמו "2+3"), תוך יצירת "שרשראות החשיבה" (Chain-of-Thought) ארוכות ומספר רב של פתרונות חלופיים, לעיתים קרובות ללא שיפור לדיוק של התשובה הסופית.

המחקר מראה אמפירית שפתרונות שבאים מאוחר יותר בשרשרת החשיבה תורמים מעט מאוד לשיפור הדיוק (לרוב, התשובה הנכונה מופיעה כבר בפתרון הראשון) ואינם מציגים גיוון משמעותי בדרכי החשיבה (פתרונות רבים חוזרים על עצמם בגישתם). התופעה בולטת במיוחד בבעיות קלות.

המחברים מגדירים מדדי יעילות חדשים:

המאמר מציג שני מדדי יעילות חדשים, שנועדו לכמת את השימוש הרציונלי במשאבי חישוב על ידי מודלים דמויי o1, מעבר למדדי הדיוק המקובלים: הראשון הוא מדד יעילות תוצאה (Outcome Efficiency): מודד את היחס בין כמות הטוקנים המינימלית הנדרשת להגעה לתשובה הנכונה הראשונה לבין סך הטוקנים שגונרטו. ערך נמוך מצביע על חשיבת יתר מבחינת תרומה לדיוק.

מדד יעילות תהליכית (ξ\_P​ - Process Efficiency): מודד את היחס בין כמות הטוקנים התורמים לגיוון בפתרונות (כלומר, טוקנים השייכים לפתרונות המציגים גישה חדשה) לבין סך הטוקנים שנוצרו. ערך נמוך מצביע על חזרתיות וחוסר גיוון בפתרונות.

פיתוח אסטרטגיות להפחתת חשיבת יתר:

המאמרב בוחן אסטרטגיות חדשניות להפחתת חשיבת היתר, המבוססות על פרדיגמת אימון עצמי (Self-training) וטכניקות אופטימיזציית העדפות (Preference Optimization), ללא צורך במידע חיצוני. החידוש מתבטא ביישום שיטות אלו לבעיה הספציפית של פישוט תגובות תוך שימור יכולות אינפרנס. המחברים השתמשו בטכניקות כמו SFT, DPO, RPO ו-SimPO כדי לאמן את המודל להעדיף תגובות קצרות ויעילות יותר (שזוהו ככאלה מתוך דגימות מרובות), תוך שימוש בתגובה הארוכה ביותר כדוגמה שלילית (נמצא יעיל יותר מתגובת ברירת המחדל).

המחברים פיתחו כמה שיטות חדשות ליצירת דאטהסט אימון יעיל יותר על ידי חיתוך מכוון של תגובות ארוכות: הראשונה היא פתרונות נכונים ראשונים (FCS - First-Correct Solutions) ששומרת רק על החלק המינימלי של התגובה עד להופעת התשובה הנכונה הראשונה. השניה היא FCS + רפלקציה (FCS+Reflection) המהווה הרחבת FCS כך שיכלול גם את הפתרון השני שהגיע לתשובה הנכונה, במטרה לשמר יכולת "חשיבה ארוכה" אך יעילה. הגישה הנוספת שנבחנה היא נקראת GDS - Greedily Diverse Solutions שהיא הרחבה חמדנית של התגובה על ידי הוספת פתרונות רק אם הם מציגים פרספקטיבה חדשה ושונה מקודמיהם.

הגישה (שילוב SimPO עם FCS+Reflection) הצליחה להפחית משמעותית את כמות הטוקנים המיוצרת (לדוגמה, הפחתה של 48.6% ב-MATH500) תוך שמירה ואף שיפור קל ברמת הדיוק במגוון מבחנים ברמות קושי שונות (GSM8K, MATH500, GPQA, AIME).

הסבר על מושגים:

שיטת (SimPO (Simple Preference Optimization: זוהי שיטת האימון ששימשה פיין טיון של המודל. המטרה שלה היא ללמד את המודל להעדיף תגובות מסוג מסוים (במקרה זה, תגובות יעילות יותר) על פני תגובות אחרות (פחות יעילות). המאמר מצא ש-SimPO הייתה היעילה ביותר מבין שיטות אופטימיזציית ההעדפות שנבדקו.

שיטת (FCS+Reflection - First-Correct Solutions + Reflection): זוהי האסטרטגיה ששימשה ליצירת דאטהסט האימון עבור SimPO. בשיטה זו, לקחו את התגובות המקוריות של המודל ו"פישטו" אותן על ידי שמירה רק על החלק המינימלי של התגובה שהוביל לתשובה הנכונה הראשונה (FCS), בתוספת הפתרון השני שהגיע לאותה תשובה נכונה (החלק של ה-Reflection). המטרה הייתה ליצור דוגמאות אימון "טובות" שהן גם יעילות (לא ארוכות מדי) וגם שומרות על יכולת ה"חשיבה הארוכה" או הרפלקטיבית של המודל.

https://arxiv.org/abs/2412.21187