⚡️🚀המאמר היומי של מייק -01.11.24: ⚡️🚀  
LLMs Are In-Context Reinforcement Learners

אני אוהב מאמרים שמשלבים כמה שיטות של ML. אסקור היום אחד כזה המציע לשדך למידת in-context עם למידה באמצעות חיזוקים או בקצרה RL. למידת in-context היא יכולת של מודל שפה ללמוד משהו חדש מכמה דוגמאות בפרופמט ללא צורך בפיין טיון. יש לא מעט הסברים ליכולת די מפתיע זו ולפעמים יכולת זו נקראה emergent capabilities.

עכשיו נשאלת השאלה: איך נוכל לבחור דוגמאת להדגמה שאנו מראים למודל שפה בפרומפט למקסום ביצועיי המודל? השאלה הזו לא מאוד טריויאלית ואין עליה כרגע תשובה חד משמעית. המחברים מציעים לגשת לבעיה זו דרך למידה עם חיזוקים (סוג של). השיטה הנאיבית היא פשוט לצבור דוגמאות עד שנגמר לנו את אורך חלון ההקשר של המודל. לכל דוגמא בהדגמה אנו שומרים בבאפר את השלישיה המכילה את הדוגמא (שאלה עצמה)ֿ, תשובת המודל ומשערך של איכות התשובה (או פשוט האם התשובה נכונה או לא). ואז באינפרנס פשוט לוקחים את הדוגמאות האלו בתור פרומפט.

לטענת המחברים הגישה הנאיבית הזו לא עובדת משתי סיבות עיקריות. קודם כל שילוב מתמשך של אותם הפרומפטים לדוגמאות שונות מוביל לשונות גדולה בפלט של LLM (לפי המחקרים הקודמים עלולה להוביל לביצועים ירודים). הסיבה השניה טמונה בכך ששלישיות (שאלה, תשובה, לא נכון) מסבכות את המודל ולא מספקות לו מספיק מידע על איך היה צריך לענות נכון (ד״א בלמידה ניגודית יש בעיה דומה המצריכה כמות מאוד גדולה של דוגמאות שליליות בכל באץ' - כתבתי על זה לא מעט בסקירותיי).

עקב כך המחברים הציעו להכניס קצת ״אקראיות״ לבניית הפרומפטים (המחברים קוראים לזה אפיזודה בהתאם לטרמינולוגיה של RL - כל אפיזודה מורכבת מכמה שלישיות של שאלה, תשובה, נכונות התשובה) וגם להשתמש באפיזודות שקיבלו ציון ״נכון״. לכל דוגמא הם הציע קודם לדגום באקראי מהבאפר של אפיזודות בצורה אקראית ולהשתמש לכל דוגמא במדגם שונה של אפיזודות. כאמור שומרים רק את האפיזודות שבהם המודל צדק. כך פרומפט לכל שאילתה הופך להיות לא קבוע ומכיל רק דוגמאות עם תשובות נכונות. זה נקרא Explorative ICRL במאמר.

כמובן ש Explorative ICRL לא יעיל חישובית כי כל פעם צריך לחשב את הפרומפט מחדש (מה שלא צריך לעשות בגישה הנאיבית אך לא עובדת). המחברים שכללו את זה עם מנגנון קאשינג המאפשר לשמור מספר קבוע של פרומפרטים (מערך של אפיזודות) ולכל אפיזודה נתונה להחליט לאלו מהם להוסיף אותה. זה מקל על העלות החישובית.

מאמר חמוד למרות שמשום מה לקח לי קצת זמן להבין אותו…

https://arxiv.org/pdf/2410.05362