⚡️🚀המאמר היומי של מייק 13.07.24: ⚡️🚀

SaySelf: Teaching LLMs to Express Confidence with Self-Reflective Rationales

בהמשך לסקירה של אתמול, מאמר קליל יותר שמציע שיטה ללמד מודלי שפה לשערך אי וודאות בתשובתם. המחברים מציעים שיטה מאוד אינטואיטיבית המורכבת משני שלבים עיקריים: יצירת דאטהסט למשימה זו (כימות אי וודאות) וטיוב (fine-tuning) של המודל על הדאטהסט הזה. בשלב השני ממשיכים לאמן את המודל עם שיטת PPO מעולם למידה באמצעות חיזוקים כדי לשיפור נוסף של ביצועיו.

בשלב הראשון לוקחים דאטהסט של שאלות ותשובות הנקרא HotpotQA ומזינים את השאלות ממנו למודל שפה ומבקשים ממנו לתת תשובה מלווה ב-reasoning. לאחר מכן מקלסטרים את תשובות המודל (יחד עם ה-reasoning) לקלסטרים לפי האמבדינג שלהם ומחשבים את יחס של גודל הקלסטר המכיל את התשובה הנכונה (מהדאטהסט) יחסית לכל התשובות. זה מדד אי הוודאות שלנו שעליו נאמן את המודל בהמשך.

לאחר מכן מפלטרים את השאלות ובסוף מבקשים מ-gpt4 לתת הסברים למה המודל היה עשוי לתת תשובות לא נכונות לשאלה (כלומר ״הסיבה״ לאי וודאות). בשלב האחרון מטייבים (מאמנים מודל שפה נתון) קודים כל לתת תשובה נכון, לדייק בממד של אי הוודאות ובנוסף לתת reasoning נכון לנוכחות של אי הוודאות. כל אלה נמצאים באופו מפורש בפונקציית הלוס.

בשלב השני ממשיכים לאמן את המודל בשיטה PPO כדי למזער (או למקסם אותה עם מינוס) את ההפרש בין נכונות התשובה (0 או 1) ורמת ה-confidence של המודל לגביה. כמו בכל שיטת PPO הדוגמאות נוצרות ״on the fly" אחרי כל עדכון של משקלי המודל.

https://arxiv.org/abs/2405.20974