המאמר היומי של מייק - 02.04.25  
SymDPO: Boosting In-Context Learning of Large Multimodal Models with Symbol Demonstration Direct Preference Optimization

היום אני עושה מעבר חד בנושא הסקירה וסוקר מאמר על אימון מודלים מולטימודליים (בפרט MLLMs). המאמר מציע שיטה לאימון מודלים למשימת למידה in-context שבא המודל מקבל כמה דוגמאות (הדגמות) שכל הדגמה מכילה תמונה, שאלה ותשובה עליה. המודל מתבקש, בהתבסס על הדאטה שקיבל (הדגמות) לענות על שאלה לגבי תמונה נוספת (עם אותם הדמויות למשל). הסקירה הולכת להיות קלילה וקצרה.

המחברים מציעים דרך לשיפור הבנת קשרים בין פיסות דאטה ממודליות שונות על ידי מודלים מולטי-מודליים. למשל למודלים התומכים בשתי מודליות, שפה ותמונות, לפעמים מתקשים במשימות שדורשות הבנת קשרים סמנטיים בין דאטה ויזואלי לשפתי למשל במשימת למידה in-context ל-MLLMs המתוארת קודם לכם. המאמר מציין כי MLLMs לפעמים מתקשים להתמודד עם משימות אלו ולמשל עונים על השאלה בלי להתחשב בהקשר כלל (שזה תמונות, שאלות ותשובות). המאמר מציע שיטת פיין טיון עבור מודל מולטי-מודלי כדי להתמודד עם כשלים כאלו.

המאמר מציע לעשות פיינטיון למודל בשיטה מעולם RLHF (שזה Reinforcement Learning with Human Feedback) הנקראת(DPO (= Direct Preference Optimization . שיטה זו נגזרת מפונקציית יעד פופולרית בעבור פיין טיון של מודל שפה (מקסום תגמול - קרבה למודל ההתחלתי) דרך מידול reward של Bradley-Terry. היתרון העיקרי של DPO מעל PPO הוא העובדה ש-DPO לא דורש אימון של מודל תגמול (reward) אלא צריך רק דאטהסט של זוגות שאלות ותשובות רצויות ותשובות לא רצויות. הרעיון העיקרי במאמר הוא להנדס דאטהסט כזה עבור יוזקייס שבנידון ולהשתמש ב-DPO לפיין טיון של מודל מולטימודלי.

בגדול המאמר מציע לשחק עם השאלות והתשובות. הוא מציע כמה טריקים כדי לאלץ את המודלי להתחשב בכל הקונטקסט שניתן לו. אחד הטריקים הוא לתת תשובה רצויה לא קשורה (מילה ללא משמעות). עוד טריק היא להחליף תשובה לא רצויה בג'בריש ועוד אחד היא למחוק את השאלה עצמה ולהשאיר את התשובות כמו שהם. יש עוד כמה טריקים מהסוג הזה ועל ידי שילובם המאמר משיג מודל יותר טוב עם שימוש ב-DPO לפיין-טיון.

כמו שהבטחתי סקירה קצרה וקלילה.

https://arxiv.org/abs/2411.11909