⚡️🚀המאמר היומי של מייק -20.11.24: ⚡️🚀

Adaptive Decoding via Latent Preference Optimization

היום סוקרים מאמר ששוב שכנע אותי שלא משנה כמה מאמרים אקרא עדיין אפספס רעיונות מעניינים גם בתחומים שאני מתמחה (סוג של) ומתעניין. כמובן מדובר בשיטות לג'נרוט דאטה ממודלי שפה? המאמר הזה מציע שיטה המתאימה את הייפר-פרמטרי הג'נרוט שלה כפונקציה של הקונטקסט. למשל המאמר שנסקור היום עוסק בהתאמה של טמפרטורת דגימה לגנרוט דאטה. אזכיר לכם שטמפרטורת הדגימה T שולטת באקראיות דגימה של טוקן הבא - ככל שהיא גדולה יותר טוקנים עם ״הסתברות דגימה״ (מותנית בהקשר) נמוכה יותר מקבלים יותר סיכוי להידגם.

מתברר שקו מחקרי זה (התאמת הייפר-פרמטרי ג'נרוט) קיים כבר איזה 4 שנים ויצאו לפחות 10 מאמרים בנושא (שלא ידעתי). אז המאמר הזה הוא המשך של כמה מאמרים שלא סקרתי בזמנו). אוקיי אז כאמור המאמר מנסה לאפטם את T בהינתן ההקשר. המחברים מניחים שאנו בוחרים T מסט טמפרטורות סגור T\_1,..., T\_k המחברים מציעים לאמן רשת M\_t(נקרא Adaptive Decoder במאמר) החוזה את T האופטימלי בהתבסס על ייצוגי טוקני ההקשר. כלומר הרשת פולטת התפלגות מעל T\_1,..., T\_k (כלומר סופטמקס).

למעשה התפלגות כזו היא ממשקלת (משנה לפי התפלגות הטמפרטורות הנוצרת על ידי M\_t) את התפלגות הסופטמקס מעל מילון הטוקנים שממנו מודל שפה מגנרט טקסט. כמובן ניתן לאמן M\_t בכמה דרכים על דאטהסט נתון במטרה למקסם את הנראות(likelihood) של הדאטה (לדעתי נעשה במאמרים קודמים). המאמר מציע לעשות את בשיטת DPO הלקוחה לעולם למידה עם חיזוקים עם RL (קראו סקירה מ 18.11.24 כדי לרענן מה זה). רק אזכיר שבשיטה זו מבצעים יישור (alignment) של מודל שפה על דאטהסט של תשובות רצויות ופחות רצויות.

אז המחברים מציעים להכליל את השיטה הזו עבור המקרה שאנו לא רק מאמנים את המודל אלא גם המודל לקביעת התפלגות טמפרטורה. הדאטהסט של תשובות וטמפרטורות רצויות נבנה על ידי דגימה של מודל שפה בטמפרטורות שונות ובחירה של התשובה הטובה ביותר והגרועה ביותר או עלי ידי מודל אחר או על ידי מתייגים אנושיים. ואז בדומה ל-DPO בונים פונקציית לוס שמעדכנת את מודל השפה וגם M\_t יחד. הרי ניתן לראות ב-M\_t מודל דגימה ממילון הטוקנים כאשר כל טוקן הוא טמפרטורה T\_k. אז זה הכללה די מתבקשת. המחברים גם מציעים פונקציית לוס שמעדכנת רק את M\_t באותה הצורה.

לבסוף המאמר מציע פונקצית לוס המאפטמת מודל שפה יחד עם M\_t כאשר התפלגות של הטוקנים (של מודל השפה) מבוטאת דרך מרגינליזציה שלה מעל התפלגות הטמפרטורות דרך נוסחת בייס. כלומר מיישרים את המודל לתעדף רק תשובות רצויות באופן ישיר אבל יחד עם זאת גם M\_t מתעדכן.

https://arxiv.org/abs/2411.09661