⚡️🚀המאמר היומי של מייק 22.07.24: ⚡️🚀

TRAINING DIFFUSION MODELS WITH REINFORCEMENT LEARNING

אוקיי, בסקירה הקודמת סקרתי מאמר בנושא מודלי דיפוזיה גנרטיביים וקיבלתי תיאבון בלסקור עוד כמה כאלו. אז בחרתי במאמר המגניב הזה שאחד ממחבריו הוא סרגיי לווין האגדי (בנוסף למאמרים הרבים יש לו קורס די מטורף מבחינת העומק בנושא deep reinforcement learning). באופן לא מפתיע המאמר שנסקור קשור ללמידה עם חיזוקים (או RL בקצרה) אבל יחד עם זאת מופיע בשמו גם מודלי דיפוזיה.

לדעתי בעבר כבר סקרתי אחד המאמרים שלו המשלב גישות מעולם ה-RL לאימון מודלי דיפוזיה. מתברר שניתן לאפיין אימון במודל דיפוזיה עם כלים מעולם ה-RL כלומר ניתן לבנות תהליך החלטה מרקובי (MDP) מאד אינטואיטיבי עבור מודל דיפוזיה.

כמו שאתם זוכרים אימון של מודל דיפוזיה מסתכם בניית מודל שמשערך את הרעש שהתווסף לדאטה באיטרציה t של התהליך הקדמי (של ההרעשה ההדרגתית של דאטה). אם יש לנו את האומדן של הרעש שהתווסף לפיסת דאטה באיטרציה t אנו יכולים לאמוד את הדאטה המורעש באיטרציה הקודמת t-1. כלומר אנו מאמנים מודל denoising לבנייה של דאטה מרעש טהור.

המאמר למעשה מצא פריימוורק מעולם RL (כלומר MDP) למידול של אימון מודל דיפוזיה גנרטיבי. בשביל כך נגדיר את כל הפרמטרים של ה- MDP באיטרציה t באופן פורמלי:

המצב (state): השלישיה {פרומפר, מספר איטרציה t, הדאטה המורעש x\_t}

הפעולה (action) היא x\_t-1

הפוליסי היא הסתברות לקבל x\_t-1 מ- x\_t ומהפרומפט c

המצב ההתחלתי מוגדר על ידי השלישיה: {רעש גאוסי סטנדרטי (ממנו מתחילים denoising), הסתברות על מרחב הפרומפטים, האיטרציה האחרונה T}

פונקציית תגמול (reward) שהמאמר מגדיר בכמה צורות. היא מחושבת באיטרציה האחרונה (על התמונה המשוחזרת).

עכשיו אחרי שיש לנו הגדרת RL של אימון מודלי דיפוזיה אנו יכולים להשתמש בשיטות RL קלאסית כמו REINFORCE או PPO למקסום של פונקציית התגמול.

לגבי פונקציית התגמול המאמר מציע כמה אופציות. האופציה הראשונה היא לחשב את מה שנקרא BERT Score שבודק כמה התמונה מתאימה לפרומפט שלה (שווים את האמבדינגס שלהם). האופציה השניה היא להשתמש במה שנקרא LAION aesthetics predictor שאומן לשערך עד כמה התמונה היא אסתטית (זו למעשה שכבה לינארית על האמבדינג של CLIP המאומן עד דאטהסט של תמונות המתויגות על ידי בני אדם.

מאמר מעניין ויחסית לא קשה לקריאה.

https://arxiv.org/pdf/2305.13301