⚡️🚀המאמר היומי של מייק 20.07.24: ⚡️🚀

Consistency Models

המאמר הזה חיכה את תורו די הרבה זמן, קצת פחות משנה וחצי המאמר הזה נכתב על ידי Yang Song האגדי (זה שכתב מאמרים חזקים מאוד בתחום הדיפוזיה) ואחד המחברים הוא איליה סלוצקבר שאני מניח שאתם מכירים היטב. המאמר נכתב עוד בתקופה ששני המדענים הדגולים אלו עבדו ב-openai. דרך אגב שני המחבריםה אחרים גם תרמו לא מעט לתחום המודלים הגנרטיבים ושניהם עבדו ב-openai לפחות נכון למרץ 2023.

המאמר מציג גישה חדשה לאימון מודלי דיפוזיה גנרטיביים. מודל דיפוזיה גנרטיבי סטנדרטי מורכב מתהליך קדמי ומתהליך האחורי (forward & backward). בתהליך הקדמי אנו מוסיפים רעש (בבד״כ גאוסי) לדאטה באופן הדרגתי עד שפיסת דאטה הופכת להיות רעש. בתהליך האחורי אנו מאמנים את המודל להסיר רעש בצורה הדרגתית גם כן. כלומר המודל לומד מה הרעש צריך להחסיר מהדאטה המורעש באיטרציה t כדי לקבל את הדאטה באיטרציה t-1. אחרי שהמודל מאומן לעשות זאת אנו יכולים להשתמש בו ולבנות פיסת דאטה מרעש טהור על ידי הסרה של רעש בצורה הדרגתית.

מה הבעיה בתהליך הזה? הוא עלול להיות די ארוך (צריך להריץ מודל כמספר האיטרציות) ויצאו לא מעט מחקרים שניסו להקטין את מספר האיטרציות בלי לפגוע באיכות הדאטה המגונרט. מודלים קונסיסטנטיים(consistency models) זה עוד ניסיון לתקוף את הבעיה המעניינת הזו. בגדול הרעיון כאן הוא שעבור פיסת דאטה נתונה x\_0 שלא משנה מאיזו איטרציה t (=דאטה מורעש x\_t) נתחיל את הסרת הרעש בסופו של דבר אנו חייבים לחזור לדאטה הנקי x\_0.

המאמר מציע שתי שיטות לאמן מודל דיפוזיה קונסיסטנטי. השיטה הראשונה מניחה שיש לנו ביד מודל דיפוזיה מאומן (consistency distillation) והשני מאמן את המודל מאפס(consistency training). כדי להסביר את השיטה הראשונה צריך טיפה לצלול למתמטיקה אבל נעשה את זה לאט ובזהירות.

נתחיל מזה התהליך הקדמי המודל דיפוזיה מתואר על ידי משוואה דיפרנציאלית סטוכסטית המתארת את היצירה הדרגתית של הדאטה המורעש. ניתן להראות שמשוואה דיפרנציאלית רגילה (ODE) ל x\_t. מעניין ש-ODE הזה מכיל לוגריתם של פונקציית ההסתברות של ההדאטה המורעש x\_t (נקרא score function או SF). המשוואה מתארת את בתהליך האחורי (הסרה הדרגתית של רעש). אז אם יש בידינו שערוך של SF אנו נוכל לשחזר (הדרגתית) את הדאטה שלנו על ידי הפתרון הנומרי (באיטרציות) של ה-ODE הזה (נגיד Euler-Maruyama).

הדבר הכי מגניב שאם יש לנו מודל דיפוזיה מאומן (שערוך של הריש באיטרציה t) אז ניתן בקלות לקבל שערוך של SF (בתנאי של רעש גאוסי).

אבל איך כל זה קשור למודלים קונסיסטנטיים שצריכים להיכנס לאותה הנקודה לא משנה מאיזו איטרציה של הרעשה מתחילים. אנו מאמנים את המודלי באופן הבא: לוקחים נקודה מורעשת t, עושים איטרציה אחת של הפתרון הנומרי של ODE (עם SF) כדי לקבל את הדאטה באיטרציה t-1. תזכרו שהמטרה שלנו היא לאמן את המודל לשחזר את הדאטה הנקי מכל איטרציה של הרעשה. אז מאמנים מודל למזער את ההפרש בין הדאטה המשוחזר מאיטרציה t לזה של האיטרציה t-1. בגדול יש כאן שני מודלים (בדומה לשיטה של למידת הייצוג הנקראת BUOL). המודל הראשון המודל המוחלק שהפרמטרים שלו הם ממוצע מעריכי של המשקלים של המודלים מהאיטרציות אימון הקודמות(לא מאומן - יש stop gradient) והוא נקרא target והמודל השני שהוא למעשה מאומן עם מורד הגרדיאנט.

ניתן גם לאמן מודל ללא מודל דיפוזיה מאומן ובמקרה הזה יוצרים את x\_t-1 מ- x\_t על ידי הורדת הרעש. ברגע שאימנו מודלי קונסיסטנטי ניתן ליצור דאטה נקי מרעש טהור באיטרציה אחת אך זה לא תמיד אופטימלי. ניתן לבצע מה שנקרא במאמר Multistep Consistency Sampling. להתחיל מרעש טהור, ליצור דאטה נקי, להוסיף רעש, שוב ליצור דאטה נקי ולחזור עד שאיכות הדאטה הוא לשביעת רצוננו. המאמר טוען שנדרש משמעותית פחות איטרציות בתהליך זה מאשר במודלי דיפוזיה סטנדרטיים.

סיימנו, מקווה שלא איבדתי אותכם כאן…

https://arxiv.org/abs/2303.01469