המאמר היומי של מייק - 30.12.24:  
Graph Diffusion Policy Optimization

לפני יומיים סקרתי מאמר על מודלי דיפוזיה המאומנים באמצעות שיטות מעולם למידה עם חיזוקים או RL, אתמול סקרתי מאמר על רשתות נוירונים על גרפים והיום החלטתי לסקור מאמר שמאחד את 3 הדברים האלו (כמעט). המאמר המסוקר היום מציע שיטה לאימון מודל המגנרט גרפים באמצעות מודלי דיפוזיה המאומנים עם שיטות RL (נכון אין כאן GNN בצורתם הטהורה אבל לפחות יש גרפים…

קודם כל אנו צריכים להבין איך ניתן למנף מודלי דיפוזיה לגנרוט גרפים. האמת זה די פשוט ודומה לגנרוט תמונות. אתם זוכרים מודלי דיפוזיה מאומנים לגנרט תמונה מרעש טהור (בד״כ) על ידי הורדה הדרגתית של הקומפוננטה הרועשת שלו עד להפיכתו לפיסת דאטה המפלגות לפי ההתפלגות של דאטהסט אימון. זה ממש בגדול ויש גישות חדשות יותר שעושות את זה טיפה אחרת למשל כמו Consistency Models שדיברנו עליהם באחת הסקירות הקודמות.

האם אנחנו יכולים לעשות משהו דומה עם גרפים? מתברר שכן. אנו יכולים להתחיל מלדגום גרף באקראי (כלומר הצמתים והקשתות שלו) ולאמן מודל לשנות את הערכים בצמתים ובקשתות כך שהגרף יהפוך להיות ״דומה״ לאחד הגרפים מדאטהסט האימון וגם יקבל ערך גבוה לפי איזה פונקציית תגמול(המאמר גם על RL, זוכרים). ד״א, יש כאן הנחה סמויה שצומת יכול לקבל מספר סופי של ערכים (נגיד מ 0 עד a) וכל קשת יכולה להיות מכמה סוגים (כלומר מ- 0 עד b). כלומר ההתפלגויות שאנו דוגמים מהם הם קטגוריאליות וזה שונה ממה שאנו רגילים לראות במודלי דיפוזיה גנרטיביים עבור התמונות.

כמובן מיד עולות כמה שאלות בנוגע לתהליך הזה?

איך דוגמים גרף באקראי במהלך האינפרנס (זה נושא עתיק ונחקר רבות עלי ידי מתמטיקאים ובפרט על ידי ארדוש, המאמר לא מתעמק בזה יותר מדי). דרך אגב במהלך האימון אנו לוקחים גרף מהדאטהסט ומרעישים אותו עלי ידי ״שינוים אקראיים״ בערכי הצמתים ובסוגי הקשתות

איך משווים גרפים, כלומר איך מבינים שגרף שקיבלנו במהלך הגנרוט הוא דומה לגרף מהדאטהסט? יש מספר רב גישות להשוות גרפים על ידי השוואה של התת-גרפים שלהם או להשוות את הלפלסיאן שלהם למשל.

בחירה של פונקצית reward בדומיין הגרפים לא טריוויאלית בכלל. למשל למשימות גנרוט גרפים למולקולות חדשות אחד המדדים לאיכות הגרף המגונרט הוא חדשנותו יחסית לדברים הקיימים, יעילותו בטיפול במחלה מסוימות או פיזיביליות של סינטוזו (synthetic accessibility). ניתן לבחור reward גם בתור פונקצית דמיון לגרפים הקיימים.

אוקיי, אז יש לנו פונקציה להשוואת הגרפים C ופונקצית תגמול לשערוך איכות הגרף r - איך אנו מאמנים מודל דיפוזיה. האמת בצורה די דומה לזו שתיארתי בסקירת של לפני 3 ימים של המאמר: RL for Consistency Models: Faster Reward Guided Text-to-Image Generation.

קודם כל אנו צריכים להגדיר את Markov Decision Process עבור אימון מודל דיפוזיה על גרפים. ומתברר שהוא ממש דומה למאמר שהזכרתי:

המצב s\_t בתור זוג של גרף מגונרטת באיטרציה T-t וגם ערך T-t  
הפעולה a\_t היא הגרף באיטרציה T−t−1

הפוליסי (הסתברות של a\_t בהינתן s\_t) היא זו פונקצית התפלגות מותנית של גרף מאיטרציה T−t−1 בהינתן גרף באיטרציה T-t

המצב ההתחלתי הוא גרף אקראי באיטרציה T ופונקציית תגמול r שנתונה לנו המחושבת על הגרף הסופי באיטרציה 0

המאמר מציע שתי שיטות לאימון של מודל דיפוזיה לגנרוט גרפים: הראשונה היא REINFORCE הקלאסי שהיא למעשה שיטת policy gradient הממקסת פוליסים בעלי תגמול גבוה. מעשה אנו דוגמים K איטרציה בין 1 ל T וממקסמים מכפלה ממוצעת (על K דגימות) של פונקציית הפוליסו (ונקצית התפלגות מותנית של גרף מאיטרציה T−t−1 בהינתן גרף באיטרציה T-t) והתגמול עבור הגרף המגונרט (באיטרציה 0).

השיטה השנייה המוצעת היא Policy Optimization כאשר במקום למקסם את הפוליסי בצורתו הטהורה אנו ממקסמים הסתברות גנרוט גרף G\_0 מהדאטהסט (שאותו מרעישים והמודל ״מסיר״ ממנו את הרעש) מוכפלת בתגמול עבור הגרף הנוצר. גם כאן יש מיצוע על K איטרציות שמהם נבנה שערוך של G\_0.

זהו זה - סקירה קצת כבדה, מקווה שהצלחתם להבין משהו ממנה…

https://arxiv.org/abs/2402.16302