סקירה זו היא חלק מפינה קבועה בה אני סוקר מאמרים חשובים בתחום ה-ML/DL, וכותב גרסה פשוטה וברורה יותר שלהם בעברית. במידה ותרצו לקרוא את המאמרים הנוספים שסיכמתי, אתם מוזמנים לבדוק את העמוד שמרכז אותם תחת השם deepnightlearners.

לילה טוב חברים, היום אנחנו שוב בפינתנו deepnightlearners עם סקירה של מאמר בתחום הלמידה העמוקה. היום בחרתי לסקירה את המאמר שנקרא:

VAEBM: A symbiosis between autoencoders and energy-based models

---

פינת הסוקר:

המלצת קריאה ממייק: מומלץ לאוהבי מודלים גנרטיביים כמו VAE ו-Energy-Based Models להרחבת אופקים, אך לא חובה.

בהירות כתיבה: בינונית.

רמת היכרות עם כלים מתמטיים וטכניקות של ML/DL הנדרשים להבנת מאמר: נדרש רקע טוב בשיטות דגימה מתקדמות (דינמיקה של Langevin) והבנה טובה במודלים גנרטיביים.

יישומים פרקטיים אפשריים: יצירה תמונות באיכות טובה יותר מ-StyleGAN אך עדיין זה לא נראה באופק עקב מורכבותה.

פרטי מאמר:

לינק למאמר: זמין להורדה.

לינק לקוד: לא נמצא בארקיב.

פורסם בתאריך: 09.02.21, בארקיב.

הוצג בכנס: ICLR2021.

תחומי מאמר:

- מודלים גנרטיביים.
- variational autoencoder (VAE).
- energy-based models (EBM).

כלים מתמטיים במאמר:

- Reparameterization trick.
- Langevin dynamics.
- markov chain monte-carlo - MCMC.
- התפלגות גיבס.

תמצית מאמר:

המאמר מציע מודל גנרטיבי המשלב VAE עם EBM בשביל ליהנות מהיתרונות של שניהם:

- היכולת של EBM לייצג התפלגויות מורכבות בצורה מדויקת.
- היכולת של VAE לגנרט דגימות בצורה מהירה ויעילה.

השילוב של VAE ו-EBM נותן מענה לחולשות העיקריות של שתי השיטות האלו:

- EBM: דגימה מאוד איטית המגבילה שימוש בגישה זו רק לגינרוט תמונות בגודל קטן.
- VAE: יכולת מידול לא מדויקת של התפלגות הדאטה המתבטא ביצירה של תמונות מטושטשות.

המאמר מציע ארכיטקטורה, הנקראת VAEBM, המורכבת משני מרכיבים עיקריים: VAE ו-EBM. ארכיטקטורה זו מנצלת את היכולת של רכיב ה-VAE בשביל ללמוד את המבנה הכללי של המרחב הלטנטי מחד, כאשר רכיב ה-EBM בא "לתקן" את אי-הדיוקים של רכיב ה-VAE ב"אזורים שאין בהם דאטה אמיתי". במאמר טוענים שמכיוון ש-VAE מצליח לבנות קירוב יחסית טוב של התפלגות הדאטה, לא נדרש מספר צעדים גבוה עבור עדכון הפרמטרים של EBM. בנוסף, שימוש ב-VAE מאפשר להאיץ את יכולת הדגימה של EBM עי" רפרמטריזציה של במרחב הלטנטי. ולבסוף, מאחר ו-VAE כופה על המרחב הלטנטי להיות מפולג עם התפלגות רציפה, הוא "משרה" התפלגות "חלקה" יותר גם של הדאטה שהוא יוצר, שגורם לדגימה יותר יעילה עם MCMC.

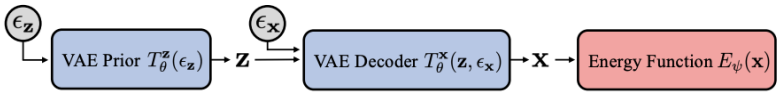


Figure 1: Our VAEBM is composed of a VAE generator (including the prior and decoder) and an energy function that operates on samples $\mathbf{x}$ generated by the VAE. The VAE component is trained first, using the standard VAE objective; then, the energy function is trained while the generator is fixed. Using the VAE generator, we can express the data variable $\mathbf{x}$ as a deterministic function of white noise samples $\boldsymbol{\epsilon}_{\mathbf{z}}$ and $\boldsymbol{\epsilon}_{\mathbf{x}}$. This allows us to reparameterize sampling from our VAEBM by sampling in the joint space of $\boldsymbol{\epsilon}_{\mathbf{z}}$ and $\boldsymbol{\epsilon}_{\mathbf{x}}$. We use this in the negative training phase (see Sec. 3.1).

רעיון בסיסי:

המאמר מגדיר את ההתפלגות של המודל המגנרט h(**x**,**z**) כמכפלה של p_vae(**x**,**z**), ההתפלגות הרגילה של VAE ו-p_ebm(**x**) - פונקצית ההתפלגות הסטנדרטית של EBM (כלומר התפלגות גיבס). כאן **x** היא דגימה מהדומיין המקורי (למשל תמונות) ו-**z** הוא וקטור לטנטי ממרחב בעל מימד נמוך. כמובן שלכל אחד מהמודלים VAE ו-EBM הפרמטרים משלהם, והם המאומנים יחד במטרה למקסם את ה-log של h(**x**) על סט האימון. h(**x**) היא פונקצית ההתפלגות של דגימה **x** מהמרחב המקורי המתקבלת לאחר מרגינליזציה של המשתנה הלטנטי **z** מ-h(**x**). המאמר מראה כי המיקסום הישיר של ביטוי זה מחייב דגימות מההתפלגות האפוסטריורית של **z|x**, כאשר **x** מגונרט עם VAE, ויש לזה סיבוכיות חישובית גבוהה (עקב שימוש ב-MCMC לדגימה מהתפלגות זו).
נציין כי פונקצית המטרה של VAEBM היא סכום של פונקציית המטרה הסטנדרטית של VAE, המסומנת ב-f_vae, וזו של EBM, המסומנת ב-f_ebm. לאור זה המאמר מציע לבצע את המקסום של בשני שלבים:

- מקסום של f_vae, כאשר הפרמטרים של f_ebm מוקפאים.
- מקסם את רכיב הלוס של EBM כאשר הפרמטרים של VAE מוקפאים.

בנוסף בשביל להקל על הדגימה מ-h(**x,z**), הנחוצה בשביל שערוך הגרדיאנט של רכיב ה-EBM, המאמר מציע לעשות רפרמטריזציה משותפת של **x** ושל המשתנה הלטנטי **z**. בעצם גישה זו מונעת דגימה דו שלבית (דוגמים את **z** ואז את **x|z**) שעלולה להיות מאוד בעייתית מאחר ו-p(**x|z**) עשויה להיות מרוכזת באיזור מאוד קטן ו-MCMC בדרך כלל מתקשה להתמודד עם המצב הזה.

תקציר מאמר:

קודם כל בואו נרענן את זכרוננו ונזכר מה זה בעצם VAE ו- EBM.

אוטו אנקודר וריאציוני (VAE):

VAE הינו מודל גנרטיבי שהומצא ב-2014 ומהווה הכללה של אוטו-אנקודר רגיל (AE - AutoEncoder), המשמש להורדת מימד של הדאטה. להבדיל מ-AE, ב-VAE ההתפלגות על המרחב הלטנטי מוגדרת מראש (למשל בתור התפלגות גאוסית סטנדרטית).
VAE מורכב משתי רשתות: רשת האנקודר ורשת הדקודר. המטרה של האנקודר היא לחשב את הפרמטרים של הייצוג הלטנטי עבור פיסת דאטה **x** מהמרחב המקורי. המטרה של הדקודר הינה לשחזר את הדוגמא מייצוג הלטנטי שלה **z**.
אז איך עובד VAE? קודם כל מחשבים את הפרמטרים של הוקטור הלטנטי של דוגמא **x** באמצעות האנקודר, דוגמים מהתפלגות, המוגדרת על ידי פרמטרים הנ״ל, את הוקטור הלטנטי **z**. לאחר מכן מזינים את **z** לדקודר לשחזור של הדוגמא המקורית **x**. פונקציית הלוס של VAE מורכבת משני מרכיבים:

- לוס השחזור: דיוק השחזור של הקלט **x**, הנמדד באמצעות מרחק L2 בין **x** לקלט של הדקודר.

- מרחק בין ההתפלגות המטרה על המרחב הלטנטי לבין ההתפלגות, המחושבת באמצעות האנקודר (במרחק KL).

נציין כי לא ניתן לבצע גזירה ישירה של פונקציית הלוס של VAE לפי הפרמטרים של הווקטור הלטנטי, המחושבים על ידי האנקודר (גזירה זו נחוצה במהלך אופטימיזצית המשקלים (backprop) של רשת האנקודר). כדי להתגבר על קושי זה, משתמשים בטריק של רפרמטריזציה. במקום לדגום מההתפלגות המוגדרת על ידי הפלט של האנקודר **o_enc**, דוגמים מהתפלגות קבועה ולא תלויה בפרמטרים (בדרך כלל גאוסית סטנדרטית). לאחר מכן מפעילים טרנספורמציה (לינארית במקרה הגאוסי), המוגדרת על ידי **o_enc** על דגימה זו כדי לדמות דגימה מהתפלגות בעלת פרמטרים **o_enc.**

<u>מודלים מבוססי אנרגיה (EBM)</u>: גם EBM הוא מודל גנרטיבי, אך להבדיל מ-VAE, הוא משערך את פונקציית ההתפלגות של הדאטה בצורה מפורשת (כלומר ממדל אותה על יד רשת נוירונים). בשביל לגנרט דוגמאות חדשות באמצעות EBM, צריך לדגום מפונקצית ההתפלגות, המשוערכת על ידה (נסמן אותו ב-p_ebm). בדרך כלל דגימה זו מתבצעת באמצעות אחד הוריאנטים של MCMC. עקב הסיבוכיות הגבוהה של שיטת דגימה זו, כרגע ניתן לגנרט עם EBM רק תמונות קטנות (עד 64x64).

אימון של EBM מתבצע באמצעות מקסום של פונקציית מטרה, שהיא התוחלת של הלוג של p_ebm על סט האימון (נראות מירבית). p_ebm מוגדרת ע"י התפלגות גיבס שהיא אקספוננט שלילי של פונקציית האנרגיה E(**x**), מוכפלת בקבוע נרמול (בשביל לאלץ את p_ebm להיות פונקציית התפלגות). EBM מאומן באמצעות GD, כאשר הגרדיאנט של פונקציית הלוס מורכב מהפרש של ערכי פונקציית האנרגיה E עבור דגימות מ-p_ebm (השלב השלילי), ושל על ערכי E עבור דוגמאות מסט האימון (השלב החיובי). מכיוון שלא ניתן לדגום מ-p_ebm ישירות, משתמשים באחד הסוגים של MCMC - בדרך כלל בדינמיקה של לנגווין (LD). LD הוא תהליך איטרטיבי הבונה דגימות של p_ebm ע"י הזזת דגימות בכיוון הגרדיאנט השלילי (לפי הדגימות **x**) של E במטרה לדגום מ-E במקומות שבהם ל-E ערכים נמוכים - כלומר באזורים בהם לפונקציית ההתפלגות ערכים גבוהים. נציין כי שלב זה לוקח את רוב הזמן באימון של EBM.

<u>איך בעצם משלבים את EBM ו- VAE</u>: המאמר מראה כי ניתן לחסום את פונקצית הלוס של VAEBM מלמטה על ידי הסכום של הלוס הסטנדרטי של VAE והלוס של EBM, המסומן על ידי L_ebm. נציין כי האופטימיזציה של L_ebm כוללת שיערוך של גרדיאנט של L_ebm לפי הפרמטרים של VAE (כי הדאטה **x** מגונרט על ידי VAE). שערוך זה מאוד כבד מבחינה חישובית כי הוא כרוך בדגימה מההתפלגות הפוסטריורית של דגימות מ-VAE. עקב כך המאמר מציע לבצע אופטימיזציה של L_ebm ו- L_vae לסירוגין, דבר שמונע את הצורך לגזור את L_ebm לפי הפרמטרים של רשת VAE. הרעיון השני של המאמר זה שימוש בטריק של רפרמטריזציה על **x** ו-**z** (המשתנה הלטנטי) <u>בו זמנית.</u> מהלך זה מונע את הצורך לדגום מההתפלגות המותנת **x|z** כי דגימה כזו עלולה להיות בעייתית (הוסבר בהרחבה בפרק "רעיון בסיסי").

כאחד ההרחבות של תהליך האימון המאמר מציע לבצע כמה איטרציות GD בשביל לקרב את פונקציית ההתפלגות של **x** לפונקצית האנרגיה E אחרי שלב האופטימיזציה של VAE (מנסים להביא למינימום את מרחק KL ביניהם). זה מזכיר את העדכון של הגנרטור ב-Wasserstein GAN.

הישגי מאמר:

המאמר משווה את ביצועיו של VAEBM מול מודלים גנרטיביים רבים מסוגים שונים ומראה את עליונותו של VAEBM במונחי inception score(IS) ו- freshet inception distance (FID) על רובם (האמת שהיחיד שמנצח אותם ב-FID על CIFAR10 זה StyleGAN2 שיש לו ארכיטקטורה מורכבת בהרבה). נציין לחיוב את הביצועים החזקים של VAEBM על StackedMNIST. כל תמונה ב-StackedMNIST היא שילוב של 3 תמונות של MNIST המקורי אז יש 1000 מודים. VAEBM מצליח לשחזר את כל המודים להבדיל מכמה מודלים עדכניים של GAN.

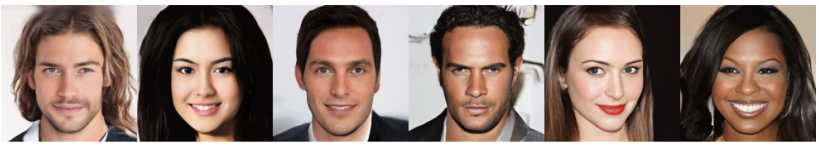
המאמר גם משווה את יעילות דגימה של VAEBM מול מודל גינרוט חזק denoising score matching ומציין כי VAEBM יעיל יותר מפי 12 ממנו בהיבט זה כאשר איכות התמונות המגונרטות באמצעות שני מודלים אלו היא די קרובה (לפחות ויזואלית).

Table 1: IS and FID scores for unconditional generation on CIFAR-10.

	Model	IS↑	FID↓
Ours	VAEBM w/o persistent chain	8.21	12.26
	VAEBM w/ persistent chain	8.43	12.19
EBMs	IGEBM (Du & Mordatch, 2019)	6.02	40.58
	EBM with short-run MCMC (Nijkamp et al., 2019b)	6.21	-
	F-div EBM (Yu et al., 2020a)	8.61	30.86
	FlowCE (Gao et al., 2020)	-	37.3
	FlowEBM (Nijkamp et al., 2020)	-	78.12
	GEBM (Arbel et al., 2020)	-	23.02
	Divergence Triangle (Han et al., 2020)	-	30.1
Other Likelihood Models	Glow (Kingma & Dhariwal, 2018)	3.92	48.9
	PixelCNN (Oord et al., 2016b)	4.60	65.93
	NVAE (Vahdat & Kautz, 2020)	5.51	51.67
	VAE with EBM prior (Pang et al., 2020)	-	70.15
Score-based Models	NCSN (Song & Ermon, 2019)	8.87	25.32
	NCSN v2 (Song & Ermon, 2020)	-	31.75
	Multi-scale DSM (Li et al., 2019)	8.31	31.7
	Denoising Diffusion (Ho et al., 2020)	9.46	3.17
GAN-based Models	SNGAN (Miyato et al., 2018)	8.22	21.7
	SNGAN+DDLS (Che et al., 2020)	9.09	15.42
	SNGAN+DCD (Song et al., 2020)	9.11	16.24
	BigGAN (Brock et al., 2018)	9.22	14.73
	StyleGAN2 w/o ADA (Karras et al., 2020a)	8.99	9.9
Others	PixelIQN (Ostrovski et al., 2018)	5.29	49.46
	MoLM (Ravuri et al., 2018)	7.90	18.9

(a) CelebA 64

(b) LSUN Church 64



(c) CelebA HQ 256

Figure 3: Qualitative results on CelebA 64, LSUN Church 64 and CelebA HQ 256. For CelebA HQ 256, we initialize the MCMC chains with low temperature NVAE samples ($t = 0.7$) for better visual quality. On this dataset samples are selected for diversity. See Appendix H for additional qualitative results and uncurated CelebA HQ 256 samples obtained from higher temperature initializations. Note that the FID in Table 3 is computed with full temperature samples.

דאטהסטים: SVHN, CIFAR100, CelebA, StackedMNIST

נ.ב.

המאמר מציע הרעיון די חזק בתחום המודלים הגנרטיביים המשלב VAE ו-EBM ומראה תוצאות מבטיחות. בינתיים עוד לא ברור האם רעיון זה יוכל לסכן את שליטתם של GANs בתחום.

#deepnightlearners

הפוסט נכתב על ידי מיכאל (מייק) ארליכסון, PhD, Michael Erlihson.

מיכאל עובד בחברת סייבר Salt Security בתור Principal Data Scientist. מיכאל חוקר ופועל בתחום הלמידה העמוקה, ולצד זאת מרצה ומנגיש את החומרים המדעיים לקהל הרחב.