

RI-GAN

Kosta Grujić, Momčilo Knežević

Matematički fakultet

Sažetak

1 Uvod

2 *Related Work*

U ovom poglavlju ćemo opisati princip rada GAN-ova, GAN arhitekturu koju koristimo i metrika kojima ćemo kvantifikovati obučenost modela.

2.1 Generativne suparničke mreže

Prvo dajemo neformalnu ideju. Osnovni cilj je naučiti raspodelu podataka tako da je moguće vršiti uzorkovanje, čime se dobija mogućnost generisanja podataka koji zapravo ne postoje. GAN arhitektura podrazumeva upotrebu dva modela. Prvi je zadužen za učenje pomenute raspodele, dok drugi treba da razlikuje stvarne i generisane podatke. Suparničkim učenjem ova dva modela se stvarna raspodela podataka sve bolje aproksimira.

Neka je $\mathcal{X} \subset \mathbb{R}^{c \times d \times d}$ prostor c -kanalnih slika dimenzija $d \times d$. Bilo koji skup slika možemo predstaviti nekom raspodelom p_r nad \mathcal{X} . Skup nad kojim obučavamo model možemo definisati kao prost slučajaj uzorak iz raspodele p_r . Cilj je naučiti parametrizovanu raspodelu p_g koja aproksimira p_r .

GAN se sastoji od *generatora* i *diskriminatora*. Diskriminator posmatramo kao preslikavanje $D : \mathcal{X} \rightarrow [0, 1]$, dok generator kao $G : \mathcal{Z} \rightarrow \mathcal{X}$, gde je \mathcal{Z} latentni prostor. Pomenuti latentni prostor služi za uzorkovanje šuma na osnovu kojeg generator pokušava da generiše elemente raspodele p_r . Ako takvo uzorkovanje šuma vršimo iz raspodele p_z definisane nad \mathcal{Z} , tada raspodelu p_g možemo definisati kao $G(p_z)$.

Napomenimo da male vrednosti diskriminatora odgovaraju generisanim podacima, dok veće stvarnim. Prema tome, generator nastoji generisanju podataka za koje diskriminator daje vrednosti bliske 0, dok diskriminator za takve ulaze nastoji da se sve viši približi vrednosti 1.

Obučavanje diskriminatora i generatora se vrši uporedo, usled čega se može videti kao min-max igra. Takav proces se može formalizovati:

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_r} [\log D(x)] + \mathbb{E}_{z \sim p_z} [\log(1 - D(G(z)))]. \quad (1)$$

Primetimo da ne postoji ograničenje za izbor modela D i G . Međutim, pretpostavljajući njihovu diferencijabilnost i parametrizovanost, obučavanje se može vršiti propagacijom unazad i time u velikoj meri automatizovati.

U početnim iteracijama obučavanja može doći do zasićenja gradijenata [1]. Usled lošeg generatora, diskriminator je u stanju vrlo precizno da razlikuje lažne i prave podatke. S tim u vezi, optimizacija iz formule 1 se može zapisati u numerički pogodnijem obliku:

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_r} [\log D(x)] - \mathbb{E}_{z \sim p_z} [\log D(G(z))]. \quad (2)$$

2.2 DCGAN

Generativne suparničke mreže ne podrazumevaju nikakav konkretan model ni za generator ni za diskriminator. S tim u vezi, zavisnosti od domena primene se mogu dizajnirati odgovarajući modeli. Kako se CelebA sastoji od slika, to se opredeljujemo za arhitekturu baziranu na konvolutivnim neuronskim mrežama. Koristićemo DCGAN arhitekturu [3], pre svega zbog povoljnog vremena obučavanja.

Prvo ćemo objasniti strukturu diskriminatora ove arhitekture. Sastoji se od 5 blokova koji vrše decimaciju prostornih dimenzija ulaznog tenzora za faktor 2. Svaki blok se sastoji od konvolutivnog sloja praćenog BN slojem [12], osim u poslednjem bloku gde ne postoji sloj normalizacije. Aktivaciona funkcija u svakom konvolutivnom sloju je ReLU [13], osim u poslednjem gde je sigmoid. Konvolutivni slojevi imaju redom 128, 256, 512, 1024, 1 kanala. Decimacija je postignuta postavljanjem pomeraja konvolutivnog sloja na 2. Ulazni tenzor je trokanalna RGB slika, koja se prethodno normalizuje sa $\hat{\mu} = 0.5$ i $\hat{\sigma} = 0.5$.

Generator u odnosu na diskriminator ima inverznu strukturu. Koristi 4 ekspanziona bloka i transponovane konvolutivne slojeve umesto običnih. Broj kanala u konvolutivnim slojevima je redom 512, 256, 128, 3. Pre nego što se latentni šum prosledi prvom konvolutivnom sloju, vrši se projektovanje potpuno povezanim slojem u prostor $\mathbb{R}^{1024 \times 4 \times 4}$. Aktivaciona funkcija poslednjeg sloja je hiperbolički tangens.

2.3 Lipšic neprekidnost generatora

Pretpostavimo da je diskriminator D neuronska mreža koja za ulaz $x \in \mathcal{X}$ ima sledeći oblik:

$$f(x, \theta) = \left(\prod_{i=1}^L a_i(\beta_i + W_i) \right) (x), \quad (3)$$

gde je $\theta = \{W_i\}_{i=1}^L$ skup parametara koji se obučavaju, $W_i \in \mathbb{R}^{d_i \times d_{i-1}}$ i $\{a_i\}_{i=1}^L$ aktivacione funkcije koje se primenjuju član po član. Napomenimo da d_0 predstavlja broj kanala ulaza, dok je $d_L = 1$.

Dokazano je [1] da za fiksiran generator G , optimalan diskriminator ima oblik:

$$D_G^*(x) = \frac{p_r(x)}{p_r(x) + p_g(x)} = \sigma(f^*(x)), \quad (4)$$

gde je σ sigmoid, a $f^*(x) = \log p_r(x) - \log p_g(x)$. Tada dobijamo:

$$\nabla_x D_G^*(x) \propto \nabla_x f^*(x) = \frac{1}{p_r(x)} \nabla_x p_r(x) - \frac{1}{p_g(x)} \nabla_x p_g(x). \quad (5)$$

Primetimo da gradijent $\nabla_x f^*(x)$ nije ograničen. Iz tog razloga dolazi do nepredvidivih oscilacija tokom obučavanja i česte divergencije. Zato je neophodno nametnuti dodatne uslove za f . Na osnovu [14], diskriminator obučavamo u prostoru K -Lipšic neprekidnih funkcija¹. Tada je gradijent iz formule 5 ograničen. Formula 1 postaje:

$$\min_G \max_{\|f\|_{\text{Lip}} \leq K} V(D, G), \quad (6)$$

gde je $\|f\|_{\text{Lip} \leq K}$ skup svih K -Lipšic neprekidnih funkcija u metričkom prostoru (\mathcal{X}, l_2) .

Spektralna normalizacija (SN) [4] se zasniva na uslovu $\|W_i\|_2 = 1$. Kako je $\|W_i\|_2 = \sqrt{\lambda_{\max}(W_i^T W_i)}$, to je neophodno odrediti najveću sopstvenu vrednost matrice $W_i^T W_i$. Pomenuti postupak je moguće sprovesti numerički.

U odnosu na standardnu DCGAN implementaciju, u našim eksperimentima je SN korišćena umesto BN i to isključivo kod diskriminatora.

2.4 IS

Adekvatno obučan GAN poseduje sigurnost i varijabilnost. To znači da generator može dati svojstvene podatke, kao i da ravnomerno pokriva čitav kombinatorni prostor.

¹ $(\exists K > 0)(\forall x, y \in \mathcal{D}) \|f(x) - f(y)\| \leq K \cdot \|x - y\|$

Prepostavimo da svakoj instanci skupa podataka x možemo pridružiti oznaku y . Skup podataka takve oznake ne mora da sadrži, ali ih je moguće uvesti dodatnim modelom \mathcal{M} (kao što je Inception [10]) koji je treniran kao klasifikacioni model na nekom drugom skupu podataka (kao što je ImageNet [11]). Zato možemo posmatrati uslovnu raspodelu $p_{\mathcal{M}}(y|x)$. Marginalna raspodela $p_{\mathcal{M}}(y)$ se dobija kao $\int_x p_{\mathcal{M}}(y|x)dp_g(x)$.

Prema tome, $p_g(y|x)$ treba da odgovara Dirakovoj δ funkciji, a $p_g(y)$ ravnomernoj raspodeli. Konačno, IS² definišemo:

$$\text{IS}(p_g) = \exp(\mathbb{E}_{x \sim p_g}[KL(p_g(y|x)||p_g(y))]), \quad (7)$$

gde je KL Kulbek-Lajblerova divergencija.

Ispostavlja se da ova metrika dobro korelira sa ljudskom odlukom [7]. Međutim, IS ni na koji način ne koristi p_r . Takođe, skup podataka nad kojim je \mathcal{M} treniran možda nema ni približno sličnu raspodelu kao p_r usled čega GAN model biva proglašen lošim.

2.5 FID

Kao što smo već rekli, ideja obučavanja GAN-a je što bolja aproksimacija p_r raspodelom p_g . Ukoliko p_r predstavlja raspodelu podataka iz realnog sveta (poput slika ljudi), tada možemo pretpostaviti ograničenost njenog nosača. U tom slučaju važi:

$$p_r \stackrel{s.s.}{=} p_g \iff (\forall k \in \mathbb{N}) \int_{\mathcal{X}} p_r(x)x^k dx = \int_{\mathcal{X}} p_g(x)x^k dx. \quad (8)$$

Drugim rečima, jednakost momenata povlači skoro sigurnu jednakost raspodela. U praksi je desnu stranu ekvivalencije 8 skoro nemoguće pokazati. Zato se ona aproksimira za vrednosti do nekog k . Mi ćemo aproksimaciju vršiti za $k = 2$.

Kako raspodele p_r i p_g nisu eksplicitno poznate, uvodimo operator ϕ tako da jednakost $\phi(p_r)$ i $\phi(p_g)$ odgovara jednakosti odgovarajućih raspodela. Kao i kod IS, koristimo Inception model čiji će neki unutrašnji sloj predstavljati operator ϕ . Inception model obučen nad ImageNet skupu podataka poseduje kvalitetnu aproksimaciju raspodele slika iz stvarnog sveta čiji duboki slojevi poseduju vizuelne karakteristike visokog nivoa. Iz tog razloga, za očekivati je da forsiranjem sličnosti $\phi(p_r)$ i $\phi(p_g)$ postizemo sličnost p_r i p_g .

Međutim, raspodele $\phi(p_r)$ i $\phi(p_g)$ takođe nisu eksplicitno poznate. Poznato je da za fiksirane momente prva dva reda normalna raspodela ima

²eng. *Inception score*

maksimalnu entropiju. Zato pretpostavljamo: $\phi(p_r) = \mathcal{N}(\mu_r, \Sigma_r)$ i $\phi(p_g) = \mathcal{N}(\mu_g, \Sigma_g)$. Razliku ovih raspodela merimo Freševim rastojanjem:

$$F(\phi(p_r), \phi(p_g)) = \|\mu_r - \mu_g\| + \text{tr} \left(\Sigma_r + \Sigma_g - 2(\Sigma_r \Sigma_g)^{1/2} \right), \quad (9)$$

gde je tr operator traga matrice.

Konačno, definišemo $\text{FID}(p_r, p_g)$ ³ kao $F(\phi(p_r), \phi(p_g))$ [8]. Primitimo da za razliku od IS, FID koristi p_r .

3 Kvantifikovanje obučenosti

4 Eksperimenti

5 Zaključak

³eng. *Fréchet Inception distance*

Literatura

- [1] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio. Generative Adversarial Nets. *arXiv preprint arXiv:1406.2661*, 2014.
- [2] Qiantong Xu, Gao Huang, Yang Yuan, Chuan Guo, Yu Sun, Felix Wu, Kilian Weinberger. An empirical study on evaluation metrics of generative adversarial networks. *arXiv preprint arXiv:1806.07755*, 2018.
- [3] Alec Radford, Luke Metz, Soumith Chintala. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *arXiv preprint arXiv:1511.06434*, 2016.
- [4] Takeru Miyato, Toshiki Kataoka, Masanori Koyama, Yuichi Yoshida. Spectral Normalization for Generative Adversarial Networks. *arXiv preprint arXiv:1802.05957*, 2018.
- [5] Martin Arjovsky, Soumith Chintala, Léon Bottou. Wasserstein GAN. *arXiv preprint arXiv:1701.07875*, 2017.
- [6] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, Aaron Courville. Improved Training of Wasserstein GANs. *arXiv preprint arXiv:1704.00028*, 2017.
- [7] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, Xi Chen. Improved Techniques for Training GANs. *arXiv preprint arXiv:1606.03498*, 2016.
- [8] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernard Nessler. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. *NIPS*, 2017.
- [9] Ziwei Liu, Ping Luo, Xiaogang Wang, Xiaoou Tang. Deep Learning Face Attributes in the Wild. *Proceedings of International Conference on Computer Vision (ICCV)*, 2015.
- [10] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich. Going Deeper with Convolutions. *arXiv preprint arXiv:1409.4842*, 2014.
- [11] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael

- Bernstein, Alexander C. Berg, Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *arXiv preprint arXiv:1409.0575*, 2014.
- [12] Sergey Ioffe, Christian Szegedy. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv preprint arXiv:1502.03167*, 2015.
- [13] Vinod Nair, Geoffrey E. Hinton. Rectified linear units improve restricted boltzmann machines. *ICML: Proceedings of the 27th International Conference on International Conference on Machine Learning*, 2010.
- [14] Guo-Jun Qi. Loss-Sensitive Generative Adversarial Networks on Lipschitz Densities. *arXiv preprint:arXiv:1701.06264*, 2017.