چگونه می توان ساختار مولد و تفکیک کننده را کوانتمی کرد؟

مقاله‌ی پروژه‌ی پایانی محاسبات کوانتمی ترم 4031
نذر بکسری و آلانی (گرو و نک نیلم)، شمار دانشجویی: 402442297

در این خلاصه نویسی از مقاله $QGAN$ به توضیح اینکه چگونه می توان ایده‌ی GAN را با استفاده از کامپیوتر کوانتمی پیاده سازی کرد، می پردازیم. در ابتدا خود GAN ها را توصیف می کنیم یک شبکه مولد تخاصمی متشکل از یک مولد و یک تفکیک کننده می باشد که مولد سعی می کند نمونه های واقعی بسازد که مشابه نمونه های مطلق از دیتاست B باشد و هدف تفکیک این است که نمونه های جعلی را از نمونه های واقعی تمایز گذاری کند. برای مثال در مخزن پروژه‌ی یک GAN کلاسیک QGAN ترکیبی با یک کوانتمی و یک کلاسیک پیاده سازی شده است.

حالا اگر بخواهیم این ساختار این مدار را کوانتمی کنیم می توانیم فرض کنیم که مولد G و دیتاست (R) های کوانتمی پارامتری هستند که به ورودی میگیرند و یک خروجی برجسب دار، توضیح انحصاراتی می دهد و تفکیک کننده یک نمونه را ورودی می گیرد و یکی از دو خروجی زیر را می دهد.

$|real\rangle, |fake\rangle$

برای ساختار G فرض کنیم علاوه بر پارامتری که دریافت می کند یک برچسب لامبدا و یک بردار اضافی زد دریافت می کند و یک ماتریس چگالی مدنظر زیر خروجی می دهد.

$|z\rangle, |\lambda\rangle$

برای ساختار R فرض کنیم علاوه بر پارامتری که دریافت می کند یک برچسب لامبدا کند و یک ماتریس چگالی مثلث زیر خروجی می دهد.

$$R(|\lambda\rangle) = \rho_R^{|\lambda\rangle}$$

G علاوه بر پارامتر ثنای جی که دریافت می کند یک برچسب لامبدا و یک بردار اضافی زد دریافت می کند و یک ماتریس چگالی مانند زیر خروجی می دهد، بردار نویز

$|z\rangle$ می تواند دو نقش مفاوت بازی کند، یکی اینکه می نطمی مطلوبی را درون نمونه های ساخته شده به ازای یک برچسب ثابت لامبدا فراهم کند یکی هم اینکه اگر z

در طول تزین کردن مدل ها، اکثرا گفتیم کلیم می توانیم جریانی که لامبدا درون خود ندارد را به z بدهیم تا نمونه های واقعی و مثلو تر تولید کنیم در این روش خصوصیاتی از نمونه ها که تابعل برچسب لامبدا ثبت نشده اند را کنترل می کنیم.

$$G(\theta_G, |\lambda, z\rangle) = \rho_G^{|\lambda\rangle}(\theta_G, |z\rangle)$$

حالا به مدار تفکیک کننده علاوه بر پارامتر اصلی اش هم یک دیفی بدون تغییر از لامبدا (برای اینکه تفکیگر مولد بیشتر وابسته به لامبدا نشود) را می دهیم و ماتریس چگالی ساخته شده از یکی از مدار R یا G را ورودی می گیرد و یکی دیگر را می دهد.

$$D(\theta_D, |\lambda\rangle, R(|\lambda\rangle))$$

$$D(\theta_D, |\lambda\rangle, G(\theta_G, |\lambda, z\rangle))$$

در این مسئله تفکیک کننده می خواهد خطا را به کمترین مقدار برساند و مولد می خواهد خطا را به بیشترین مقدار برساند. و یک بهینه سازی تخاصمی داریم همین برای همین ضابطه زیر و ضوابطی که مزیت یا شناسایی نشن نمونه جعلی و حقیقی هست می تواند ساخت باشد.

$$V(\theta_G, \theta_D) = \frac{1}{2}\left(P(D(\theta_D, |\lambda\rangle, R(|\lambda\rangle)) = |real\rangle) + P(D(\theta_D, |\lambda\rangle, G(\theta_G, |\lambda, z\rangle)) = |fake\rangle)\right)$$

در توابع هزینه‌ی GAN ها مرسوم است که از توابع *log-likehood* استفاده نمود اما در اینجا ساده تر است که از توابع خطای ساخته شده از تجانبی که هر دو توابع محدب هستند لفظا بهینه شان مشابه است و فرقی در نتیجه نهایی حاصل نمی کند، با فرض اینکه تعدادی متفاهی و قابل شمارش برچسب داریم می توان توابع هزینه بر حسب پارامتر های اصلی مولد و تفکیک کننده را به صورت زیر نوشت:

$$V(\theta_G, \theta_D) = \frac{1}{2}\sum_{\lambda=1}^{\Lambda} P\left(D(\theta_D, |\lambda\rangle, R(|\lambda\rangle)) = |real\rangle \cap D(\theta_D, |\lambda\rangle, G(\theta_G, |\lambda, z\rangle)) = |fake\rangle\right)$$

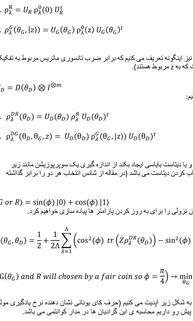
مولد تلاش می کند پارامتری را انتخاب کند که نتایج هزینه که رابطه مستقیم با نقاط تفکیک کننده دارد را کمینه کند و تفکیک کننده تلاش می کند این مقدار را با به روز رسانی پارامتر مربوطه به خود بیشینه کند.

$$\min_{\theta_G} \max_{\theta_D} V(\theta_G, \theta_D)$$

برای پیاده سازی مدار های مربوطه شکل بالی مدار به صورت زیر خواهد بود که رجیستر ها رو به شش بخش تقسیم می کنیم:

1. OUT D که شامل یک کیوبیت است و اندازی یک قرار می گیرد تا در نهایت خروجی تفکیک کننده را به ما نشان می دهد.
2. Both D شامل تعدادی کیوبیت برای محاسبات درون تفکیک کننده است.
3. Label D که جزئی برچسب است.
4. Both R|G شامل تن کیوبیت که ماتریس چگالی ساخته شده توسط G یا R را در آن ذخیره می کنیم.
5. Label R|G شامل برچسب است.
6. Both R|G شامل پارامتر اصلی z است

ساختار کلی مدار را می توانید در شکل زیر می توانید ببینید. فرض کنید که R به صورت یونیتری خالص سازی شده (operator purified unitary) هست که یک حالت کوانتمی را به یک حالت کوانتمی دیگر می برد و $G(\theta_G)$ هم همینطور در هر بار آموزش این شبکه مولد تا به روز کردن پارامتر خودش سعی می کند به R نزدیکتر شود و $D(\theta_D)$ هم سعی می کند که حالاتی که میگیرد را به دو نسله افراز کند به طوری که نمونه های واقعی در نسله $|real\rangle$ و نمونه های ساختگی در نسله $|fake\rangle$ باشند.



$$Z \equiv |real\rangle\langle real| - |fake\rangle\langle fake|$$

در ابتدای آموزش مدل از یک وضعیت تصادفی و با دلخواه مانند زیر شروع می کنیم.

$$\left(\theta_D^0, \theta_G^0\right) \rightarrow \text{initial arbitairy state}$$

و ابتدای هر مرحله آموزش ساخت کلی مدار مطابق زیر خواهد بود:

$$\rho_L^0 = (|0\rangle\langle 0|)^{\otimes d+1} \otimes |\lambda\rangle\langle\lambda| \otimes (|0\rangle\langle 0|)^{\otimes n} \otimes |\lambda\rangle\langle\lambda| \otimes |z\rangle\langle z|$$

ابتدا فرض کنید در حل ساخت یک نمونه به به وسیله دیتاست یا به وسیله مولد هستیم برای هر دو یک عملگر یونیتری در نظر می گیریم(بر اساس تعداد رجیستر های برچسب است).

$$U_R = I^{\otimes(1+d+z)} \otimes R$$

$$U_G(\theta_G) = I^{\otimes(1+d+z)} \otimes G(\theta_G)$$

حال بسته به اینکه کام یک از بین مولد و دیتاست انتخاب شده باشد بعد از اعمال عملگر ها وضعیت به یکی از دو حالت اول یا دوم تغییر پیدا می کند.

1. $\rho_L^R = U_R\, \rho_L^0 (0)\, U_R^\dagger$
2. $\rho_L^G(\theta_G, |z\rangle) = U_G(\theta_G)\, \rho_L^0(z)\, U_G(\theta_G)^\dagger$

حال نمونه ساخته شده وارد تفکیک کننده می شود. عملگر مربوط به تفکیک کننده را نیز اینگونه تعریف می کنیم که برابر ضرب تانسوری ماتریس مربوط به تفکیک کننده و یک ماتریس همانی به ابعاد مشخص شده(m برابر تعداد رجیستر هایی هست که به مربوط هستند).

$$U_D = D(\theta_D) \otimes I^{\otimes m}$$

پس از اینکه نمونه به تفکیک کننده رسید عملگر تفکیک کننده را بر آن اعمال می کنیم.

1. $\rho_L^{DR}(\theta_D) = U_D(\theta_D)\, \rho_L^R\, U_D(\theta_D)^\dagger$
2. $\rho_L^{DG}(\theta_D, \theta_G, z) = U_D(\theta_D)\, \rho_L^G(\theta_G, |z\rangle)\, U_D(\theta_D)^\dagger$

در طول تمرین مدل ها برای هاز گوری از اینکه وجود یک الگو در برگزاین مولد و یا دیتاست بایاسی ایجاد نکن از اندازه گیری یک سوپرپوزیشن نقاط زیر استفاده می کنیم که برای مثال D نقاط دهنده انتخاب کردن مولد و 1 نشان دهنده انتخاب کردن دیتاست می باشد. از مقدار β شانس انتخاب هر دو را برابر گذاشته است).

$$|G\ or\ R\rangle = \sin(\phi)\,|0\rangle + \cos(\phi)\,|1\rangle$$

بنابر این ما می توانیم تابع هزینه را به این شکل حساب بگیریم در ادامه روش گرادیان نزولی را برای به روز کردن پارامتر ها پیاده سازی می‌خواهم‌کرد.

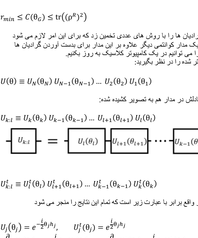
$$V(\theta_G, \theta_D) = \frac{1}{2} + \frac{1}{2\Lambda}\sum^{\Lambda}\left(\cos^2(\phi)\ tr\left(Z\rho_L^{DR}(\theta_D)\right) - \sin^2(\phi)\ tr\left(Z\rho_L^{DG}(\theta_D, \theta_G, z)\right)\right)$$

$$\left(G(\theta_G)\ and\ R\ will\ chasen\ by\ a\ fair\ coin\ so\ \phi = \frac{\pi}{4}\right) \rightarrow \min_{\theta_G}\max_{\theta_D} \frac{1}{2} + \frac{1}{4\Lambda}\sum^{\Lambda} tr\left(Z\left(\rho_L^{DR}(\theta_D) - \rho_L^{DG}(\theta_D, \theta_G, z)\right)\right)$$

در روش گرادیان نزولی بعد از مرحله قانو دام پارامتر های تفکیک کننده و مولد را به شکل زیر اپدیت می کنیم (هرچه کای بوتانی نشان دهنده نرخ یادگیری مولد و یا تفکیک کننده می باشد که در حالت کلی می تواند تابعی از β باشد، چاشی که پیش رو داریم محاسبه ی این گرادیان ها در مدار کوانتمی می باشد.

$$\theta_D^{k+1} = \theta_D^k + \chi_D^k\, \nabla_{\theta_D} V(\theta_D^k, \theta_G^k)$$

$$\theta_G^{k+1} = \theta_G^{k+1} + \chi_G^k\, \nabla_{\theta_G} V(\theta_D^k, \theta_G^k)$$



در حالت ایده آل ما به یک مرحله‌ای می رسیم که در آن مولد کاملا با دیتاست برابر می شود و یعنی عبارت 1 که برابر احتمال این است که تفکیک کننده با پارامتر بتواند نمونه های واقعی و نمونه های ساختگی که توسط مولد با پارامتر ساخته شده اند بر ابر یک نود باشد در این مرحله هیچ یک از تفکیک کننده و مولد نمی توانند استراتژی خود را بهتر کنند و مدل همگرا می شود و آموزش به پایان می رسد.

1. $P(Success\ D(\theta_D))|\theta_G\rangle$

$$G(\theta_G^*) = R, P(Success\ D(\theta_D^*))|\theta_G^*\rangle = \frac{1}{2} \rightarrow \nabla_{\theta_G} V(\theta_D^*, \theta_G^*) = 0, \nabla_{\theta_G} V(\theta_D^*, \theta_G^*) = 0$$

$$\rightarrow end\ of\ the\ training\ (model\ is\ converged)$$

اما مقادیر ممکن احتمال این پیشامد با تابع خطوی زیر کران دار می شود:

$$\frac{1}{2}\, tr\left(\rho^R \rho^G(\theta_G)\right)$$

$$\frac{1}{2}\, C(\theta_G) \leq P(Success\ D(\theta_D))|\theta_G\rangle \leq 1 - \frac{1}{2}\, C(\theta_G)$$

که خود از این تابع خلوص هر طبق طبیعت هدف دیتاست کران دار می شود:

$$r_{min} \leq C(\theta_G) \leq tr\left((\rho^R)^2\right)$$

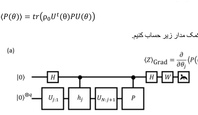
که r_{min} کوچک ترین مقدار روند ماتریس چگالی دکر شده هست.

حال از انجایی که گرادیان ها مشابهت با خلوص ماتریس چگالی هستند می توان گرادیان ها را با روش های عددی تخمین زد که برای این امر لازم می شود نمونه های زیادی در اطراف پارامتر های مولد و تفکیک کننده داشته باشیم، ما از یک مدار کوانتمی دیگر علاوه بر این مدار برای ساخت آورمن گرادیان ها استفاده می کنیم و بعد از آن با نرم به به دست می به گرادیان ها را برای مدار های مول ما اعمال می کنیم تو در یک کامپیوتر کلاسیک به روز بگیریم. یک تبدیل خطی یونیتری که توسط یک درایو متشکل از تعدادی متفاهی عضو پارامتر شده و از نظر یکی به:

$$U(\theta) \equiv U_N(\theta_N)\, U_{N-1}(\theta_{N-1}) \ldots U_2(\theta_2)\, U_1(\theta_1)$$

نمد گذاری زیر را چون در محاسبات پیش رو استفاده می شود معرفی می کنم معادلش در مدار هم به تصویر کشیده شده.

$$U_{E,l} \equiv U_N(\theta_N)\, U_{N-1}(\theta_{N-1}) \ldots U_{l+1}(\theta_{l+1})\, U_l(\theta_l)$$



و همچنین طبق خواص ماتریس های یونیتری داریم:

$$U_{E,l}^\dagger \equiv U_l^\dagger(\theta_l)\, U_{l+1}^\dagger(\theta_{l+1}) \ldots U_{N-1}^\dagger(\theta_{N-1})\, U_N^\dagger(\theta_N)$$

فرض کنید هر عضو توسط یک ماتریس همیلتونی به h_l ساخته شده پس عملگر ما در واقع برابر با عبارت زیر است که تمام این نتایج را منعدر می کند.

$$U_l(\theta_l) = e^{-\frac{i}{2}\theta_l h_l}, \qquad U_l^\dagger(\theta_l) = e^{\frac{i}{2}\theta_l h_l}$$

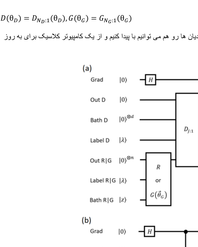
$$\frac{\partial}{\partial\theta_l} U_l(\theta_l) = -\frac{i}{2} h_l\, U_l(\theta_l), \qquad \frac{\partial}{\partial\theta_l} U_l^\dagger(\theta_l) = \frac{i}{2} h_l\, U_l^\dagger(\theta_l)$$

$$\frac{\partial}{\partial\theta_l} U(\theta) = -\frac{i}{2} U_{N:l+1}\, h_l\, U_{l:1}, \qquad \frac{\partial}{\partial\theta_l} U^\dagger(\theta) = \frac{i}{2} U_{l:1}^\dagger\, h_l\, U_{l+1:N}$$

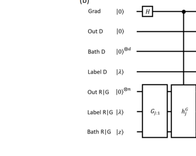
اگر یک شدت اولیه برای کیوبیت ها تعریف بکنیم امید ریاضی ازروزول برحسب لئا می شود:

$$\langle P(\vec{\theta})\rangle = tr\left(\rho_0 U^\dagger(\vec{\theta}) P U(\vec{\theta})\right)$$

حل با استفاده از دانسته های مان می توانیم گرادیان امید ریاضی ازروزول را به کمک مدار زیر حساب کنیم.



برگردیم به مسئله اصلی حالا که این ساختار چندلایه U را توانستیم گرادیان بگیریم می توانیم فرض کنیم که ساختار مولد و تفکیک کننده هم مشابه این ساختار است مشابه زیر:

$$D(\theta_D) = D_{N_D:1}(\theta_D), G(\theta_G) = G_{N_G:1}(\theta_G)$$

بنابر این با این مدار می توانیم هم مدل مولد و تفکیک کننده را بدین شکل بپیاده کنیم و گرادیان ها را هم می توانیم با پیتا کلیپور تو کلاسیک برای به روز رسانی پارامتر ها استفاده کنیم.