بنام خدا

علی قاسم زاده ۴۰۱۱۰۶۳۳۹

Subject

Date : Year: Month: Day:

① ابتدا فرضیات و اطلاعاتی که داریم را می نویسیم:

$P_t(x)$ ⟵ توزیع x در زمان t

$P_{0t}(x'|x)$ ⟵ احتمال رفتن از x_0 به x_t

forward:

$$dx = f(x,t)\,dt + g(t)\,d\bar{w}$$

Inverse:

$$dx = [f(x,t) - g(t)^2 \nabla_x \log P_t(x)]\,dt + g(t)\,d\bar{w}$$

$$J_{SM}(\theta;\lambda(\cdot)) = \frac{1}{2}\int_0^T E_{P_t(x)}\left[\lambda(t)\left\|\nabla_x \log P_t(x) - s_\theta(x,t)\right\|_2^2\right] dt$$

$$J_{DSM}(\theta;\lambda(\cdot)) = \frac{1}{2}\int_0^T E_{P(x)P_t(x'|x)}\left[\lambda(t)\left\|\nabla_{x'} \log P_{0t}(x'|x) - s_\theta(x',t)\right\|_2^2\right] dt$$

انتظار داریم در نهایت: $P_T(x) \approx \pi(x)$

پس از انتشار از $\hat{x}_\theta(T) \sim \pi$ شروع می کنیم و به $\hat{x}_\theta(0) \sim P_\theta^{SDE}$ می رسیم.

$$d\hat{x} = [f(\hat{x},t) - g(t)^2 s_\theta(\hat{x},t)]\,dt + g(t)\,d\bar{w}, \quad \hat{x}_\theta(T) \sim \pi$$

$$\frac{dx}{dt} = f(x,t) - \frac{1}{2} g(t)^2 \nabla_x \log P_t(x) \quad \leftarrow \text{ODE}$$

$$\frac{d\tilde{x}}{dt} = f(\tilde{x},t) - \frac{1}{2} g(t)^2 s_\theta(\tilde{x},t) \quad \leftarrow \text{تقریب ما}$$

parsian

Subject

Date : Year: Month: Day:

$$D_{KL}(P \,\|\, P_\theta^{SDE}) \le J_{SM}(\theta; g(\cdot)) + D_{KL}(P_T \,\|\, \pi)$$ قضیه ۱:

$$H(P) = H(P_T(x)) - \frac{1}{2}\int_0^T E_{P_t(x)}\left[2\nabla \cdot f(x,t) + g(t)^2 \,\|\nabla_x \log P_t(x)\|_2^2\right] dt$$

5

10

15

20

Subject

Date : Year Month Day

الف) KL بین p و p_θ^{SDE} به ما نشان می‌دهد که مدل‌ما چقدر از توزیع داده‌ی واقعی فاصله دارد.

ایده‌آل ما این است که $p \approx p_\theta^{SDE}$. یعنی واگرایی ما کمینه شود

Jsm هزینه‌ی اصلی تخمین score است و هرچه آنرا کوچکتر کنیم یعنی بهتر score

را یاد گرفته‌ایم.

KL بین p_T و π در زمان نهایی T، توزیع نویز شده p_T چقدر با توزیع هدف π تفاوت دارد.

این کسان می‌گوید که: اگر می‌خواهید کل واگرایی KL بین p و p_θ^{SDE} کم شود، کافی است که:

① تابع score را خوب یادگرفته باشید تا Jsm کوچک شود، ② مطمئن باشید که فرایند پیش‌رو در زمان T به توزیع p_T که از نظر KL فاصله‌ی چندانی با π نداشته باشد، می‌رسد.

در مقاله هم داریم که اگر π را ثابت fixed prior در نظر بگیریم:

$$-E_{p(x)}[\log p_\theta^{SDE}] \le Jsm(\theta; g(\cdot)^2) + C_1$$

هنگامی که اسکور به خوبی یاد گرفته شود، فرایند معکوس می‌تواند نمونه‌های نمونه‌ی واقعی را قابل قبولی بسازد، پس از نظر کیفیت بهبود داریم، از طرفی کوچک شدن Jsm باعث

parsian

Subject

Date : Year: Month: Day:

نزدیک شدن P_θ^{SDE} و P درفضای توزیع‌ها می‌شود. بنابراین اگر با معیارهایی مثل KL, FID یا حتی تحت شرایطی likelihood ارزیابی کنیم، بهبود در JSM باعث بهبود در این معیارها نیز خواهد شد.

← JSM و در کل این کران بالا از نظر کیفی و کمی همسو عمل می‌کند. وقتی این عبارت کوچک شود، هم از بعد بهینه و هم از بعد آماری، احتمالاتی نمونه‌های نزدیک به واقعیت و توزیع واقعی خواهیم داشت.

ب) حکم:

$$-E_{P(x)}\left[\log P_\theta^{SDE}(x)\right] \le -E_{P_T(x)}\left[\log \pi(x)\right] + \frac{1}{2}\int_0^T E_{P_t(x'|x)P(x)}\left[g(t)^2 \left\| s_\theta(x',t) - \nabla_{x'}\log P_t(x'|x)\right\|_2^2 - g(t)^2\left\|\nabla_{x'}\log P_t(x'|x)\right\|_2^2 - 2\nabla_{x'} f(x',t)\right] dt$$

$$D_{KL}(P \,\|\, P_\theta^{SDE}) = E_{P(x)}\left[\log \frac{P(x)}{P_\theta^{SDE}(x)}\right] = E_{P(x)}[\log P(x)] - E_{P(x)}\left[\log P_\theta^{SDE}(x)\right]$$

$$= -H(P) - E_{P(x)}\left[\log P_\theta^{SDE}(x)\right] \Rightarrow$$

$$-E_{P(x)}\left[\log P_\theta^{SDE}(x)\right] \le H(P) + JSM(\theta; g(\cdot)) + D_{KL}(P_T \,\|\, \pi)$$

parsian

Subject

Date : Year: Month: Day:

$$D_{KL}(P_T \| \pi) = E_{P_T(x)}[\log P_T(x)] - E_{P_T(x)}[\log \pi(x)] =$$

$$- H(P_T(x)) - E_{P_T(x)}[\log \pi(x)]$$

$$\rightarrow H(P) + D_{KL}(P_T \| \pi) = H(P_T(x)) - \frac{1}{2}\int_0^T E_{P_t(x)}[2\nabla \cdot f(x,t) + g(t)^2 \|\nabla_x \log P_t(x)\|_2^2]\, dt$$

$$- H(P_T(x)) - E_{P_T(x)}[\log \pi(x)]$$

$$\rightarrow - E_{P(x)}[\log P_\theta^{SDE}(x)] \leq - E_{P_T(x)}[\log \pi(x)] - \frac{1}{2}\int_0^T E_{P_t(x)}[2\nabla \cdot f(x,t)$$

$$+ g(t)^2 \|\nabla_x \log P_t(x)\|_2^2]\, dt + \frac{1}{2}\int_0^T E_{P_t(x)}[g(t)^2 \|\nabla_x \log P_t(x) - s_\theta(x,t)\|_2^2]\, dt$$

$$\rightarrow - E_{P(x)}[\log P_\theta^{SDE}(x)] \leq - E_{P_T(x)}[\log \pi(x)] + \frac{1}{2}\int_0^T E_{P_t(x)}[g(t)^2 \|s_\theta(x,t)$$

$$- \nabla_x \log P_t(x)\|_2^2 - g(t)^2 \|\nabla_x \log P_t(x)\|_2^2 - 2\nabla \cdot f(x,t)]\, dt \quad ***$$

$$\int_0^T E_{P_{0t}(x'|x)P(x)}[g(t)^2 \| s_\theta(x',t) - \nabla_{x'} \log P_{0t}(x'|x)\|_2^2 - g(t)^2 \| \nabla_{x'} \log$$

$$P_{0t}(x'|x)\|_2^2]\, dt = \int_0^T E_{P_{0t}(x'|x)P(x)}[g(t)^2 (\| s_\theta(x',t)\|_2^2 - 2\nabla_{x'} \log P_{0t}(x'|x)^T$$

$$s_\theta(x',t))]\, dt = \int_0^T \iint_{x\, x'} P_{0t}(x'|x) P(x)\, g(t)^2 (\|s_\theta(x',t)\|_2^2 - 2\nabla_{x'} \log P_{0t}(x'|x)^T$$

$$s_\theta(x',t))\, dx'\, dx\, dt = \int_0^T g(t)^2 \int_{x'} \|s_\theta(x',t)\|_2^2 \underbrace{\int_x P(x) P_{0t}(x'|x)\, dx}_{P_t(x')}\, dx'\, dt$$

parsian

Subject

Date : Year Month Day

$$-2\int_0^T g(t)^2 \int_{x'} s_\theta(x',t)^T \int_x p(x)\, p_t(x'|x)\, \nabla_{x'} \log p_{0t}(x'|x)\, dx\, dx'\, dt$$

$$= \int_0^T g(t)^2\, E_{p_t(x')}\left[\| s_\theta(x',t)\|_2^2\right] dt - 2\int_0^T g(t)^2 \int_{x'} s_\theta(x',t)^T \int_x p(x)\, p_{0t}(x'|x)$$

$$\frac{\nabla_{x'} p_{0t}(x'|x)}{p_{0t}(x'|x)}\, dx\, dx'\, dt = \qquad \nabla_{x'} \int_x p(x)\, p_{0t}(x'|x)\, dx = \nabla_{x'} p_t(x')$$

$$\int_0^T g(t)^2 E_{p_t(x')}\left[\| s_\theta(x',t)\|_2^2\right] dt - 2\int_0^T g(t)^2 \int_{x'} s_\theta(x',t)^T \nabla_{x'} p_t(x')\, dx'\, dt$$

$$p_t(x')\, \nabla_{x'} \log p_t(x') \rightarrow E_{p_t(x')}[\text{---}]$$

$$= \int_0^T g(t)^2 E_{p_t(x')}\left[\| s_\theta(x',t)\|_2^2 - 2 s_\theta(x',t)^T \nabla_{x'} \log p_t(x')\right] dt$$

$$= \int_0^T g(t)^2 E_{p_t(x')}\left[\| s_\theta(x',t) - \nabla_{x'} \log p_t(x')\|_2^2 - \|\nabla_{x'} \log p_t(x')\|_2^2\right] dt$$

$$= \int_0^T E_{p_t(x)}\left[g(t)^2 \| s_\theta(x,t) - \nabla_x \log p_t(x)\|_2^2 - g(t)^2 \|\nabla_x \log p_t(x)\|_2^2\right] dt$$

از طرفی داریم :

$$\int_0^T E_{p_t(x'|x)p(x)}\left[-2\nabla\cdot f(x',t)\right] dt = \int_0^T -2\nabla\cdot f(x',t)\left(\int p_t(x'|x)p(x)dx\right) dx'\, dt$$

$$= \int_0^T E_{p_t(x')}\left[-2\nabla\cdot f(x',t)\right] dt \Rightarrow$$

$$\frac{1}{2}\int_0^T E_{p_{0t}(x'|x)p(x)}\left[g(t)^2 \| s_\theta(x',t) - \nabla_{x'} \log p_{0t}(x'|x)\|_2^2 - g(t)^2 \|\nabla_{x'} \log p_t(x')\|_2^2 - 2\nabla\cdot f(x',t)\right] dt$$

parsian

Subject

Date : Year: Month: Day:

این عبارت هم معادل *** است.

پس حکم بخش ب اثبات می‌شود.

پ) می دانیم که حکم این بخش همان حکم بخش ب است که یک Expected روی $p(x)$ از آن حذف شده است. حالا اگر یک SDE فیلتر را در نظر بگیریم، یعنی تمامی transition kernel های $P_{ot}(x'|x)$ فیلتر باشند، قضیه قسمت ب برای تمام توزیع های $p(x)$ برقرار است. حالا با استفاده از اثبات با تناقض بست می‌آید که می‌توانیم $E_{p(x)}$ را از دو طرف رابطه حذف کنیم و به این برسیم:

$$-\log p_\theta^{SDE}(x) \le \mathcal{L}_\theta^{DSM}(x)$$

ابتدا عبارت بخش ب را بازنویسی می‌کنیم:

$$E_{p(x)}\left[-\log p_\theta^{SDE}(x)\right] \le -E_{p_T(x)}\left[\log \pi(x)\right] + \frac{1}{2}\int_0^T E_{P_{ot}(x'|x)p(x)}\Big[$$

$$g(t)^2 \left\| s_\theta(x',t) - \nabla_{x'} \log P_{ot}(x'|x)\right\|_2^2 - g(t)^2 \left\| \nabla_{x'} \log P_{ot}(x'|x)\right\|_2^2 - 2\nabla \cdot f(x',t)\Big]dt$$

عبارت سمت راست را به این صورت می‌نویسیم:

parsian

Subject

Date : Year: Month: Day:

$$\int_0^T \int_{x'} \int_{x} P_{0t}(x'|x) P(x) \left[g(t)^2 \| s_\theta(x',t) - \nabla_{x'} \log P_{0t}(x'|x) \|_2^2 - g(t)^2 \| \nabla_{x'} \log P_t(x') \|_2^2 \right.$$

$$\left. - 2 \nabla \cdot f(x',t) \right] dx\, dx'\, dt = \int_x P(x) \int_0^T \int_{x'} P_{0t}(x'|x) \left[g(t)^2 \| s_\theta(x',t) - \nabla_{x'} \log P_{0t}(x'|x) \|_2^2 \right.$$

$$\left. - g(t)^2 \| \nabla_{x'} \log P_t(x') \|_2^2 - 2 \nabla \cdot f(x',t) \right] dx'\, dt\, dx =$$

$$E_{P(x)} \left[\int_0^T P_{0t}(x'|x) \left[g(t)^2 \| s_\theta(x',t) - \nabla_{x'} \log P_{0t}(x'|x) \|_2^2 - g(t)^2 \| \nabla_{x'} \log P_t(x') \|_2^2 \right. \right.$$

$$\left. \left. - 2 \nabla \cdot f(x',t) \right] dx'\, dt \right] =$$

$$E_{P(x)} \left[\int_0^T E_{P_{0t}(x'|x)} \left[g(t)^2 \| s_\theta(x',t) - \nabla_{x'} \log P_{0t}(x'|x) \|_2^2 - g(t)^2 \| \nabla_{x'} \log P_t(x') \|_2^2 \right. \right.$$

$$\left. \left. - 2 \nabla \cdot f(x',t) \right] dt \right]$$

از طرفی برای $E_{P_T(x)}[\log \pi(x)]$ هم داریم که:

$$E_{P_T(x)}[\log \pi(x)] = \int P_T(x) \log \pi(x)\, dx = \int \log \pi(x) \int P(x') P_{0T}(x|x')\, dx'\, dx$$

$$= \iint P(x')\, P_{0T}(x|x') \log \pi(x)\, dx\, dx' = E_{P(x')} \left[\int P_{0T}(x|x') \log \pi(x)\, dx \right]$$

$$= E_{P(x')} \left[E_{P_{0T}(x|x')} [\log \pi(x)] \right] = E_{P(x)} \left[E_{P_{0T}(x'|x)} [\log \pi(x')] \right]$$

پس داریم که:

Subject

Date : Year: Month: Day:

$$E_{p(x)}\left[-\log p_\theta^{SDE}(x)\right] \le E_{p(x)}\left[-E_{p_T(x'|x)}[\log \pi(x')]\right]$$

$$+ \; E_{p(x)}\left[\frac{1}{2}\int_0^T E_{p_t(x'|x)}\left[g(t)^2 \| s_\theta(x',t) - \nabla_{x'}\log p_t(x'|x)\|_2^2\right] dt\right.$$

$$\left. -\frac{1}{2}\int_0^T E_{p_t(x'|x)}\left[g(t)^2 \| \nabla_{x'}\log p_t(x'|x)\|_2^2 + 2\nabla_{x'}\cdot f(x',t)\right] dt\right]$$

که یعنی داریم:

$$* \quad E_{p(x)}\left[-\log p_\theta^{SDE}(x)\right] \le E_{p(x)}\left[\mathcal{L}_\theta^{DSM}(x)\right]$$

دقت شود که اگر همه را به ازای $E_{p(x)}$ بگیریم رابطه بالا بدست می‌آید.

حالا می دانیم که به ازای هر توزیع $p(x)$ ای این رابطه برقرار است به ازای ساختار فیکس SDE، اثبات با تناقض.

فرض کنیم که نقطه x^* ای داریم که در آن $-\log p_\theta^{SDE}(x^*) > \mathcal{L}_\theta^{DSM}(x^*)$ باشد،

آنگاه خواهیم داشت که: به ازای هر توزیع $p(x)$ رابطه * برقرار است، پس

برای توزیع $p(x)$ به صورتی که $p(x^*)=1$ و در سایر نقاط صفر هم باید این رابطه برقرار باشد که به وضوح برقرار نیست ← همواره باید داشته باشیم که:

$-\log p_\theta^{SDE}(x) \le \mathcal{L}_\theta^{DSM}(x)$ ← بخش ب هم اثبات می‌شود.

parsian

Subject

Date : Year Month Day

۳_

الف) در سری های زمانی ممکن است یه سری از دیتاها پاک شده باشند یا دچار خطا شده باشند، از این رو ما دچار مشکل می شویم در یادگیری آنها، بدین منظور از مدل های CSDI استفاده می کنیم، ایده آن هم این است که به جای تولید داده های جدید (مانند تصاویر)، از این مدل ها برای بازسازی مقادیر گمشده با استفاده از اطلاعات موجود در داده های مشاهده شده بهره گیرند.

روش آموزش:

در این روش، داده های مشاهده شده به دو بخش تقسیم می شوند. همچنین یه بخشی را برای جایگزینی و یه بخشی را برای اطلاعات شرطی در نظر می گیریم. همچنین یه mask تعریف می کنیم که برای دیتا از دست رفته ی x_{ij}، $mask_{ij}=0$ است وگرنه ۱ است. گفتیم که بخشی را به عنوان جایگزینی target : x_0^{tar} ، بخشی به عنوان اطلاعات شرطی : x_0^{co} استفاده می شود.

← در واقع از x_0^{co} به عنوان راهنما استفاده می کنیم، احتمالات به شرط آنرا حساب می کنیم و سعی می کنیم x_0^{tar} را پیش بینی کنیم. در گام t فرایند هم نسخه ای از x_0^{tar} با نویز ترکیب شده است،

$$x_t^{tar} = \sqrt{\alpha_t}\, x_0^{tar} + \sqrt{1-\alpha_t}\, \varepsilon$$

parsian

Subject

Date : Year: Month: Day:

برای آموزش مدل، مدل باید یاد بگیرد نویز X_t^{tar} را با توجه به X_0^{co} پیش‌بینی کند ←

تابع لاسی با: $\mathcal{L}(\theta) = \mathbb{E}\|\varepsilon - \varepsilon_\theta(X_t^{tar}, t \mid X_0^{co})\|^2$ تعریف می‌کنیم و سعی

می‌کنیم آنرا کمینه کنیم.

مراحل معکوس هم به این صورت است که: در زمان استنتاج، مدل از نویز تصادفی شروع

می‌کند و تکرار حذف نویز، مقادیر گمشده را بازسازی می‌کند.

ب) استراتژی‌ها عبارتند از:

① رندوم: درصدی از داده‌های مشاهده شده به صورت تصادفی به‌عنوان هدف انتخاب می‌شوند

مزایایش انعطاف‌پذیری برای نسبت‌های گم‌شده است، معایبش این است که ممکن است

الگوهای ساختاریافته را یاد نگیرد.

② تاریخی: از الگوهای گم‌شده در داده‌های آموزشی استفاده می‌کند. این استراتژی برای داده‌هایی

با الگوهای ساختاریافته (مانند مقادیر گمشده/ متوالی) مناسب است.

مزایا: برای داده‌های تکرارشونده مناسب است. معایب، خطر overfitting ممکن است

اتفاق بیفتد parsian

Subject

Date : Year Month Day

③ ترکیبی : ترکیبی از استراتژی‌های رندوم و تاریخی است تا هم از تعمیرپذیری و هم از الگوهای ساختار یافته بهره ببرد. این استراتژی ترکیبی تعادل مناسبی بین تعمیرپذیری و یادگیری الگوهای ساختار یافته ایجاد می‌کند و احتمالاً بهبود بیشتری در عملکرد نهایی مدل دارد.

④ الگوی آزمون : زمانی استفاده می‌شود که الگوهای کمشده در داده‌های آزمون از قبل مشخص باشند (مثل پیش‌بینی آینده) این استراتژی هم فقط زمانی کاربرد دارد که الگوهای کمشده / آزمون مشخص باشند.

ج) معیار های عبارتنداز:

CRPS (continuous Ranked probability Score):

هدفش ارزیابی دقت پیش‌بینی‌های احتمالاتی است، هرچقدر CRPS کمتر باشد، یعنی عملکرد مدل بهتر است.

CSDI ، در حدود ۴۰-۹۵٪ مقدار CRPS را بهبود داده است نسبت به روش‌های احتمالاتی موجود.

parsian

Subject

Date : Year: Month: Day:

MAE(Mean absolute Error):

هدفش ارزیابی دقت پیش بینی ها / قطعی است.

مکانیزم آن به این صورت است که، میانگین اختلاف قدرمطلق بین مقادیر واقعی

و پیش بینی شده را محاسبه می کند.

state_of_art deterministic method نسبت به CSDI مقدار MAE را بین ۵-۲۰ ٪

کاهش داده است.

با پیاده سازی

parsian