Paper summary

AI VISION Lab

1. 공부한 논문의 제목, 게재된 학회 혹은 저널 등 논문 기본 정보를 적으세요.
   A. 이름: A Connection Between Score Matching and Denoising Autoencoders
   B. 저널: Neural Computation
   C. 도메인: Score matching, DAE(Denoising Autoencoder)
   D. 출판연도: 2010
   E. 저자: Pascal Vincent

2. 논문에서 제안한 알고리즘 및 프레임워크에 대해 본인이 이해한대로 다이어그램을 그려보세요. **논문 Figure를 그대로 따라 그리면 안됩니다.**
   A. [선행연구] Score Function의 역할: gradient of log density에 대해서 model($p(x;\theta)$ 과 data($q(x)$)의 차이를 줄이도록 설계됨. 이때 가장 중요한 부분은, $p(x;\theta)$를 표현하는데 $z(x)$를 구하는 것이 불가능하다는 점인데, 이는 gradient + $\delta\theta$처리를 해주는 과정에서 삭제된다는 장점이 있음.

   B. [선행연구] 이러한 개념을 가지고, 이전 연구에서는, ISM(Implicit Score Matching) 방식과, ESM(Explicit Score Matching) Objective Function을 제안함. 이는 Denoising 관점에서 일반적인 X: input 만 처리할 수 있다는 한계를 가짐. 또한, 기존 연구에서 계속해서 X: input이 아닌, $\hat{x}$: input으로 사용할 때 성능이 높게 나온다는 점에서, X에 Noise를 추가한 $\hat{x}$에도 적용가능한 Function의 필요성이 높아짐.

   C. 이 논문에서는 Denoising Autoencoder에서 X에 Noise를 추가하는 점에서 아이디어를 가져와서 목적함수를 설계하는데 이용함

   D. 사실, 구조상 다른 점은 $q(x)$가 아닌, $q_\sigma(\hat{x}|x)$에 대해서 처리해주고 있다는 점인데, σ가 여기서 Gaussian을 따른다고 가정하고 전개하고 있기 때문에 상대적으로 쉽게 처리할 수 있음(기존의 DAE에서는 Noise를 X에 Add해주는 형식이 아닌, Random하게 선택된 x에 대해서 0으로 바꿔주는 형태였음)

   E. 뿐만 아니라, Energy Function을 적용한 $p(\hat{x};\theta)$에 대해서도 수식을 증명함에 따라, Energy Based Model 과 Score Function을 결합한 형태의 목적함수를

제안함

F. 최종적으로 본 논문에서는 $ESM_{q_\sigma} \approx ISM_{q_\sigma} \approx DSM_{q_\sigma} \approx DAE_{q_\sigma}$ 임을 입증함. (노트 내용 참고)

3. 본인이 생각하는 이 논문의 장점이 무엇이라고 생각하나요? **논문 Contribution bullet을 그대로 따라 적으면 안됩니다.**
   A. 이 논문이 등장하기 전, 존재하던 SOTA에 대해서 (Kingma and LeCun et al.2010) 단점을 파악하고, 주장하는 근거의 장점을 부각하는 점이 인상깊었음. (Kingma and LeCun 논문의 한계: 미분 2번, Sample(X)에 대해서만 적용 가능. 본 논문이 가지는 우수함: 미분 한 번으로 처리 가능, General한 Objective Function ($\sigma = 0, X. \sigma > 0, \hat{X}.$ *즉* $\sigma = 0$ *은 특수한 상황*)

4. 이 논문을 읽으면서 느낀 점, 혹은 배운 점이 있으면 적어보세요.
   A. Score matching에 대한 이해도 향상, 수식 증명과정.
   B. SOTA에 해당하는 연구와, 이 논문에서 증명하고자 했던 목표가 달랐으나, 결과적으로 같은 이야기를 하고 있다는 점에서 흥미로웠음.

5. 이 논문의 한계점이 있다면 무엇이라고 생각하나요?
   A. Gaussian kernel이 없다면 -> Conditional Distribution을 미분했을 때 처리하는데 상당히 까다로울 수 있음. ($q(x, \hat{x}) = q(\hat{x}|x) * q(x)$, 해당 논문에서는 $N(0, I\sigma^2)$를 따르는 식에 대해서 처리함. 따라서, 미분 결과가 명확했다는 장점이 있음)

6. 본인의 연구에 접목시켜볼 점이 있을지 생각하고 적어보세요.
   A. Diffusion 선행연구

7. 본 Summary를 작성하는 과정에서 생성형AI를 사용했나요?
   A. 아니요

날짜: 2025-06-26

이름: 신준원

2025. 06. 26 목요일 A Connection Between Score Matching & DAE.

Energy based Model (EBM)

softmax → image 사용↓

probabilistic x → Energy based.

목표 → observed ⇒ 낮은 E (확률 ↑)
→ wrong ⇒ 높은 E (확률 ↓)

그렇다면, Energy based 가 왜 필요한가?

$\boxed{\text{Energy} = \text{존재하지 않을 가능성}}$

기존의 문제 확률모델 (Generative model).

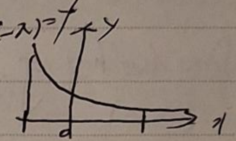
$$p = \frac{1}{Z} \cdot q \Rightarrow$$

Z = 구하는 것이 불가능

$$q = \exp(\boxed{-E_\theta(x)})$$

→ 새로 정의.

$E_\theta(x)$ 가 작을수록 가능성↑



Boltzman Distribution (= Gibbs Distribution)

$$p_i \propto \exp\left(-\frac{\varepsilon_i}{kT}\right)$$

→ $\boxed{Z \text{ 우회}}$ → $\boxed{\text{Contrasive Divergence}}$

Introduction

- Score matching의 발전.
  - ① Continuous 뿐만 아니라, discrete 한 분포에서도 적용가능.
- Denoising Autoencoder의 발전 (with SM - Score Matching)

like DAE.

추가설명 → input x (vanilla) 의 성능보다 input (x+noise) 성능 ↑.

→ RBM + SM ( ) ⇒ AE 성능과 동일. (SM 사용X)

SM.

저자 ⇒ $\boxed{\text{DAE 로 SM을 사용한다면?}}$

결합

성능향상.

notation

discrete distribution → Dirac - deltas δ.

$J_1 \sim J_2$, J_1, J_2 서로 동일.

$\langle u, v \rangle = \sum_i u_i v_i$ (dot product)

DAE + Score Matching

DAE ⇒ Vanilla version ⇒ Data Blank 처리 (Random)

↳ 본 연구에서는, $X + \varepsilon \sim N(0, I\cdot\sigma)$ (noise Additional)

Score Matching = true distribution 과 일치(유사) 하도록 하는 θ 찾기.

Energy Based 형태. $p(x;\theta) = \frac{1}{Z(\theta)} \exp(-\underline{E}(x;\theta))$

energy function

Score Matching (ESM) $J(\theta) = E_{q(x)}\left[\frac{1}{2}\left\| \psi(x;\theta) - \frac{\partial \log q(x)}{\partial (x)} \right\|^2\right]$

$\psi(x;\theta) = \frac{\partial \log p(x;\theta)}{\partial x}$ ~~true : q~~ ~~model : p~~

data

infinite sample 사용

Score Matching (ISM) $J(\theta) = \underline{E_{q(x)}}\left[\frac{1}{2}\|\psi(x;\theta)\|^2 + \sum_{i=1}^{d} \frac{\partial \psi_i(x;\theta)}{\partial x_i}\right] + C.$

$\boxed{q \approx p}$ 따라서 implicit $\boxed{J_{ESM,q} \smile J_{ISM,q}}$

이 논문에서는 finite 개수의 Data를 다루고 있으므로 변형이 필요함.

Score Matching (ISM q^0) $J(\theta) = \frac{1}{n}\sum_{t=1}^{n}\left(\frac{1}{2}\|\psi(x;\theta)\|^2 + \sum_{i=1}^{d} \boxed{\frac{\partial \psi_i(x^{(t)};\theta)}{\partial x_i}}\right)$ (Score) $\frac{d}{dx}$ 미분 2번 단점

극한 → 이론적 ← 즉. 극한 취하면, $\boxed{\text{ISM } J \smile \lim_{n\to\infty} \text{ISM } q^0}$

결과적으로. (이론적인 ISM q^0 가 아닌, $\boxed{\text{ISM reg}(\theta) \text{ 사용}}$. $\boxed{\text{Kingma \& LeCun (2010)}}$

$J_{ISM\,reg(\theta)} = \boxed{J_{ISM\,q_0}(\theta) + \lambda \sum_{i=1}^{d}\left(\frac{\partial \psi_i(x^{(t)};\theta)}{\partial x_i}\right)^2}$

Connecting DAE + SM

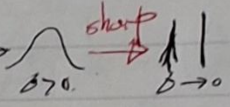
Parzen window density estimator (with UM)

(non-parametric)

observed Data sample을 기반으로 $\boxed{p(x)}$ 추정방법임.

$$p(x) = \frac{1}{n}\sum_{i=1}^{n} k_\sigma(x - x_i) \qquad (k_\sigma = \text{kernel 함수})$$

→ Score function 적용. objective $J_{ESMq_\sigma}(\theta) = E\left[\frac{1}{2}\left\|\psi(\tilde{x};\theta) - \frac{\partial \log q_\sigma(\tilde{x})}{\partial \tilde{x}}\right\|^2\right]$

($\sigma > 0$, 미분가능 $\sigma \to 0$, 재산분포 = 분포분포 0) → 

noise.

Denoising Score Matching

model. target → DAE 기반 $x \to \tilde{x}$

$$J_{DSMq_\sigma}(\theta) = E_{q_\sigma(x,\tilde{x})}\left[\frac{1}{2}\left\|\psi(\tilde{x};\theta) - \frac{\partial \log q_\sigma(\tilde{x}|x)}{\partial \tilde{x}}\right\|^2\right]$$

$x \to \tilde{x}$ 가 아닌 $\boxed{\tilde{x}}$ 로

→ Gaussian kernel ~~Parzen~~

$$\frac{\partial \log q_\sigma(\tilde{x}|x)}{\partial \tilde{x}} = \boxed{\frac{1}{\sigma^2}(x - \tilde{x})}$$

~~σ>0 설정~~ ~~Parzen 설정 → 사용가능~~

$\boxed{J_{ESMq_\sigma} \sim J_{DSMq_\sigma}}$

noise 주입 방법가 Gaussian 따라서 가능

J_{ESM}의 q_σ $\quad x \leftrightarrow \tilde{x}$ J_{ESMq_σ}

$\boxed{J_{ESM} = J_{DSM} \text{ 증명}}$

$$J_{ESM q_\sigma} = E_{q_\sigma(\tilde{x})}\left[\frac{1}{2}\left\|\psi(\tilde{x};\theta) - \frac{\partial \log q_\sigma(\tilde{x})}{\partial \tilde{x}}\right\|^2\right]$$ 전개시 (제곱)

$$= E_{q_\sigma(\tilde{x})}\left[\frac{1}{2}\|\psi(\tilde{x};\theta)\|^2\right] - S(\theta) + C_2$$ (notation 간소화)

이때 $C_2 = E_{q(\tilde{x})}\left[\frac{1}{2}\left\|\frac{\partial \log q_\sigma(\tilde{x})}{\partial \tilde{x}}\right\|^2\right] \rightarrow \theta$와 무관한 상수

$$S(\theta) = \frac{1}{2}\cdot 2 \cdot E_{q_\sigma(\tilde{x})}\left[\psi(\tilde{x};\theta)\cdot\frac{\partial \log q_\sigma(\tilde{x})}{\partial(\tilde{x})}\right]$$ $E_{q(x)} f(x) = \int q(x) f(x)\, dx$ 이므로

$$= \int q_\sigma(\tilde{x})\left[\psi(\tilde{x};\theta)\cdot\frac{\partial \log q_\sigma(\tilde{x})}{\partial(\tilde{x})}\right] d\tilde{x},$$ $\frac{d}{d\tilde{x}}\log q_\sigma(\tilde{x}) = \frac{\frac{\partial}{\partial \tilde{x}} q_\sigma(\tilde{x})}{q_\sigma(\tilde{x})}$

$$= \int\left[\psi(\tilde{x};\theta)\frac{\frac{\partial}{\partial\tilde{x}} q_\sigma(\tilde{x})}{q_\sigma(\tilde{x})}\cdot q_\sigma(\tilde{x})\right] d\tilde{x}$$ 이때 & $q_\sigma(\tilde{x}) = \int q_0(x)\cdot q_\sigma(\tilde{x}|x)\, dx$

$\rightarrow q(\tilde{x}|x) \rightarrow q(\tilde{x},x)$ marginal distribution

$$= \int \psi(\tilde{x};\theta)\frac{\partial}{\partial\tilde{x}}\left(\int q_0(x)\cdot q_\sigma(\tilde{x}|x)\, dx\right) d\tilde{x}$$

$\rightarrow \frac{q_\sigma(\tilde{x}|x)\partial}{\partial\tilde{x}}$ | $\Rightarrow \frac{\partial \log q_\sigma(\tilde{x}|x)}{\partial x} = \frac{1}{q_\sigma(\tilde{x}|x)}\cdot\frac{\partial\cdot q_\sigma(\tilde{x}|x)}{\partial x}$

$\Rightarrow \frac{q_\sigma(\tilde{x}|x)\partial}{\partial\tilde{x}} = \boxed{q_\sigma(\tilde{x}|x)\cdot\frac{\partial \log q_\sigma(\tilde{x}|x)}{\partial x}}$ $\quad \psi(\tilde{x};\theta) \Rightarrow$ 상수 $\Rightarrow$ 이

$$= \int_{\tilde{x}}\int_x q_\sigma(\tilde{x},x)\left(\psi(\tilde{x};\theta), \frac{\partial \log q_\sigma(\tilde{x}|x)}{\partial\tilde{x}}\right) dx\, d\tilde{x}$$

$$= E_{q_\sigma(\tilde{x},x)}\left(\psi(\tilde{x};\theta), \frac{\partial \log q_\sigma(\tilde{x}|x)}{\partial\tilde{x}}\right) = S(\theta)$$ 해당 값을 J식에 넣으면

$$J_{ESM q_\sigma} = E_{q_\sigma(\tilde{x})}\left[\frac{1}{2}\|\psi(\tilde{x};\theta)\|^2\right] - E_{q_\sigma(\tilde{x},x)}\left(\psi(\tilde{x};\theta), \frac{\partial \log q_\sigma(\tilde{x}|x)}{\partial\tilde{x}}\right) + C_2.$$

J_{DSM}도 동일하게 전개하면 (상수항 C_3) 수식. $\boxed{J_{ESM} = J_{DSM} + C_2 - C_3}$

(X)

→ Energy Based로 푼 처리, SM으로 처리 둘의 관계파악하기

Energy Based, Denoising Autoencoder

$\theta = W, b, c$

b → encoder의 bias

c → Decoder의 bias

→ $p(x;\theta) = \frac{1}{Z(\theta)} \cdot \exp(-E(x;\theta))$

$$E(\underbrace{x; W, b, c}_{\theta}) = -\frac{\langle c, x\rangle - \frac{1}{2}\|x\|^2 + \sum_{j=1}^{d_h} \mathrm{softplus}(\langle W_j, x\rangle + b)}{\sigma^2}$$

softplus $= \log(1+e^x)$

적용

$$\psi_i(x;\theta) = \frac{\partial E}{\partial x_i} = \frac{\partial \log p(x;\theta)}{\partial x_i}$$

c = linear attraction

Energy 함수 = 확률높이기 항

→ $\frac{d}{dx} s.p = \underbrace{\frac{e^x}{1+e^x}}_{\text{sigmoid}}$

→ $\log(1+e^{f(x)}) \Rightarrow \frac{e^{f(x)}}{1+e^{f(x)}} \cdot f'(x)$

모든 항에 대해서 x_i 미분시

$$\frac{1}{\sigma^2}\left(c_i - \frac{1}{2} \cdot 2 \cdot x_i + \sum_{j=1}^{d_h} \mathrm{sigmoid}(\langle W_j, x\rangle + b) W_{ji}\right)$$

$\boxed{\frac{d}{dx}\langle c, x\rangle = c \rightarrow \langle W_j, x\rangle + b \text{ 로 동일}}$

→ $\frac{1}{\sigma^2}(c_i - x_i + \langle W_{\cdot i}, \mathrm{sigmoid}(Wx+b)\rangle) = \psi_i(x;\theta)$

해당 $\psi_i(x;\theta)$ 를 J_{DSM}에 대입. 이때, E, $\frac{\partial \log q_\sigma(\tilde{x}|x)}{\partial \tilde{x}} = \frac{1}{\sigma^2}(x-\tilde{x})$

$$J_{DSM} = E_{q_\sigma(\tilde{x},x)}\left[\frac{1}{2}\left\|\frac{1}{\sigma^2}(W^T \mathrm{sigmoid}(W\tilde{x}+b) + c - \tilde{x}) - \frac{1}{\sigma^2}(x-\tilde{x})\right\|^2\right]$$

temperature (= noise controller)

$$= E_{q_\sigma(\tilde{x},x)}\left[\boxed{\frac{1}{2}\frac{1}{\sigma^4}}\left\|W^T \mathrm{sigmoid}(W\tilde{x}+b) + c - \cancel{\tilde{x}} - x + \cancel{\tilde{x}}\right\|^2\right]$$

$J_{DAE} = \frac{d}{dx} DAE$ — input 형태

$\boxed{J_{DSM} \simeq J_{DAE}}$

Conclusion

$\delta > 0$ 에 대해서.

$$\boxed{J_{IGM\,36} \smile J_{EGM\,36} \smile J_{DGM\,36} \smile J_{DAZ\,36}}$$

① equivalent 증

② 기본 관련 항 x J 제안 → 한번.

③ Energy Based 제안 (Energy J.).