Paper summary

AI VISION Lab

1. 공부한 논문의 제목, 게재된 학회 혹은 저널 등 논문 기본 정보를 적으세요.
   A. 이름: Extracting and Composing Robust Features with Denoising Autoencoders
   B. 저널: ICML
   C. 도메인: Denoising Autoencoder
   D. 출판 연도: 2008
   E. 저자: Pascal Vincent et al.

2. 논문에서 제안한 알고리즘 및 프레임워크에 대해 본인이 이해한대로 다이어그램을 그려보세요. **논문 Figure를 그대로 따라 그리면 안됩니다.**

   A. [선행연구] Deep Architecture를 구성할 때, 성능이 떨어지는 문제가 발생함 (원인: Gradient Vanishing/Exploding)

   B. [선행연구] 이를 해결하는 좋은 방법은 Unsupervised 방식으로 layer wise하게 initialization하는 것. (e.g., X – H 가 있다면, X의 다음 Layer인 H만을 가지고 X를 초기화) -> 결과적으로 학습이 불안정한 문제가 해결되는 결과를 가져왔음.

   C. 해당 과정에서 논문의 저자는, 중간 표현 (intermediate Representation)의 존재에 대한 고민을 함 (?)

   D. 대체 강건한 중간 표현이란 무엇이며, 중간 표현을 평가하는 기준으로는 어떤 점이 있을까?

   E. 저자는 중간표현의 기준을 제시하고
      i. Input에 대한 정보 유지
      ii. X를 그대로(identity)하게 받아오는 것이 아닌, 제약(Constraint)이 존재
      iii. 희소표현 (0 or 1)

   F. 강건한 중간표현을 구하기 위한 실험을 설계함 = DAE(Denoising AE)

G. X: Input, $f_{\theta}(X) = Y = W * X + b$, $g_{\theta'}(Y) = Z = W^T * Y + b'$. 이때, g의W^T는 f의$W\ Transpose(tied\ weight)$ 가 기본적인 AE의 구조라면, DAE는 X를 그대로 Input으로 이용하는 것이 아닌, $\hat{X}$를 Input으로 이용함 (이때, $\hat{X}$는 X에 noise를 더해주는 구조가 아닌, random하게 선택한 X의 element를 0으로 변경해주는 식으로 처리함)

H. DAE model의 확률적 표현방식은 다음과 같음

I. $q^0(X, \hat{X}, Y)$이고 이는 $q^0(x) * q_D(\hat{X}|X) * q^0(Y|\hat{X})$로 전개됨. 이때, $q^0(Y|\hat{X}) = \delta_{f_{\theta(\hat{x})}}(Y)$로 수정가능. ($\delta$의 경우, $Y = \hat{X}\ Return\ 1\ else\ 0$)
(Y = label, 결정적인 존재이므로 수정가능)

J. 이처럼, 표현할 수 있다는 점에서 Robust하다고 볼 수 있으며, 효과적인 denoising 과정을 보여줬다는 점에서 의의가 있음.

K. 추가적으로 Manifold Learning과의 상관관계를 확인했을 때, $\hat{X}$는 noisy하므로, Manifold 밖, 혹은 매우 낮은 확률을 가진 값이라고 볼 수 있음.
따라서, $q_D(\hat{X}|X)$를 처리해주는 과정은 Manifold Learning과 유사하다고 볼 수 있음.

3. 본인이 생각하는 이 논문의 장점이 무엇이라고 생각하나요? **논문 Contribution bullet을 그대로 따라 적으면 안됩니다.**
   A. Robust함을 확인하려고 시작해서, AE의 한계였던 X의 중간표현으로의 전달을 Robust하게 표현하는데 성공했다는 것 또한 우수하다고 할 수 있으나, 개인적으로는 Denoising, Manifold로의 확장까지 기여했다는 점이 이 논문의 가장 큰 장점이라고 생각함

4. 이 논문을 읽으면서 느낀 점, 혹은 배운 점이 있으면 적어보세요.
   A. 새로운 가설에 대해서 다양한 이론에 적용하는 과정. 또한, DAE에 대해서 다루다가 갑자기 등장하는 Manifold에서 Diffusion뿐만 아니라, 넓은 범위에서의 학습의 필요성에 대해서 다시한번 배울 수 있었음.

5. 이 논문의 한계점이 있다면 무엇이라고 생각하나요?
   A. 수식을 간단하게 표현할 수 있었던 이유이자, q^0에 대해서 전개할 때, z를 condition으로 사용하지 않을 수 있었던 이유는 tied weight기 때문임. 즉

이는 y -> z과정을 deterministic(+ 간결함)하게 처리할 수 있도록 만들었으나, Generative model의 설계 측면에서 생각해보자면, Probability하지 않기 때문에 다양성 측면에서 한계를 가지고 있음이 분명함 (이후 나온 VAE에서는 Z를 Condition으로 사용해서 한계를 해결)

6. 본인의 연구에 접목시켜볼 점이 있을지 생각하고 적어보세요.
   A. Diffusion 선행 논문

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

날짜: 2025-06-25

이름: 신준원

2025.6.25 수요일 ~~A connection Between Score matching & Denoising Autoencoder~~

Denoising Autoencoder 소개

Previous work 에서는. (class : Y , data : X)

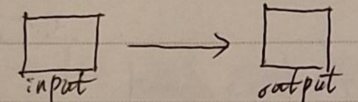
차이를 줄이는!

Generative (ex. GPT) → 전체파악을 목적으로함. → p_data → p_model

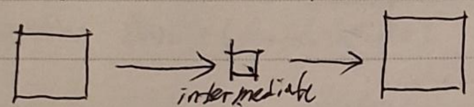
Discriminative (ex. SVM) → p(Y/X)

→ 초기 unsupervised 관계에서, input 을 중간표현(intermediate representation)

즉,

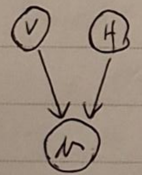


가 아닌



중간 단계의 존재 = 학습의 난이도 ↓

Deep Architecture 의 어려움.

① 

→ V, n 의 relation 을 정확히 파악 가능하다면

H, n 의 관계를 모르더라도 알아낼수있음 (explaining away)

그런데, 해당 graphical model은 깊게 쌓게한다면 난이도 ↑

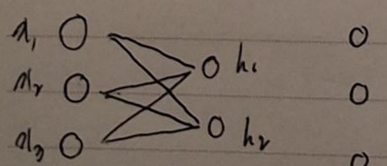
② hidden layer 1~2개 를 넘어서는경우 → optimization의 어려움

unsupervised pretrain

①, ② 문제점 → 중간표현 (차원축소)의 등장으로 해결

ex). training belief Net , stacked Autoencoder

+ local optimization (fully-connected 라고 가정)



input layer 특성치 → hidden layer.

end-to-end ⇒ gradient vanishing 존재

→ local optimization 존재.

x_1, x_2, x_3 → [h_1, h_2] $\boxed{x}$ 를 기반으로!

→ local! → fine-tune 진행 (최종)

unsupervised

with label

poor solution 방지

무엇이 해당결과를 좋게 만드는 것인가?

Unsupervised 가 좋은 결과 (중간 표현)를 만들어내지만 $\boxed{\text{"어떻게?"}}$

↳ RBM, Autoencoder 의 학습 알고리즘에서 동작 중

이 논문에서는 $\boxed{\text{좋은 중간 표현}}$에 대한 명시적인 기준을 이야기함.

→ 최소한의 정보 보유 → $\boxed{\text{정보 유지}}$ ①

→ $\boxed{\text{제약}}$ ② (상수 vector 같은...)

) ①,② 고려 → 좋은 중간표현

→ ③ $\boxed{\text{희소 표현}}$ (sparsity representation) – 추가 기준

⇒ 중간표현이 - input에 대한 정보를 가진상태라면. with AE.

→ input + noise(조금) 에 대한 정보를 입력시 비슷한 결과가 나올것이다.

$\boxed{\text{Algorithm (DAE)}}$

notation

Expectation: $E_{p(x)}[f(x)] = \int p(x)f(x)dx$

Entropy : $H(X) = H(p) = E_{p(x)}[-\log p(X)]$

Conditional Entropy : $H(X|Y) = E_{p(X,Y)}[-\log p(X|Y)]$

KL - divergence : $D_{KL}(p\|q) = E_{p(x)}\left[\log \frac{p(x)}{q(x)}\right]$

Cross entropy : $H(p\|q) = E_{p(x)}[-\log q(x)] = H(p) + D_{KL}(p\|q)$

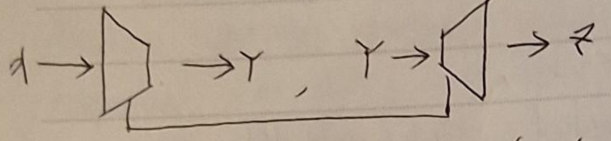
Mutual information : $I(X;Y) = H(X) - H(X|Y)$

Sigmoid : $s(x) = \frac{1}{1+e^{-x}}$, $(\Rightarrow s(X) = (s(x_1), \ldots s(x_n))^T$

Bernoulli distribution : μ : $B_\mu(x) = (B_{\mu_1}(x_1), \ldots B_{\mu_d}(x_d))$

~~① 독립, $p(x,t) = p(x)\cdot p(t)$ 성립~~

Basic Autoencoder



[가중치 공유
tied weight

$\underline{W^T = W'}$

S = sigmoid.

$y = \underbrace{s(Wx + b)}_{f_\theta}$, $z = \underbrace{s(W'y + b')}_{g_{\theta'}}$,

$\theta = [W, b]$ $\quad$ $\theta' = [W', b']$

$$\text{optimizer} = \arg\min_{\theta, \theta'} \frac{1}{n} \sum_{i=1}^{n} \boxed{L(x^{(i)}, z^{(i)})}$$

$z^{(i)} \rightarrow g_{\theta'}(f_\theta(x))$

$$\text{Loss} = L(x^i, z^i) = \| x - z \|^2$$

(general 한 상황) . ┌ 연속분포

$$\text{Loss}_H = -\sum_{k=1}^{d} [x_k \log z_k + (1 - x_k) \log(1 - z_k)]$$

(베르누이 , (0,1)) ┌ 이산분포 .

$\boxed{\text{identity mapping}}$ → 한계점.

X 그 자체를 받아와 $\boxed{\text{전달}}$ 하는 문제가 발생하기도 함.

→ Robust 하지 않음 . → $\tilde{X}$ 를 (X를 망가뜨려 복원을 강제함)
(DAE)

비율 v → 각 x에 대해서 random하게 0으로 처리
나머지는 그대로 유지

Denoising Autoencoder

x 입력의 AE 와 달리 $\tilde{x}$ ~ $q_D(\tilde{x}|x)$ 이용
→ 부분 noisy

이후 x의 예측값 = 빈칸 (0으로 전환된 부분) 채우기
↳ input $\tilde{x}$ / $Loss_H(x, z)$

DAE: $X \longrightarrow \tilde{X} \xrightarrow{f} Y \xrightarrow{g} z$
AE: $X \xrightarrow{f} Y \xrightarrow{g} z$
] $\boxed{\text{차이점}}$

Additional notation. $q^0(X)$ → 모든 t (test_target) input에 대한 분포
$\delta_{f(\theta)(\tilde{x})}(Y)$ → deterministic한 경우 → 1 or 0 return
↳ [$f(\theta) \neq Y$ return 0
else return 1

Corruption process $q_D(\tilde{X}|X)$

$$q^0(X, \tilde{X}, Y) = q^0(Y, \tilde{X}, X) = q^0(Y|\tilde{X}, X) \cdot q^0(\tilde{X}|X) \cdot q^0(X)$$

① $q^0(Y|\tilde{X}, X) = q^0(Y|\tilde{X})$ 임. X는 영향X (관계X)
② $q^0(\tilde{X}|X) = q_D(\tilde{X}|X)$
→ $q^0(Y|\tilde{X})$ 에 대해서, $\delta_{f(\theta)(\tilde{x})}(Y)$ 로 표현가능

결과적으로 $q^0(X, \tilde{X}, Y) = \underline{q^0(X)} \cdot \underline{q_D(\tilde{X}|X)} \cdot \underline{\delta_{f_\theta(\tilde{x})}(Y)}$ 로 표현가능
input → $\tilde{X}$ → Y

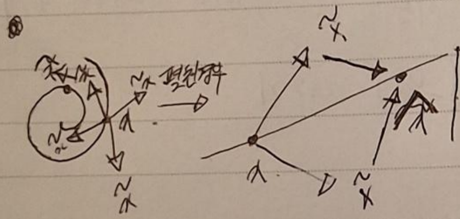
→ $\underline{z}$ explicit 하게 나와있지 않음. = 한계점 ✓

→ Generative modeling 측면에서.
($z \to Y \to \bar{x} \to x$ 의 과정 필요. 그러나 현재 과정은 Y 부터.
즉, $z \to Y$ 관계 X.) → Deterministic하므로 ~~설명X~~ 불가능
→ 즉, 다양성 X
→ 다양성 부여 필요
→ VAE의 등장
+S probability

(5)

Manifold Learning Perspective

(이 부분에서는 blank)



x(uncorrupted)에 corrupt 추가 = $\tilde{x}$

$\tilde{x}$는 존재자체가 매우 낮은확률
(manifold 밖에 있다)

즉, $p(x|\tilde{x})$를 구하는 과정은 여러 step이 필요.

⇒ manifold learning 과정과 유사함. (복원)

Top-down Generative model Perspective

$X \to \bar{X} \to Y$ 에 대해서 [illegible]함수.

$Y \to \bar{X} \to X$: p explicit 하므로 문제없이 입증가능함