MORE ABOUT AUTO-ENCODER

Hung-yi Lee 李宏毅
Auto-encoder

As close as possible

Embedding, Latent Representation, Latent Code

• More than minimizing reconstruction error
• More interpretable embedding
What is good embedding?

• An embedding should represent the object.
**Beyond Reconstruction**

How to evaluate an encoder?

Loss of the classification task is $L_D$

Train $\phi$ to minimize $L_D$

$\phi^* = \min \phi L_D$

Small $L_D^*$ The embeddings are representative.

Say “Yes”

Say “No”
**Beyond Reconstruction**

How to evaluate an encoder?

Loss of the classification task is $L_D$

Train $\phi$ to minimize $L_D$

$$L_D^* = \min_{\phi} L_D$$

Small $L_D^*$ ➞ The embeddings are representative.

Large $L_D^*$ ➞ Not representative

---

Say "Yes"

Say "No"
Beyond Reconstruction

How to evaluate an encoder?

loss of the classification task is $L_D$

Train $\phi$ to minimize $L_D$

\[ L^*_D = \min_{\phi} L_D \]

Small $L^*_D \implies$ The embeddings are representative.

Large $L^*_D \implies$ Not representative

Train $\theta$ to minimize $L^*_D$

\[ \theta^* = \arg \min_{\theta} L^*_D \]

\[ = \arg \min_{\theta} \min_{\phi} L_D \]

Train the encoder $\theta$ and discriminator $\phi$ to minimize $L_D$

Deep InfoMax (DIM)

(c.f. training encoder and decoder to minimize reconstruction error)
**Typical auto-encoder is a special case**

As close as possible

![Diagram of auto-encoder process](image)

Typical auto-encoder is a special case where the encoder and decoder are used to compress and reconstruct the input data. The encoder takes the input and compresses it into a lower-dimensional vector, which is then passed to the decoder to reconstruct the original data. The reconstruction error is calculated and used by the discriminator to improve the performance of the auto-encoder.

Discriminator

score (reconstruction error)
Sequential Data

**Skip thought**

A document is a sequence of sentences.


**Quick thought**

Sequential Data

• Contrastive Predictive Coding (CPC)

Auto-encoder

As close as possible

- More than minimizing reconstruction error
- More interpretable embedding
Feature Disentangle

- An object contains multiple aspect information

**Input Audio**
- **Encoder**
- **Decoder**
- Input audio
- Reconstructed
  - Include phonetic information, speaker information, etc.

**Input Sentence**
- **Encoder**
- **Decoder**
- Input sentence
- Reconstructed
  - Include syntactic information, semantic information, etc.
Feature Disentangle

input audio → Encoder → Decoder → reconstructed

Encoder 1 → Decoder → reconstructed

Encoder 2 → Decoder → reconstructed

phonetic information

speaker information
Feature Disentangle
- Voice Conversion

How are you?
Encoder

How are you?
Decoder

Hello
Encoder

Hello
Decoder
Feature Disentangle - Voice Conversion

Encoder

Decoder

How are you?

Hello

How are you?

Hello
Feature Disentangle - Voice Conversion

- The same sentence has different impact when it is said by different people.

Do you want to study a PhD?

Go away!

Do you want to study a PhD?

Student

Student

新垣結衣（Aragaki Yui）
Feature Disentangle - Adversarial Training

Speaker classifier and encoder are learned iteratively

Encoder

Speaker Classifier (Discriminator)

Decoder

Learn to fool the speaker classifier

How are you?

How are you?
Feature Disentangle - Designed Network Architecture

- Designed Network Architecture

How are you?

Encoder 1

Encoder 2

IN

Decoder

How are you?

= instance normalization (remove global information)
Feature Disentangle
- Designed Network Architecture

Encoder 1

Encoder 2

IN

Decoder

AdaIN

AdaIN

IN

How are you?

IN = instance normalization  (remove global information)

AdaIN = adaptive instance normalization

(only influence global information)
Thanks Ju-chieh Chou for providing the results.
https://jjery2243542.github.io/voice_conversion_demo/
Discrete Representation

- Easier to interpret or clustering

One-hot

Non differentiable

Discrete Representation

• Vector Quantized Variational Auto-encoder (VQVAE)

For speech, the codebook represents phonetic information

https://arxiv.org/abs/1711.00937


Sequence as Embedding

https://arxiv.org/abs/1810.02851

This is a **seq2seq2seq auto-encoder**.
Using a sequence of words as latent representation.

Only need a lot of documents to train the model.

![Diagram of seq2seq auto-encoder](https://arxiv.org/abs/1810.02851)
Sequence as Embedding

Human written summaries

Let Discriminator considers my output as real

Real or not

Discriminator

document

word sequence

Readable

Summary?

Seq2seq

document
Sequence as Embedding

- **Document**: 澳大利亚今天与13个国家签署了反兴奋剂双边协议，旨在加强体育竞赛之外的药品检查并共享研究成果 ......

- **Summary**:
  - **Human**: 澳大利亚与13国签署反兴奋剂协议
  - **Unsupervised**: 澳大利亚加强体育竞赛之外的药品检查

- **Document**: 中华民国奥林匹克委员会今天接到一九九二年冬季奥运会邀请函，由于主席张丰绪目前正在中南美洲进行友好访问，因此尚未决定是否派队赴赛 ......

- **Summary**:
  - **Human**: 一九九二年冬季奥运会函邀我参加
  - **Unsupervised**: 奥委會接獲冬季奧運會邀請函
Sequence as Embedding

• **Document**: 據此間媒體27日報道，印度尼西亞蘇門答臘島的兩個省近日來連降暴雨，洪水泛濫導致塌方，到26日為止至少已有60人喪生，100多人失蹤 ......

• **Summary**:
  • Human: 印尼水災造成60人死亡
  • Unsupervised: 印尼門洪水泛濫導致塌雨

• **Document**: 安徽省合肥市最近為領導幹部下基層做了新規定: 一律輕車簡從，不準搞迎來送往、不準搞層層陪同 ......

• **Summary**:
  • Human: 合肥規定領導幹部下基層活動從簡
  • Unsupervised: 合肥領導幹部下基層做搞迎來送往規定: 一律簡
Tree as Embedding

https://arxiv.org/abs/1806.07832

Concluding Remarks

- More than minimizing reconstruction error
  - Using Discriminator
  - Sequential Data
- More interpretable embedding
  - Feature Disentangle
  - Discrete and Structured

As close as possible