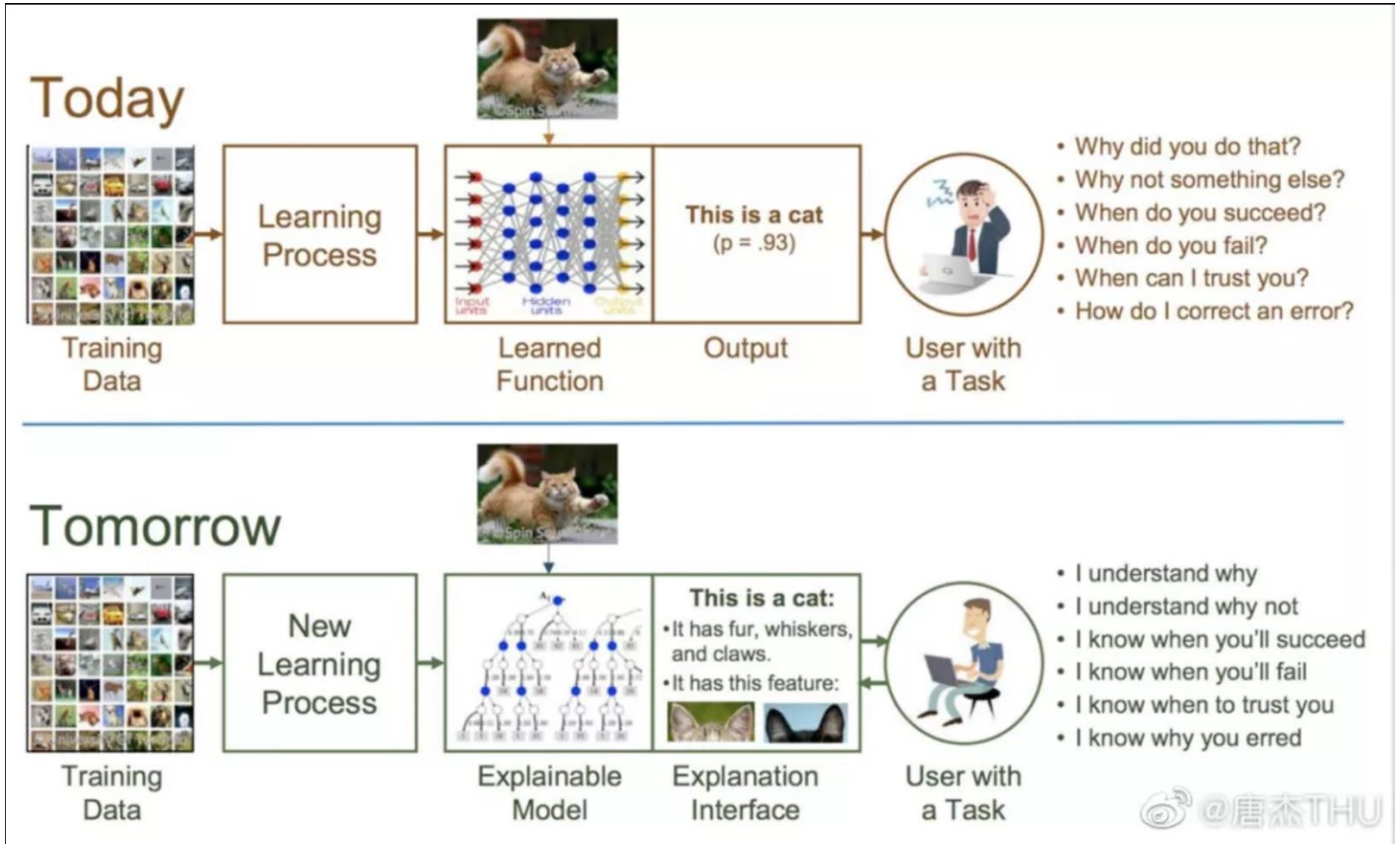
Disentangled Representation Learning

Statistical Learning

Causal Learning



Disentangled Representation Learning

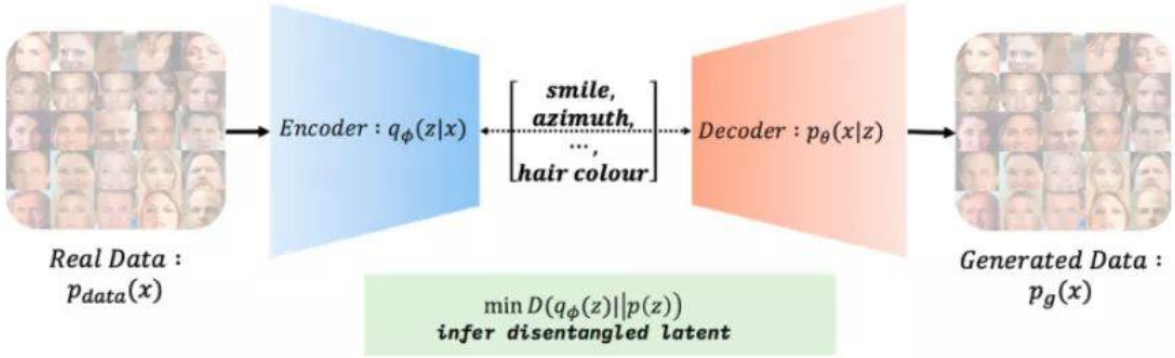
- A disentangled representation should separate the distinct, informative factors of variations in the data.
- Single latent units are sensitive to changes in single generative factors, while being relatively invariant to changes in other factors (Y. Bengio et al., 2013).
- A change in a single underlying factor of variation z_i should lead to a change in a single factor in the learned representation $r(x)$ (Locatello et al., 2019).

Disentangled Representation Learning

Disentangled Representation Learning

Representation Learning	从高维数据中抽取有用的表示，使得习得的表示不仅起到降维而且可以对下游的任务有所帮助
Disentangled Representation	从数据中学习到的表示能够自行分割为几块可解释的独立部分

将原始数据空间中纠缠着的数据变化，变换到一个好的表征空间中，在这个空间中，不同要素的变化是可以彼此分离的

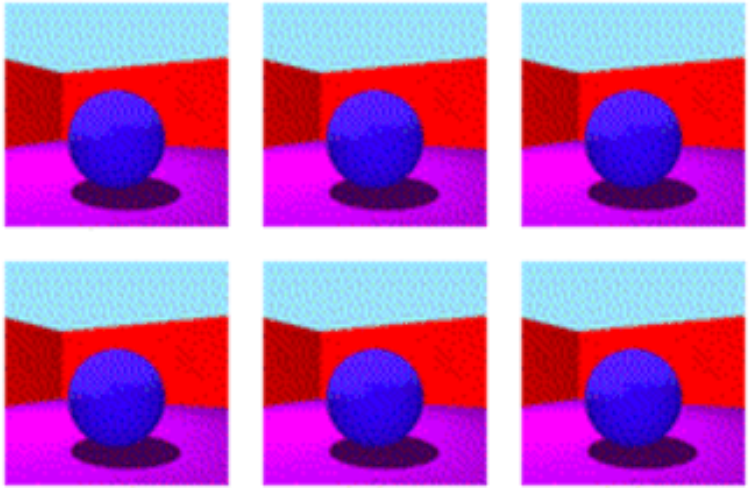


Disentangled Representation Learning

It aims to learn **factorized** representations that uncover and disentangle the latent causal factors hidden in the observed data (Bengio et al., 2013).

Existing work mostly focuses on image data, while we focus on the user behavior data collected in a recommender system.

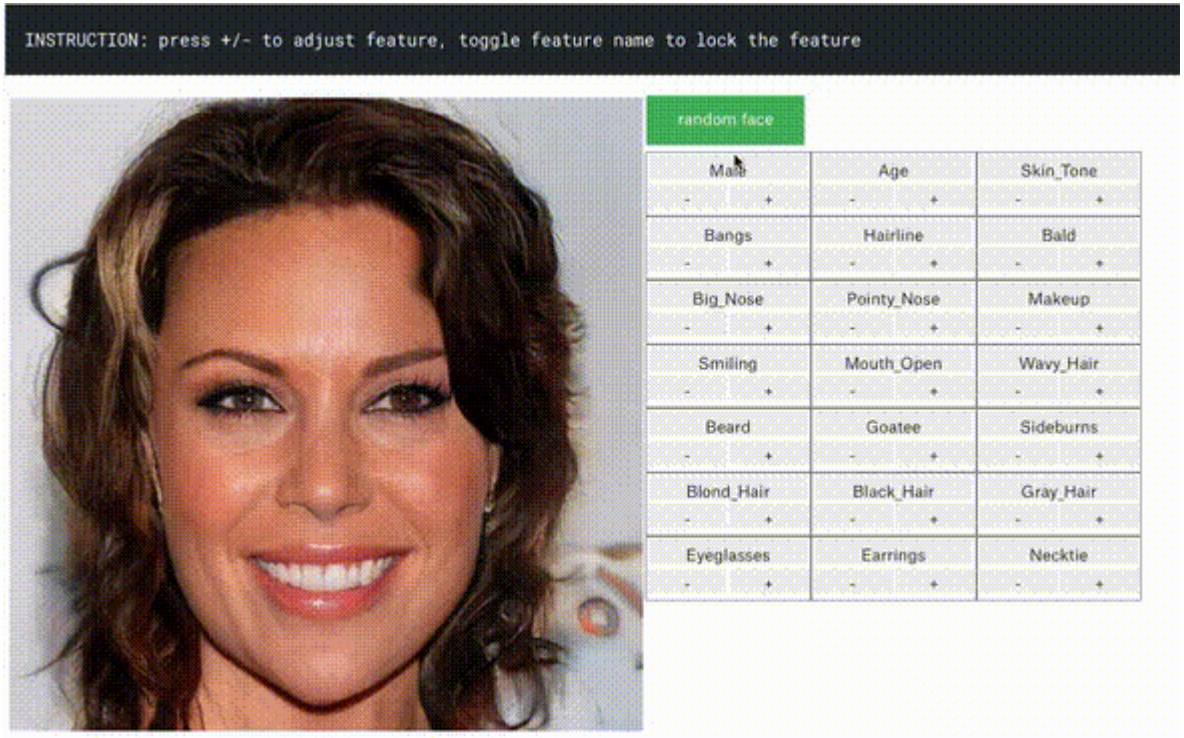
Unsupervised learning of disentangled representations 认为：现实世界的数据是由一些可解释的独立因子不同组合产生的可以通过unsupervised learning的方式找到这些独立因子。例如在shapes3D数据集中，每个图像有6个独立因子控制，分别是物体形状、物体大小、相机角度、地板颜色、墙壁颜色、物体颜色



disentanglement_lib

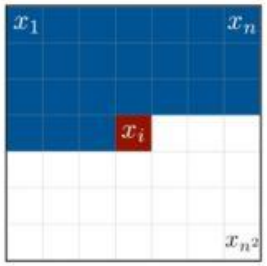
Transparent Latent-space GAN

使用 TL-GAN 模型进行受控人脸图像合成的示例



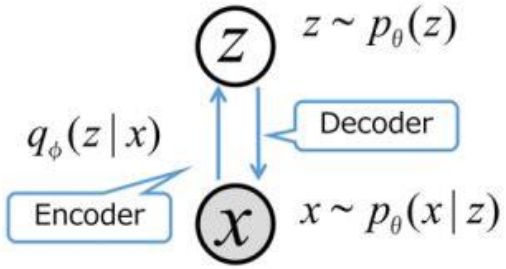
TL-GAN

Autoregressive Models

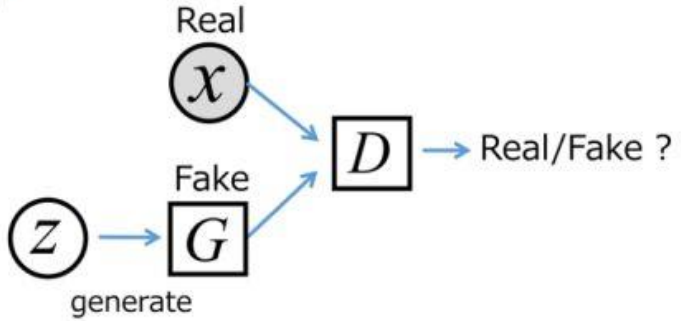


$$p(\mathbf{x}) = \prod_{i=1}^{n^2} p(x_i|x_1, ..., x_{i-1})$$

Variational AutoEncoders (VAE)



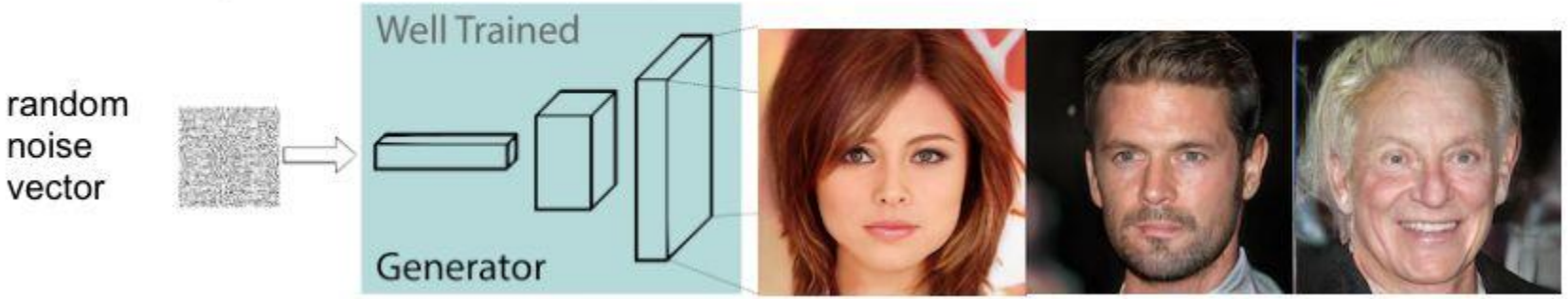
Generative Adversarial Networks (GAN)



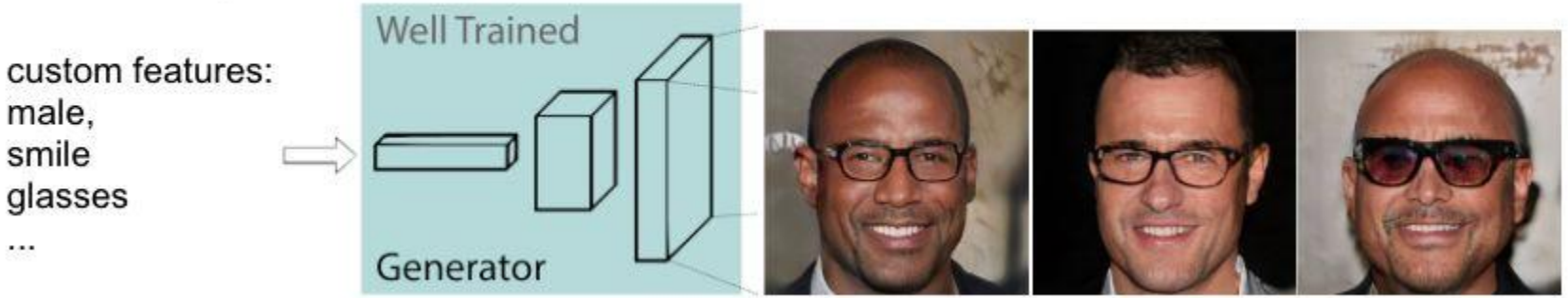
$$\min_G \max_D V(D,G)$$
$$= \mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))].$$

控制GAN模型的输出

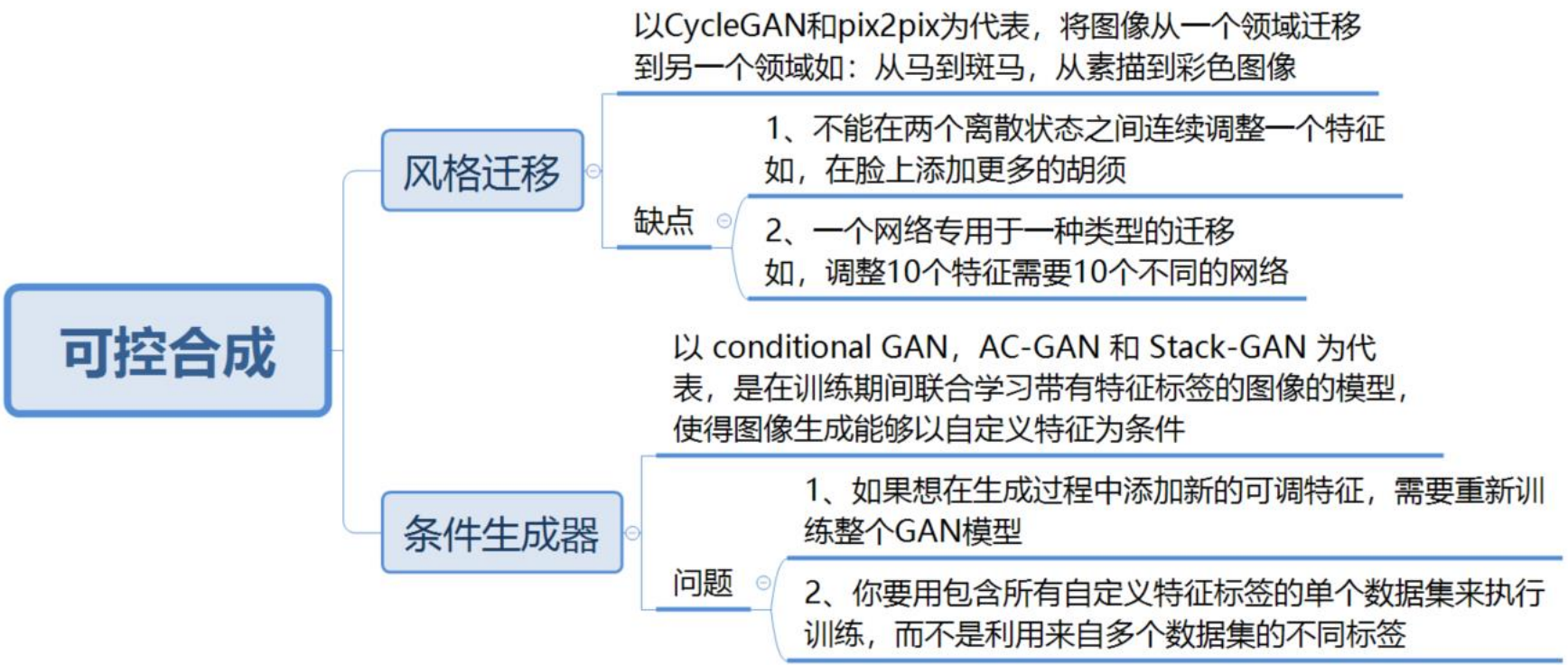
Random generation of high quality images



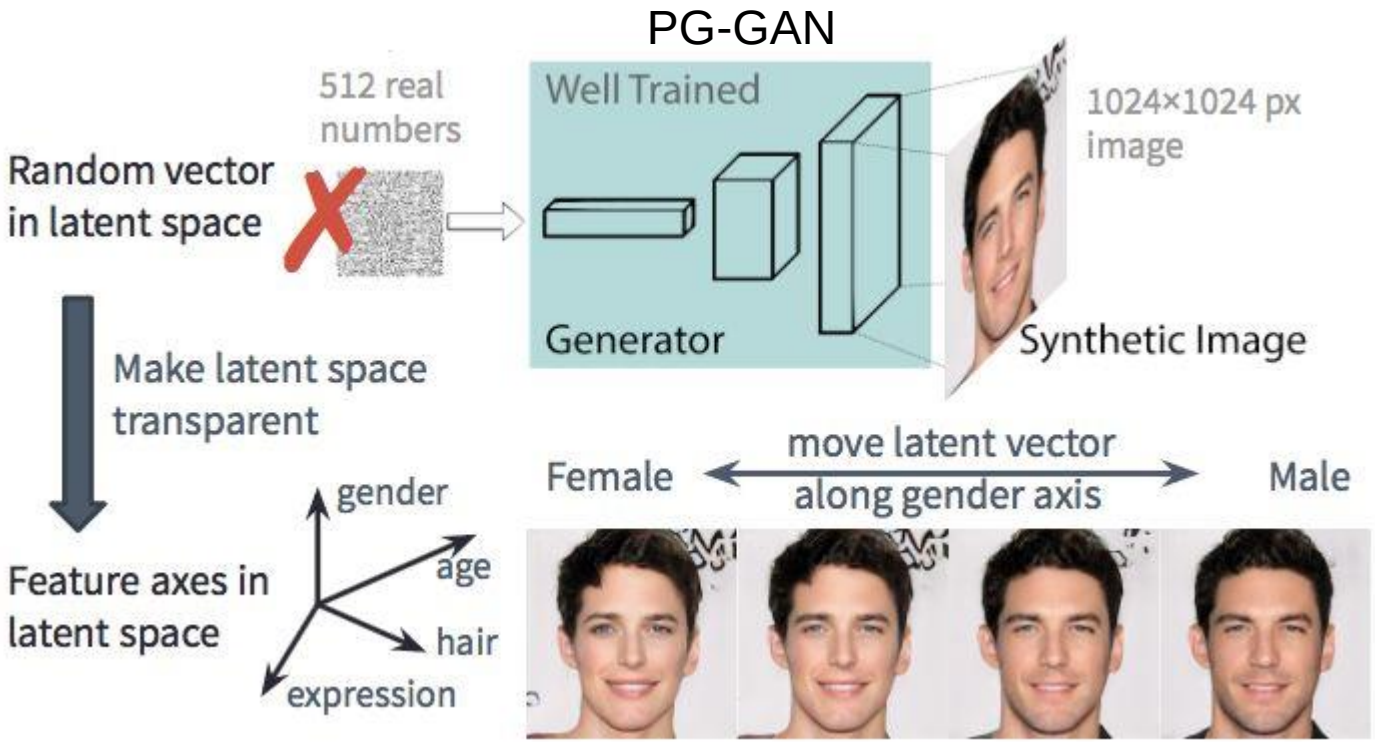
Controlled image generation according to custom features



TL-GAN


可控合成
风格迁移
以CycleGAN和pix2pix为代表，将图像从一个领域迁移到另一个领域如：从马到斑马，从素描到彩色图像
缺点
1、不能在两个离散状态之间连续调整一个特征
如，在脸上添加更多的胡须
2、一个网络专用于一种类型的迁移
如，调整10个特征需要10个不同的网络
条件生成器
以 conditional GAN，AC-GAN 和 Stack-GAN 为代表，是在训练期间联合学习带有特征标签的图像的模型，使得图像生成能够以自定义特征为条件
问题
1、如果想在生成过程中添加新的可调特征，需要重新训练整个GAN模型
2、你要用包含所有自定义特征标签的单个数据集来执行训练，而不是利用来自多个数据集的不同标签

TL-GAN


PG-GAN
512 real numbers
Random vector in latent space
Well Trained
Generator
1024×1024 px image
Synthetic Image
Make latent space transparent
Feature axes in latent space
gender
age
hair
expression
move latent vector along gender axis
Female
Male

Axis: smile

Non smile

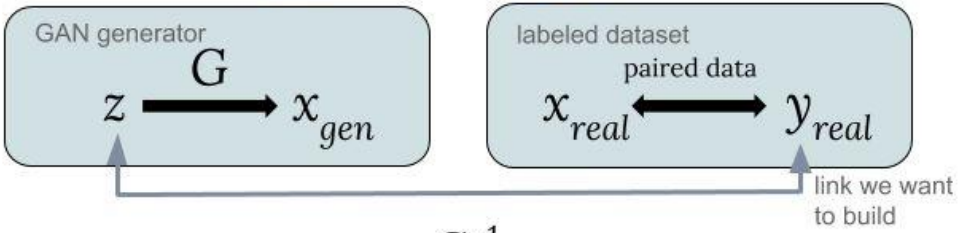
Smile

TL-GAN

连接潜在向量 z 和特征标签 y 的方法

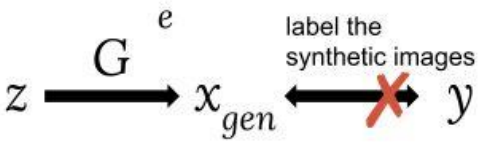
male
not young
non-smile
...

z: latent vector
x: image
y: feature label



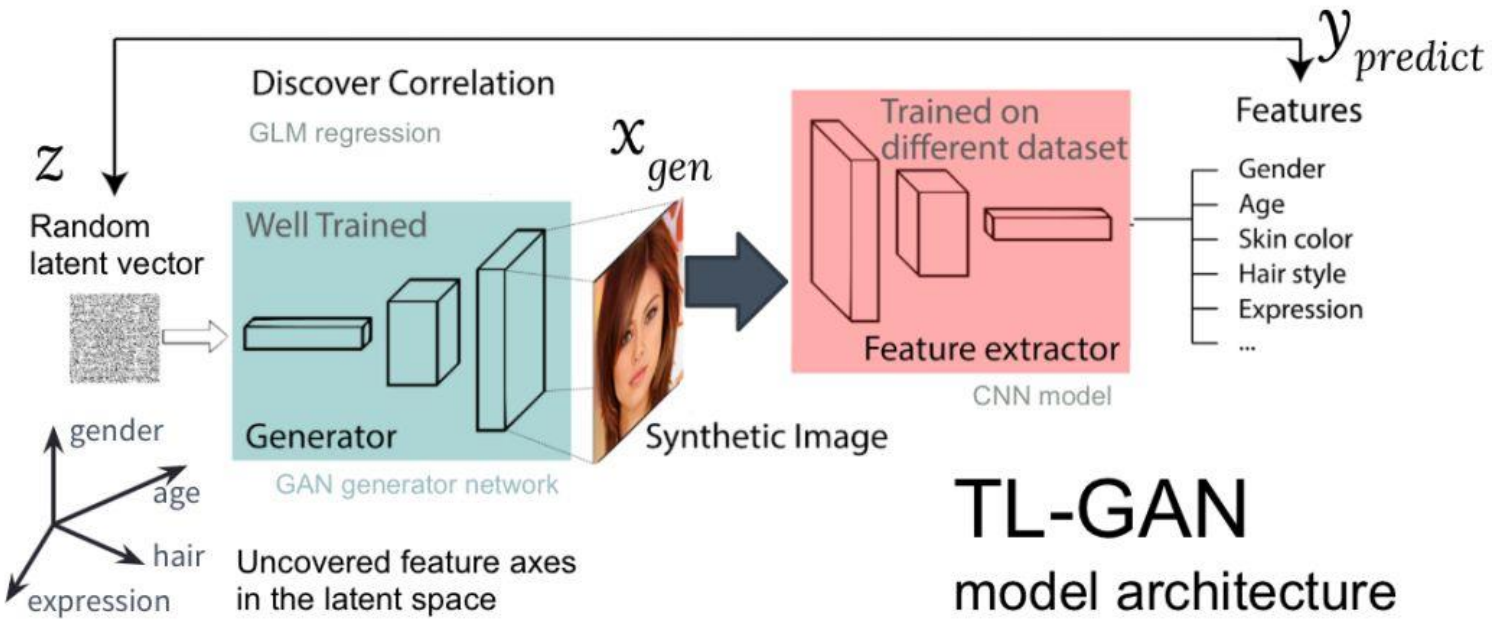
Potential approach 1:
Computing the latent vector for images in the labeled dataset

Potential approach 2:
label the features of synthetic images manually



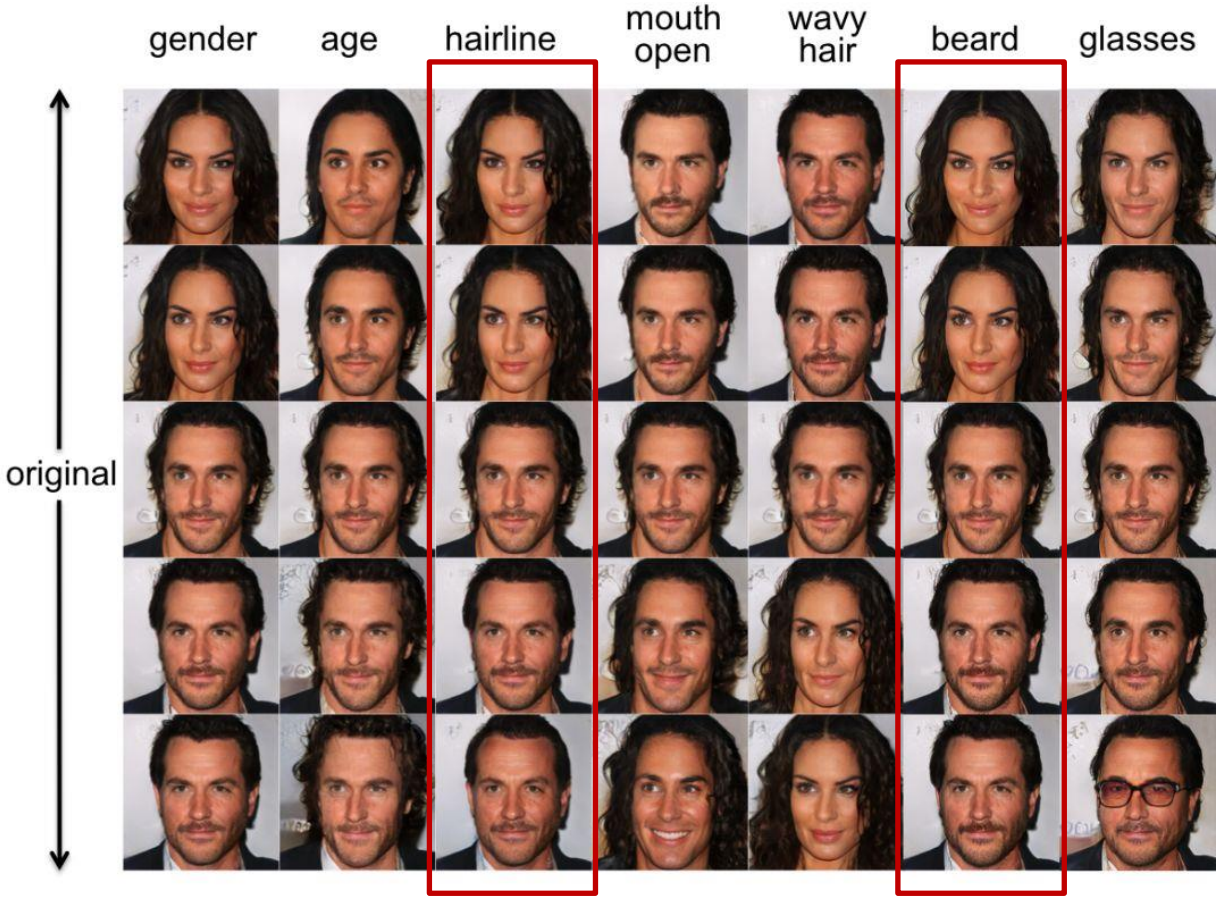
Approach of TL-GAN:
Use a separately trained feature extractor network F to produce feature labels

F is trained on labelled data (x_{real}, y_{real})



- TL-GAN
  - 1、学习分布：选择一个训练好的GAN模型作为生成网络 — PG-GAN
  - 2、分类：选择一个预训练特征提取器模型(CNN)
  - 3、生成：生成大量Z向量，传输到训练好的GAN中合成图像，使用特征提取器为每张图像生成特征
  - 4、关联：使用广义线性模型（GLM）执行Z向量和特征之间的回归任务
  - 5、探索：从一个Z向量开始，沿着一或多个特征轴移动，并检测对生成图像的影像

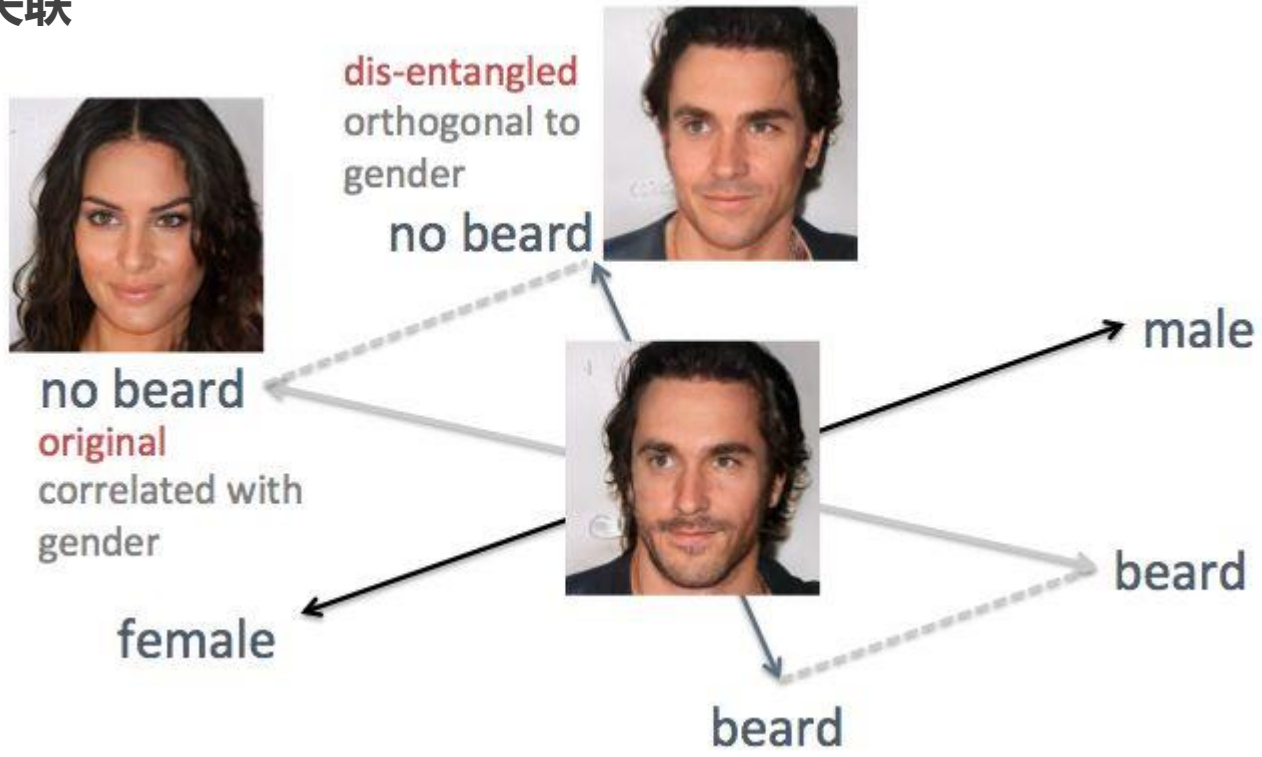
TL-GAN


gender
age
hairline
mouth open
wavy hair
beard
glasses
original

TL-GAN

解除相关特征轴之间的关联

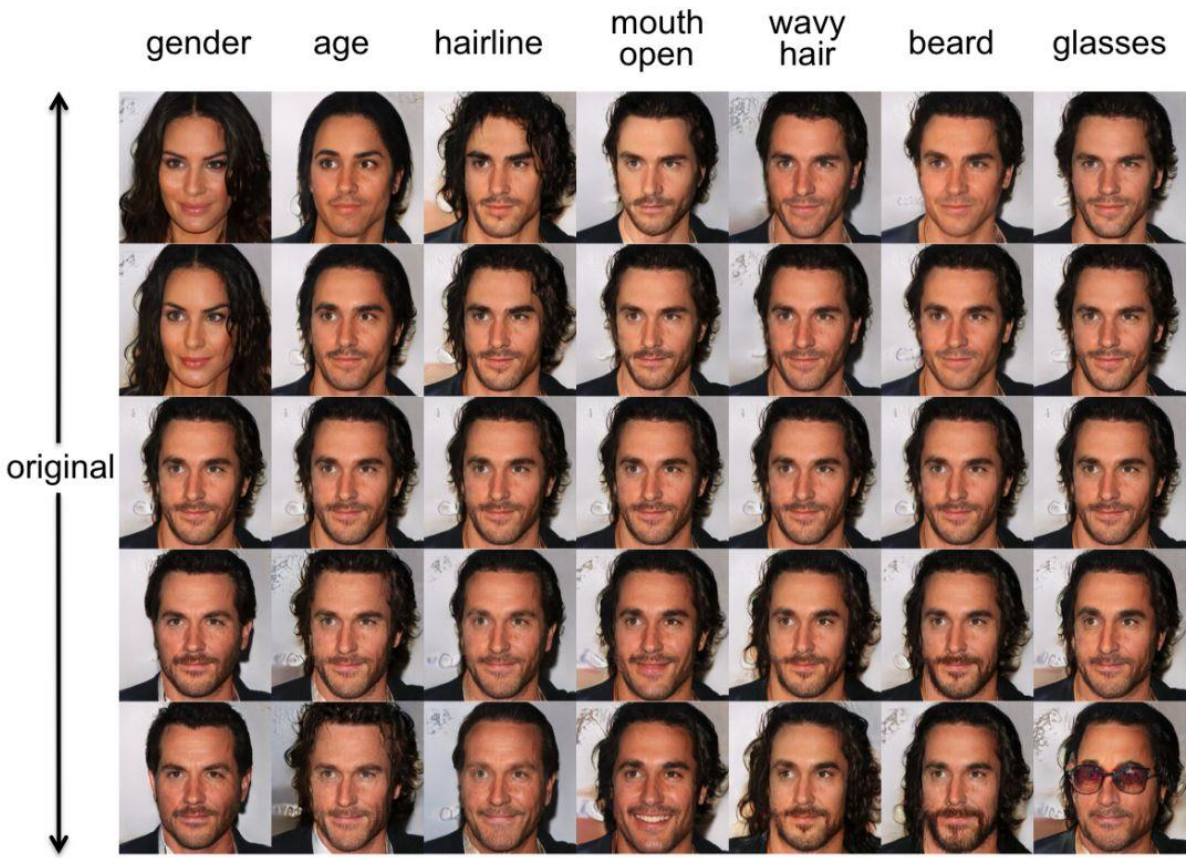
使用线性代数技巧解除相关特征轴之间的关联



TL-GAN

After disentangle

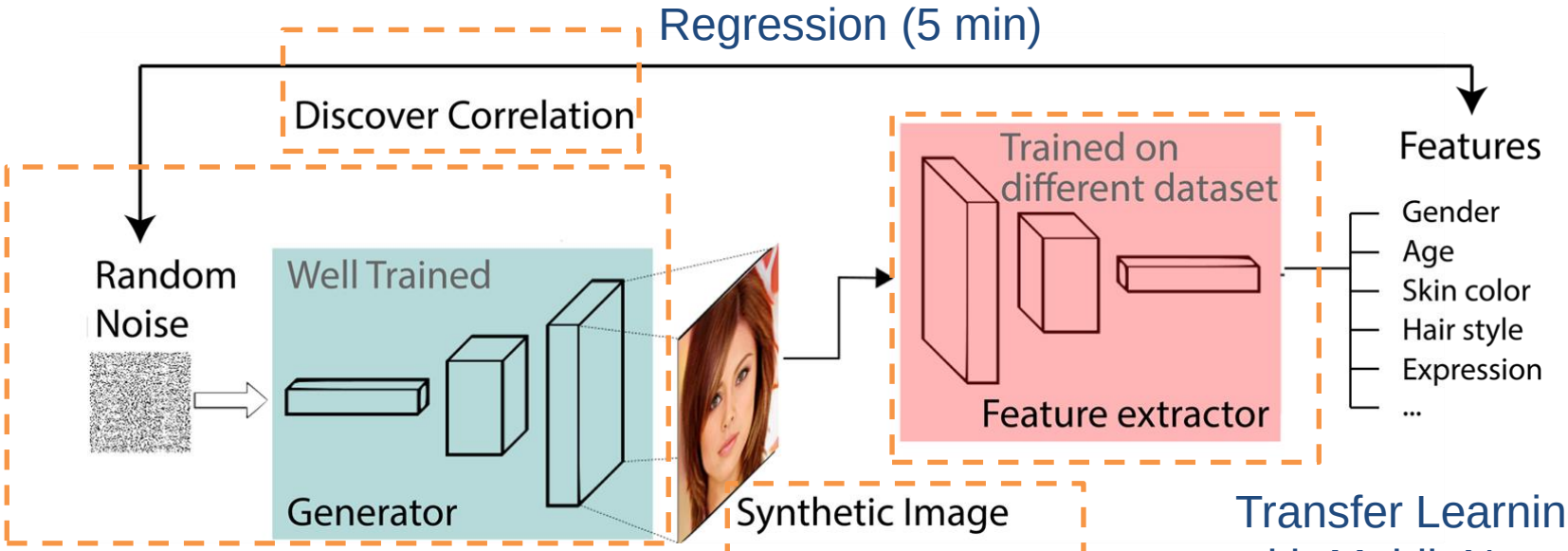
(make all other features orthogonal to gender and age)



TL-GAN

西安交通大学 XI'AN JIAOTONG UNIVERSITY

Simple and efficient workflow, no need to retrain GAN



Well-trained GAN,
(2 weeks Nvidia's time)

Pre-generate 20000 images (8h)

Transfer Learning with MobileNet on low-resolution labeled images (1h)

Recommendation

2019NIPS

NeurIPS 2019

Learning Disentangled Representations for Recommendation

Jianxin Ma[1]*, Chang Zhou[2]*, Peng Cui[1], Hongxia Yang[2], Wenwu Zhu[1]
[1]Tsinghua University, [2]Alibaba Group

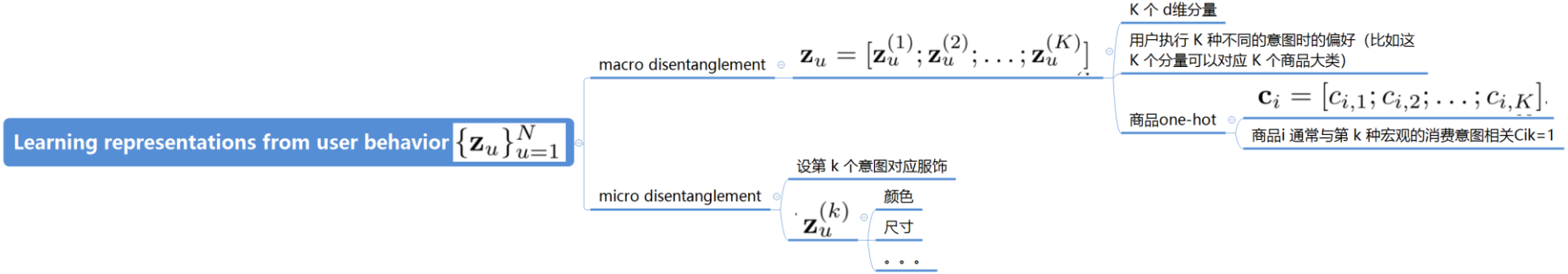
认知因素，并不是商品固有的细粒度的属性、品类，而是一种从人的角度理解商品的可传播可解释的概念

如何挖掘潜在的认知概念，并以合理的方式，将潜在可接受的认知概念传递给用户

思路

- 1、把所有商品分为K个大类
  - 电子产品
  - 家居
  - 服饰
  - 食品
  - 图书
- 2、构建一个user-user的Auto Encoder
  - 输入历史序列（商品、点击、加购或购买）
- 3、把user的历史序列里的商品按K个大类进行分类，分别通过各自的auto encoder（共K个）
  - 得到K个user embedding
  - $\mathbf{z}_u = [\mathbf{z}_u^{(1)}; \mathbf{z}_u^{(2)}; \ldots; \mathbf{z}_u^{(K)}]$
- 4、使用的beta-VAE

Recommendation



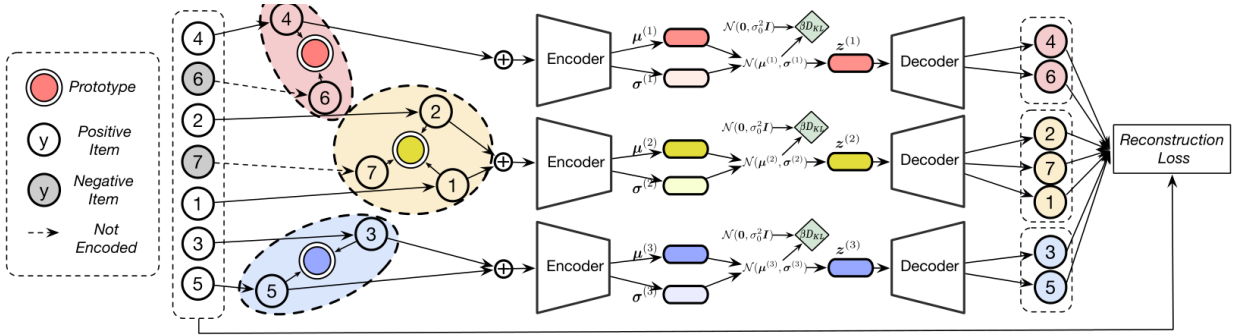


Figure 1: Our framework. Macro disentanglement is achieved by learning a set of prototypes, based on which the user intention related with each item is inferred, and then capturing the preference of a user about the different intentions separately. Micro disentanglement is achieved by magnifying the KL divergence, from which a term that penalizes total correlation can be separated, with a factor of β.

Recommendation

Table 1: Collaborative filtering. All methods are constrained to have around $2Md$ parameters, where M is the number of items and d is the dimension of each item representation. We set $d = 100$.

Dataset	Method	Metrics		
		NDCG@100	Recall@20	Recall@50
AliShop-7C	MultDAE	0.23923 (±0.00380)	0.15242 (±0.00305)	0.24892 (±0.00391)
	β-MultVAE	0.23875 (±0.00379)	0.15040 (±0.00302)	0.24589 (±0.00387)
	Ours	**0.29148** (±0.00380)	**0.18616** (±0.00317)	**0.30256** (±0.00397)
ML-100k	MultDAE	0.24487 (±0.02738)	0.23794 (±0.03605)	0.32279 (±0.04070)
	β-MultVAE	0.27484 (±0.02883)	0.24838 (±0.03294)	0.35270 (±0.03927)
	Ours	**0.28895** (±0.02739)	**0.30951** (±0.03808)	**0.41309** (±0.04503)
ML-1M	MultDAE	0.40453 (±0.00799)	0.34382 (±0.00961)	0.46781 (±0.01032)
	β-MultVAE	0.40555 (±0.00809)	0.33960 (±0.00919)	0.45825 (±0.01039)
	Ours	**0.42740** (±0.00789)	**0.36046** (±0.00947)	**0.49039** (±0.01029)
ML-20M	MultDAE	0.41900 (±0.00209)	0.39169 (±0.00271)	**0.53054** (±0.00285)
	β-MultVAE	0.41113 (±0.00212)	0.38263 (±0.00273)	0.51975 (±0.00289)
	Ours	**0.42496** (±0.00212)	**0.39649** (±0.00271)	0.52901 (±0.00284)
Netflix	MultDAE	0.37450 (±0.00095)	0.33982 (±0.00123)	0.43247 (±0.00126)
	β-MultVAE	0.36291 (±0.00094)	0.32792 (±0.00122)	0.41960 (±0.00125)
	Ours	**0.37987** (±0.00096)	**0.34587** (±0.00124)	**0.43478** (±0.00125)

1从用户行为学习解离化表征 — 学习到一定可解释性的解离化的用户商品表征

2用户可控制的交互式推荐 — 解离化表征带来了一定的可控制性。比如说，表征维度关联的是不同的商品属性，把用户的表征向量提供给用户，允许用户自行固定绝大部分维度（比如对应的是衣服的风格、价格、尺寸等）、然后单独调整某个维度的取值（比如颜色对应的维度），系统再根据这个反馈调整推荐结果。这将帮助用户更加精准地表达自己想要的、并检索得到自己想要的。

在无监督的情况下，训练出可解释的模型仍然是需要运气的，在加了 beta-VAE 的约束后，获得可解释模型的概率相比普通 VAE 大大提高，但仍然避免不了「反复训练多个模型，然后挑出最好的模型」这一陷阱

问题

并不是所有的维度都有人类可以理解的语义 — 建议未来的研究者们多多关注（弱/半）监督方法，引入标签信息

更多用户体验方面的问题被摆在了决策者的眼前，比如为什么买了又推，为什么都是点过的商品，如何创造真正增量的价值

Challenging

西安交通大学
XI'AN JIAOTONG UNIVERSITY

ICML2019 best p

Challenging Common Assumptions in the Unsupervised Learning of Disentangled Representations

Francesco Locatello [1 2] Stefan Bauer [2] Mario Lucic [3] Gunnar Rätsch [1] Sylvain Gelly [3] Bernhard Schölkopf [2]
Olivier Bachem [3]

1、理论证明，无先验知识的无监督解耦表示学习是不可能的

2、模型的超参数和初始随机种子对最后的disentangle效果影响极大，想要好的解耦效果，调参是关键

3、对下游任务的正面影响微乎其微

4、当前各种用于检验无监督解耦表示学习效果的metrics互相之间不一致

Challenging

1.解耦表示学习需要引入先验知识或弱监督学习来指引建模解耦表示。

2.需要进一步探究究竟解耦表示有没有对下游任务的帮助或者其他有用的地方。

3.未来的解耦的工作需要在各种数据集上都跑一下效果，不然可能一个好，另一个不好。实验的初始设置问题也需要进一步讨论。