

条件概率图产生式对抗网络*

李崇轩, 朱军, 张钊



(清华大学 计算机科学与技术系, 北京 100084)

通讯作者: 李崇轩, E-mail: chongxuanli@mail.tsinghua.edu.cn; 朱军, E-mail: dcszj@mail.tsinghua.edu.cn

摘要: 产生式对抗网络(generative adversarial networks, 简称 GANs)可以生成逼真的图像, 因此最近被广泛研究. 值得注意的是, 概率图生成对抗网络(graphical-GAN)将贝叶斯网络引入产生式对抗网络框架, 以无监督的方式学习到数据的隐藏结构. 提出了条件概率图生成对抗网络(conditional graphical-GAN), 它可以在弱监督环境下, 利用粗粒度监督信息来学习到更精细而复杂的结构. 条件概率图生成对抗网络的推理和学习遵循与 graphical-GAN 类似的方法. 提出了条件概率图生成对抗网络的两个实例. 条件高斯混合模型(conditional Gaussian mixture GAN, 简称 cGMGAN)可以在给出粗粒度标签的情况下从混合数据中学习细粒度聚类. 条件状态空间模型(conditional state space GAN, 简称 cSSGAN)可以在给定对象标签的情况下学习具有多个对象的视频的动态过程.

关键词: 深度生成模型; 产生式对抗网络; 概率图模型; 弱监督学习; 条件模型

中图法分类号: TP181

中文引用格式: 李崇轩, 朱军, 张钊. 条件概率图产生式对抗网络. 软件学报, 2020, 31(4): 1002-1008. <http://www.jos.org.cn/1000-9825/5924.htm>

英文引用格式: Li CX, Zhu J, Zhang B. Conditional graphical generative adversarial networks. Ruan Jian Xue Bao/Journal of Software, 2020, 31(4): 1002-1008 (in Chinese). <http://www.jos.org.cn/1000-9825/5924.htm>

Conditional Graphical Generative Adversarial Networks

LI Chong-Xuan, ZHU Jun, ZHANG Bo

(Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China)

Abstract: Generative adversarial networks (GANs) have been promise on generating realistic images and hence have been studied widely. Notably, graphical generative adversarial networks (graphical-GAN) introduce Bayesian networks to the GAN framework to learn the underlying structures of data in an unsupervised manner. This study proposes a conditional version of graphical-GAN, which can leverage coarse side information to enhance the graphical-GAN and learn finer and more complex structures, in weakly-supervised learning settings. The inference and learning of conditional graphical-GAN follows a similar protocol to graphical-GAN. Two instances of conditional graphical-GAN are presented. The conditional Gaussian mixture GAN can learn fine clusters from mixture data given a coarse label. The conditional state space GAN can learn the dynamics of videos with multiple objects given the labels of the objects..

Key words: deep generative model; generative adversarial network; graphical model; weakly-supervised learning; conditional model

如何在无监督或者弱监督环境中学习到可解释的抽象特征是机器学习乃至人工智能领域非常重要的课题. 产生式对抗网络(generative adversarial networks, 简称 GANs)^[1]可以无监督地建模图像数据的分布^[2,3], 因此

* 基金项目: 国家自然科学基金(61620106010, 61621136008); 博士后创新人才拔尖计划

Foundation item: National Natural Science Foundation of China (61620106010, 61621136008); Chinese Postdoctoral Innovative Talent Support Program

本文由“非经典条件下的机器学习方法”专题特约编辑高新波教授、黎铭教授、李天瑞教授推荐.

收稿时间: 2019-05-30; 修改时间: 2019-07-29; 采用时间: 2019-09-20; jos 在线出版时间: 2020-01-10

CNKI 网络优先出版: 2020-01-14 09:53:21, <http://kns.cnki.net/kcms/detail/11.2560.TP.20200114.0953.008.html>

最近被广泛加以研究.进一步地,一些工作^[4,5]希望引入额外的正则化项和先验知识,提高 GAN 模型学习到的特征的可解释性,并控制生成图像的语义.传统的深度神经网络的层次化特征提取过程是完全“黑盒”的,这是由于这种网络完全忽略了特征之间的结构化先验知识,特征的不同维度之间相对于模型来说是等价的.解决这一问题,往往需要引入监督信号或者结构化归纳偏见(inductive bias),其中一种可行的方法便是利用图(graph)的拓扑结构来刻画特征之间的依赖关系,进而学习到可解释特征.基于这种思想,概率图生成对抗网络(graphical-GAN)^[5]将贝叶斯网络(Bayesian network)和产生式对抗网络相结合,可以灵活地表达随机变量之间的依赖结构,同时拟合变量之间的复杂函数.概率图生成对抗网络提出了结构化的变分推理网络和基于期望传播算法^[6]的局部对抗学习方法,可以用无监督的方式学习到数据的隐藏结构.但是,概率图生成对抗网络没有形式化地考虑如何利用粗粒度的额外监督信号,因此无法直接应用在更加复杂的数据上.

为了解决这个问题,本文考虑存在粗粒度监督信号的弱监督学习环境,提出一种新型的条件概率图生成对抗网络(conditional graphical-GAN),可以在任何贝叶斯网络中利用粗粒度信号信息来学习到更精细而复杂的结构.为了解决隐藏变量的推理问题和模型参数的学习问题,我们提出了基于粗粒度监督信号信息的条件结构化推理网络和期望传播启发的局部对抗学习方法.其主要思想和概率图生成对抗网络(graphical-GAN)的推理和学习方法相一致,但是能够利用额外监督信号,因此结果优于概率图生成对抗网络.

为了解释如何使用条件概率图生成对抗网络(conditional graphical-GAN)进行学习和推理,我们提出了两个实例,分别处理混合数据和时序数据.其中条件高斯混合模型(conditional Gaussian mixture GAN,简称 cGMGAN)可以在给出粗粒度标签的情况下从混合数据中学习细粒度聚类,条件状态空间模型(conditional state space GAN,简称 cSSGAN)可以在给定对象标签的情况下学习具有多个对象的视频的动态过程.

本文第 1 节形式化描述一般的条件概率图生成对抗网络理论框架,包括产生式模型和推理模型的定义以及局部对抗学习算法.第 2 节列出两个实例,包括应用场景描述和具体的变量和网络定义.第 3 节展示实验,包括具体的实验环境和两个实例对应的结果.第 4 节总结本文的贡献.

1 理论框架

我们考虑弱监督学习环境,假设 (X, Y) 表示可观测变量,其中 X 表示数据, Y 表示粗粒度的监督信息, Z 表示隐藏变量,其语义与 Y 无关.这里, X 、 Y 和 Z 均可能为高维随机变量.我们希望在给定 Y 的情况下,学习一个条件产生式模型(也就是一个从 Z 到 X 的映射)能够产生和隐藏语义相对应的数据.与贝叶斯网络类似,这个产生式模型中语义的对应关系是通过一个有向无环图 G 引入的.这里的条件产生式模型的严格定义为

$$P(X, Z | Y) = \prod_{i=1, \dots, |Z|} P(z_i | \pi_G(z_i)) \prod_{j=1, \dots, |X|} P(x_j | \pi_G(x_j), Y),$$

其中, $\pi_G(z_i)$ 表示节点 z_i 在图 G 中的所有父亲节点的集合.需要注意的是,这里任意两点之间的关系都通过神经网络来连接.同时,每个可观测变量 x_j 的概率分布都与 GAN 一样通过确定性映射变换而来,从而保证了模型的拟合能力.

在上述生成式模型中,由于神经网络和高维随机变量的存在,隐藏变量 Z 的后验分布是难以计算的.为了解决推理问题,我们提出了条件结构化推理网络模型(也就是一个从 X 到 Z 的映射),能够在给定 Y 的情况下,推断出数据对应的隐藏变量.这种结构也是通过一个有向无环图 H 引入的. H 的结构由概率图模型中的经典方法^[7,8]基于产生式模型的图 G 得出,具体构建方法参见文献[5].这里的条件推理模型的严格定义为

$$Q(X, Z | Y) = Q(X)Q(Z | X, Y) = Q(X) \prod_{i=1, \dots, |Z|} Q(z_i | \pi_H(z_i), Y),$$

其中, $Q(X)$ 表示训练数据的经验分布.在推理模型中,两点之间的依赖关系也可以用神经网络建模,网络大小与产生式模型中对应因子的网络相当.我们在图 1 中对比了 graphical-GAN 和 conditional graphical-GAN 的混合模型实例(详细内容见第 2 节)的图模型.Conditional graphical-GAN 可以利用粗粒度弱监督信息.

为了同时训练条件产生式模型和条件推理模型,我们最小化两者之间的琴生-香依散度(JS-divergence).这个优化问题可以严格地描述为 $\min_{P, Q} J(P(X, Z | Y) \| Q(X, Z | Y))$, 直接利用对抗训练方法^[9,10]优化上述目标函数在

结构复杂的情况下可能无法收敛^[5].因此,我们基于期望传播算法^[6]开发了局部对抗训练方法,其思想与概率图产生式对抗网络一致^[5].具体地,我们按照 G 中的拓扑关系,考虑一组因子集合 F_G ,其中每一个因子 A 表示 $\{z_i, \pi_G(z_i)|Y\}$ 或者 $\{x_j, \pi_G(x_j)|Y\}$,即枚举图 G 去掉条件变量 Y 之后的所有节点和父亲的因子.我们将上述全局优化问题转化为若干个局部因子优化问题,其形式化描述如下:

$$\min_{P,Q} \frac{1}{|F_G|} \sum_{A \in F_G} J(P(A|Y) \| Q(A|Y)).$$

我们不知道某些因子 A 对应的条件概率分布 $P(A|Y)$ 和 $Q(A|Y)$,所以上述局部因子的琴生-香依散度并不能直接优化.利用对抗学习的方法,我们引入了若干局部条件因子判别器来判别两个模型产生的因子采样,以估计上述局部琴生-香依散度 $J(P(A|Y) \| Q(A|Y))$.最终的优化问题如下:

$$\min_{P,Q} \max_{D_A, \forall A \in F_G} E_Q \left[\sum_{A \in F_G} \log D_A(A|Y) \right] + E_P \left[\sum_{A \in F_G} \log(1 - D_A(A|Y)) \right],$$

其中, $D_A(A|Y)$ 表示因子 A 对应的局部判别器, E_Q 表示依概率分布 Q 取期望.

2 两个实例

我们考虑两种典型的结构化数据以展示如何使用上述框架进行弱监督结构化数据建模.第 1 种情况,我们考虑混合图像数据,在给定的粗粒度类别的情况下,假设数据内还有细粒度的簇(cluster),且没有细粒度监督信号.第 2 种情况,我们考虑多物体的视频数据,给定物体的类别信号,但是物体的位置等信息均为未知.我们提出条件概率图产生式对抗网络的两个重要实例,即条件高斯混合产生式对抗网络(conditional Gaussian mixture GAN,简称 cGMGAN)和条件状态空间产生式对抗网络(conditional state space GAN,简称 cSSGAN),以分别处理上述这两种情况.

2.1 条件高斯混合产生式对抗网络

在条件高斯混合模型中,我们假设每个粗粒度类别下的图像有 K 个细粒度簇,因此引入高斯混合模型作为先验分布,为了让隐藏变量的语义信息独立于粗粒度监督信号,我们假设两者相互独立,共同生成数据.形式化的生成过程如下:

$$k \sim \text{Cat}(K, \theta), \quad h | k \sim N(\mu_k, \Sigma_k), \quad x | h, y = G_{net}(h, y).$$

$\text{Cat}(\cdot)$ 表示范畴分布, $N(\cdot, \cdot)$ 表示高斯分布, G_{net} 表示产生器,是一个确定性神经网络映射.对应于上述理论框架,我们有: $X=x, Y=y, Z=(k, h)$.

在条件高斯混合模型中,我们首先用条件推理网络推理隐藏变量 h ,然后用贝叶斯公式直接得到 k 的后验分布.形式化的推理过程如下:

$$h | x, y = I_{net}(x, y), \quad k | h = \frac{\theta_k N(\mu_k, \Sigma_k)}{\sum_k \theta_k N(\mu_k, \Sigma_k)}.$$

I_{net} 表示条件推理网络.模型示意图如图 1 所示.其中,左侧所示为概率图生成对抗网络的混合模型实例对应的产生式模型和推理模型,右侧所示为条件概率图生成对抗网络的混合模型实例对应的两个模型.在局部因子对抗训练中,我们引入两个局部条件判别器,分别处理因子 (x, h, y) 和 (h, k) .

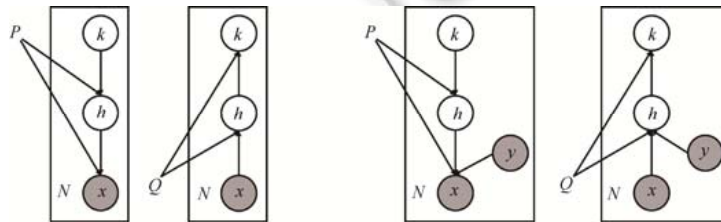


Fig.1 Graphical models to compare graphical-GAN and conditional graphical GAN

图 1 概率图生成对抗网络和条件概率图生成对抗网络的图模型对比

2.2 条件状态空间产生式对抗网络

在条件状态空间模型中,我们假设每个视频有两个物体,并且物体类别给定,视频的长度为 T . 视频的隐藏变量分为两种,其中一种为语义不随时间变化而变化,用 h 表示;另一种语义随时间变化,用 v_t 表示. 我们仍然假设隐藏变量的语义信息独立于粗粒度监督信号,共同生成数据. 形式化的生成过程如下:

$$v_1 \sim N(0, I), h \sim N(0, I), \varepsilon_t \sim N(0, I), v_{t+1} | v_t = O_{net}(v_t, \varepsilon_t), x_t | v_t, h, y_1, y_2 = G_{net}(h, v_t, y_1, y_2).$$

$N(0, I)$ 表示标准高斯分布, ε_t 为辅助随机变量, O_{net} 表示时间序列产生器, G_{net} 表示视频帧的生成器,二者均为确定性神经网络映射. 对应于上述理论框架,我们有: $X=(x_1, \dots, x_T), Y=(y_1, y_2), Z=(h, v_1, \dots, v_T)$.

在条件高斯混合模型中,我们引入两个条件推理网络,分别得到不变隐藏变量和变化隐藏变量. 形式化的推理过程如下:

$$h | x_1, \dots, x_T, y_1, y_2 = I_{net}(x_1, \dots, x_T, y_1, y_2), v_t | x_t, y_1, y_2 = I'_{net}(x_t, y_1, y_2).$$

注意,在上述过程中,时间序列产生器 O_{net} 、视频帧的生成器 G_{net} 和条件推理网络 I'_{net} 均在时间序列上共享. 在局部因子对抗训练中,我们引入两个局部条件判别器,分别处理因子 (x_t, h, v_t, y_1, y_2) 和 (v_t, v_{t+1}) .

3 实验

我们使用 TensorFlow 库实现条件概率图产生式对抗网络的两个实例. 采用无结构的产生式对抗网络 ALI^[9,10], 与概率图生成对抗网络(graphical-GAN)^[5]作为基准模型. 为了进行公平比较,我们延续基准模型的设定,在所有实验中使用经典模型 DCGAN^[2]的网络结构和 ADAM 优化器^[11]. 以 SVHN 数据的条件高斯混合产生式对抗网络为例,具体网络结构如下:(1) 生成器的隐变量为 128 维,级联 10 维的类别标注信息,经过全联接层变换到 4 096 维,形状变换为 $4 \times 4 \times 256$,通过 3 个 5×5 的反卷积操作依次变为 $8 \times 8 \times 128$ 、 $16 \times 16 \times 64$ 、 $32 \times 32 \times 3$;(2) 推理网络使用和生成器相反的网络结构;(3) 判别器将输入图像通过 3 层 5×5 卷积映射到 4 096 维,然后级联隐变量和类别标注信息,通过两层 512 维的全连接网络映射到 1 维二值输出. 生成器和推理网络的非线性为 RELU,判别器的非线性为 Leaky-RELU,斜率为 0.2. 其余超参数若未加说明也均与 graphical-GAN^[5]一致(具体参数可以参考 graphical-GAN 附录 D). 我们在 MNIST、SVHN 和 CIFAR10 数据集上测试条件高斯混合产生式对抗网络(cGMGAN),在移动 MNIST 数据集上测试条件状态空间产生式对抗网络(cSSGAN).

我们希望通过实验验证如下假设.

(1) 条件高斯混合产生式对抗网络,可以在混合数据中学习到细粒度的数据簇,与已有模型相比,可以避免类别混淆和不平均的问题,同时提高推理模型的准确程度.

(2) 条件状态空间产生式对抗网络,可以在多物体视频中学习到动态信息和固定信息,完成数据生成和动态模仿等任务.

3.1 条件高斯混合产生式对抗网络

我们首先展示条件高斯混合产生式对抗网络(cGMGAN)在 MNIST、SVHN 和 CIFAR10 数据集上的结果. 其中, MNIST 为手写数字,粗粒度类别为 0~9, SVHN 为街景拍摄门牌号,粗粒度类别为数字 0~9, CIFAR10 为自然场景图像,粗粒度类别为汽车、鸟等 10 类.

在给定类别的情况下, cGMGAN 可以学习到细粒度的数据簇表示. 如图 2 所示,我们展示了 cGMGAN 在数字 2 和数字 6 上的结果. 以数字 2 为例,最左侧的图像展示了聚类中心对应的代表字体,我们发现对应了数字 2 的两种写法,即左下角有没有环. 随后我们随机产生了 5×2 张采样,其中每一列的采样对应同一个聚类,即 k 值,我们发现其类型与聚类中心的代表字体相同,这表示 cGMGAN 可以在混合数据中学习到细粒度的数据簇. 随后,我们可视化了聚类中心周围邻域内的隐藏变量,发现其隐藏空间是连续的,并且处处对应于合理的图片. cGMGAN 在 SVHN 和 CIFAR10 上的结果如图 3 和图 4 所示. 我们可视化了第 1 个聚类中心周围的隐藏空间. 在这两个数据集上,聚类的语义主要表现在色彩和物体姿态上. 与 GMGAN^[5]的结果相比, cGMGAN 的细粒度数据簇没有类别混淆和类间不平均的问题.

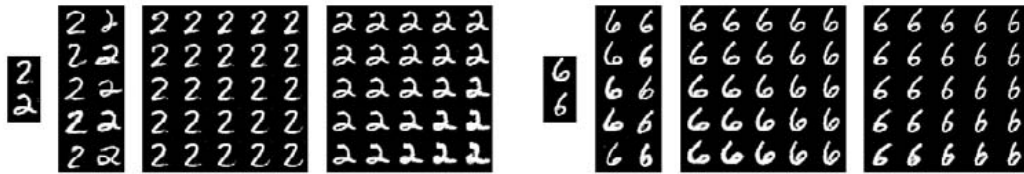


Fig.2 cGMGAN results on the MNIST dataset, $K=2$. Left panel: Digit 2, right panel: Digit 6

图 2 条件高斯混合模型在 MNIST 数据上的结果, $K=2$. 左侧: 数字 2, 右侧: 数字 6



Fig.3 cGMGAN results on the SVHN dataset, $K=5$. Left panel: Digit 2, right panel: Digit 7

图 3 条件高斯混合模型在 SVHN 数据上的结果, $K=5$. 左侧: 数字 2, 右侧: 数字 7



Fig.4 cGMGAN results on the CIFAR10 dataset, $K=5$. Left panel: Horse, right panel: Bird

图 4 条件高斯混合模型在 CIFAR10 数据上的结果, $K=5$. 左侧: 马, 右侧: 鸟

我们在 MNIST 数据集上定量测试了 cGMGAN 的推理模型的准确程度.用 cGMGAN 和两个基准模型对测试数据进行了重建,即先推理隐藏变量,再把这些变量给产生式模型恢复原始输入,并计算了重建图像和原始输入的均方误差.表 1 给出了与无结构的产生式对抗网络和高斯混合产生式对抗网络进行比较的情况,cGMGAN 取得了最好的结果.为了公平比较,我们使用了无结构的产生式对抗网络^[9]中提出的全局学习算法.

Table 1 RMSE results of reconstruction on the MNIST dataset (the smaller, the better)

表 1 MNIST 数据集上重建数据和输入的均方误差结果(越小越好)

名称	结果
无结构的产生式对抗网络(ALI)	0.071
概率图产生式对抗网络(GMGAN)	0.056
本文提出的条件概率图产生式对抗网络(cGMGAN)	0.046

3.2 条件状态空间产生式对抗网络

我们随后来展示条件空间状态产生式对抗网络(cSSGAN)在移动 MNIST 数据集上的结果.该数据是从 MNIST 数据合成而来.首先随机采样两个 MNIST 数据,放在一个 64×64 大小的画布上,然后随机采样速度大小和方向,让数据在画布上进行移动,数据遇到边界会反弹,我们截取了长度为 16 的视频作为数据.为了更好地可视化结果,我们将两个数字放在了不同的信道上.

图 5 展示了随机生成的 3 个视频,可以发现,cSSGAN 可以非常好地建模多物体的时序关系,并且保证物体的固有特征不变,完成像素级别的生成.

图 6 展示了动作模仿的结果.在图中,奇数行表示一个输入视频,我们将输入视频的变化特征即 v_t 利用推理

模型推理出来,然后随机产生一个不变特征 h ,并令 $y_1=y_2=1$,结合推理出来的 v_t ,生成对应偶数行的视频,我们发现,偶数行视频可以生成不同形态的两个数字 1,并且保留了与输入视频一致的变化信息,即每一帧的数字未知.这表明,cSSGAN 可以在无位置监督信号的情况下,自动解耦合不变特征和变化特征,可以用可解释的因子合成符合需求的视频.

上述图 5 和图 6 的结果表明,在引入弱监督信号的情况下,cSSGAN 可以建模多物体视频,而 SSGAN^[5]的结果仅局限在单物体视频上,这体现了本文提出的条件概率图产生式对抗网络的优势.

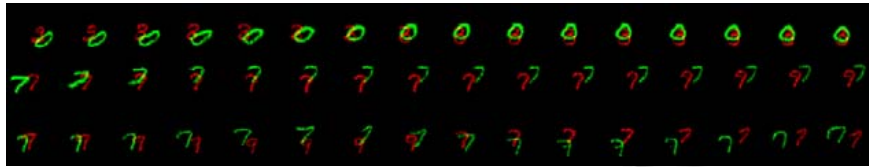


Fig.5 cSSGAN results on the moving MNIST dataset. Each row presents the frames of a generated video

图 5 条件空间状态模型在移动 MNIST 数据上的结果.每一行表示一个生成的视频的所有帧

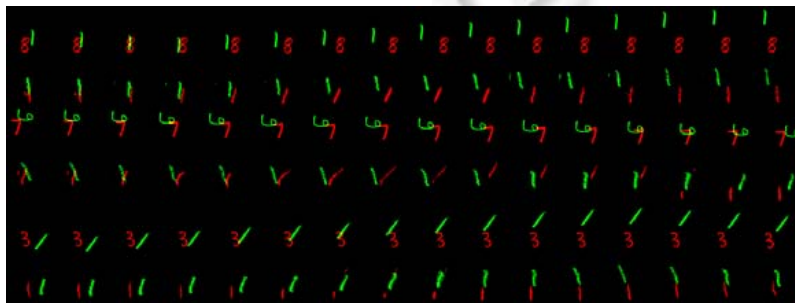


Fig.6 cSSGAN motion analogy results on the moving MNIST dataset

图 6 条件空间状态模型在移动 MNIST 数据集上动作模仿的结果

4 结 论

综上所述,本文的主要贡献如下.

(1) 提出了一种新型普适的面向弱监督学习的条件概率图产生式对抗网络,旨在利用粗粒度监督信号,进一步刻画数据中潜在的细粒度无标定结构信息,提高产生式对抗网络提取特征和生成图像的可解释性.

(2) 提出了条件高斯混合产生式对抗网络,可以在混合数据中学习细粒度的数据簇,避免了类别混淆和不平均的问题,同时提高了推理模型的准确率.

(3) 提出了条件状态空间产生式对抗网络,可以在多物体视频中学习动态信息和固定信息,完成数据生成和动态模仿等任务.

References:

- [1] Goodfellow I, Mirza M, Xu B, Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial nets. In: Advances in Neural Information Processing Systems. Montreal: MIT Press, 2014. 2672–2680.
- [2] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. In: Proc. of the Int'l Conf. on Learning Representations. San Diego: Journal of Machine Learning Research, 2015.
- [3] Arjovsky M, Chintala S, Bottou L. Wasserstein GAN. In: Proc. of the Int'l Conf. on Machine Learning. Sydney: Journal of Machine Learning Research, 2015. 214–223.

- [4] Chen X, Duan Y, Houthoofd R, Schulman J, Sutskever I, Abbeel P. InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets. In: Advances in Neural Information Processing Systems. Barcelona: MIT Press, 2016. 2172–2180.
- [5] Li C, Welling M, Zhu J, Zhang B. Graphical generative adversarial networks. In: Advances in Neural Information Processing Systems. Montreal: MIT Press, 2018. 6073–6083.
- [6] Minka T. Expectation propagation for approximate Bayesian inference. In: Proc. of the 17th Conf. on Uncertainty in Artificial Intelligence. Washington: Morgan Kaufmann Publishers Inc., 2001. 362–369.
- [7] Jordan M, Ghahramani Z, Jaakkola T, Saul L. An introduction to variational methods for graphical models. Machine Learning, 1999,37(2):183–233.
- [8] Stuhlmüller A, Taylor J, Goodman N. Learning stochastic inverses. In: Advances in Neural Information Processing Systems. Lake Tahoe: MIT Press, 2013. 3048–3056.
- [9] Donahue J, Henbühl P, Darrell T. Adversarial feature learning. In: Proc. of the Int'l Conf. on Learning Representations. Toulon: Journal of Machine Learning Research, 2017.
- [10] Dumoulin V, Belghazi I, Poole B, Lamb A, Arjovsky M, Mastropietro O, Courville A. Adversarially learned inference. In: Proc. of the Int'l Conf. on Learning Representations. Toulon: Journal of Machine Learning Research, 2017.
- [11] Kingma DP, Ba J. Adam: A method for stochastic optimization. In: Proc. of the Int'l Conf. on Learning Representations. San Diego: Journal of Machine Learning Research, 2015.



李崇轩(1991—),男,河北保定人,博士,助理研究员,CCF 学生会员,主要研究领域为深度生成模型.



朱军(1983—),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为机器学习.



张钊(1935—),男,博士,教授,博士生导师,主要研究领域为人工智能.