

# 基于实值 RBM 的深度生成网络研究\*

张健<sup>1,2</sup>, 丁世飞<sup>1,2</sup>, 丁玲<sup>3</sup>, 张成龙<sup>1</sup>

<sup>1</sup>(中国矿业大学 计算机科学与技术学院, 江苏 徐州 221116)

<sup>2</sup>(矿山数字化教育部工程研究中心, 江苏 徐州 221116)

<sup>3</sup>(中国矿业大学 徐海学院, 江苏 徐州 221008)

通讯作者: 丁世飞, E-mail: dingsf@cumt.edu.cn; 张健, E-mail: 597409675@qq.com



**摘要:** 受限玻尔兹曼机(restricted Boltzmann machine, 简称 RBM)是一种概率无向图,传统的 RBM 模型假设隐藏层单元是二值的,二值单元的优势在于计算过程和采样过程相对简单,然而二值化会对基于隐藏层单元的特征提取和数据重构过程带来信息损失.因此,将 RBM 的可见层单元和隐藏层单元实值化并保持模型训练的有效性,是目前 RBM 理论研究的重点问题.为了解决这个问题,将二值单元拓展为实值单元,利用实值单元建模数据并提取特征.具体而言,在可见层单元和隐藏层单元之间增加辅助单元,然后将图正则化项引入到能量函数中,基于二值辅助单元和图正则化项,流形上的数据有更高的概率被映射为参数化的截断高斯分布;同时,远离流形的数据有更高的概率被映射为高斯噪声.由此,模型的隐层单元可以被表示为参数化截断高斯分布或高斯噪声的采样实值.该模型称为基于辅助单元的受限玻尔兹曼机(restricted Boltzmann machine with auxiliary units, 简称 ARBM).在理论上分析了模型的有效性,然后构建了相应的深度模型,并通过实验验证模型在图像重构任务和图像生成任务中的有效性.

**关键词:** 受限玻尔兹曼机;神经网络;概率图模型;深度学习

**中图分类号:** TP18

中文引用格式: 张健,丁世飞,丁玲,张成龙.基于实值 RBM 的深度生成网络研究.软件学报,2021,32(12):3802–3813. <http://www.jos.org.cn/1000-9825/6126.htm>

英文引用格式: Zhang J, Ding SF, Ding L, Zhang CL. Deep generative neural networks based on real-valued RBM with auxiliary hidden units. Ruan Jian Xue Bao/Journal of Software, 2021,32(12):3802–3813 (in Chinese). <http://www.jos.org.cn/1000-9825/6126.htm>

## Deep Generative Neural Networks Based on Real-valued RBM with Auxiliary Hidden Units

ZHANG Jian<sup>1,2</sup>, DING Shi-Fei<sup>1,2</sup>, DING Ling<sup>3</sup>, ZHANG Cheng-Long<sup>1</sup>

<sup>1</sup>(School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China)

<sup>2</sup>(Engineering Research Center of Mine Digitization of Ministry of Education, Xuzhou 221116, China)

<sup>3</sup>(Xuhai College, China University of Mining and Technology, Xuzhou 221008, China)

**Abstract:** Restricted Boltzmann machine (RBM) is a probabilistic undirected graph, and most traditional RBM models assume that their hidden layer units are binary. The advantage of binary units is their calculation process and sampling process are relatively simple. However, binarized hidden units may bring information loss to the process of feature extraction and data reconstruction. Therefore, a key research point of RBM theory is to construct real-valued visible layer units and hidden layer units, meanwhile, maintain the effectiveness of model training. In this study, the binary units are extended to real-valued units to model data and extract features. To achieve this, specifically, an auxiliary unit is added between the visible layer and the hidden layer, and then the graph regularization term is introduced into the energy function. Based on the binary auxiliary unit and graph regularization term, the data on the manifold has a higher probability to be mapped as a parameterized truncated Gaussian distribution, simultaneously, the data far from the manifold has a higher

\* 基金项目: 国家自然科学基金(61976216, 61672522)

Foundation item: National Natural Science Foundation of China (61976216, 61672522)

收稿时间: 2020-04-14; 修改时间: 2020-06-05; 采用时间: 2020-08-07

probability to be mapped as Gaussian noises. The hidden units can be sampled as real-valued units from the parameterized Gaussian distribution and Gaussian noises. In this study, the resulting RBM based model is called restricted Boltzmann machine with auxiliary units (ARBM). Moreover, the effectiveness of the proposed model is analyzed theoretically. The effectiveness of the model in image reconstruction task and image generation task is verified by experiments.

**Key words:** restricted Boltzmann machines; neural nets; probabilistic undirected graphs; deep learning

近年来,深度学习引起了广泛关注.基于概率图和神经网络的受限玻尔兹曼机(restricted Boltzmann machine,简称 RBM)、变分自编码器(variational autoencoder,简称 VAE)和生成对抗网络(generative adversarial network,简称 GAN)被广泛应用于图像分类和图像生成任务中<sup>[1,2]</sup>.与此同时,近年来对 RBM 的研究遇到了一些困难<sup>[3,4]</sup>,其原因在于,其他方法、特别是 GAN 在实践中效果更好.事实上,RBM 及其衍生模型一般都有足够的特征表达能力来学习复杂的数据分布,其应用和拓展上的困难往往出现在训练过程中,RBM 训练需要显式的激活概率和梯度表达式.因此对于 RBM 而言,在保证显式梯度前提下构建有效的方法来学习数据分布是比较困难的.为了建模实值数据,人们提出了多种实值 RBM 模型<sup>[5-7]</sup>.但是 RBM 模型的特征表达问题和深度拓展问题一直没有得到很好的解决.具体而言,RBM 最常用的训练算法是基于 Gibbs 采样的对比散度算法(contrastive divergence,简称 CD)算法,而在 CD 算法中,可见层单元的状态  $\tilde{x}$  总是从条件概率  $p(x|h)$  中采样得到,因此,一个充分参数化且易学习的条件概率  $p(x|h)$  对于建模数据分布而言是非常重要的.通常我们希望可见层单元的边缘概率和条件概率具有足够的表达能力,从而最大程度地拟合数据分布.然而在大多数 RBM 中,隐藏层单元是二值的,二值隐藏层单元限制了 RBM 对可见层单元条件概率的参数化能力,因为可见层单元的条件概率是由基于二值隐藏层单元的高斯分布的组合确定的.

为了有效地参数化可见层单元的激活概率并建模数据中的流形结构,本文在可见层单元和隐藏层单元之间增加二值的辅助单元从而实值化可见层单元和隐藏层单元,并在 RBM 的能量函数中引入了 Laplacian 正则化项来构建特征化的隐藏层单元表达,由此提出了基于辅助单元的受限玻尔兹曼机(restricted Boltzmann machine with auxiliary units,简称 ARBM).具体而言,假设 ARBM 中的可见层单元和隐藏层单元服从截断高斯分布,模型参数化截断高斯分布中的期望和协方差.基于实值隐藏层单元,可见层单元的条件激活概率可以看作是无限数量的截断高斯分布的加权和,从而能够拟合复杂的数据分布.在 ARBM 中,基于二值辅助单元和图正则化项,当样本靠近数据流形时,它有更高的概率被映射为参数化的截断高斯分布;反之,远离流形的样本更可能被映射为高斯噪声.辅助单元的另一个作用是缓解过拟合问题,过拟合在神经网络和 RBM 的训练中普遍存在,而 dropout 方法是缓解过拟合问题的常用方法.在基于 dropout 方法的 RBM 模型中,可见层和隐藏层中的部分单元被随机屏蔽,这可能会影响 RBM 的特征提取能力和图像重构能力<sup>[8-10]</sup>.而在 ARBM 中,辅助单元控制可见层单元和隐藏层单元的激活概率,并在一定程度上保留了单元的随机性.当辅助单元被激活时,ARBM 的可见层单元和隐藏层单元以参数化的条件概率形式被激活.当辅助单元灭活时,可见层单元和隐藏层单元的条件概率退化为高斯噪声.此时类似于 dropout 方法,基于 ARBM 的神经网络的训练过程可以也看作是一个去噪过程.为了建立深度生成模型并获得锐利的生成图像,本文基于 ARBM 提出了基于附加单元的实值深度置信网(real-valued deep belief net with auxiliary units,简称 ADBN).本文首先验证了 ARBM 的性能优于常用的 RBM 模型,然后利用 ADBN 提取的深度特征作为条件,生成对抗网络(conditional generative adversarial net,简称 CGAN)的条件输入,构建了一个深度生成神经网络.我们称该模型为基于辅助深度特征的条件生成对抗网络(conditional GAN with auxiliary deep features,简称 ACGAN).本文的主要贡献可以概括如下:

(1) 在可见层和隐藏层之间引入二值附加单元,提出了 ARBM 模型.基于能量函数中的图正则化项和附加单元,靠近数据流形的样本有更高的概率被参数化为截断高斯分布的形式,远离流形的样本有更高的概率被映射为高斯噪声,因此,ARBM 可以有效地学习数据中的流形结构,并将其表示为隐藏层的参数化特征;

(2) 为了构建有效的深度生成模型,本文将 ADBN 提取的深度特征用作 CGAN 的附加输入,提出了 ACGAN 模型.与传统的随机噪声相比,这些深度特征能为 GAN 提供更稳定有效的初始状态,从而生成高质量的图像、缓解 GAN 模型训练不稳定的问题.

本文第 1 节介绍基础理论,包括 RBM 和 CGAN.第 2 节详细介绍 ARBM 及其相应的深生成模型 ADBN 和 ACGAN.在第 3 节中,本文通过实验验证所提出模型的有效性.最后一部分是结论和展望.

## 1 相关工作

### 1.1 受限玻尔兹曼机

典型的 RBM 模型由可见层  $x$  和隐藏层  $h$  组成,RBM 的联合概率可以通过能量函数来定义,RBM 模型及其基于 Gibbs 采样的训练过程示意图可以表示为图 1 的形式.

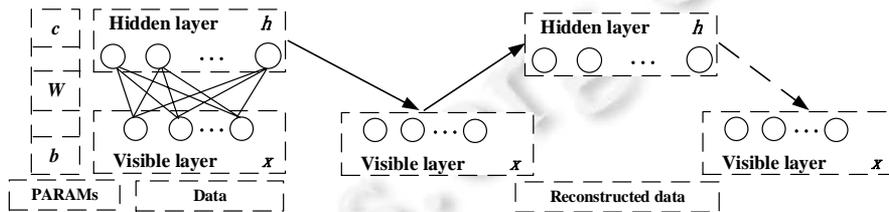


Fig.1 Topology of an RBM and its training process

图 1 RBM 的拓扑结构及其训练过程

在图 1 中, $W$  是可见层和隐藏层之间的权值矩阵.如果可见层单元和隐藏层单元是二值的,那么能量函数可以定义如下:

$$E(x, h) = -\sum_{i=1}^{n_v} a_i x_i - \sum_{j=1}^{n_h} b_j h_j - \sum_{i=1}^{n_v} \sum_{j=1}^{n_h} h_j \times W_{ji} \times x_i \quad (1)$$

其中, $a$  和  $b$  是 RBM 的偏置, $x$  表示可见层向量, $h$  表示隐藏层向量, $W$  是权值矩阵.基于能量函数  $E(x, h)$ ,联合分布可以表示为  $p(x, h) = Z^{-1} \exp(-E(x, h))$ ,可见层单元和隐藏层单元的激活函数可以表示如下:

$$p(h_k = 1 | x) = \text{sigmoid}(b_k + \sum_{i=1}^{N_v} w_{ki} x_i) \quad (2)$$

$$p(x_k = 1 | h) = \text{sigmoid}(a_k + \sum_{j=1}^{N_h} h_j w_{kj}) \quad (3)$$

典型 RBM 模型的目标是最大化其边缘分布  $p(x)$  的 log 似然,这个目标函数可以表示为

$$L_s = \ln(p(x)) = \ln \left( \sum_h p(x, h) \right) = \ln \left( \sum_h e^{-E(x, h)} \right) - \ln(Z) \quad (4)$$

根据极大似然估计,似然函数关于参数的梯度可以表示如下:

$$\frac{\partial \ln p(x)}{\partial \theta} = -\sum_h p(h | x) \frac{\partial E(x, h)}{\partial \theta} + \sum_{x, h} p(x, h) \frac{\partial E(x, h)}{\partial \theta} \quad (5)$$

将公式(5)表示为期望的形式,可以得到:

$$\frac{\partial L_s}{\partial \theta} = E_{p(x, h)} \left[ \frac{\partial E(x, h)}{\partial \theta} \right] - E_{p(h | x)} \left[ \frac{\partial E(x, h)}{\partial \theta} \right] \quad (6)$$

如公式(6)所示:等式右边的第 1 项称为模型期望,第 2 项称为数据期望,两个期望的差值决定了似然函数关于参数的梯度.直观上看,数据期望给出了参数迭代的起始条件,模型期望提供了迭代的终止条件.随着迭代的进行,数据期望和模型期望逐渐接近,RBM 的训练随迭代趋于稳定.此时,RBM 模型建模了输入样本的分布特性.然而在大样本下,精确地计算这两个期望是非常困难的,尤其是模型期望.因此,为了降低 RBM 训练的复杂度,需要对似然函数的梯度做近似.为此,Hinton 等学者提出了对比散度算法(contrastive divergence,简称 CD)来近似似然函数的梯度.除了对比散度算法,还有其他一些类型的方法可以用于近似计算 RBM 的梯度<sup>[11-13]</sup>.

### 1.2 条件生成对抗网络

作为近些年最受关注的生成模型,生成对抗网络(generative adversarial net,简称 GAN)得到了广泛的应用.GAN 的优势在于可以回避难以直接计算的基于模型分布的积分函数,只使用反向传播就可以根据梯度训练生

成模型,不需要推理.此外,当生成器和判别器在某些额外的条件  $y$  的约束下,可以将 GAN 扩展到条件模型:条件生成对抗网络(conditional generative adversarial net,简称 CGAN).其中, $y$  可以是任何类型的辅助信息,例如类标签或来自其他模态的数据.可以通过将  $y$  作为附加输入传递给判别器和生成器来执行该条件<sup>[14]</sup>.CGAN 的目标函数可以用公式(7)表示:

$$\min_D \max_G V(D, G) = E_{x \sim p_{data}} [\log D(x | y)] + E_{z \sim p_z} [\log(1 - D(G(z | y)))] \tag{7}$$

图 2 是一个典型的 CGAN 的网络结构.

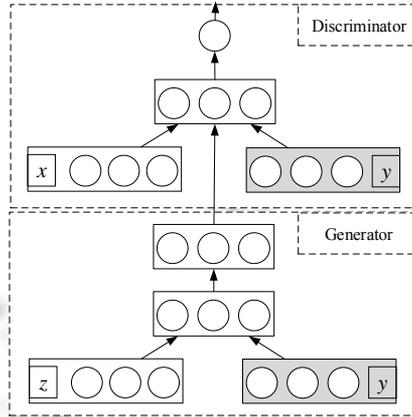


Fig.2 Structure of a CGAN

图 2 CGAN 的结构图

作为一个 CGAN 模型,其输入的条件  $y$  是非常重要的.模型可以根据条件  $y$  所包含的信息完成特定的任务,不仅可以将 CGAN 用于监督和半监督任务,本文还使用深度特征作为条件输入,从而增强模型在图像生成任务中的有效性.

## 2 基于附加单元的受限玻尔兹曼机及其深度神经网络扩展

### 2.1 基于附加单元的受限玻尔兹曼机

首先,我们通过直观的方式分析二值单元在 RBM 特征表达方面存在的不足.为了方便表述,假设模型有一个隐藏层单元,且只考虑单个样本的情形.根据模型的独立性假设,在此基础上,可以很方便地推广到多样本多隐藏层单元的情形.二值 RBM 不能有效地建模实值数据,是因为对实值数据二值化的过程中,数据中一些重要的相关性信息可能会丢失.为了建模实值数据,传统的 Gaussian-binary RBM 假设其可见层单元的条件概率服从高斯分布,其隐藏层单元为二值的.基于极大似然估计,目标函数可以表示为  $L = \log p(x) = \log \sum_i p(x, h_i)$ ,其中,  $h$  是二值单元.基于二值隐藏层单元  $h$ ,可见层单元的条件概率  $p(x|h)$  可以表示为两种形式:  $p(x|h=0)$  和  $p(x|h=1)$ .因此,RBM 可见层单元的边缘概率可以进一步分解为  $p(x) = p(x|h=0)p(h=0) + p(x|h=1)p(h=1)$ ,其中,  $p(x|h=0)$  和  $p(x|h=1)$  为高斯分布.基于二值的隐藏层单元,可见层单元的边缘分布可以看作是  $p(x|h=0)$  和  $p(x|h=1)$  的加权和.这个过程可以表示为图 3 的形式.

如果一个分布的密度函数  $p(y)$  可以写成  $p(y) = \sum_i p(y_i)$ ,当  $i \rightarrow \infty$  且  $p(y_i)$  是高斯分布的时候,  $p(y)$  可以用于拟合任何连续分布函数.然而在传统的实值 RBM 中,图 3 中  $p(x|h=0)$  和  $p(x|h=1)$  的组合(两个高斯分布的加权和)很难拟合一个复杂数据分布;而当 RBM 的隐藏层单元也是实值单元时,可见层单元的边缘概率可以看作是无数高斯分布的加权和  $p(x) = \int p(x|h)p(h)dh$ ,理论上可以拟合一个复杂连续的数据分布.因此,本文将 RBM 的隐藏层单元由二值单元拓展为实值单元.然而,高斯分布形式的激活函数在 BP 算法中具有较高的计算复杂度且不稳定,因此我们假设可见层单元和隐藏层单元服从截断高斯分布,并由此近似神经网络中的 ReLU 激活函数.为了

建模嵌入在数据中的流形结构,本节在可见层和隐藏层之间增加辅助单元,然后将图正则化项引入到能量函数中.基于二值辅助单元和图正则化项,流形上的数据有更高的概率被映射为参数化的截断高斯分布;同时,远离流形的数据更可能被映射为高斯噪声.ARBM 的结构如图 4 所示(其中,可见层单元和附加单元之间的连接值为权值  $W$ ,二值辅助单元和实值隐藏层单元之间是一一对应的关系).

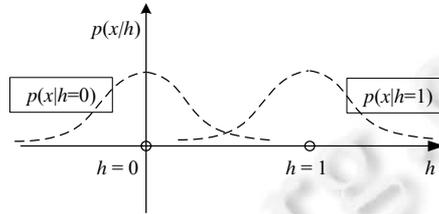


Fig.3 A sketch map of Unweighted marginal distribution of the visible units in a Gaussian-binary RBM

图 3 Gaussian-binary RBM 中可见层单元未加权的边缘分布示意图

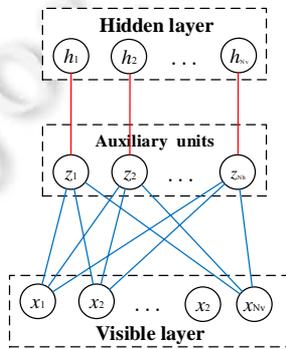


Fig.4 Structure of the ARBM

图 4 ARBM 的结构图

ARBM 的能量函数可以表示如下:

$$E(x, z, h) = x^T \text{diag}(\alpha)x + h^T \text{diag}(\beta)h - 2c^T x - 2\varepsilon^T z - 2x^T W(h \times z) \tag{8}$$

其中,  $x$  和  $h$  被截断在区间  $[0, +\infty]$ ,  $\text{diag}(\alpha)$  和  $\text{diag}(\beta)$  是对角矩阵,  $c$  是偏置,  $\varepsilon$  是当前样本与其所在 mini-batch 内其他样本之间的图相似性系数.图相似性系数可以使用高斯扩散核(Gaussian diffusion kernel)的形式计算:

$$\Sigma(i, j) = \sigma \exp(-d(i, j)) \tag{9}$$

其中,  $d(i, j)$  是欧氏距离;  $\sigma$  是用于平衡数量级的超参数;  $\Sigma$  是  $N \times N$  的矩阵, 且  $\varepsilon(i) = (1/N) \sum_j \Sigma(i, j)$ . 基于公式(9), 联合概率和条件概率可以表示为

$$p(x, z, h) = \frac{1}{Z} \exp(-E(x, z, h)) I(x \geq 0) I(h \geq 0) \tag{10}$$

$$p(z_i = 1 | x) = \text{sigmoid}((x^T W)^T \beta_i^{-1} (x^T W) + 2\varepsilon) \tag{11}$$

$$p(h_i | x, z) \sim N_T(\beta_i^{-1} (x^T W z), \beta_i^{-1}) \tag{12}$$

$$p(x_i | h, z) \sim N_T(\alpha_i^{-1} (c_i + W(h \times z)), \alpha_i^{-1}) \tag{13}$$

其中,  $I(\cdot)$  是示性函数,  $N_T$  表示截断高斯分布, 辅助单元用于控制隐藏层单元和可见层单元的激活概率. 基于公式(12)和公式(13), 可见层单元和隐藏层单元的条件概率可以表示为截断高斯分布, 实值隐藏层单元可以为可见层单元的条件概率提供更灵活的期望. 接下来, 本文从理论上讨论流形学习和辅助单元的作用.

**命题 1.** 在 ARBM 中, 可见层单元和隐藏层单元都是条件独立的, 为了方便表述, 我们假设样本和特征都是一维的. 给定一个 mini-batch 的样本, 截断高斯分布的期望是  $\beta^{-1}(xWz) + \beta^{-1}(f(0)/S(0))$ , 协方差是  $\beta^{-1} - \beta^{-1}(xWz)(f(0)/S(0)) - \beta^{-2}(f(0)/S(0))^2$ , 其中  $f(h) = p(h|z, x)$ ,  $S(h) = 1 - P(h|z, x)$ . 对于一个样本而言, 当它靠近数据流形时, 其对应的

隐式表达具有更高的概率密度,并且该样本以近似为  $p(z=1|x)$  的概率被映射到以  $\beta^{-1}vW$  为中心、 $3\sqrt{\beta^{-1}}$  为邻域的区间内;反之,当样本点远离流形时,它以  $p(z=0|x)$  的近似概率被映射为高斯噪声  $N(0,\beta^{-1})$ .

证明:假设一个 mini-batch 包含  $N$  个样本,基于公式(9),如果样本  $x^{(i)}$  位于流形附近,其相似性系数  $\alpha(i)$  则会相对较高.与  $x^{(i)}$  对应的附加单元  $z^{(i)}$  以概率  $p(z^{(i)}=1|x)$  激活,与相似性系数  $\alpha(i)$  正相关.因此,流形上的样本对应于  $p(z^{(i)}=1|x)$  的高条件概率.当  $z^{(i)}=1$ ,隐藏层单元的条件概率可以表达为  $N_T(\beta^{-1}(xW),\beta^{-1})$ .根据高斯分布置信区间的概念,从高斯分布中获得的样本以超过 99.74% 的概率位于区间  $[\mu-3\sigma,\mu+3\sigma]$ .因此在本文中,对于截断高斯分布,这个概率大于 99.74%,本节假设这个概率可以近似为 1.因此,样本以近似概率  $p(z=1|x)$  被映射到区间  $[\mu-3\sigma,\mu+3\sigma]$ ,且  $\mu=\beta^{-1}xW$ .如果  $x^{(i)}$  不在流形附近,其相似性系数  $\alpha(j)$  会偏小,其对应的附加单元  $z^{(i)}$  更可能以  $p(z=0|x)$  激活.当  $z=0$ ,隐藏层单元的条件概率可以表示为  $N_T(0,\beta^{-1})$ .因此,基于公式(12),样本  $x^{(i)}$  在近似概率  $p(z^{(i)}=0|x^{(i)})$  下被映射到高斯噪声  $N_T(0,\beta^{-1})$ .截断高斯分布的期望可以表示为

$$\begin{aligned} E(h|h>0) &= \int_0^{+\infty} h(p(h|x,z)/S(0))dh \\ &= \frac{\mu}{S(0)} \int_0^{+\infty} p(h|x,z)dh + \frac{-\sigma^2}{S(0)} \int_0^{+\infty} \frac{-(h-\mu)}{\sigma^2} p(h|x,z)dh \\ &= \mu + \sigma^2(p(h=0|x,z)/S(0)) \\ &= \mu + \sigma^2(f(0)/S(0)) \end{aligned} \quad (14)$$

其中,  $\mu=\beta^{-1}(xWz)$ ,  $f(0)=p(h=0|z,x)$ ,  $\sigma^2=\beta^{-1}$ .根据期望和协方差的关系可知,  $h$  的协方差是  $Var(h|h>0)=E(h^2|h>0)-E(h|h>0)^2$ ,期望  $E(h^2|h>0)$  可以表示为

$$\begin{aligned} E(h^2|h>0) &= \int_0^{+\infty} h^2 \frac{p(h|x,z)}{S(0)} dh = \int_0^{+\infty} h(x-\mu+\mu) \frac{p(h|x,z)}{S(0)} dh \\ &= \frac{-\sigma^2}{S(0)} \int_0^{+\infty} h \frac{-(x-\mu)}{\sigma^2} p(h|x,z)dh + \frac{\mu}{S(0)} \int_0^{+\infty} hp(h|x,z)dh \\ &= \mu^2 + \sigma^2 + \mu\sigma^2 \frac{p(h=0|x,z)}{S(0)} \\ &= \mu^2 + \sigma^2 + \mu\sigma^2 \frac{f(0)}{S(0)} \end{aligned} \quad (15)$$

因此,  $Var(h|h>0)=\sigma^2-\mu\sigma^2(f(0)/S(0))-(\sigma^2(f(0)/S(0)))^2$ .  $\square$

基于命题 1,在 ARBM 中,如果样本位于流形附近,则更有可能映射到参数化概率,并且可以学习到流形特征保留在隐藏层中;反之,如果样本远离流形,则更有可能退化为隐藏层的高斯噪声.

ARBM 可以用 CD 算法训练,似然函数关于参数的梯度可以表示为

$$\frac{\partial \ln p(x)}{\partial \theta} = E_{p(h,z,x)} \left[ \frac{\partial E(x,z,h)}{\partial \theta} \right] - E_{p(h,z,x)} \left[ \frac{\partial E(x,z,h)}{\partial \theta} \right] \quad (16)$$

能量函数的梯度可以表示为

$$\partial E(x,h,z)/\partial w_{ij} = -2(x_i)(z^*h)_j, \partial E(x,h,z)/\partial c_i = -2(x)_i, \partial E(x,h,z)/\partial \alpha_j = x_j^2, \partial E(x,h,z)/\partial \beta_i = h_i^2.$$

在 ARBM 中,为了计算截断高斯分布下的期望值,借鉴高斯图模型,本章采用文献[15,16]中的方法.ARBM 还可以经过逐层堆叠构建深度置信网,本文将其称为基于附加单位的深度置信网(deep belief net an auxiliary units deep belief net,简称 ADBN).

## 2.2 基于 ARBM 的深度生成模型

ARBM 本身作为一种概率图模型,既可以用于判别任务,也可以用于生成任务.本文关注的重点是基于 ARBM 的生成模型.由于 ARBM 本身是一种单隐藏层神经网络结构,其特征表达能力有限,因此我们需要根据 ARBM 模型构建相应的深度神经网络结构.

首先,我们构建了一个基于 ARBM 的深度置信网,该神经网络含有 2 个隐藏层,本文称其为基于附加单元的

实值深度置信网(real-valued deep belief net with auxiliary units,简称 ADBN),该 ADBN 的网络结构如图 5 所示.

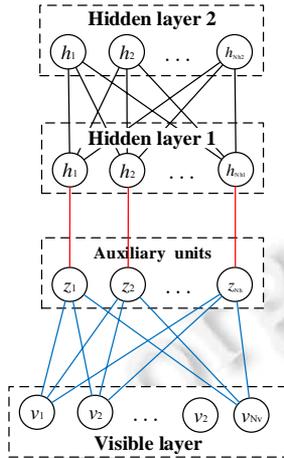


Fig.5 Structure of an ADBN  
图 5 ADBN 模型的结构图

ARBM 和 ADBN 可以直接用于图像生成和图像分类任务,但由于 RBM 训练过程中的标准卷积运算和最大化运算是不可逆的,因此 ARBM 很难与卷积神经网络相结合构建深度神经网络结构.为了建立更强大的深度神经网络,我们利用 ADBN 提取的深层特征作为条件 GAN(CGAN)的条件输入数据,建立深层生成神经网络. GAN 被训练以最小化数据分布与模型分布之间的距离,而不是最大化似然函数.然而 GAN 训练是不稳定的,通常观察到具有相似结构和超参数的生成器在不同的训练批次中表现出显著不同的行为.虽然 RBM 训练需要使用基于采样的算法,但 RBM 的优势在于不存在模式丢失问题且 RBM 训练是稳定的.为了充分利用 ARBM 的优势,并结合卷积操作生成清晰锐利的图像,本文利用 ADBN 提取的深度特征作为 CGAN 的条件输入  $y$ ,建立卷积深度生成神经网络,与传统的随机噪声相比,这些深层特征能为 GAN 提供更稳定的初始状态.我们称这种深度生成神经网络为基于辅助深度特征的条件生成对抗网络(conditional GAN with auxiliary deep features,简称 ACGAN),其模型结构如图 6 所示.

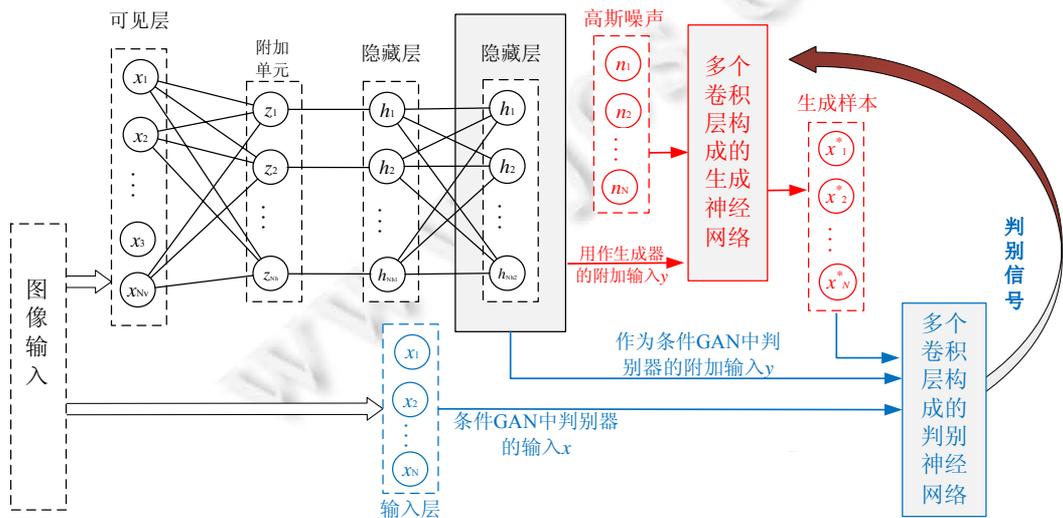


Fig.6 Structure of an ACGAN  
图 6 ACGAN 模型的结构图

图 6 是 ACGAN 的模型结构图,模型的训练分为两部分:首先,根据输入图像训练一个 ADBN 模型,在训练好的 ADBN 模型基础上,将输入图像传递给 ADBN 得到特征;该特征用作 CGAN 的附加输入  $y$  传递到 CGAN 的生成器和判别器,同时,图像输入也用作 CGAN 判别器中的真实样本.模型的目标函数可以表示为

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}} [\log D(x | ADBN(x)) + E_{z \sim p_z} [\log(1 - D(G(z | ADBN(x))))]] \quad (17)$$

其中,  $x$  来自于数据集.  $ADBN(x)$  是由 ADBN 得到的特征,该特征作为 CGAN 的附加输入.接下来,本文通过实验验证 ARBM, ADBN 和 ACGAN 的有效性.

### 3 实验分析

在实验中,我们尝试验证基于 ARBM 的模型在图像生成和图像重构任务中的有效性.实验主要分为两部分:一部分是图像重构,另一部分是图像生成.为了验证 ARBM 能够学习数据分布的流形结构,我们首先在 3 个人工数据集上进行对比实验,其中使用的基于 RBM 的模型具有完全相同的网络结构.结果如图 7 所示.

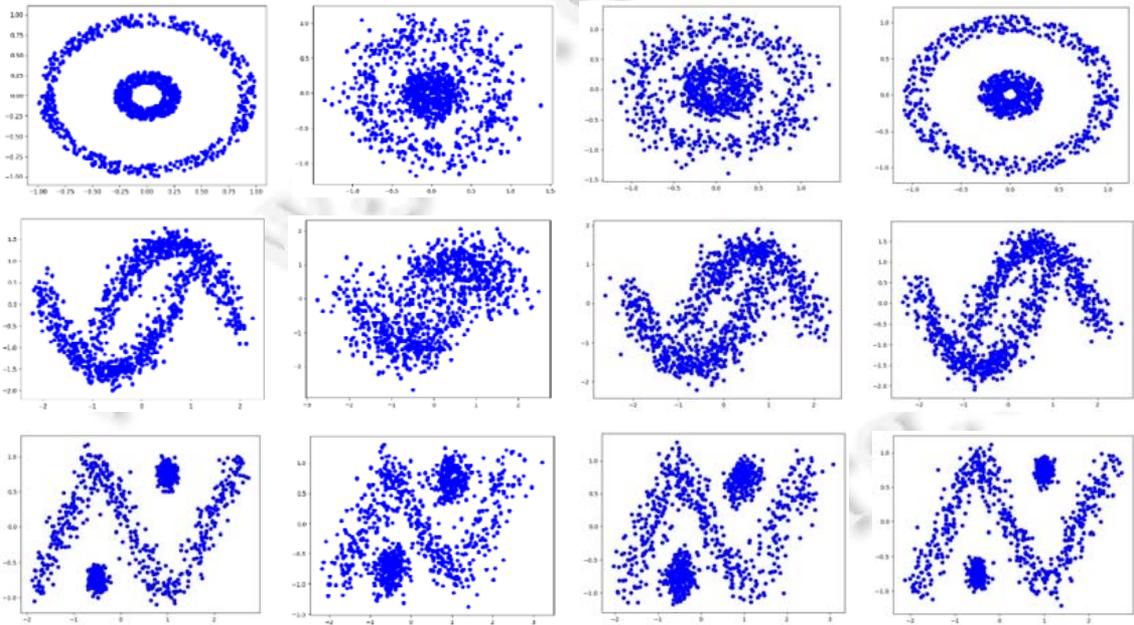


Fig.7 Comparison of ARBM with other models on three artificial datasets

图 7 3 个人工数据集上 ARBM 与其他模型比较

图 7 由 4 列样本组成,其中,

- 第 1 列样本是原始数据;
- 第 2 列样本是由 Gaussian-binary RBM 重构得到的,Gaussian-binary RBM 是一种可见层单元为实值、隐藏层单元为二值的 RBM 模型;
- 第 3 列样本由受限截断高斯图模型(RTGMM)重构得到.RTGMM 是在 2017 年提出的一种无向概率图,其可见层单元和隐藏层单元都是实值的.不同于 ARBM,RTGMM 由一个可见层和一个隐藏层组成,其可见层单元和隐藏层单元是直接连接的;
- 最后一列样本由 ARBM 重构得到.

如图 7 所示,标准的 Gaussian-binary RBM 更倾向于把模型密度散布在其支撑数据集上.而不同于 Gaussian-binary RBM,ARBM 能够学习这 3 个数据集的分布特征,并识别出数据分布上的流形结构.

接下来,本实验在真实数据集上测试 ARBM 和 ADBN 的图像重构能力.数据集有 MNIST,small

Norb, Fashion 和 Cifar-10 数据集的属性见表 1.

Table 1 Attributes of data sets

表 1 数据集的属性

数据集	训练样本	测试样本	属性	标签
MNIST	50 000	10 000	28×28	10
small Norb	24 300	24 300	32×32	5
Fashion	50 000	10 000	28×28	10
Cifar-10	50 000	10 000	32×32	10

MNIST 数据集是由 250 名来自不同人群的实验人员完成的手写数字数据集,其中包括 MNIST 基本数据集、MNIST back-rand 数据集和 MNIST back image 数据集等.Norb 是一个由灰度化的双视角玩具图像组成的数据集,包含 5 类玩具(人、动物、汽车、飞机、卡车),在不同的光照条件下,由相机系统从不同的角度进行成像.本节实验将两幅立体图像的原始分辨率从  $108 \times 108 \times 2$  降低到  $32 \times 32 \times 2$ ,以加快实验速度.Fashion 数据集包含了 10 个类别的图像,分别是 t-shirt, trouser, pullover, dress, coat, sandal, shirt, sneaker, bag, ankle boot. Cifar-10 是常用于目标识别的计算机视觉数据集,它是 8 000 万个微型图像数据集的一个子集,由 60 000 个  $32 \times 32$  分辨率的彩色图像组成,包含 10 对象类,每个类有 6 000 个图像.

本实验在 MNIST, small Norb 和 Cifar-10 上测试了这些模型的图像重构能力,原始图像如图 8 所示.

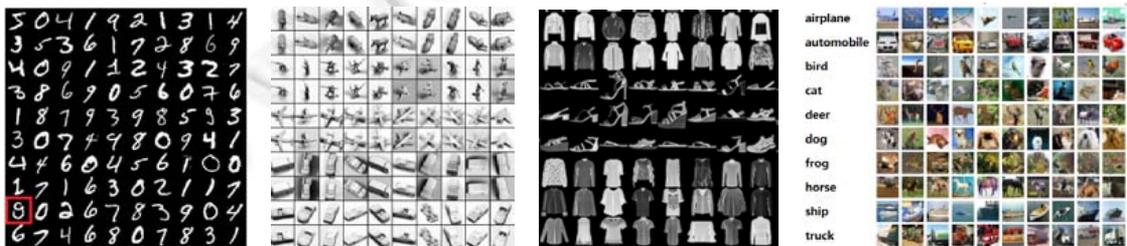


Fig.8 Original images of MNIST, small Norb, and Cifar-10

图 8 MNIST, small Norb 和 Cifar-10 的原始图像

与 GAN 不同, ARBM 和 ADBN 的模式塌陷问题并不严重.然而,似然函数过平滑的问题阻碍了模型生成锐利的图像.为了保持多模态特征并产生清晰的图像,本文将提取的 ADBN 特征作为 CGAN 的条件输入,测试了 ACGAN 的生成能力.第 1 个实验在 MNIST 数据集上测试了所提出的 ARBM, ADBN 的图像重构能力以及 ACGAN 的图像生成能力.所用的 ARBM 具有 1 000 个隐藏层单元,重构和生成的图像如图 9 所示.



Fig.9 Reconstructed images and generated images on MNIST

图 9 在 MNIST 上的重构图像和生成的图像

在图 9 中,左数第 1 张图像是由 ARBM 重构的图像,第 2 张图像是由 ADBN 重构的图像.如图 9 所示,本文提出的 ARBM 算法能够提取手写体图像的边缘特征,得到平滑的图像.此外,ADBN 重构的图像相比于 ARBM 而言与原始图像更为相似.第 3 张图像是由深度卷积生成对抗网络(deep convolutional generative adversarial network,简称 DCGAN)生成的<sup>[17]</sup>,DCGAN 是一种引入了深度卷积结构的深度对抗生成网络,其目标函数是最小

化数据分布和模型分布之间的 JS 散度.第 4 张图像是由 WGAN-GP<sup>[18]</sup>生成的,WGAN-GP 将 JS 散度替换为 Wasserstein 距离并优化求解方式,从而解决了 JS 散度中距离度量问题,是目前非常常用的图像生成算法.第 5 张图像是由 ACGAN 生成,可以看出,ACGAN 生成的图像在细节上更接近原始图像.

接下来,本文测试了所提出的模型在 small Norb 数据集上的图像重构能力和图像生成能力.实验中,ACGAN 中的生成器有 2 个全连接层和 2 个反卷积层,判别器有 2 个卷积层和 2 个全连接层.生成器和判别器均采用 leakyReLU 激活函数以及 Batch-Norm.DCGAN,LSGAN 和 WGAN-GP 具有与 ACGAN 相同的网络结构.生成器和判别器的初始学习率为  $8e-5$ .图 10 展示了模型在 small Norb 上的重构图像.

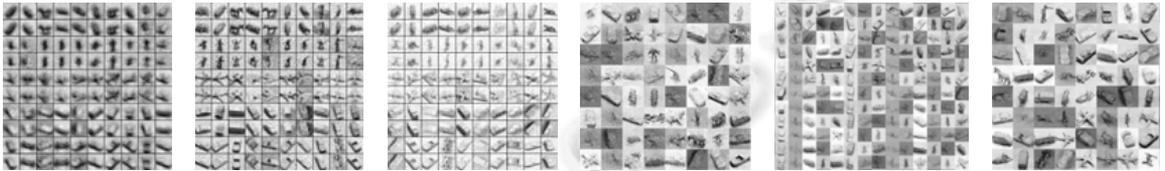


Fig.10 Reconstructed images and generated images on small Norb

图 10 small Norb 上的重构图像和生成图像

在图 10 中,左数第 1 张图像由 RTGMM 重构得到.RTGMM 可以看作是一种实值单元的 RBM 模型,能够很好地结合到神经网络中.第 2 张图像由 ARBM 重构得到,第 3 张图像由 ADBN 重构得到.可以看出:相比于传统的 RBM 方法,ARBM 和 ADBN 在 small Norb 数据集上重构的图像更清晰.第 4 张图像是由 DCGAN 生成的,第 5 张图像是由 WGAN-GP 生成的,最后一张图像由 ACGAN 生成.可以看出,ACGAN 生成的图像相对于其他常用的以 GAN 为基础的生成模型更为清晰、更接近于原始的图像.接下来,我们测试了模型在 Fashion 数据集上的图像重构和图像生成能力,得到的图像如图 11 所示.



Fig.11 Reconstructed images and generated images on Fashion

图 11 Fashion 上的重构图像和生成图像

在图 11 中,第 1 张图像是由 ADBN 重构得到的.第 2 张图像由 DCGAN 生成.第 3 张图像由 LSGAN 生成,LSGAN 是一种深度对抗生成网络,其中,目标函数被替换为均方差的形式.第 4 张图像由 WGAN-GP 生成.最后一张图像由 ACGAN 生成.可以看出,ACGAN 生成的 Fashion 图像更清晰、更接近原始图像.

Cifar-10 对于  $32 \times 32$  分辨率的图像生成任务而言是困难的,因为 Cifar-10 是彩色的,其背景也更复杂.特别是对于没有卷积结构的浅层网络(如 ARBM),很难生成 Cifar-10 数据.本节尝试使用两个隐藏层的 ADBN 来学习 Cifar-10 数据集.在 ADBN 中,第 1 隐藏层有 5 000 个单元,第 2 隐藏层有 1 500 个单元.ADBN 重构的图像和生成的图像如图 12 所示.

如图 12 所示:尽管 ADBN 提取输入图像能够可视化边缘特征,但生成的图像过于平滑和模糊.我们认为,导致图像模糊的主要原因是卷积运算难以加入 RBM 模型.因为在 RBM 训练过程中,卷积和池化运算是不可逆的.为了建立一个有效的基于 ADBN 的深度卷积生成神经网络,我们利用 ADBN 提取的深度特征作为 CGAN 的条件输入数据提出 ACGAN 模型,我们将 ACGAN 生成的图像与现有模型(例如 DCGAN,WGAN-GP 和 Real-NVP 模型)生成的图像进行比较<sup>[19]</sup>.Real-NVP 模型是一个非对抗性可逆生成神经网络.在本实验中,为了生成彩色图

像,我们首先利用 ADBN 模型生成 Cifar-10 图像,然后通过 max-pooling 操作将生成的图像下采样到  $16 \times 16, 8 \times 8$  和  $4 \times 4$  分辨率的特征图像.这些特征被用作不同分辨率 ACGAN 的条件输入.从图 12 可以看出:尽管 ACGAN 与其他模型相比生成相对清晰的图像,但是生成图像的颜色和部分轮廓受到条件输入的影响,并且所有生成的图像都不够清晰.如何在 Cifar-10 上生成清晰的图像,仍然是我们未来研究的重点.最后,我们定量分析了 ACGAN 模型生成图像的质量,所使用的指标是 FID,结果见表 2.



Fig.12 Generated images of DBM, real NVP, DCGAN, WGAN-GP, and ACGAN models on cifar-10 dataset

图 12 DBM, Real NVP, DCGAN, WGAN-GP, 以及 ACGAN 模型在 Cifar-10 数据集上的生成图像

**Table 2** FID of ACGAN and commonly used generative models

表 2 ACGAN 和常用生成模型的 FID

数据集	LSGAN	WGAN	WGAN-GP	ACGAN
MNIST	7.8±0.6	6.7±0.4	6.4±0.4	6.2±0.2
small Norb	47.2±2.4	39.7±2.4	39.1±2.3	38.2±2.1
Fasion	30.7±2.2	21.5±1.6	20.4±1.5	20.2±1.0
Cifar-10	53.9±2.8	41.3±2.0	40.7±1.8	41.1±1.6

从表 2 可以看出:相较于常用的生成模型,ACGAN 生成的图像具有相对更低的 FID.

#### 4 总结和展望

本文将辅助变量和流形正则项引入到能量函数中,并使用实值隐藏层单元来模拟可见单位的概率.我们验证了 ARBM 和 ADBN 在图像生成任务中的有效性.为了融合卷积深度神经网络生成清晰锐利的图像,本文提出了一种 ACGAN 模型.实验验证了该模型的有效性.但是,ARBM 仍然存在一些问题,传统的算法(PCD 算法和改变 Gibbs 采样)在训练 ARBM 时是非常耗时的.目前,无向图的训练算法主要有变分法、马尔可夫链和基于对抗损失的训练方法.优化无向图的训练算法仍然是我们未来研究的重点.

#### References:

- [1] Kingma D, Welling M. Auto-encoding variational bayes. In: Proc. of the Int'l Conf. on Learning Representations. 2014.
- [2] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks. In: Proc. of the Advances in Neural Information Processing Systems. 2014.
- [3] Kuleshov V, Ermon S. Neural variational inference and learning in undirected graphical models. In: Proc. of the Advances in Neural Information Processing Systems. 2017.
- [4] Fisher C, Smith A, Walsh J. Boltzmann encoded adversarial machines. arXiv: 1804.08682, arXiv, 2018.
- [5] Ranzato M, Krizhevsky A, Hinton G. Factored 3-way restricted Boltzmann machines for modeling natural images. Journal of Machine Learning Research, 2010,9:621–628.
- [6] Courville A, Desjardins G, Bergstra J, et al. The spike-and-slab RBM and extensions to discrete and sparse data distributions. IEEE Trans. on Pattern Analysis & Machine Intelligence, 2014,36(9):1874–1887.
- [7] Sohn K, Zhou G, Lee C, et al. Learning and selecting features jointly with point-wise gated Boltzmann machines. In: Proc. of the Int'l Conf. on Machine Learning. 2013.
- [8] Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 2014,15(1):1929–1958.

- [9] Blundell C, Cornebise J, Kavukcuoglu K. Weight uncertainty in neural networks. In: Proc. of the Int'l Conf. on Machine Learning. 2015.
- [10] Zhang N, Ding S, Zhang J, *et al.* Research on point-wise gated deep networks. Applied Soft Computing, 2017,52:1210–1221.
- [11] Huang H, Toyozumi T. Advanced mean-field theory of the restricted Boltzmann machine. Physical Review E Statistical Nonlinear & Soft Matter Physics, 2015.
- [12] Zhang N, Ding S, Zhang J, *et al.* An overview on restricted Boltzmann machines. Neurocomputing, 2018,275:1186–1199.
- [13] Zhang J, Ding S, Zhang N. An overview on probability undirected graphs and their applications in image processing. Neurocomputing, 2018,321:156–168.
- [14] Isola P, Zhu J, Zhou T, *et al.* Image-to-image translation with conditional adversarial networks. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2017.
- [15] Cho K, Raiko T, Ilin A. Gaussian-bernoulli deep Boltzmann machine. In: Proc. of the IEEE Int'l Joint Conf. on Neural Networks. 2014.
- [16] Su Q, Liao X, Chen C, *et al.* Nonlinear statistical learning with truncated gaussian graphical models. In: Proc. of the Int'l Conf. on Machine Learning. 2016.
- [17] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv: 1511.06434, arXiv, 2018.
- [18] Gulrajani I, Ahmed F, Arjovsky M, *et al.* Improved training of wasserstein GANs. In: Advances in Neural Information Processing Systems. 2017.
- [19] Dinh L, Sohl-Dickstein J, Bengio S. Density estimation using Real NVP. In: Proc. of the Int'l Conf. of Learning Research. 2016.



张健(1990—),男,博士,讲师,CCF 专业会员,主要研究领域为机器学习,模式识别.



丁玲(1994—),女,讲师,主要研究领域为机器学习,数据挖掘.



丁世飞(1963—),男,博士,教授,博士生导师,CCF 杰出会员,主要研究领域为智能信息处理,人工智能与模式识别,机器学习与数据挖掘,粗糙集与软计算,大数据分析 with 云计算.



张成龙(1992—),男,博士生,主要研究领域为机器学习,模式识别.