

## 基于噪声数据与干净数据的深度置信网络\*

张楠<sup>1</sup>, 丁世飞<sup>1,2,3</sup>, 张健<sup>1</sup>, 赵星宇<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(中国矿业大学), 江苏 徐州 221116)

<sup>3</sup>(中国科学院 计算技术研究所 智能信息处理重点实验室, 北京 100190)

通讯作者: 丁世飞, E-mail: dingsf@cumt.edu.cn

**摘要:** 建立以受限玻尔兹曼机(restricted Boltzmann machine, 简称 RBM)为基石的深度网络模型, 是深度学习研究的热点领域之一. Point-wise Gated 受限玻尔兹曼机(point-wise gated RBM, 简称 pgRBM)是一种 RBM 的变种算法. 该算法能够在含噪声的数据中自适应地找到数据中与分类有关的部分, 从而实现较好的分类结果. 假设一组数据中有噪声数据和干净数据, 如何应用不含噪声的数据提升 pgRBM 的性能, 是一个重要的研究问题. 针对这一问题, 首先, 在传统的 pgRBM 基础上提出一种基于随机噪声数据与干净数据的 Point-wise Gated 受限玻尔兹曼机(pgRBM based on random noisy data and clean data, 简称 pgrncRBM)方法, 其网络中与分类有关权值的初值是通过不含噪声的数据学习得到的, 所以 pgrncRBM 在处理随机噪声数据时可以学习到更为“干净”的数据. 在 pgrncRBM 中, 与分类有关的数据与噪声都是使用 RBM 建模. 如果噪声是图片, pgrncRBM 就不能很好地去除噪声. Spike-and-Slab RBM(ssRBM)是一种处理实值数据的 RBM 变种模型, 其定义两种不同类型的隐层用来学习实值数据的分布特性. 因此, 将 ssRBM 与 pgRBM 相结合, 提出一种基于图像噪声数据与干净数据的 Point-wise Gated 受限玻尔兹曼机(pgRBM based on image noisy data and clean data, 简称 pgincRBM)方法. 该方法使用 ssRBM 对噪声建模, 其在处理图像噪声数据时可以学习到更为“干净”的数据. 然后, 通过堆叠 pgrncRBM、pgincRBM 和传统的 RBM 构建出深度网络模型, 并探讨了权值不确定性方法在提出网络模型中的可行性. 最后, 在含噪声的手写数据集上进行 MATLAB 仿真实验. 实验结果表明, pgrncRBM 和 pgincRBM 都是有效的神经网络学习方法.

**关键词:** 受限玻尔兹曼机; 深度置信网络; 权值不确定性; 特征选择

**中图法分类号:** TP183

中文引用格式: 张楠, 丁世飞, 张健, 赵星宇. 基于噪声数据与干净数据的深度置信网络. 软件学报, 2019, 30(11): 3326-3339. <http://www.jos.org.cn/1000-9825/5574.htm>

英文引用格式: Zhang N, Ding SF, Zhang J, Zhao XY. Deep belief network based on noisy data and clean data. Ruan Jian Xue Bao/Journal of Software, 2019, 30(11): 3326-3339 (in Chinese). <http://www.jos.org.cn/1000-9825/5574.htm>

## Deep Belief Network Based on Noisy Data and Clean Data

ZHANG Nan<sup>1</sup>, DING Shi-Fei<sup>1,2,3</sup>, ZHANG Jian<sup>1</sup>, ZHAO Xing-Yu<sup>1</sup>

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

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

<sup>3</sup>(Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China)

\* 基金项目: 国家自然科学基金(61672522, 61379101); 国家重点基础研究发展计划(973)(2013CB329502)

Foundation item: National Natural Science Foundation of China (61672522, 61379101); National Program on Key Basic Research Project of China (973) (2013CB329502)

收稿时间: 2017-10-23; 修改时间: 2017-12-25; 采用时间: 2018-04-02

**Abstract:** Stacking restricted Boltzmann machines (RBM) to create deep networks, such as deep belief networks (DBN), has become one of the most important research fields in deep learning. Point-wise gated restricted Boltzmann machines (pgRBM), an RBM variant, can effectively find the task-relevant patterns from data containing irrelevant patterns and thus achieves satisfied classification results. Given that train data is composed of noisy data and clean data, how the clean data is applied to promote the performance of the pgRBM is a problem. To address the problem, this study first proposes a method, named as pgRBM based on random noisy data and clean data (pgrncRBM). The pgrncRBM makes use of RBM and the clean data to obtain the initial values of the task-relevant weights, so it can learn the “clean” data from the data containing random noisy. In the pgrncRBM, the general RBM is used to pre-train the weights of task-relevant patterns from data and irrelevant patterns. If the noise is an image, the pgrncRBM cannot learn the task-relevant patterns from the noisy data. Spike-and-Slab RBM, an RBM variant, uses two types of hidden layers to determine the mean and covariance of each visible unit. Therefore, this study combines ssRBM with pgRBM and proposes a method, named as pgRBM based on image noisy data and clean data (pgincRBM). The pgincRBM uses the ssRBM to model the noise, so it can learn the “clean” data from the data containing image noisy. And then, this study stacks pgrncRBM, pgincRBM, and RBMs to create deep networks, and discusses the feasibility that the weight uncertainty method is developed to prevent overfitting in the proposed networks. Experimental results on MNIST variation datasets show that pgrncRBM and pgincRBM are effective neural networks learning methods.

**Key words:** restricted Boltzmann machine (RBM); deep belief network; weight uncertainty; feature selection

玻尔兹曼机(Boltzmann machine,简称 BM)是一种概率图模型.该模型能够学习到输入数据的统计特性<sup>[1]</sup>.Smolensky 在 BM 的基础上提出了受限玻尔兹曼机(restricted Boltzmann machine,简称 RBM)<sup>[2]</sup>.RBM 是一种产生式无监督学习方法,它不仅能够从数据中学习到有效的特征,而且能够利用学习到的特征重构出原始数据.2006 年,Hinton 等人在《SCIENCE》上提出一种有效的深度网络学习方法——深度置信网(deep belief network,简称 DBN),掀起了深度学习的研究热潮<sup>[3]</sup>.其实,以受限玻尔兹曼机为基石的深度网络模型除了深度置信网以外,还有深度 Sigmoid 置信网<sup>[4]</sup>、深度玻尔兹曼机<sup>[5]</sup>等.

与 RBM 相比,以 RBM 为基石的深度网络模型会表现出较好的分类能力.但是在处理噪声图像时,这些深度网络并没有展现出优于 RBM 的学习能力.如果数据中存在噪声,那么 RBM 学习到的特征也会蕴含噪声信息.特征选择方法可以从蕴含噪声信息的特征中找到与分类有关的特征<sup>[6]</sup>.Sohn 等人结合受限玻尔兹曼机与特征选择方法提出了 Point-wise Gated 受限玻尔兹曼机(point-wise gated RBM,简称 pgRBM)<sup>[7]</sup>.pgRBM 不仅能够得到与分类有关的特征,还能在噪声数据中找到数据中与分类有关的部分.Zhang 等人结合 pgRBM 提出了 Point-wise Gated 深度置信网(point-wise gated deep belief network,简称 pgDBN)和 Point-wise Gated 深度玻尔兹曼机<sup>[8]</sup>.这表明以 RBM 为基石的深度网络在处理噪声时,传统的 RBM 学习到的特征中蕴含的噪声信息影响了深度网络的性能.但是当一组数据中有噪声数据和干净数据时,如何应用干净数据提升 pgRBM 在噪声数据中的学习能力,从而进一步提高深度网络模型处理噪声的能力,这是一个重要的研究问题.

本文在传统的 pgRBM 基础上提出一种基于随机噪声数据与干净数据的 Point-wise Gated 受限玻尔兹曼机(pgRBM based on random noisy data and clean data,简称 pgrncRBM)方法.它可以利用干净数据提升算法在随机噪声数据上的学习能力.pgRBM 把隐层节点分为与分类有关的和与分类无关的两个部分,其连接权值的初值是用特征选择的方法对 RBM 学习的权值处理得到的.pgrncRBM 就是在传统的 pgRBM 基础上对 pgRBM 学习到的数据二次去噪,其与分类无关的隐层节点相连权值的初值是用特征选择的方法对 RBM 对一次降噪的数据学习到的权值处理得到的,但是其与分类有关的隐层节点相连权值的初值是用 RBM 对不含噪声的数据学习得到的.这样,pgrncRBM 在处理随机噪声数据时可以学习到更为“干净”的数据.如果噪声是图片时,pgrncRBM 不能很好地去除噪声.传统的 RBM 一般只适于处理二值图像.为了更好地处理实值图像,提出了一系列的 RBM 的变种算法,如均值与协方差受限玻尔兹曼机(mcRBM)<sup>[9]</sup>和 Spike-and-Slab 受限玻尔兹曼机(ssRBM)<sup>[10]</sup>.mcRBM 和 ssRBM 都表现出较好的实质图像数据建模能力,但不同的是,当隐层节点状态保持固定时,mcRBM 求取数据在可见层的联合概率分布用的是混合蒙特卡洛算法(HMC),而 ssRBM 采用的是简单而有效的 Gibbs 采样方法.因此,本文将 ssRBM 与 pgRBM 相结合,提出了一种基于图像噪声数据与干净数据的 Point-wise Gated 受限玻尔兹曼机(pgRBM based on image noisy data and clean data,简称 pgincRBM)方法.该方法使用 ssRBM 对噪声建模,其

在处理图像噪声数据时可以学习到更为“干净”的数据.然后,本文可以在 pgrncRBM 和 pgincRBM 的基础上堆叠出深度置信网络模型,如基于随机噪声数据与干净数据的 Point-wise Gated 受限玻尔兹曼机(pgDBN based on random noisy data and clean data,简称 pgrncDBN)以及基于图像噪声数据与干净数据的 Point-wise Gated 受限玻尔兹曼机(pgDBN based on image noisy data and clean data,简称 pgincDBN).

训练神经网络时常常会遇到过拟合问题,以 RBM 为基石的深度网络也是如此.目前,RBM 中常用的解决过拟合问题有权值衰减、Dropout<sup>[11]</sup>、Dropconnect<sup>[12,13]</sup>、权值不确定性<sup>[14-16]</sup>.Zhang 等人将权值不确定性引入 pgRBM 和 pgDBN,验证了权值不确定性在这两种网络中的有效性<sup>[8]</sup>.本文将权值不确定性引入 pgrncRBM 和 pgincRBM 中,把 pgrncRBM 和 pgincRBM 中与分类有关的权值看作一组符合高斯分布的变量而不是固定值.同样地,本文还将权值不确定性引入以 pgrncRBM 和 pgincRBM 为基石的深度网络 pgrncDBN 和 pgincDBN 中,并探究了这几种算法的可行性.

本文第 1 节简述受限玻尔兹曼机算法、深度置信网算法和 Point-wise Gated 受限玻尔兹曼机算法.第 2 节详述本文提出的基于噪声数据与干净数据的 Point-wise Gated 受限玻尔兹曼机算法和以它们为基石的 Point-wise Gated 深度置信网算法.第 3 节介绍权值不确定性方法在 pgncRBM 和 pgncDBN 两种网络模型中的应用.第 4 节用含噪声的手写数据集测试本文提出的算法的性能.实验结果表明,pgrncRBM 和 pgincRBM 都是有效的神经网络学习方法.

## 1 相关工作

### 1.1 受限玻尔兹曼机与深度置信网算法

受限玻尔兹曼机(RBM)是一种生成式随机网络,由可见层和隐层组成.RBM 网络的权值  $\theta$  由可见层和隐层的连接权值矩阵  $\mathbf{W}=(W_{ij}) \in R^{D \times J}$ 、可见层的偏置向量  $\mathbf{c} = \{c_i\}_{i=1}^D$  和隐层的偏置向量  $\mathbf{b} = \{b_j\}_{j=1}^J$  组成.当给定一组可见层状态  $\mathbf{v} = \{v_i\}_{i=1}^D$  和隐层状态  $\mathbf{h} = \{h_j\}_{j=1}^J$  时,RBM 的能量函数和似然函数分别表述为

$$E(\mathbf{v}, \mathbf{h}; \theta) = -\sum_{i=1}^D \sum_{j=1}^J v_i W_{ij} h_j - \sum_{j=1}^J b_j h_j - \sum_{i=1}^D c_i v_i \quad (1)$$

$$P(\mathbf{v}; \theta) = \sum_{\mathbf{h}} P(\mathbf{v}, \mathbf{h}; \theta) = \frac{1}{Z(\theta)} \sum_{\mathbf{h}} \exp(-E(\mathbf{v}, \mathbf{h}; \theta)) \quad (2)$$

其中,  $v_i \in \{0, 1\}; h_j \in \{0, 1\}; Z(\theta) = \sum_{\mathbf{v}} \sum_{\mathbf{h}} \exp(-E(\mathbf{v}, \mathbf{h}; \theta))$  是配分函数;  $P(\mathbf{v}; \theta)$  对应  $P(\mathbf{v}, \mathbf{h}; \theta)$  的边缘分布,又称为似然函数.当可见层和隐层其中之一状态固定时,RBM 的条件概率分布可以表述为

$$P(v_i = 1 | \mathbf{h}) = \sigma \left( \sum_j W_{ij} h_j + c_i \right) \quad (3)$$

$$P(h_j = 1 | \mathbf{v}) = \sigma \left( \sum_i v_i W_{ij} + b_j \right) \quad (4)$$

其中,  $\sigma(x) = \frac{1}{1 + \exp(-x)}$ . 给定训练样本集  $\mathbf{X} = \{\mathbf{v}^k\}_{k=1}^N$  时,通过使对数似然函数  $\sum_{k=1}^N \ln P(\mathbf{v}^k)$  最大化,RBM 用梯度上升算法去更新网络的权值  $\theta$  并且,我们可以用对比散度(CD)<sup>[17]</sup>等算法去估计这些梯度.

深度置信网(DBN)是一个概率生成模型.它是由 RBM 堆叠构成的深度网络模型.首先,DBN 通过利用不带标签数据,用 RBM 算法自底向上逐层训练得到深度网络的初值.在预训练之后,DBN 通过利用带标签数据,用 BP 算法对网络的权值进行调整.

### 1.2 Point-wise Gated受限玻尔兹曼机算法

与 RBM 不同,pgRBM 把隐层节点分为与分类有关的和与分类无关的两个部分.pgRBM 的网络结构如图 1 所示.此时,pgRBM 的能量函数可表述为

$$\left. \begin{aligned} E(\mathbf{v}, \mathbf{z}, \mathbf{h}^1, \mathbf{h}^2; \theta) &= -\sum_{i=1}^D \sum_{j=1}^J \sum_{r=1}^2 (z_i^r v_i) W_{ij}^r h_j^r - \sum_{j=1}^J \sum_{r=1}^2 b_j^r h_j^r - \sum_{i=1}^D \sum_{r=1}^2 c_i^r (z_i^r v_i) \\ \text{s.t. } \sum_{r=1}^2 z_i^r &= 1, i = 1, \dots, D \end{aligned} \right\} \quad (5)$$

其中,与分类有关的隐层  $\mathbf{h}^1$  对应的权值为  $\{\mathbf{W}^1, \mathbf{c}^1, \mathbf{b}^1\}$ ,与分类无关的隐层对应的权值为  $\{\mathbf{W}^2, \mathbf{c}^2, \mathbf{b}^2\}$ .

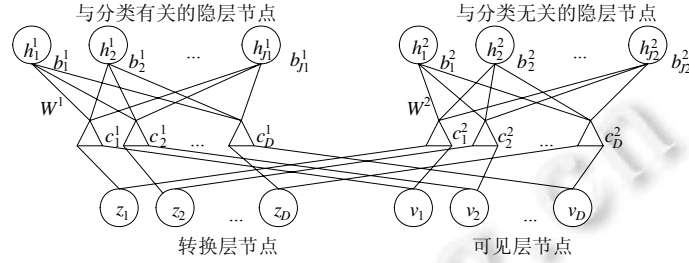


Fig.1 Network structure of pgRBM

图1 pgRBM 的网络结构

当可见层、转换层和隐层任意两层状态固定时,pgRBM 的条件概率分布可以表述为

$$P(h_i^r = 1 | \mathbf{z}, \mathbf{v}) = \sigma \left( \sum_i (z_i^r v_i) W_{ij}^r + b_j^r \right) \quad (6)$$

$$P(v_i = 1 | \mathbf{z}, \mathbf{h}^1, \mathbf{h}^2) = \sigma \left( \sum_r z_i^r \left( \sum_j W_{ij}^r h_j^r + c_i^r \right) \right) \quad (7)$$

$$P(z_i^r = 1 | \mathbf{v}, \mathbf{h}^1, \mathbf{h}^2) = \frac{\exp \left( v_i \left( \sum_j W_{ij}^r h_j^r + c_i^r \right) \right)}{\sum_s \exp \left( v_i \left( \sum_j W_{ij}^s h_j^s + c_i^s \right) \right)} \quad (8)$$

其中,  $\sigma(x) = \frac{1}{1 + \exp(-x)}$ .文献[7]发现,pgRBM 用 RBM 学习的结果对网络中的参数赋初值会比随机赋初值取得更好的结果.如图 1 所示,与分类有关的隐层和与分类无关的隐层的节点数分别为  $J_1$  和  $J_2$ ,当用 RBM 的结果赋初值时,RBM 的隐层节点数要略大于  $J_1$  与  $J_2$  之和.RBM 学习得到的参数后,需要用特征选择方法,例如 t-test 来判断隐层节点对分类的贡献,从而确定 pgRBM 网络权值的初始值.

Zhang 等人将权值不确定性引入 pgRBM,提出了权值不确定性 Point-wise Gated 受限玻尔兹曼机(weight uncertainty pgRBM,简称 pgwRBM)算法.与 pgRBM 不同的是,pgwRBM 将与分类有关的隐层与可见层、转换层的连接权值  $\mathbf{W}^1$  看作符合高斯分布的变量,其均值和标准差分别是  $\boldsymbol{\mu}^1$  和  $\boldsymbol{\sigma}^1 = \log(\mathbf{1} + \exp(\boldsymbol{\rho}^1))$ .此时,pgwRBM 的能量函数可表述为

$$\left. \begin{aligned} E(\mathbf{v}', \mathbf{z}', \mathbf{h}^1, \mathbf{h}^2; \theta) &= -\sum_{i=1}^D \sum_{j=1}^J (z_i^1 v_i) (\mu_{ij}^1 + \log(1 + \exp(\rho_{ij}^1))) \odot \epsilon_{ij}^1 h_j^1 - \sum_{i=1}^D \sum_{j=1}^J (z_i^2 v_i) W_{ij}^2 h_j^2 - \sum_{j=1}^J \sum_{r=1}^2 b_j^r h_j^r - \sum_{i=1}^D \sum_{r=1}^2 c_i^r (z_i^r v_i) \\ \text{s.t. } \sum_{r=1}^2 z_i^r &= 1, i = 1, \dots, D \end{aligned} \right\} \quad (9)$$

其中,  $\odot$  表示矩阵中的元素对位相乘.

pgwRBM 的学习过程详见文献[8].和 RBM 一样,pgRBM 和 pgwRBM 也可以使用 CD 等算法去计算权值的梯度,从而更新网络的权值.

## 2 基于噪声数据与干净数据的 Point-wise Gated 深度置信网络模型

### 2.1 基于随机噪声数据与干净数据的 Point-wise Gated 受限玻尔兹曼机算法

当数据中存在少量噪声时,添加噪声在某些情况下可以增加分类器的泛化能力.但是,如果数据中存在大量噪声,噪声的添加就降低了训练数据的分类准确率,并且还不能获得更好的泛化能力.Point-wise Gated 受限玻尔兹曼机就是针对这种情况提出来的.它能够在噪声数据中找到数据中与分类有关的部分,从而提升分类器的分类准确率.但是,当一组训练数据中同时出现噪声数据和干净数据时,如何应用干净数据提升传统的 pgRBM 在噪声数据上的学习能力,是一个重要的研究问题.针对这一问题,本文首先在传统的 pgRBM 基础上提出一种基于随机噪声数据与干净数据的 Point-wise Gated 受限玻尔兹曼机方法.它可以利用不含噪声的数据提升算法在噪声数据上的学习能力.和传统的 pgRBM 一样,pgRncRBM 的网络结构同样可以用图 1 来表示.当给定一组可见层状态和隐层状态时,pgRncRBM 的能量函数也可以表述为

$$\left. \begin{aligned} E(\mathbf{v}', \mathbf{z}', \mathbf{h}^1, \mathbf{h}^2; \theta) &= -\sum_{i=1}^D \sum_{j=1}^J \sum_{r=1}^2 (z_i^{r'} v_i') W_{ij}^r h_j^r - \sum_{j=1}^J \sum_{r=1}^2 b_j^r h_j^r - \sum_{i=1}^D \sum_{r=1}^2 c_i^r (z_i^{r'} v_i') \\ \text{s.t. } \sum_{r=1}^2 z_i^{r'} &= 1, i = 1, \dots, D \end{aligned} \right\} \quad (10)$$

其中,  $\mathbf{v}'$  是 pgRBM 对噪声数据  $\mathbf{v}$  处理之后得到的与分类有关的数据<sup>[7]</sup>.我们可以看出,pgRncRBM 就是在传统的 pgRBM 基础上对 pgRBM 学习到的数据二次去噪.并且,pgRncRBM 网络权值的赋初值的方法也与传统的 pgRBM 不同.pgRncRBM 同样把隐层节点分为与分类有关的和与分类无关的两个部分,那么其与分类有关的隐层相连接权值的初值是用 RBM 对不含噪声的数据学习得到的,并且其与分类无关的隐层相连接权值的初值还是用特征选择的方法对 RBM 对一次降噪的数据  $\mathbf{v}'$  学习到的权值处理得到的.pgRncRBM 预训练完成后,通过利用带标签的噪声数据与 BP 算法对网络的权值进行调整.在 pgRBM 中,网络的输入数据是 pgRBM 对噪声数据  $\mathbf{v}$  处理之后得到的“干净”数据,输出是与数据对应的标签.同理,在 pgRncRBM 中,网络的输入数据是 pgRncRBM 对一次降噪的数据  $\mathbf{v}'$  处理之后得到的更为“干净”的数据,输出是与数据对应的标签.

### 2.2 基于图像噪声数据与干净数据的 Point-wise Gated 受限玻尔兹曼机算法

pgRncRBM 在处理随机噪声数据时可以学习到更为“干净”的数据,但当噪声是图片时,它不能很好地去除噪声.针对这一问题,本文又将 ssRBM 与 pgRBM 相结合,提出一种基于图像噪声数据与干净数据的 Point-wise Gated 受限玻尔兹曼机方法.该方法使用 ssRBM 对噪声建模,其在处理图像噪声数据时可以学习到更为“干净”的数据,可以利用不含噪声的数据,提升算法在图像噪声数据上的学习能力.当给定一组可见层状态和隐层状态时,pgRncRBM 的能量函数也可以表述为

$$\left. \begin{aligned} E(\mathbf{v}', \mathbf{z}', \mathbf{h}^1, \mathbf{h}^2, \mathbf{s}^2; \theta) &= -\sum_{i=1}^D \sum_{r=1}^2 c_i^r (z_i^{r'} v_i') - \sum_{i=1}^D \sum_{j=1}^J (z_i^1 v_i') W_{ij}^1 h_j^1 - \sum_{i=1}^D \sum_{j=1}^J (z_i^2 v_i') W_{ij}^2 h_j^2 s_j^2 - \sum_{j=1}^J \sum_{r=1}^2 b_j^r h_j^r + \frac{1}{2} \sum_{j=1}^J \alpha_j (s_j^2)^2 \\ \text{s.t. } \sum_{r=1}^2 z_i^{r'} &= 1, i = 1, \dots, D \end{aligned} \right\} \quad (11)$$

其中,  $\mathbf{v}'$  是 pgRBM 对噪声数据  $\mathbf{v}$  处理之后得到的与分类有关的数据,  $\alpha_j$  是对  $\|s_j^2\|_2^2$  惩罚项的系数.我们可以看出,pgRncRBM 中与分类无关的隐层相连的可见层节点并不像 ssRBM 那样使用对角矩阵对其加以惩罚,而是使用偏置.这是因为当转换层和隐层两层状态固定时,只能推导出与分类有关的隐层相连的可见层节点被激活的概率,而不能得到其概率分布情况,这样就只能推倒得到与分类无关的隐层相连的可见层节点被激活的概率.这样,当可见层、转换层和隐层任意两层状态固定时,剩余一层节点被激活的概率(或者概率分布)可以表述为

$$P(h_j^1 = 1 | \mathbf{z}, \mathbf{v}) = \sigma \left( \sum_i (z_i^1 v_i) W_{ij}^1 + b_j^2 \right) \quad (12)$$

$$P(h_j^2 = 1 | z, v) = \sigma \left( \frac{1}{2} (\alpha_j)^{-1} \left( \sum_i (z_i^2 v_i) W_{ij}^2 \right)^2 + b_j^2 \right) \quad (13)$$

$$P(s_j^2 | z, v, h_j^2) = N \left( (\alpha_j)^{-1} h_j^2 \sum_i (z_i^r v_i) W_{ij}^2, (\alpha_j)^{-1} \right) \quad (14)$$

$$P(v_i = 1 | z, h^1, h^2, s^2) = \sigma \left( z_i^1 \left( \sum_j W_{ij}^1 h_j^1 + c_i^1 \right) + z_i^2 \left( \sum_j W_{ij}^2 h_j^2 s_j^2 + c_i^2 \right) \right) \quad (15)$$

$$P(z_i^1 = 1 | v, h^1, h^2, s^2) = \sigma \left( v_i \left( \sum_j W_{ij}^1 h_j^1 + c_i^1 \right) - v_i \left( \sum_j W_{ij}^2 h_j^2 s_j^2 + c_i^2 \right) \right) \quad (16)$$

和 pgrncRBM 一样,pgincRBM 也可以使用 CD 等算法去计算权值的梯度,从而更新网络的权值.pgincRBM 预训练完成后,通过利用带标签的噪声数据和 BP 算法对网络的权值进行调整.在 pgincRBM 中,网络的输入数据是 pgincRBM 对一次降噪的数据  $v'$  处理之后得到的更为“干净”的数据,输出是与数据对应的标签.

### 2.3 基于噪声数据与干净数据的 Point-wise Gated 深度置信网算法

在处理含噪声的图像时,DBN 并没有展现出优于 RBM 的学习能力.Zhang 等人结合 pgRBM 提出了 Point-wise Gated 深度置信网(pgDBN),并且 pgDBN 展示出优于 pgRBM 的分类能力.我们得出,以传统的 RBM 为基石的深度网络在处理噪声时,传统的 RBM 学习到的特征中蕴含的噪声信息影响了深度网络的性能.当一组数据中同时出现噪声数据和干净数据时,就可以利用干净数据进一步提高深度网络模型处理噪声的能力.因此,本文堆叠 pgrncRBM、pgincRBM、传统的 RBM,构建出两种基于噪声数据与干净数据的 Point-wise Gated 深度置信网 (pgDBN based on noisy data and clean data,简称 pgncDBN),包含基于随机噪声数据和干净数据的 Point-wise Gated 深度置信网,以及基于图像噪声数据和干净数据的 Point-wise Gated 深度置信网.以两隐层 pgrncDBN 为例,图 2 给出了 pgrncDBN 的预训练过程.首先,pgrncDBN 和 pgincDBN 分别用 pgrncRBM 或者 pgincRBM 对一次降噪的数据  $v'$  预训练,得到更为“干净”的数据;然后,它们都利用 RBM 预训练隐层间的连接权值;最后,它们都随机确定最后一层隐层与输出层的连接权值,随后用 BP 算法微调整个网络的权值.

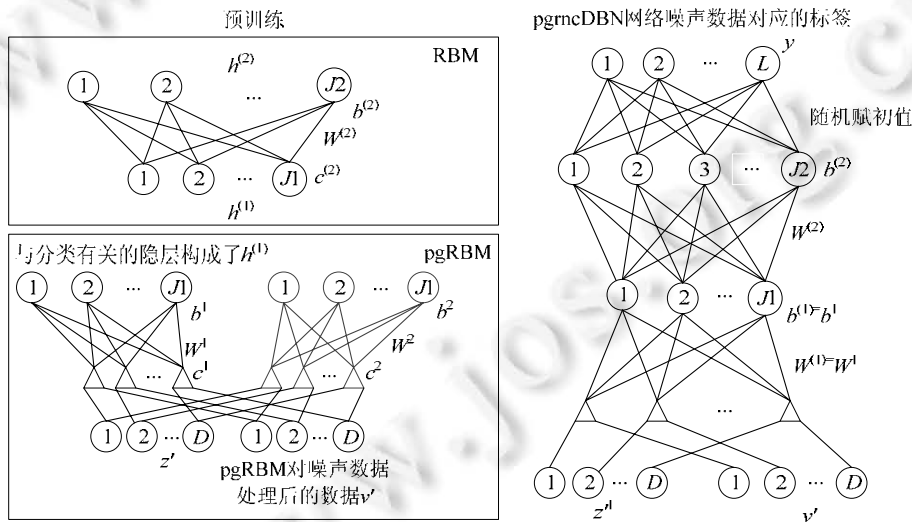


Fig.2 Pre-training process of pgrncDBN

图 2 pgrncDBN 的预训练过程

### 3 权值不确定性方法在 pgrncDBN 与 pgincDBN 深度网络中的应用

权值不确定性方法是一种神经网络中常见的解决过拟合现象的方法.它把神经网络中的每个权值看作一个可能值的概率分布,而不是以前的单一的固定值,这样学习到的特征更为鲁棒.Zhang 等人把权值不确定性方法引入 RBM,提出了权值不确定性受限玻尔兹曼机(weight uncertainty RBM,简称 wRBM),其将每个可见层与隐层间的连接权值看作一个可能值的概率分布.同时,Zhang 等人将权值不确定性引入 pgRBM,提出了权值不确定性受限玻尔兹曼机,其将与分类有关的隐层节点相连的权值看作一个可能值的概率分布.wRBM 能够有效地解决 RBM 的过拟合问题,但它在处理含噪声的数据时并不是都能达到理想的效果.权值不确定性中权值的波动也可以理解为训练数据的变化,这种数据中噪声的波动可能在一定程度上影响了算法的性能,因此,pgwRBM 中与分类无关的隐层节点相连的权值还是实值.也就是说,我们可以将 pgwRBM 中与分类有关的数据看作是变化的,将与分类无关的噪声看作是不变的.本文同样将权值不确定性引入 pgrncRBM 和 pgincRBM,提出了基于随机噪声数据与干净数据的 Point-wise Gated 权值不确定性受限玻尔兹曼机(weight uncertainty pgrncRBM,简称 pgwrncRBM)以及基于图像噪声数据与干净数据的 Point-wise Gated 权值不确定性受限玻尔兹曼机(weight uncertainty pgincRBM,简称 pgwincRBM).为了对比两种模型引入权值不确定性前后的性能,pgwrncRBM 和 pgwincRBM 也是在传统的 pgRBM 基础上对 pgRBM 学习到的数据二次去噪.

在 pgwrncRBM 算法中,与分类有关的隐层与可见层、转换层的连接权值  $\mathbf{W}^1$  被看作符合高斯分布的变量,其均值和标准差分别表述为  $\boldsymbol{\mu}^1$  和  $\boldsymbol{\sigma}^1 = \log(\mathbf{1} + \exp(\boldsymbol{\rho}^1))$ .与 pgwRBM 类似,当给定一组可见层状态和隐层状态时,pgwrncRBM 的能量函数可表述为

$$E(\mathbf{v}', \mathbf{z}', \mathbf{h}^1, \boldsymbol{\theta}) = - \sum_{i=1}^D \sum_{j=1}^J (z_i^1 v_i) (\mu_{ij}^1 + \log(\mathbf{1} + \exp(\rho_{ij}^1))) \odot \varepsilon_{ij}^1 h_j^1 - \sum_{i=1}^D \sum_{j=1}^J (z_i^2 v_i) W_{ij}^2 h_j^2 - \sum_{j=1}^J \sum_{r=1}^2 b_j^r h_j^r - \sum_{i=1}^D \sum_{r=1}^2 c_i^r (z_i^r v_i^r) \quad (17)$$

$$\text{s.t. } \sum_{r=1}^2 z_i^r = 1, i = 1, \dots, D$$

其中,与分类有关的隐层  $\mathbf{h}^1$  对应的权值为  $\{\boldsymbol{\mu}^1, \boldsymbol{\rho}^1, \mathbf{c}^1, \mathbf{b}^1\}$ ,与分类无关的隐层对应的权值为  $\{\mathbf{W}^2, \mathbf{c}^2, \mathbf{b}^2\}$ .

pgwrncRBM 与 pgwRBM 的不同之处有:一是赋初值的方法不同,二是学习的数据不同.pgwrncRBM 与分类无关的隐层相连权值  $\{\mathbf{W}^2, \mathbf{c}^2, \mathbf{b}^2\}$  的初值还是用特征选择的方法对一次降噪的数据  $\mathbf{v}'$  学习到的 RBM 权值处理得到的,但其与分类有关的隐层相连权值  $\{\boldsymbol{\mu}^1, \mathbf{c}^1, \mathbf{b}^1, \boldsymbol{\rho}^1\}$  的初值有两种方式:一是用 wRBM 对不含噪声的数据学习得到的,二是权值  $\{\boldsymbol{\mu}^1, \mathbf{c}^1, \mathbf{b}^1\}$  的初值用 RBM 对不含噪声的数据学习得到的,权值  $\boldsymbol{\rho}^1$  的初始值是随机赋值的.pgwrncRBM 是对 pgRBM 学习到的数据  $\mathbf{v}'$  二次去噪,而 pgwRBM 是对原始的噪声数据去噪.pgwrncRBM 的连接权值矩阵  $\mathbf{W}^1$  可以表述为  $\mathbf{W}^1 = \boldsymbol{\mu}^1 + \log(\mathbf{1} + \exp(\boldsymbol{\rho}^1)) \odot \boldsymbol{\varepsilon}^1$  (其中,  $\boldsymbol{\varepsilon}^1 \sim N(0, \mathbf{I})$ ),则当可见层、转换层和隐层任意两层状态固定时,pgwrncRBM 的条件概率分布可以用公式(6)~公式(8)表示.当 pgwrncRBM 用 CD- $k$  算法调整权值时,参数  $\{\boldsymbol{\mu}^1, \boldsymbol{\rho}^1, \mathbf{W}^2\}$  的梯度分别为

$$\Delta \mu_{ij}^1 = \alpha_w (P(h_j^1 | \mathbf{z}^{1(0)}, \mathbf{v}^{1(0)})(z_i^{1(0)} v_i^{1(0)}) - P(h_j^1 | \mathbf{z}^{1(k)}, \mathbf{v}^{1(k)})(z_i^{1(k)} v_i^{1(k)})) \quad (18)$$

$$\Delta \rho_{ij}^1 = \Delta \mu_{ij}^1 \times \frac{\varepsilon_{ij}^1}{1 + \exp(-\rho_{ij}^1)} \quad (19)$$

$$\Delta W_{ij}^2 = \alpha_w (P(h_j^2 | \mathbf{z}^{2(0)}, \mathbf{v}^{2(0)})(z_i^{2(0)} v_i^{2(0)}) - P(h_j^2 | \mathbf{z}^{2(k)}, \mathbf{v}^{2(k)})(z_i^{2(k)} v_i^{2(k)})) \quad (20)$$

用 CD- $k$  算法调整权值时,pgwrncRBM 的  $\{\mathbf{c}^1, \mathbf{b}^1, \mathbf{W}^2, \mathbf{c}^2, \mathbf{b}^2\}$  的梯度的计算方法与 pgrncRBM 一致.

pgwincRBM 算法的能量函数与 pgwrncRBM 和 pgwRBM 略有不同,其使用 ssRBM 对噪声建模.假设 pgwincRBM 与分类有关的隐层与可见层、转换层的连接权值  $\mathbf{W}^1$  是符合高斯分布变量,其均值和标准差同样可表述为  $\boldsymbol{\mu}^1$  和  $\boldsymbol{\sigma}^1 = \log(\mathbf{1} + \exp(\boldsymbol{\rho}^1))$ .当给定一组可见层状态和隐层状态时,pgwincRBM 的能量函数可表述为

$$\begin{aligned}
E(\mathbf{v}', \mathbf{z}', \mathbf{h}^1, \mathbf{h}^2, \mathbf{s}^2; \theta) = & - \sum_{i=1}^D \sum_{j=1}^J (z_i^1 v_j) (\mu_{ij}^1 + \log(1 + \exp(\rho_{ij}^1)) \odot \varepsilon_{ij}^1) h_j^1 - \sum_{i=1}^D \sum_{j=1}^J (z_i^2 v_j') W_{ij}^2 h_j^2 s_j^2 \\
& - \sum_{j=1}^J \sum_{r=1}^2 b_j^r h_j^r + \frac{1}{2} \sum_{j=1}^J \alpha_j (s_j^2)^2 - \sum_{i=1}^D \sum_{r=1}^2 c_i^r (z_i^r v_i') \\
\text{s.t. } & \sum_{r=1}^2 z_i^r = 1, i = 1, \dots, D
\end{aligned} \quad (21)$$

其中,与分类有关的隐层  $\mathbf{h}^1$  对应的权值为  $\{\boldsymbol{\mu}^1, \boldsymbol{\rho}^1, \mathbf{c}^1, \mathbf{b}^1\}$ ,与分类无关的隐层对应的权值为  $\{\mathbf{W}^2, \mathbf{c}^2, \mathbf{b}^2, \boldsymbol{\alpha}\}$ .

与 pgincRBM 相比,pgwincRBM 与分类无关的隐层相连权值  $\{\mathbf{W}^2, \mathbf{c}^2, \mathbf{b}^2, \boldsymbol{\alpha}\}$  的初值还是用特征选择的方法对一次降噪的数据  $\mathbf{v}'$  学习到的 ssRBM 权值处理得到的,但其与分类有关的隐层相连权值  $\{\boldsymbol{\mu}^1, \mathbf{c}^1, \mathbf{b}^1, \boldsymbol{\rho}^1\}$  的赋初值方式和 pgwincRBM 一样.在 pgwincRBM 算法中,当可见层、转换层和隐层任意两层状态固定时,剩余一层节点被激活的概率(或者概率分布)可以用公式(12)~公式(16)表示.当 pgwincRBM 用 CD- $k$  算法调整权值时,其参数  $\{\boldsymbol{\mu}^1, \boldsymbol{\rho}^1, \mathbf{W}^2\}$  的梯度分别为

$$\Delta \mu_{ij}^1 = \alpha_w (P(h_j^1 | \mathbf{z}^{1(0)}, \mathbf{v}'^{1(0)}) (z_i^{1(0)} v_j^{1(0)}) - P(h_j^1 | \mathbf{z}^{1(k)}, \mathbf{v}'^{1(k)}) (z_i^{1(k)} v_j^{1(k)})) \quad (22)$$

$$\Delta \rho_{ij}^1 = \Delta \mu_{ij}^1 \times \left( \frac{\varepsilon_{ij}^1}{1 + \exp(-\rho_{ij}^1)} \right) \quad (23)$$

$$\begin{aligned} \Delta W_{ij}^2 = & \alpha_w (P(h_j^2 | \mathbf{z}^{2(0)}, \mathbf{v}'^{2(0)}) \times P(s_j^2 | \mathbf{z}^{2(0)}, \mathbf{v}'^{2(0)}, h_j^{2(0)} = 1) \times (z_i^{2(0)} v_j^{2(0)}) - \\ & P(h_j^2 | \mathbf{z}^{2(k)}, \mathbf{v}'^{2(k)}) \times P(s_j^2 | \mathbf{z}^{2(k)}, \mathbf{v}'^{2(k)}, h_j^{2(k)} = 1) \times (z_i^{2(k)} v_j^{2(k)})) \end{aligned} \quad (24)$$

用 CD- $k$  算法调整权值时,pgwincRBM 的  $\{\mathbf{c}^1, \mathbf{b}^1, \mathbf{W}^2, \mathbf{c}^2, \mathbf{b}^2\}$  的梯度的计算方法与 pgincRBM 一致.

和 pgrncDBN 与 pgincDBN 类似,可以通过堆叠 pgwincRBM、pgwincRBM 和 wRBM 分别构造出包含基于随机噪声数据与干净数据的 Point-wise Gated 权值不确定性深度置信网(pgwDBN based on random noisy data and clean data,简称 pgwincDBN)以及基于图像噪声数据与干净数据的 Point-wise Gated 权值不确定性深度置信网(pgwDBN based on image noisy data and clean data,简称 pgwincDBN).pgwincDBN(pgwincDBN)首先用 pgwincRBM(pgwincRBM)对一次降噪的数据  $\mathbf{v}'$  预训练,得到更为“干净”的数据.和 pgrncDBN(pgincDBN)一样,pgwincDBN(pgwincDBN)网络的输入层与隐层,以及隐层与隐层间的连接权值的初值也有两种赋初值方式:一是利用 pgwincRBM(pgwincRBM)对一次降噪的数据  $\mathbf{v}'$  预训练得到转换层、可见层与第 1 层隐层的连接权值,随后用 wRBM 预训练隐层间的连接权值;二是利用 wRBM 通过对不含噪声的数据学习得到转换层、可见层与第 1 层隐层的连接权值,随后同样用 wRBM 预训练隐层间的连接权值.pgwincDBN(pgwincDBN)最后随机确定最后一层隐层与输出层的连接权值,随后用 BP 算法微调整个网络的权值.

## 4 实验与分析

### 4.1 实验设置与数据集

为了测试提出的算法的性能,将其与 RBM、wRBM、pgRBM、pgwRBM、DBN、wDBN、pgDBN、pgwDBN 和卷积神经网络(convolutional neural network,简称 CNN)进行比较.以上几种方法都是在 Intel(R) Xeon(R) CPU E4500 0@3.6GHZ 处理器、18GB 内存、Windows 7 64 位操作系统和 MATLAB 2015B(其中,CNN 是通过 Python3.5+ TensorFlow 框架实现的)的环境中运行的.本文实验所用的随机噪声数据集是依据文献[18]的方法对 MNIST basic 和 MNIST rotated 处理得到的数据集.设置不同邻域相关度值  $\{0, 0.2, 0.4, 0.6, 0.8, 1\}$ ,使 MNIST basic 中每个像素点的边缘分布在(0,1),可以得到 6 种不同的数据集 MNIST basic-back-random-a/b/c/d/e/f.同样可以对 MNIST rotated 处理得到数据集 MNIST rotated-back-random-a/b/c/d/e/f.本文实验所用的图像数据集是文献[18]中的 MNIST back-image 和 MNIST rotated-back-image.在所有的数据集中,训练样本都有 10 000 个噪声数据和 10 000 个干净数据,验证和测试样本分别为 2 000 个和 50 000 个噪声数据,样本维数是 784,标签数目为 10.图 3 给出了若干数据集的部分样本.



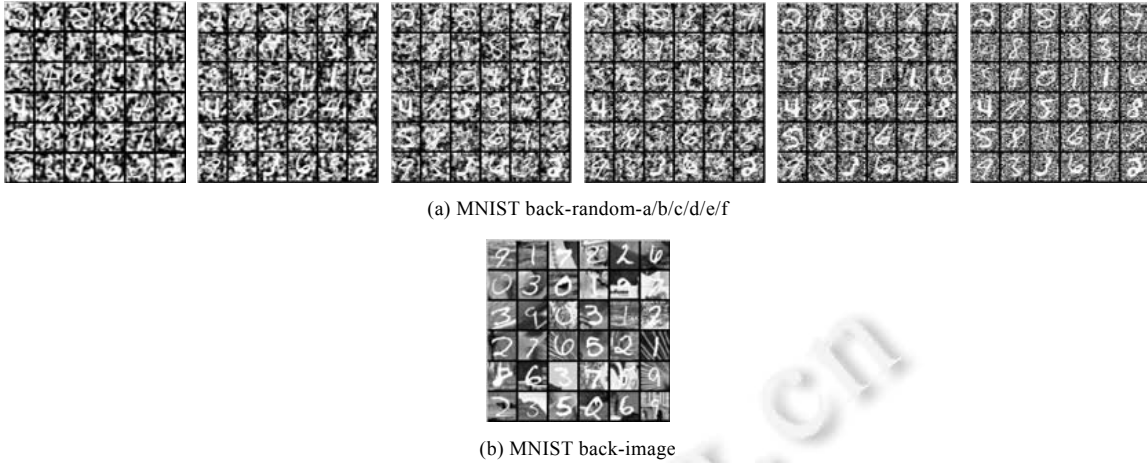


Fig.3 Legends of benchmark data sets

图3 基准数据集的图例

在本文中,所有算法都采用 mini-batch 方法学习,并且批量的大小均为 100.RBM 和 wRBM 隐层节点数为 500 或 1 000.pgRBM 和 pgwRBM 都是通过特征选择的方法对隐层节点数为 1 200 的 RBM 学习到的权值处理得到初值,并且它们的与分类有关的隐层和与分类无关的隐层节点数均为 500.

pgnrcRBM、pgincRBM、pgwrncRBM 和 pgwincRBM 都是在传统的 pgRBM 基础上对 pgRBM 学习到的数据二次去噪,并且它们的与分类有关的隐层和与分类无关的隐层节点数均为 500.

pgnrcRBM、pgincRBM、pgwrncRBM 和 pgwincRBM 与分类有关的隐层相连接权值的初值是用 RBM 或 wRBM 对不含噪声的数据学习得到的,其与分类无关的隐层相连接权值的初值都是用特征选择的方法对隐层节点数为 1 200 的 RBM 或者 ssRBM 对一次降噪的数据  $v'$  学习到的权值处理得到的.

DBN、wDBN、pgDBN、pgwDBN、pgnrcDBN、pgincDBN、pgwrncDBN 和 pgwincDBN 的隐层结构均为 500-500-2 000.CNN 是五隐层网络结构,依次为卷积层(32 个  $5 \times 5$  卷积核)、池化层(过滤器大小为  $2 \times 2$ )、卷积层(64 个  $5 \times 5$  卷积核)、池化层(过滤器大小为  $2 \times 2$ )和全连接层(1 024 个隐层节点).所有算法最后均使用梯度下降算法分类,最大迭代次数为 200(其中,CNN 的最大的迭代数目为 500),并且依据验证数据集的错误率采用提前终止方法.

#### 4.2 算法性能比较与分析

表 1 给出了 RBM、wRBM、pgRBM、pgwRBM、pgnrcRBM 和 pgwrncRBM 算法在含随机噪声的手写数据集上的错误率.

Table 1 Error rates of shallow algorithms on random noisy data sets (%)

表 1 浅层算法在随机噪声数据集上的错误率 (%)

数据集	算法					
	RBM	wRBM	pgRBM	pgwRBM	pgnrcRBM	pgwrncRBM
MNIST back-random-a	32.32	32.21	28.10	28.15	24.74	<b>24.54</b>
MNIST back-random-b	28.56	28.57	25.13	25.07	20.47	<b>20.32</b>
MNIST back-random-c	24.71	24.46	21.80	21.27	16.45	<b>16.36</b>
MNIST back-random-d	21.30	20.88	19.25	18.76	12.98	<b>13.33</b>
MNIST back-random-e	17.24	16.86	15.28	14.95	9.60	<b>9.57</b>
MNIST back-random-f	12.33	11.11	10.57	10.41	5.03	<b>4.97</b>
MNIST rotated -random-a	72.31	69.43	68.28	68.07	60.83	<b>60.40</b>
MNIST rotated -random-b	66.83	66.66	63.73	63.82	55.42	<b>55.14</b>
MNIST rotated -random-c	63.54	63.53	60.86	60.22	50.31	<b>49.76</b>
MNIST rotated -random-d	59.32	59.16	56.08	56.28	44.61	<b>44.18</b>
MNIST rotated -random-e	54.24	52.66	51.61	51.56	37.66	<b>37.13</b>
MNIST rotated -random-f	47.77	46.50	45.81	45.36	21.49	<b>20.53</b>

我们可以看出,pgwrcRBM 在所有数据集上都表现出优于其他浅层学习算法的分类效果。

通过对比 pgRBM 和 pgrncRBM,我们可以看出,pgrncRBM 在所有数据集上的分类性能都优于 pgRBM.这说明利用不含噪声的数据可以提升 pgrncRBM 在随机噪声数据上的学习能力.通过对比表 1 中的算法,我们还可以发现:

- 1) wRBM 基本上在所有随机噪声数据集上都表现出优于 RBM 的分类性能.
- 2) pgwRBM 仅在 MNIST back-random-a、MNIST rotated-back-b 和 MNIST rotated-back-d 这 3 个数据集上的错误率略高于 pgRBM.
- 3) pgwrcRBM 在所有数据集上都表现出优于 pgrncRBM 的分类性能.
- 4) 随机噪声数据中噪声结构越简单,与 pgRBM 相比,pgrncRBM 提升的分类效果就越明显.

表 2 给出了 RBM、wRBM、pgRBM、pgwRBM、pgincRBM 和 pgwincRBM 算法在含图像噪声的手写数据集上的错误率.同样地,pgwincRBM 在所有图像噪声数据集上都表现出优于其他浅层学习算法的分类效果。

**Table 2** Error rates of shallow algorithms on image noisy data sets (%)

**表 2** 浅层算法在图像噪声数据集上的错误率 (%)

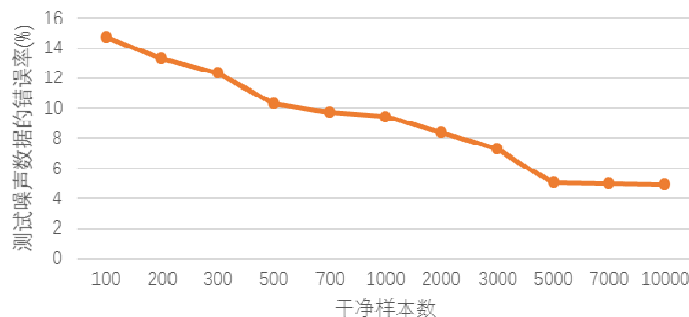
数据集	算法					
	RBM	wRBM	pgRBM	pgwRBM	pgincRBM	pgwincRBM
MNIST back-image	15.83	15.83	15.01	14.65	14.58	<b>14.34</b>
MNIST rotated-back-image	47.47	47.34	44.32	44.34	42.10	<b>41.86</b>

通过对比表 2 中的算法,我们还可以发现:

- 1) wRBM 在所有图像噪声数据集上的分类性能略优于 RBM.
- 2) pgwRBM 在 MNIST rotated-back-image 数据集上的错误率略高于 pgRBM.
- 3) pgwincRBM 在所有图像噪声数据集上都表现出优于 pgincRBM 的分类性能.

从表 1 和表 2 我们可以得出:1) 权值不确定性方法能够有效地解决 RBM、pgRBM、pgrncRBM 和 pgincRBM 在处理含噪声的数据时出现的过拟合问题;2) 利用不含噪声的数据,可以提升 pgrncRBM 和 pgincRBM 在噪声数据上的学习能力.

图 4 展示了 pgrncRBM 在 MNIST back-random-f 上的分类性能随干净样本数目变化曲线。



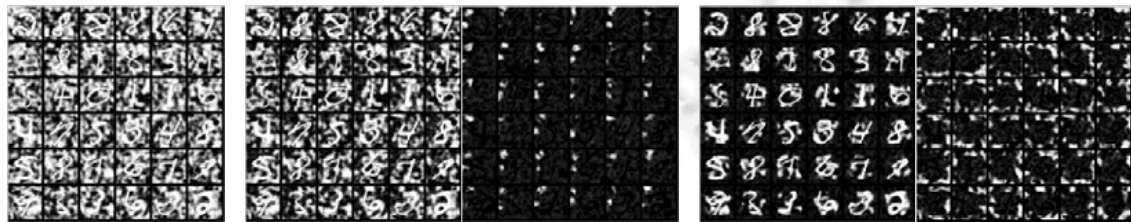
**Fig.4** Variations in the performance of pgrncRBM on MNIST back-random-f, as we increase the number of the clean data

**图 4** pgrncRBM 在 MNIST back-random-f 上的分类性能随干净样本数目变化曲线

我们可以从图 4 中看出,当干净样本数目增大时,pgrncRBM 分类效果越来越好,这和我们的直观理解是一致的.pgrncRBM 中与分类有关的隐层节点相连接权的初值是用 RBM 对不含噪声的数据学习得到的,干净样本越多,RBM 学习到的不含噪声数据的信息就越多,pgrncRBM 也能更好地将噪声数据中与分类有关的信息表示出来.图 5 给出了 pgRBM 和 pgrncRBM 在 MNIST back-random-a/f 两种数据集上的学习结果,图 6 给出了 pgRBM 和 pgincRBM 在 MNIST back-image 数据集上的学习结果.pgRBM、pgrncRBM 和 pgincRBM 都把隐层节点分

为与分类有关的和与分类无关的两个部分,并且都能在含噪声的数据中自适应找到数据中与分类有关的部分.但是,我们可以从图 5 和图 6 中看出:

- 1) pgRBM 学习得到的与分类无关的图像(也就是噪声)中包含的噪声信息很少,pgRBM 学习得到的与分类有关的图像与原始的含噪声图像相差不是很大.
- 2) pgrncRBM 是在 pgRBM 基础上对其学习到的与分类有关的图像二次去噪,但是 pgrncRBM 可以很好地把图像中与分类有关的信息(也就是手写数字)和噪声信息分别学习出来,特别是 pgncRBM 在 MNIST back-random-f 这个数据集上学习得到的与分类有关的图像基本上没有噪声.
- 3) pgincRBM 是在 pgRBM 基础上对其学习到的与分类有关的图像二次去噪,pgncRBM 可以把图像中与分类有关的信息(也就是手写数字)和图像噪声信息分别学习出来,但是有时也只能学习到部分与分类有关的信息,如第 1 列中的数字 2.

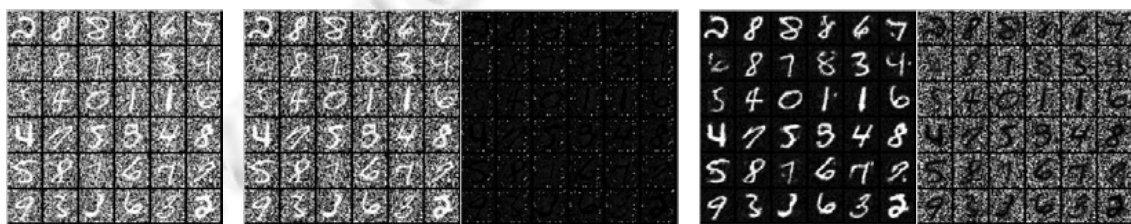


原始图像

pgRBM 学习到的与分类有关/无关的图像

pgrncRBM 学习到的与分类有关/无关的图像

(a) pgRBM 和 pgrncRBM 在 MNIST back-random-a 上学习到的与分类有关/无关图像对比



原始图像

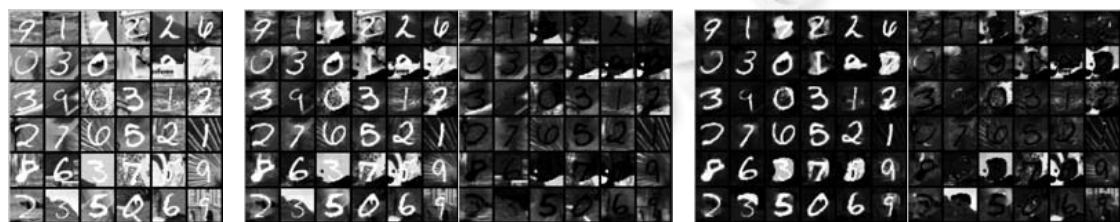
pgRBM 学习到的与分类有关/无关的图像

pgrncRBM 学习到的与分类有关/无关的图像

(b) pgRBM 和 pgrncRBM 在 MNIST back-random-f 上学习到的与分类有关/无关图像对比

Fig.5 Learning results of pgRBM and pgrncRBM on MNIST back-random-a/f data sets

图 5 pgRBM 和 pgrncRBM 在 MNIST back-random-a/f 数据集上的学习结果



原始图像

pgRBM 学习到的与分类有关/无关的图像

pgrncRBM 学习到的与分类有关/无关的图像

Fig.6 Learning results of pgRBM and pgrncRBM on MNIST back-image data set

图 6 pgRBM 和 pgrncRBM 在 MNIST back-image 数据集上的学习结果

表 3 给出了含噪声的原始图像、pgRBM 学习到的与分类有关的图像和 pgrncRBM/pgncRBM 学习到的与分类有关的图像与干净数据的信噪比.我们可以得出,pgrncRBM 和 pgncRBM 可以在传统的 pgRBM 基础上对

pgRBM 学习到的数据二次去噪得到蕴含噪声信息较少的图像,从而达到更好的分类效果.pgrncRBM 在 MNIST back-random-f 和 MNIST rotated-random-f 上的错误率分别是 5.03%和 21.49%,而 pgRBM 在这两个数据集的错误率是 pgrncRBM 的 2 倍以上.并且,特别是在文献[18]中,RBM 在 MNIST basic 和 MNIST rotated 的错误率分别是 3.94%和 14.69%,这与 pgrncRBM 在 MNIST back-random-f 和 MNIST rotated-random-f 上的错误率还是比较接近的.

**Table 3** Signal-noise ratios of learning results of pgRBM and pgincRBM/pgrncRBM on different valid data sets

**表 3** pgRBM 和 pgincRBM/pgrncRBM 学习结果在验证数据集上的信噪比

数据集	原始噪声图像	pgRBM 学习到的 与分类有关的图像	pgincRBM/pgrncRBM 学习到的 与分类有关的图像
MNIST back-random-a	5.567 0	7.070 7	<b>11.935 6</b>
MNIST back-random-b	5.581 3	6.826 4	<b>12.726 2</b>
MNIST back-random-c	5.580 3	6.906 5	<b>13.235 2</b>
MNIST back-random-d	5.555 1	6.400 3	<b>13.661 8</b>
MNIST back-random-e	5.551 1	6.255 9	<b>15.110 0</b>
MNIST back-random-f	5.554 0	6.361 2	<b>18.949 7</b>
MNIST back-image	6.662 1	10.659 7	<b>15.484 3</b>
MNIST rotated -random-a	5.618 2	6.911 6	<b>11.110 2</b>
MNIST rotated -random-b	5.600 5	7.556 0	<b>12.096 0</b>
MNIST rotated -random-c	5.612 4	6.871 5	<b>12.586 5</b>
MNIST rotated -random-d	5.596 6	6.550 9	<b>12.386 4</b>
MNIST rotated -random-e	5.573 0	6.278 9	<b>12.429 1</b>
MNIST rotated -random-f	5.579 6	6.341 6	<b>18.648 2</b>
MNIST rotated-back-image	6.585 5	10.342 8	<b>14.235 5</b>

表 4 给出了 DBN、wDBN、pgDBN、pgwDBN、pgncDBN、pgwncDBN 和 CNN 算法含随机噪声的手写数据集上的错误率.我们从表 4 可以看出,pgwncDBN 在绝大多数数据集上的性能都表现出优于其他深度学习算法的分类效果.通过对比 pgDBN 和 pgncDBN,我们可以看出,pgrncDBN 在所有随机噪声数据集上的分类性能都优于 pgDBN.实验结果表明,pgrncDBN 还是一种有效的神经网络学习算法.我们又可以发现以下结果.

- 1) wDBN 在所有数据集上都要表现出优于 DBN 的分类性能.
- 2) pgwDBN 在所有数据集上都要表现出优于 pgDBN 的分类性能.
- 3) pgwncDBN 仅在 MNIST rotated-back-f 数据集上的错误率略高于 pgrncDBN.
- 4) CNN 仅在 MNIST back-random-a、MNIST back-random-b 和 MNIST rotated-random-a 这 3 个数据集上的错误率低于 pgrncDBN 和 pgwncDBN.

**Table 4** Error rates of deep algorithms related to DBN on random noisy data sets (%)

**表 4** 与 DBN 相关的深度算法在随机噪声数据集上的错误率 (%)

数据集	算法						
	DBN	wDBN	pgDBN	pgwDBN	pgncDBN	pgwncDBN	CNN
MNIST back-random-a	43.21	35.95	28.01	27.73	23.56	23.55	<b>12.32</b>
MNIST back-random-b	40.57	30.40	24.41	23.82	20.31	19.65	<b>15.96</b>
MNIST back-random-c	35.34	26.37	20.41	19.32	15.94	<b>15.86</b>	16.89
MNIST back-random-d	34.78	21.90	16.45	15.81	12.79	<b>12.22</b>	14.41
MNIST back-random-e	32.98	17.67	12.54	11.76	8.60	<b>7.91</b>	12.52
MNIST back-random-f	25.29	12.56	8.54	7.41	4.04	<b>3.88</b>	8.31
MNIST rotated -random-a	72.31	71.13	67.48	67.41	57.61	56.83	<b>46.42</b>
MNIST rotated -random-b	71.04	67.96	62.07	62.24	51.67	<b>51.33</b>	59.20
MNIST rotated -random-c	67.21	63.65	57.42	56.24	45.63	<b>45.44</b>	57.04
MNIST rotated -random-d	67.42	60.33	48.98	48.81	39.23	<b>38.96</b>	55.49
MNIST rotated -random-e	66.25	53.71	42.55	40.39	31.29	<b>31.27</b>	44.40
MNIST rotated -random-f	63.43	43.47	33.02	31.16	<b>15.18</b>	15.45	32.70

表 5 给出了 DBN、wDBN、pgDBN、pgwDBN、pgincDBN、pgwincDBN 和 CNN 算法在含图像噪声的手写数据集上的错误率.除了在 MNIST back-image 上的错误率略高于 CNN 以外,pgwincDBN 在所有图像噪声数据集上都表现出优于深度浅层学习算法的分类效果.通过对比表 5 中的算法,我们还可以发现:

- 1) wDBN 在所有图像噪声数据集上的分类性能略优于 DBN.

- 2) pgwDBN 在所有图像数据集上的错误率略高于 pgDBN.
- 3) pgwincRBM 在所有图像噪声数据集上都表现出优于 pgncRBM 的分类性能.
- 4) CNN 仅仅在 MNIST back-image 数据集上的错误率略低于 pgincDBN 和 pgwincDBN.

**Table 5** Error rates of deep algorithms related to DBN on image noisy data sets (%)

**表 5** 与 DBN 相关的深度算法在图像噪声数据集上的错误率 (%)

数据集	算法						CNN
	DBN	wDBN	pgDBN	pgwDBN	pgincDBN	pgwincDBN	
MNIST back-image	21.65	16.97	13.17	13.49	13.44	12.98	<b>12.50</b>
MNIST rotated-back-image	62.44	46.42	39.03	39.52	37.37	<b>37.09</b>	40.60

从表 4 和表 5 中我们可以得出:1) 权值不确定性方法有效地解决了 pgrncDBN 和 pgincDBN 两种深度网络中出现的过拟合问题;2) pgrncDBN/pgwrcDBN 和 pgincDBN/pgwincDBN 可以在绝大多数数据集上的性能超过 CNN 的主要原因是 pgrncRBM/pgwrcRBM 和 pgincRBM/pgwincRBM 可以学习得到更为“干净”的数据.通过对比 RBM、DBN 和 wDBN,我们可以得出,DBN 在处理含噪声数据时出现过拟合现象,并没有展现出比 RBM 更好的学习能力,并且权值不确定性方法有效地解决了 DBN 深度网络中出现的过拟合问题.通过对比 pgRBM 与 pgDBN、pgrncRBM 与 pgrncDBN、pgincRBM 与 pgincDBN,我们又可以发现,堆叠 pgRBM/pgrncRBM/pgincRBM 和 RBM 构造出的深度网络(pgDBN、pgrncDBN 和 pgincDBN)展现出优于浅层网络(pgRBM、pgrncRBM 和 pgincRBM)的学习能力.

## 5 结束语

Point-wise Gated 受限玻尔兹曼机是一种针对噪声数据的浅层学习算法.本文在此基础上利用干净数据提升其在噪声数据上的学习能力,提出了两种基于噪声数据与干净数据的 Point-wise Gated 受限玻尔兹曼机(pgrncRBM 和 pgincRBM)方法,然后将它们学习到的与分类有关的数据子集用到深度置信网中,提出了两种基于噪声数据与干净数据的 Point-wise Gated 深度置信网(pgrncDBN 和 pgincDBN).pgrncRBM 和 pgincRBM 在绝大多数手写数据集上都表现出优于 pgRBM 的学习能力.同样,以 pgrncRBM 和 pgincRBM 为基石的 pgrncDBN 和 pgincDBN 一般都优于 pgDBN.然后,本文将权值不确定性方法用在所提出的 4 种算法中,并将这几种算法称为基于噪声数据与干净数据的 Point-wise Gated 权值不确定性受限玻尔兹曼机(pgwrncRBM 和 pgwincRBM)以及基于噪声数据与干净数据的 Point-wise Gated 权值不确定性深度置信网(pgwrcDBN 和 pgwincDBN).实验结果表明,权值不确定性方法能够有效地解决 pgrncRBM、pgincRBM、pgrncDBN 和 pgincDBN 在处理含噪声数据时出现的过拟合问题.我们同时发现,pgRBM、pgrncRBM 和 pgincRBM 并不能在部分噪声数据集(数据中与分类有关的部分被噪声损坏时)上取得理想的结果.如何把它们应用到更多的噪声数据上,也是我们下一步的研究方向.

## References:

- [1] Ravanbakhsh S, Póczos B, Schneider J. Stochastic neural networks with monotonic activation functions, In: Gretton A, Robert CC, eds. Proc. of the 19th Int'l Conf. on Artificial Intelligence and Statistics. Cadiz: MIT Press, 2016. 809–818.
- [2] Salakhutdinov R, Hinton GE. An efficient learning procedure for deep Boltzmann machines, Neural Computation, 2012,24(8): 1967–2006. [doi: 10.1162/NECO\_a\_00311]
- [3] Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. Science, 2006,313(5786):504–507. [doi: 10.1126/science.1127647]
- [4] Zhao S, Muraoka Y, Fujimaki R, Carin L. Scalable model selection for belief networks. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R, eds. Proc. of the Advances in Neural Information Processing Systems 30. Long Beach: MIT Press, 2017. 4612–4622.
- [5] Zhang N, Ding SF, Zhang J, Xue Y. An overview on RESTRICTED Boltzmann machines. Neurocomputing, 2018,275:1186–1199. [doi: 10.1016/j.neucom.2017.09.065]

- [6] Guyon I, Elisseeff A. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 2003,3:1157–1182.
- [7] Sohn BK, Zhou G, Lee C, Lee H. Learning and selecting features jointly with point-wise gated Boltzmann machines. In: *Proc. of the 30th Int'l Conf. on Machine Learning*. Atlanta: MIT Press, 2013. 217–225. <https://icml.cc/2013/>
- [8] Zhang N, Ding SF, Zhang J, Xue Y. Research on point-wise gated deep networks. *Applied Soft Computing*, 2017,52:1210–1221. [doi: 10.1016/j.asoc.2016.08.056]
- [9] Ranzato M, Hinton GE. Modeling pixel means and covariances using factorized third-order Boltzmann machines. In: *Proc. of the 23rd IEEE Conf. on Computer Vision and Pattern Recognition*. San Francisco: IEEE Press, 2010. 2551–2558. <http://www.cvpapers.com/cvpr2010.html> [doi: 10.1109/CVPR.2010.5539962]
- [10] Courville A, Desjardins G, Bergstra J, Bengio Y. The spike-and-slab RBM and extensions to discrete and sparse data distributions. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2014,36(9):1874–1887. [doi: 10.1109/TPAMI.2013.238]
- [11] Srivastava N, Hinton GE, Krizhevsky A, Sutskever I, Salakhutdinov RR. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 2014,15:1929–1958. [doi: 10.1214/12-AOS1000]
- [12] Wan L, Zeiler M, Zhang S, Le Cun Y, Fergus R. Regularization of neural networks using dropconnect. In: *Proc. of the 30th Int'l Conf. on Machine Learning*. Atlanta: MIT Press, 2013. 1058–1066. <https://icml.cc/2013/>
- [13] Iosifidis A, Tefas A, Pitas I. DropELM: Fast neural network regularization with dropout and DropConnect. *Neurocomputing*, 2015, 162:57–66. [doi: 10.1016/j.neucom.2015.04.006]
- [14] Blundell C, Cornebise J, Kavukcuoglu K. Weight uncertainty in neural networks. In: Bach FR, Blei DM, eds. *Proc. of the 32nd Int'l Conf. on Machine Learning*. Lille: MIT Press, 2015. 1613–1622.
- [15] Zhang J, Ding SF, Zhang N, Xue Y. Weight uncertainty in Boltzmann machine. *Cognitive Computation*, 2016,8(6):1064–1073. [doi: 10.1007/s12559-016-9429-1]
- [16] Ding SF, Zhang N, Shi ZZ. Laplacian multi layer extreme learning machine. *Ruan Jian Xue Bao/Journal of Software*, 2017,28(10): 2599–2610 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/5128.htm> [doi: 10.13328/j.cnki.jos.005128]
- [17] Hinton GE. Training products of experts by minimizing contrastive divergence. *Neural Computation*, 2002,14(8):1711–1800. [doi: 10.1162/089976602760128018]
- [18] Larochelle H, Erhan D, Courville A, Bergstra J, Bengio Y. An empirical evaluation of deep architectures on problems with many factors of variation. In: Ghahramani Z, ed. *Proc. of the 24th Int'l Conf. on Machine Learning*. New York: MIT Press, 2007. 473–480. [doi: 10.1145/1273496.1273556]

#### 附中文参考文献:

- [16] 丁世飞,张楠,史忠植.拉普拉斯多层极速学习机. *软件学报*,2017,28(10):2599–2610. <http://www.jos.org.cn/1000-9825/5128.htm> [doi: 10.13328/j.cnki.jos.005128]



张楠(1991—),男,安徽萧县人,博士,主要研究领域为机器学习,玻尔兹曼机.



张健(1990—),男,博士生,主要研究领域为深度学习,玻尔兹曼机.



丁世飞(1963—),男,博士,教授,博士生导师,CCF 杰出会员,主要研究领域为人工智能,模式识别,机器学习,数据挖掘.



赵星宇(1994—),男,硕士,主要研究领域为机器学习.