

# 基于 MIC 的深度置信网络研究

曾 安 郑齐弥

(广东工业大学计算机学院 广州 510000)

**摘 要** 传统的深度置信网络(DBNs)训练过程采用重构误差作为 RBM 网络的评价指标,它能在一定程度上反映网络对训练样本的似然度,但它并不是可靠的。而最大信息系数(MIC)能反映两个属性间的相关度,保留相关度较大的属性,且 MIC 较稳健,不易受异常值的影响,可作为网络评价指标。故提出一种基于最大信息系数(MIC)的深度置信网络方法,一方面用 MIC 对数据进行降维预处理,提高数据与网络的拟合度,降低网络分类误差;另一方面将 MIC 作为网络评价标准,改进重构误差的不可靠性。分别利用传统方法与基于 MIC 的深度置信网络方法对手写数据集 MNIST 和 USPS 进行分类实验,结果表明,基于 MIC 的深度置信网络方法能有效地提高识别率。

**关键词** 深度置信网络,最大信息系数,重构误差,降维

**中图分类号** TP183 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.8.050

## Deep Belief Networks Research Based on Maximum Information Coefficient

ZENG An ZHENG Qi-mi

(School of Computer Science and Technology,Guangdong University of Technology,Guangzhou 510000,China)

**Abstract** The traditional deep belief networks use reconstruction error as the evaluation criteria of restricted boltzmann machine(RBM) networks in the training process, which can reflect the likelihood between RBM network and training samples to some extent. However, it is not reliable. Maximum information coefficient (MIC), based on the estimations of Shannon entropy and conditional entropy, identifies interesting relationships between pairs of variables in large data sets and captures a subset of highly related features. The MIC can be used as a criterion for evaluating a network since it is robust to outliers. In order to construct models that fit data well and reduce classification error, a deep belief networks based on MIC method was proposed. MIC is used not only in dimensionality reduction, but also in improving the unreliability of the reconstruction error. Classification experiments were performed on handwriting data sets of MNIST and USPS by several traditional methods and deep belief networks based on MIC method. The results show that the latter can improve the recognition rate effectively.

**Keywords** DBNs, MIC, Reconstruction error, Dimensionality reduction

## 1 引言

2006 年, Hinton 等人在支持向量机(SVM)<sup>[1]</sup>、BP 神经网络<sup>[2]</sup>等浅层学习基础上提出了深度学习模型,它模拟人脑的机制,将数据从底层映射到高层,是特征提取的一个学习过程<sup>[3]</sup>。其因在复杂函数表示和复杂数据分类上具有较好的效果和效率,已成为机器学习领域的一个研究热点,并已成功应用到语音识别<sup>[4]</sup>、手写数字识别<sup>[5]</sup>、图像识别<sup>[6]</sup>和自然语言处理<sup>[7]</sup>等领域。

深度学习的提出,促进了神经网络模型和算法在深层结构上的研究,其学习方法分为有监督和无监督学习。有监督学习模型包含卷积神经网络(Convolutional Neural Networks, CNNs);无监督学习模型包含自动编码器(Auto Encoders)、稀疏编码(Sparse Coding)、受限玻尔兹曼机(Re-

stricted Boltzmann Machine, RBM)、深度置信网络(Deep Belief Networks, DBNs)<sup>[8]</sup>。CNNs 模型是一个多层神经网络,且每层包含多个二维平面,而每个二维平面由多个独立神经元组成。它利用空间关系减少学习参数的数目,提高算法的训练性能<sup>[9]</sup>。自动编码器模型是利用神经网络的低维代表高维输入的性质,提取出可代表输入数据的最重要的因素<sup>[10]</sup>。稀疏编码模型是寻找一组超完备的基向量来高效地表示样本数据,超完备基能有效找出隐含在输入数据内部的结构与模式<sup>[11]</sup>。RBM 模型含有可视层与隐藏层,其层间节点相互全连接,层内节点相互独立<sup>[12,13]</sup>。DBNs 由一组 RBM 堆叠而成,是目前研究和应用较广泛的深度学习结构<sup>[14,15]</sup>。DBNs 采用无监督的学习策略获得网络参数,并初始化网络,可避免陷入局部最优;DBNs 学习过程是将上一层的输出作为下一层的输入,能够有效提取特征<sup>[16]</sup>。文献<sup>[17]</sup>提出一种判别的

到稿日期:2015-06-19 返修日期:2015-09-11 本文受国家自然科学基金项目(61300107),广东省自然科学基金项目(S2012010010212),广州市科技计划项目(201504301341059,201505031501397)资助。

曾 安(1978-),女,博士,教授,主要研究方向为智能信息处理、数据挖掘,E-mail:zengan2010@126.com;郑齐弥(1990-),女,硕士生,主要研究方向为数据挖掘。

深度置信(DBBN)网络模型,其通过优化目标函数更新参数,并成功应用到可视化数据分类上。文献[18]提出将堆稀疏自动编码SSAE与DBN的优点相结合,得到新方法SSAD,该方法利用SSAE对数据进行预处理,再用DBNs对数据进行学习,并成功应用到图像识别上。

DBNs建模中常采用贪婪策略将其分成一组RBM网络,用无监督的方法学习每一个RBM。RBM学习过程采用对比散度(Contrastive Divergence, CD)<sup>[19]</sup>算法,用重构误差(Reconstruction Error)作为其评价指标。重构误差由训练样本和重构样本来确定,能在一定程度上反映RBM对训练样本的似然度,但它并不是可靠的<sup>[20]</sup>。在现有的研究中,鲜有关于DBNs的文献对重构误差进行研究。David等人提出最大信息系数(Maximal Information Coefficient, MIC)<sup>[21]</sup>用于确定两个属性间的相关程度,由于其基于信息熵和条件熵,比较稳健,可取代重构误差作为网络的评价指标,且MIC可用于降维预处理数据,能提取出代表原始数据的重要特征<sup>[22]</sup>。因此,本文提出一种基于MIC的深度置信网络,一方面用MIC作为RBM网络的评价指标;另一方面用MIC预处理样本数据,提高分类精度。

## 2 深度置信网络

### 2.1 受限玻尔兹曼机

受限玻尔兹曼机(Restricted Boltzmann Machine, RBM)<sup>[12]</sup>是一种基于能量的生成模型,包含两层,分别为可视层和隐藏层,其层间节点相互全连接,层内节点无连接,如图1所示。

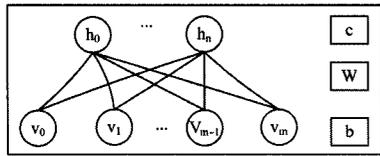


图1 RBM网络结构

能量模型是指通过能量函数 $E(v, h, \theta)$ <sup>[23]</sup>来定义变量的概率分布,其中 $v$ 代表可见层神经元状态, $h$ 代表隐藏层神经元状态,定义为:

$$E(v, h, \theta) = -c^T h - b^T v - v^T W h \quad (1)$$

其中, $\theta = \{W, b, c\}$ 是模型参数。 $W$ 代表可视层到隐藏层权重矩阵, $b$ 代表可视层的偏置向量, $c$ 代表隐藏层的偏置向量。根据能量函数 $E(v, h, \theta)$ ,得到联合概率分布 $p(v, h)$ 。

$$p(v, h) = \frac{e^{-E(v, h, \theta)}}{Z} \quad (2)$$

其中, $Z = \sum_{v, h} e^{-E(v, h, \theta)}$ 是归一化因子。

由于RBM是层内无连接、层间全连接的二部图,因此给定其中一层的状态,另一层的状态是条件独立的。即给定可视层 $v$ 向量值,隐藏层单元 $h_j$ 的状态如式(3)所示:

$$p(h_j = 1 | v) = \sigma(c_j + \sum_{i=1}^m v_i W_{ij}') \quad (3)$$

RBM是对称模型,求得可视节点的激活概率如式(4)所示:

$$p(v_i = 1 | h) = \sigma(b_i + \sum_{j=1}^n h_j W_{ij}) \quad (4)$$

其中, $\sigma(x) = \frac{1}{(1+e^{-x})}$ 是一个激活函数,具体推导过程参照文献[3, 12, 24, 25]。

模型参数 $\theta = \{W, b, c\}$ 采用梯度下降法的更新策略,即可视数据的最大似然函数对更新参数进行求导,得到更新的梯度值,具体如式(5)~式(7)所示。

$$\frac{\partial \log p(v)}{\partial W_{ij}} = \langle v_i, h_j \rangle_0 - \langle v_i, h_j \rangle_\infty \quad (5)$$

$$\frac{\partial \log p(v)}{\partial b_i} = \langle v_i \rangle_0 - \langle v_i \rangle_\infty \quad (6)$$

$$\frac{\partial \log p(v)}{\partial c_j} = \langle h_j \rangle_0 - \langle h_j \rangle_\infty \quad (7)$$

其中, $\langle \cdot \rangle_0$ 表示数据分布期望得到的数据, $\langle \cdot \rangle_\infty$ 表示训练RBM模型期望得到的数据<sup>[26]</sup>。本文采用对比散度算法(Contrastive Divergence algorithm, CD)<sup>[19]</sup>对RBM网络进行快速学习。

### 2.2 深度置信网络

深度置信网络(Deep Belief Network, DBNs)起源于人工神经网络,由一组受限玻尔兹曼机自下向上堆叠而成。传统的神经网络在训练时需要大量的有标签样本,而获得大量有标签数据是非常困难的<sup>[14]</sup>,文献[15]指出DBNs的无监督学习方法可以解决这个问题。基于贪心策略将DBNs拆分成一组RBM网络,再利用大量无标签数据对RBM进行无监督学习,得到初始参数,然后利用有标签数据微调整个DBNs网络。一方面将DBNs拆分成多个RBM训练,可降低训练整体DBNs的复杂度;另一方面初始化参数,可避免网络陷入局部最优解。

### 2.3 最大信息系数

最大信息系数(Maximal Information Coefficient, MIC)是David N. Reshef等人提出的,用于在大数据集中找到潜在的相关变量对,是确定两个变量相关度的一种标准,具备广泛性和公平性<sup>[21]</sup>。例如对于给定的有序对 $(X, Y)$ ,其数据集为 $D$ ,数据样本大小为 $n$ ,分别将 $X, Y$ 按 $x$ 值和 $y$ 值划分成 $x, y$ 维,这种划分称为 $x \times y$ 网格划分,记为 $G$ 。

数据集 $D$ 包含两个节点变量 $X, Y$ 。 $X$ 和 $Y$ 的特征矩阵是一个无限矩阵,则定义信息系数公式如下:

$$M(X, Y | D)_{i, j} = \frac{I^*(X, Y, D, i, j)}{\log \min(i, j)} \quad (8)$$

其中, $I^*(X, Y, D, i, j) = \max I(X, Y, D | G, i, j)$ 是 $G$ 被划分成 $i \times j$ 个网格中 $X, Y$ 的最大互信息。

根据式(8),节点变量 $X, Y$ 的最大信息系数定义为:

$$MIC(X, Y | D) = \max_{i \times j < B(n)} \{M(X, Y | D)_{i, j}\} \quad (9)$$

其中, $B(n) = n^{0.6}$ , $i \times j < B(n)$ 表示网格 $G$ 的划分维度的限制。

因MIC依赖于互信息,而由互信息的公式 $I(X, Y) = H(X) + H(Y) - H(X, Y)$ 可知其依赖于熵 $H(X) = -\sum_{i=1}^n p_i \log p_i$ ,而熵 $H(X)$ 依赖于变量 $X$ 的分布,是对变量 $X$ 不确定性的度量,与样本的数值大小无关<sup>[27]</sup>,因此熵 $H(X)$ 是稳健的,推理得到互信息 $I(X, Y)$ 是稳健的, MIC也具有稳健性<sup>[28-30]</sup>。

## 3 基于MIC的深度置信网络

传统的深度置信网络学习过程采用重构误差作为网络评

价指标,它能够在一定程度上反映网络对训练样本的似然度,但它并不是可靠的<sup>[20]</sup>。同时,传统的深度置信网络对数据进行学习时,对一些相关度较小的属性没有进行预处理,这在一定程度上会影响网络的分类精度。

基于 MIC 的深度置信网络如图 2 所示,其中 MIC 用来衡量两个变量间的相关度,其基于信息熵和条件熵,较稳健,不易受到异常值的影响<sup>[29,30]</sup>;而重构误差用来表示训练样本和重构样本间的差异量,较不稳定,易受异常值的影响。故将 MIC 取代重构误差作为网络评价指标,并且 MIC 的值越大,其隐藏节点表示可视节点的能力也越好,同时文献<sup>[31]</sup>中也说明了采用 MIC 作为模型选择的标准是可行的。除此之外, MIC 还可用于预处理数据,利用 MIC 能保留相关度较大属性的性质,提取出能代表输入数据的最重要属性,并且在文献<sup>[22]</sup>中得到了证明。

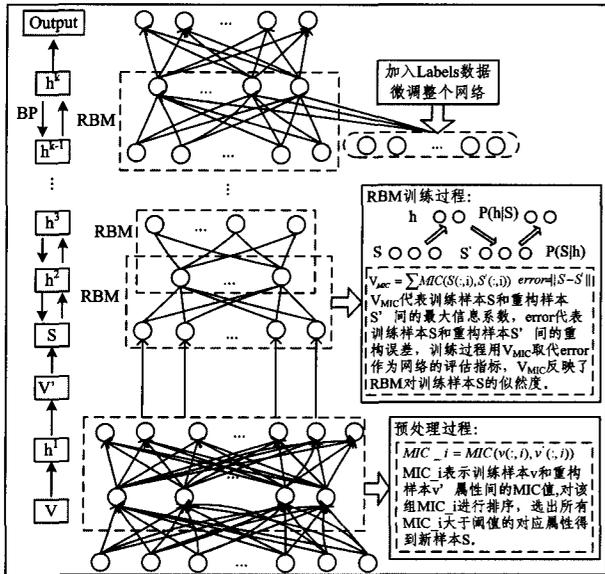


图 2 基于 MIC 的深度置信网络流程图

基于 MIC 的深度置信网络算法的主要步骤如下。

输入:训练样本  $v$ , 隐层节点数  $m$ , 隐藏层数  $L$ , 学习率  $\delta$ , 阈值  $\epsilon$ , 最大训练周期  $K$ , 权重矩阵  $W$ , 可视层偏置向量  $b$ , 隐藏层偏置向量  $c$ 。

输出:新样本  $S$ , 测试结果。

- 1) 根据训练样本  $v$  的大小构建可视层→隐层→重构层( $v \rightarrow h^1 \rightarrow v'$ ), 并利用 RBM 网络学习规则进行学习;
- 2) 初始化参数, 设可视层初始状态  $v_0 = v, W = 0, b = 0, c = 0$ ;
- 3) 根据式(3)、式(4)重构可视层, 如图 2 所示, 得到重构样本  $v'$ ;
- 4) 根据式(8)、式(9)计算出训练样本  $v$  和重构样本  $v'$  间对应属性间的最大信息系数, 记为  $MIC_i = MIC(v(:, i), v'(:, i)) (i = 1, 2, \dots, p, p$  为样本属性个数);
- 5) 预处理过程: 对第 4) 步得到的一组  $MIC_i$  进行排序, 选出值大于阈值  $\epsilon$  的所有属性, 输出新样本  $S$ , 新样本属性个数记为  $n$ ;
- 6) 拆分网络: 根据新样本  $S$  的大小  $n$ 、隐藏层数  $L$  和隐藏节点数  $m$ , 构建深度置信网络 DBNs, 采用贪心策略将 DBNs 网络拆分成一组 RBM;
- 7) 学习网络: 对上述的一组 RBM 进行无监督学习, 利用 CD 算法对网络进行 Gibbs 采样学习, 如图 2 训练过程, 采样过程采用  $V_{MIC} = \sum_{i=1}^n MIC(S(i, i), S'(i, i)) (i = 1, 2, \dots, n, n$  为新样本  $S$  属性个数) 代替

重构误差  $error = \|S - S'\|$ <sup>[20]</sup> 作为网络评价的新指标, 根据  $V_{MIC}$  的公式可知其通过 MIC 公式计算而来, 不易受到异常值的影响, 同时不会随样本容量的变化而变化<sup>[28,29,31]</sup>, 根据计算  $error$  的公式可知其通过 2-范数公式计算而来, 易受到异常值的影响并且会随着样本容量变化发生相应的变化,  $V_{MIC}$  的值越大说明 RBM 网络对训练样本  $S$  的似然度越大;

- 8) 根据式(5)一式(7)更新网络参数  $w, b, c$ ;
- 9) 将步骤 7)、8)过程迭代  $k$  次, 并依次对剩下的 RBM 进行步骤 7)–9)的学习, 直到所有的 RBM 被学习;
- 10) 微调网络并测试: 根据上述过程得到网络初始参数  $w, b, c$ , 加入带有 labels 的数据集, 利用 BP 网络误差反向传播规则, 有监督微调整个 DBNs 网络, 并利用该网络对测试样本进行分类, 输出分类结果。

基于 MIC 的深度置信网络的实现主要分为两个阶段: 预处理阶段和训练阶段。预处理阶段是利用 MIC 预处理数据, 提取相关度较高的属性, 根据公式  $MIC_i = MIC(v(:, i), v'(:, i))$  计算出对应属性间的最大信息系数, 并对得到的一组  $MIC_i$  值进行排序,  $MIC_i$  值越大说明属性间相关度越大, 对结果的影响更大<sup>[21,30]</sup>。根据预先设定的阈值选取出符合要求的所有属性, 构成新样本。训练阶段分为无监督学习和有监督学习, 无监督学习过程采用 CD 算法对 RBM 进行采样学习, 选择  $V_{MIC}$  作为网络评价的新指标, 其中  $V_{MIC}$  用来度量训练样本  $S$  与重构样本  $S'$  间的相关度,  $V_{MIC}$  的值越大说明训练样本与重构样本间的相关度越高, 隐层节点能更好地代表可视节点, 网络对训练样本的似然度也就越大。通过无监督学习方法得到网络初始参数相比随机初始化网络参数会取得更好的识别效果; 有监督学习过程采用 BP 网络误差反向传播规则微调网络。

#### 4 实验结果与分析

本文实验基于 MNIST 和 USPS 手写数字数据集, 其中数据集 MNIST 是 0–9 的手写阿拉伯数字样本, 实验中样本都是  $28 \times 28$  的标准灰度图像; 数据集 USPS 是 0–9 的手写阿拉伯数字样本, 实验中样本都是  $16 \times 16$  的标准灰度图像。为得到更高效的学习结果, 实验过程采用小批量数据模式进行学习<sup>[20]</sup>。其中, MNIST 数据集选择 60000 个训练样本和 10000 个测试数据, 小批量数据大小为 100; USPS 数据集选择 7290 个训练样本和 2000 个测试数据, 小批量数据大小为 10。

分别设定 DBNs 隐藏节点数为 100 和 500, 在 MNIST 数据集上进行一次采样学习得到重构误差与样本间最大信息系数  $V_{MIC}$  的关系, 如图 3 所示。

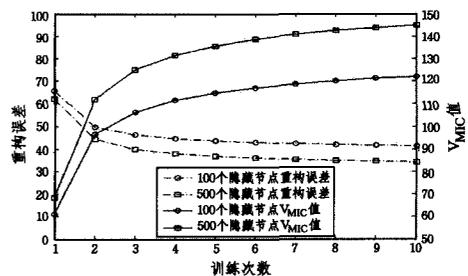
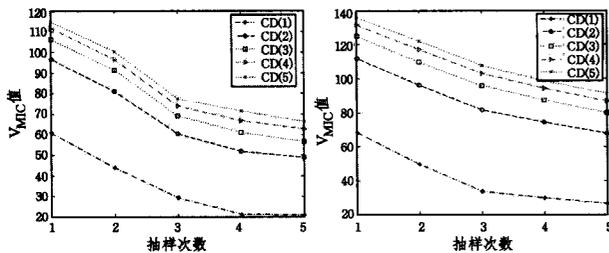


图 3 MNIST 数据集重构误差与  $V_{MIC}$  的关系

由图 3 可知,给定隐藏节点的情况下,随着训练次数的增加,重构误差在不断减小,而样本间最大信息系数  $V_{MIC}$  在不断增大。因此,重构误差和样本间最大信息系数  $V_{MIC}$  成反比,即重构误差越小, $V_{MIC}$  值越大,训练样本  $v$  和重构样本  $v'$  间的关联度越大,隐藏节点能更好地表示可视节点;重构误差越大, $V_{MIC}$  值越小,训练样本  $v$  和重构样本  $v'$  间的关联度也就越小,隐藏节点表示可视节点误差就越大。重构误差作为网络评价指标并不可靠<sup>[20]</sup>,而 MIC 基于信息熵和条件熵较稳健,不易受异常值影响<sup>[29,30]</sup>,可作为评价 RBM 的新指标。

RBM 网络的训练过程采用对比散度(Contrastive Divergence algorithm, CD)<sup>[19]</sup>算法对网络进行多次 Gibbs 采样学习,在不同隐藏节点和不同训练次数下  $V_{MIC}$  值的变化情况如图 4 所示。



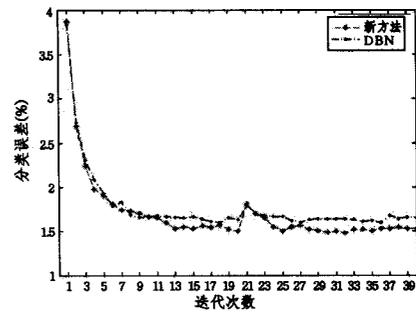
(a) 100 个隐藏节点抽样次数与  $V_{MIC}$  的关系 (b) 500 个隐藏节点抽样次数与  $V_{MIC}$  的关系图

图 4 MNIST 抽样次数与  $V_{MIC}$  的关系

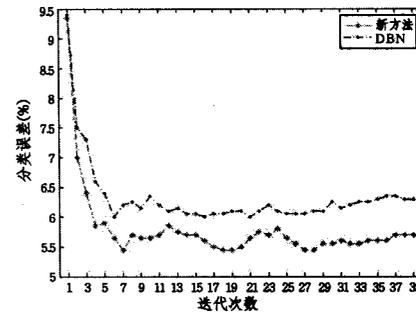
图 4(a)是隐藏节点为 100 的情况下抽样次数与  $V_{MIC}$  的关系图,图 4(b)是隐藏节点为 500 的情况下抽样次数与  $V_{MIC}$  的关系图,其中  $CD(i)$  ( $i=1,2,3,4,5$ ) 表示训练第  $i$  次的结果。由图可知,当确定隐藏节点数与抽样次数时, $V_{MIC}$  随着训练次数的增加而增大,而当确定隐藏节点数和训练次数时, $V_{MIC}$  随着抽样次数的增加而减小。因  $V_{MIC}$  值越大,网络对数据的拟合程度越好,根据图 4 可知当确定隐藏节点数和训练次数时,采样一次就可使得  $V_{MIC}$  值达到最大,即网络对输入数据的拟合程度达到最佳结果,在文献[32]中也证明了当采样一次时结果最佳。在后续实验中,对整个网络学习时,采用 CD 算法对网络进行一次采样学习,并且将  $V_{MIC}$  作为学习网络时的评价指标。

在基于 MIC 的深度置信网络中,在 MNIST 数据集上建立可视层→隐层→重构层(784→500→784)网络,采用 RBM 学习策略对该网络进行学习得到重构样本,计算训练样本与重构样本对应属性的 MIC 值,根据 MIC 值越大相关度越大,提取出相关度较高的属性组成新样本  $S_1$ 。在 USPS 数据集上建立可视层→隐层→重构层(256→160→256)网络,采用在 MNIST 数据集上相同的处理策略,提取出相关度较高的属性组成新样本  $S_2$ 。对上述两个数据集得到的新样本  $S_1$  和  $S_2$ ,分别构建新的 DBNs,并按照如图 2 所示的训练过程进行学习,最后分别将各自的测试样本应用到已学到的相应网络上进行测试,具体结果如图 5 所示。

图 5(a)示出在 MNIST 数据集上的结果,图 5(b)示出在 USPS 数据集上的结果。由图 5 可知,相比传统 DBNs,基于 MIC 的深度置信网络算法效果更好,其分类误差率更低。



(a) MNIST 数据集分类误差



(b) USPS 数据集分类误差

图 5 MNIST 和 USPS 分类误差

分别基于 MNIST 数据集和 USPS 数据集,列举了利用本文提出的基于 MIC 的深度置信网络算法和其他经典的算法如深度置信网络算法、BP 神经网络算法、支持向量机的结果比较,如表 1 和表 2 所列。

表 1 MNIST 数据集上不同方法的分类误差

| 方法             | 网络规模<br>(输入层-隐含层-输出层) | 分类误差(%) |
|----------------|-----------------------|---------|
| 基于 MIC 的深度置信网络 | 614-500-10            | 1.69    |
| 深度置信网络         | 784-500-10            | 1.78    |
| BP 神经网络        | 784-500-10            | 3.98    |
| 支持向量机          | 无                     | 5.56    |

表 2 USPS 数据集上不同方法的分类误差

| 方法             | 网络规模<br>(输入层-隐含层-输出层) | 分类误差(%) |
|----------------|-----------------------|---------|
| 基于 MIC 的深度置信网络 | 252-160-10            | 5.78    |
| 深度置信网络         | 256-160-10            | 6.32    |
| BP 神经网络        | 256-160-10            | 7.24    |
| 支持向量机          | 无                     | 5.85    |

由表 1 和表 2 的结果可知,基于 MIC 的深度置信网络算法分类误差低于其他算法,在 MNIST 数据集上相比深度置信网络降低 5.06%,相比 BP 神经网络降低 57.54%,相比支持向量机降低 69.60%;在 USPS 数据集上相比深度置信网络降低 8.54%,相比 BP 神经网络降低 20.17%,相比支持向量机降低 1.20%。综上可知,将 MIC 用于数据预处理是可行的。

**结束语** 本文借鉴了 MIC 能够表达两个变量间的相关度性质,将其引入到传统 DBNs 算法中,改进传统 DBNs 网络评价指标和预处理过程。传统 DBNs 用重构误差作为其评价指标,但它易受异常值的影响,从而影响网络评价。而变量间的 MIC 基于信息熵和条件熵不易受异常值影响,更适合作为评价网络的指标。根据属性间 MIC 值来筛选相关度较高的

属性,即对样本数据进行预处理,提取出相关度较高的特征作为新样本。因此在本文中提出一种基于 MIC 的深度置信网络,并在数据集 MNIST 和 USPS 上进行实验,结果表明 MIC 可代替重构误差用作网络的评价标准,且 MIC 预处理使基于 MIC 的深度置信网络算法能有效地降低分类误差率。接下来将研究在训练海量的数据时,如何通过 MIC 快速而精确地对海量数据提取相关度较高的属性并且在网络训练过程中能够灵活地调整参数。

## 参 考 文 献

- [1] Catanzaro B, Sundaram N, Keutzer K. Fast support vector machine training and classification on graphics processors[C]// Helsinki Finland, ACM, 2008;104-111
- [2] Pradhan B, Lee S. Regional landslide susceptibility analysis using back-propagation neural network model at Cameron Highland, Malaysia [J]. Landslides, 2010, 7(1):13-30
- [3] Yoshua B. Learning Deep Architectures for AI [J]. Foundations and Trends in Machine Learning, 2009, 2(1):1-127
- [4] George D, Marc' Aurelio R, Hinton G. Phone Recognition with the Mean-Covariance Restricted Boltzmann Machine[C]// 24th Annual Conference on Neural Information Processing Systems. 2010;1-9
- [5] Ragheb W, Ali L. Handwritten Digit Recognition using Sparse Deep Architectures[C]// 2014 9th International Conference on Intelligent Systems; Theories and Applications. 2014;1-6
- [6] Peng Bo, Zang Di. Vehicle Logo Recognition Based on Deep Learning[J]. Computer Science, 2015, 42(4):268-273 (in Chinese)  
彭博,臧笛. 基于深度学习的车标识别方法研究[J]. 计算机科学, 2015, 42(4):268-273
- [7] Sarikaya R, Hinton G, Deoras A. Application of Deep Belief Networks for Natural Language Understanding[J]. ACM Transactions on Audio, Speech and Language Processing, 2014, 22(4):778-784
- [8] Guo Li-li, Ding Shi-fei. Research Progress on Deep Learning[J]. Computer Science, 2015, 42(5):28-33 (in Chinese)  
郭丽丽,丁世飞. 深度学习研究进展[J]. 计算机科学, 2015, 42(5):28-33
- [9] Abdel O, Mohamed A, Jiang Hui, et al. Convolutional Neural Networks for Speech Recognition[J]. Audio, Speech, and Language Processing, 2014, 22(10):1533-1545
- [10] Vidya R, Nasira G M, Priyanka R. Sparse Coding: A Deep Learning Using Unlabeled Data for High-Level Representation [C]// Computing and Communication Technologies. 2014;124-127
- [11] Jiang Xiao-juan, Zhang Ying-hua, Zhang Wen-sheng, et al. A novel sparse auto-encoder for deep unsupervised learning [C]// 2013 Sixth International Conference on Advanced Computational Intelligence. 2013;256-261
- [12] Liu Jian-wei, Liu Yuan, Luo Xiong-lin. Research and Development on Boltzmann Machine[J]. Journal of Computer Research and Development, 2014, 51(1):1-16 (in Chinese)  
刘建伟,刘媛,罗雄麟. 玻尔兹曼机研究进展[J]. 计算机研究与发展, 2014, 51(1):1-16
- [13] Courville A, Bergstra J, Bengio Y. A Spike and Slab Restricted Boltzmann Machine[C]// Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics. 2011;233-241
- [14] Bengio Y, Lamblin P, Popovici D, et al. Greedy layer-wise training of deep networks[J]. Advances in Neural Information Processing System, 2007, 19:153-160
- [15] Hinton G, Osindero S, Teh Y. A fast learning algorithm for deep belief nets[J]. Neural Comput, 2006, 18(7):1527-1554
- [16] Hinton G, Salakhutdinov R. Reducing the Dimensionality of Data with Neural Networks[J]. Science, 2006, 313:504-507
- [17] Liu Yan, Zhou Shu-sen, Chen Qing-cai. Discriminative deep belief networks for visual data classification[J]. Pattern Recognition, 2011, 44(10/11):2287-2296
- [18] Zhu Ming, Wu Yan. A novel deep model for image recognition [C]// 2014 5th IEEE International Conference on Software Engineering and Service Science. 2014;373-376
- [19] Hinton G. Training products of experts by minimizing contrastive divergence[J]. Neural Comput, 2002, 14(8):1771-1800
- [20] Hinton G. A practical guide to training Restricted Boltzmann Machines[J]// Momentum, 2010, 9(1):599-619
- [21] David N, et al. Detecting Novel Associations in Large Data Sets [J]. Science, 2011, 334(6062):1518-1524
- [22] Zhao Xi, Deng Wei, Shi Yong. Feature Selection with Attributes Clustering by Maximal Information Coefficient [J]. Procedia Computer Science, 2013, 17:70-79
- [23] LeCun Y, Chopra S, Hadsell R, et al. A tutorial on energy-based learning[C]// Conference on Predicting Structured Data. 2006;191-246
- [24] Dahl G, Dong D, Li D, et al. Context-Dependent Pre-trained Deep Neural Networks for Large Vocabulary Speech Recognition[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2011(20):30-42
- [25] Larochelle H, Mande M, Pascanu R, et al. Learning Algorithms for the Classification Restricted Boltzmann Machine[J]. Journal of Machine Learning Research, 2012(13):643-669
- [26] Roux N, Bengio Y. Representational power of Restricted Boltzmann Machines and deep belief networks[J]. Neural Comput, 2008(6):1631-1649
- [27] Robert M. Entropy and Information Theory [M]. Springer-Verlag, New York, 2011
- [28] David N, Yakir A, Michael M, et al. Equitability Analysis of the Maximal Information Coefficient with Comparisons[J]. Computer Science, 2013;1-22
- [29] Justin B, Gurinder S. Equitability, mutual information, and the maximal information coefficient [C]// Proceedings of the National Academy of Sciences of the United States of America. 2014;3354-3359
- [30] Speed T. A Correlation for the 21st Century [J]. Science, 2011, 334(6062):1502-1503
- [31] Tan Qiu-heng, Jiang Hang-jin, Ding Yi-ming. Model Selection Method Based on Maximal Information Coefficient of Residual [J]. Acta Mathematica Scientia, 2014, 34(2):579-592
- [32] Liu Jian-wei, Chi Guang-hui, Luo Xiong-lin. Contrastive Divergence Learning for the Restricted Boltzmann Machine [C]// 2013 Ninth International Conference on Natural Computation. 2013;18-22