

文章编号:1001-9081(2016)09-2521-05

doi:10.11772/j.issn.1001-9081.2016.09.2521

## 基于多新息理论的深度信念网络算法

李萌, 秦品乐\*, 李传朋

(中北大学 计算机与控制工程学院, 太原 030051)

(\*通信作者电子邮箱 QPL@nuc.edu.cn)

**摘要:**针对深度信念网络(DBN)算法在采用反向传播修正网络的连接权值和偏置的过程中,容易产生梯度小、学习率低、误差收敛速度慢等问题,提出一种结合多新息理论对标准DBN算法进行改进的算法,即多新息DBN(MI-DBN)。MI-DBN算法是对标准DBN算法中反向传播的过程重新建模,使得算法在原先只利用单个新息的情况下,扩展为能够充分利用之前多个周期的新息,从而大幅提高误差收敛速度。通过实验对MI-DBN算法和其他分类算法进行了数据集分类的比较,实验结果表明,MI-DBN算法相较其他分类算法,其误差收敛速度较快,而且最终对MNIST数据集和Caltech101数据集的识别中误差结果相对更小。

**关键词:**深度信念网络算法; 误差收敛速度; 多新息理论; 反向传播

中图分类号: TP389.1 文献标志码:A

### Deep belief network algorithm based on multi-innovation theory

LI Meng, QIN Pingle, LI Chuanpeng

(School of Computer Science and Control Engineering, North University of China, Taiyuan Shanxi 030051, China)

**Abstract:** Aiming at the problem of small gradient, low learning rate, slow convergence of error during the process of using Deep Belief Network (DBN) algorithm to correct connection weight and bias of network by the method of back propagation, a new algorithm called Multi-Innovation DBN (MI-DBN) was proposed based on combination of standard DBN algorithm with multi-innovation theory. The back propagation process in standard DBN algorithm was remodeled to make full use of multiple innovations in previous cycles, while the original algorithm can only use single innovation. Thus, the convergence rate of error was significantly increased. MI-DBN algorithm and other representative classifiers were compared through experiments of datasets classification. Experimental results show that MI-DBN algorithm has a faster convergence rate than other sorting algorithms; especially when identifying MNIST and Caltech101 dataset, MI-DBN algorithm has the fewest inaccuracies among all the algorithms.

**Key words:** Deep Belief Network (DBN) algorithm; error convergence rate; multi-innovation theory; back-propagation

## 0 引言

深度学习<sup>[1]</sup>近几年来在人工智能领域(语音识别、自然语言处理、计算机视觉、图像与视觉分析、多媒体等)取得重要突破。深度学习中的模型有多种,有一种无监督的学习算法——深度信念网络(Deep Belief Network, DBN)学习算法<sup>[2-3]</sup>,这种方法可以创建一些网络层来检测特征而不使用带标签的数据,这些网络层可以用来重构或者对特征检测器的活动进行建模。通过预训练过程,深度网络的权值可以被初始化为有意义的值,然后一个输出层被添加到该网络的顶部,并且使用标准的反向传播算法进行微调,这个工作对手写数字的识别效果显著。

Hinton 等<sup>[2]</sup>将 DBN 应用于对 MNIST 数据集的数字识别任务上。他提出了网络结构为 784-500-500-2000-10 的 DBN 模型,其中第一层的 784 个特征是由 MNIST 数据集中图像的像素  $28 \times 28$  而得来的,最后一层是含有标签的 10 个单元,中间三层是函数随机二进制神经元的隐藏层单元。Hinton 在文

章中提到该算法能达到 1.25% 的分类错误率。文献[4]在 Hinton 等<sup>[2]</sup>提出的标准 DBN 基础上改进了受限玻尔兹曼机(Restricted Boltzmann Machine, RBM)的采样方法,将原始的对比散度(Contrastive Divergence, CD)算法改为持续自由能量对比散度(Free Energy in Persistent Contrastive Divergence, FEPGD)算法,最终 MNIST 数据集的识别结果达到 1.11% 的分类错误率。文献[5]中 Liu 提出了一种新的基于标准 DBN 的分类器判别式深度信念网络(Discriminative Deep Belief Network, DDBN),整合了 DBN 的抽象能力和反向传播策略的鉴别能力,其除了对 MNIST 数据集进行识别之外,还对 Caltech 101 数据集进行分类识别。

Hinton 等<sup>[6]</sup>提出的 DBN 是非线性模型,用来进行特征的提取和降维,是一种在最后一层重构产生的特征可见数据的能力模型。一般的神经网络被用于多层时,极容易被困在局部极小和其性能也较低,因此,确定神经网络的权值的初始值至关重要,而 DBN 很好地解决了权值初始化的问题。本文引进丁锋等<sup>[7]</sup>和 Ding<sup>[8]</sup>提出的多新息辨识理论,提出了一种基

收稿日期:2016-02-24;修回日期:2016-03-30。 基金项目:山西省自然科学基金资助项目(2015011045)。

**作者简介:**李萌(1990—),女,山西运城人,硕士研究生,主要研究方向:计算机视觉、数字图像处理、深度学习; 秦品乐(1978—),男,山西太原人,副教授,博士,CCF 会员,主要研究方向:大数据、机器视觉、三维重建; 李传朋(1991—),男,山东济南人,硕士研究生,主要研究方向:计算机视觉、数字图像处理、深度学习。

于多新息理论的深度信念网络学习算法,改进算法在原有的基础上提高了权值误差的收敛速度,提高了识别 MNIST 数据集的精确度。

## 1 DBN 模型

DBN 是建立一个观察数据和标签之间的联合分布的概率生成模型,而传统的神经网络则是辨别模型。DBN 结构由若干层 RBM<sup>[9]</sup>组成,在每一层中对输入向量进行非线性变换,产生的输出作为下一层的输入向量。一个 DBN 的连接过程是通过自顶向下的生成权值来确定的,RBM 更易于连接权值的学习<sup>[10]</sup>。在分层技术训练的 DBN 模型中,一个标签集被附加到顶层,通过自下向上学习到的识别权值而获得一个网络的分类面,使用反向传播技术调整权值优化分类。结构如图 1 所示。

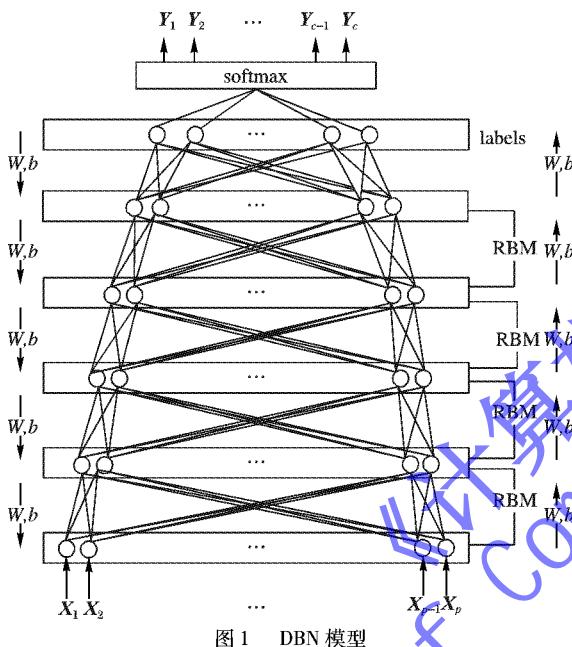


图 1 DBN 模型

DBN 模型进行特征的提取需要三个过程,分别是预训练过程、微调过程和测试过程,以下将详细介绍三个过程。

### 1.1 预训练过程

预训练是如图 1 中自底向上识别权值与偏置的过程,此过程将输入数据经过网络的每一层提取特征后获得激励响应,将其添加至输出层,具体过程如下:

1) 先单独无监督地训练每一层 RBM 网络,在确保特征向量映射到不同特征空间的同时,尽可能多地保留特征信息。

2) 在 DBN 的最顶层设置一个 logistic 分类器,用于接受 RBM 的输出特征向量,将其作为 logistic 回归网络的输入特征向量,进行有监督的训练。而且每一个 RBM 网络层能够确保自身层内的权值对该层特征向量映射达到最优,而不是对整个 DBN 的特征向量映射达到最优,所以反向传播会将错误信息自上向下传播至每一层 RBM,从而需要微调整整个 DBN。RBM 网络训练模型过程的目的是对 logistic 回归网络的权值参数的初始化,使 DBN 克服 logistic 回归网络因随机初始化权值参数而容易陷入局部最优和训练时间过长的缺点。

### 1.2 微调过程

微调是如图 1 自顶向下更新权值和偏置的过程,此过程

可以大幅度提升一个自编码神经网络的性能表现,将网络的所有层视为一个模型,在每次迭代中,网络中的所有权值和偏置都会被优化。具体过程如下:

1) 进行前馈传导计算,得到  $L_2$  层、 $L_3$  层直至输出层  $L_n$  的激活值  $a_i^{(l)}$ 。

$a_i^{(l)}$  表示第  $l$  层的第  $i$  单元的激活值(输出值),对于给定的参数集合  $W, b$ , 网络可以按照函数  $h_{w,b}(x)$  来计算输出结果,  $h_{w,b}(x)$  表示相应的期望输出结果。前向传播的计算步骤如下:

$$z^{(l+1)} = W^{(l)} a^{(l)} + b^{(l)} \quad (1)$$

$$h_{w,b}(x) = a^{(l+1)} = f(z^{(l+1)}) \quad (2)$$

$$f(z^{(l+1)}) = \text{sigmoid}(z^{(l+1)}) = \frac{1}{1 + e^{-z^{(l+1)}}} \quad (3)$$

其中用到 sigmoid 函数作为激活函数:

$$f(z) = \text{sigmoid}(z) = \frac{1}{1 + e^{-z}} \quad (4)$$

对于单个样例  $(x, y)$ ,其非线性代价函数为:

$$J(W, b; x, y) = \frac{1}{2} \| h_{w,b}(x) - y \|^2 \quad (5)$$

其中  $y$  是输出值,当计算输出节点时,网络训练规则将使  $J$  在每个训练循环按梯度下降,每次迭代对参数  $W$  和  $b$  的更新如下:

$$W_{ij}^{(l)} = W_{ij}^{(l)} - \alpha \frac{\partial J(W, b)}{\partial W_{ij}^{(l)}} \quad (6)$$

$$b_i^{(l)} = b_i^{(l)} - \alpha \frac{\partial J(W, b)}{\partial b_i^{(l)}} \quad (7)$$

$\alpha$  是学习率;参数  $W_{ij}^{(l)}$  表示第  $l$  层的第  $j$  单元与第  $l+1$  层的第  $i$  单元之间的权重; $b_i^{(l)}$  表示第  $l$  层的第  $i$  单元的偏置项。对于第  $l$  层的每一个节点  $i$ ,计算其“误差” $\delta_i^{(l)}$ ,表明了该节点对最终输出值的误差产生了多少影响。基于第  $l+1$  层节点误差的加权平均值来计算  $\delta_i^{(l)}$ ,这些节点以  $a_i^{(l)}$  作为输入。对于误差  $\delta_i^{(l)}$  有以下公式:

$$\frac{\partial J(W, b; x, y)}{\partial W_{ij}^{(l)}} = \frac{\partial J(W, b; x, y)}{\partial z_i^{(l+1)}} \frac{\partial z_i^{(l+1)}}{\partial W_{ij}^{(l)}} \quad (8)$$

$$\frac{\partial J(W, b; x, y)}{\partial b_i^{(l)}} = \frac{\partial J(W, b; x, y)}{\partial z_i^{(l+1)}} \frac{\partial z_i^{(l+1)}}{\partial b_i^{(l)}} \quad (9)$$

由式(1)可知:

$$z_i^{(l+1)} = \sum_{k=1}^{s_l} W_{ik}^{(l)} a_k^{(l)} + b_i^{(l)} \quad (10)$$

由式(7)、(9)可知:

$$\begin{cases} \frac{\partial z_i^{(l+1)}}{\partial W_{ij}^{(l)}} = a_j^{(l)} \\ \frac{\partial z_i^{(l+1)}}{\partial b_i^{(l)}} = 1 \end{cases} \quad (11)$$

由式(7)、(8)、(10)可知:

$$\frac{\partial J(W, b; x, y)}{\partial W_{ij}^{(l)}} = \frac{\partial J(W, b; x, y)}{\partial z_i^{(l+1)}} a_j^{(l)} \quad (12)$$

$$\frac{\partial J(W, b; x, y)}{\partial b_i^{(l)}} = \frac{\partial J(W, b; x, y)}{\partial z_i^{(l+1)}} \quad (13)$$

故而令:

$$\delta_i^{(l)} := \frac{\partial J(W, b; x, y)}{\partial z_i^{(l)}} \quad (14)$$

2) 由式(4)、(13) 可知对于输出层( $n_l$  层), 每个输出单元  $i$  的误差公式可推导为:

$$\delta_i^{(n_l)} = \frac{\partial J(W, b; x, y)}{\partial z_i^{(n_l)}} = -(y_i - a_i^{(n_l)}) \cdot f'(z_i^{(n_l)}) \quad (15)$$

3) 由式(4)、(13) 可知对  $l = n_l - 1, n_l - 2, n_l - 3, \dots, 2$  的各个层, 第  $l$  层的第  $i$  个节点的误差公式可推导为:

$$\delta_i^{(l)} = \left( \sum_{j=1}^{s_{l+1}} W_{ji}^{(l)} \delta_j^{(l+1)} \right) \cdot f'(z_i^{(l)}) \quad (16)$$

4) 计算所需的偏导数, 计算方法如下:

$$\frac{\partial J(W, b; x, y)}{\partial W_{ij}^{(l)}} = a_j^{(l)} \delta_i^{(l+1)} \quad (17)$$

$$\frac{\partial J(W, b; x, y)}{\partial b_i^{(l)}} = \delta_i^{(l+1)} \quad (18)$$

由式(5)、(6) 可知参数  $W$  和  $b$  的更新可以写为:

$$W_{ij}^{(l)} = W_{ij}^{(l)} - \alpha a_j^{(l)} \delta_i^{(l+1)} \quad (19)$$

$$b_i^{(l)} = b_i^{(l)} - \alpha \delta_i^{(l+1)} \quad (20)$$

### 1.3 测试过程

经过预训练和微调过程, DBN 模型已经逐层提取特征, 同时参数  $W$  和  $b$  已经优化。通过数据集的测试样本对 DBN 进行测试。

## 2 基于多新息理论的 DBN 算法

针对标准 DBN 算法中更新参数权值  $W$  和偏置  $b$  时, 均只用到当前周期的数据, 之前周期的数据将全部遗失的问题, 本文引进多新息理论。多新息辨识已经发展为一个全新的辨识领域, 能充分利用和扩展辨识新息, 通过对辨识新息的扩展, 从标量新息到新息向量。由于该方法能够提高梯度下降的速率, 所以已经广泛应用于多种模型参数的估计中。

### 2.1 多新息理论

常见的辨识算法一般都是单新息辨识方法<sup>[11]</sup>, 也就是只利用单步的新息去修正参数, 对于标量系统<sup>[12]</sup>:

$$\mathbf{y}(t) = \boldsymbol{\varphi}^T(t) \boldsymbol{\theta} + \mathbf{v}(t) \quad (21)$$

其中:  $\mathbf{y}(t)$  表示系统的输出;  $\boldsymbol{\varphi}^T(t)$  是由系统输入数据和输出数据组成的向量;  $\boldsymbol{\theta}$  是最终需要辨识的参数, 这里是向量形式;  $\mathbf{v}(t)$  是均值为零的系统随机噪声, 表示该系统存在噪声。

针对参数向量  $\boldsymbol{\theta}$ , 利用随机梯度辨识算法有以下形式:

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \mathbf{L}(t) \mathbf{e}(t) \quad (22)$$

其中:  $\mathbf{L}(t)$  为增益向量;  $\mathbf{e}(t)$  为标量新息, 即单新息。标量新息  $e(t)$  是用来描述  $t$  时刻的输出预报误差, 可表示为:

$$\mathbf{e}(t) := \mathbf{y}(t) - \boldsymbol{\varphi}^T(t) \hat{\boldsymbol{\theta}}(t-1) \in \mathbf{R} \quad (23)$$

对于  $\hat{\boldsymbol{\theta}}(t)$ , 即为  $t$  时刻的参数估计在  $\hat{\boldsymbol{\theta}}(t-1)$  的基础上递推出来的, 依靠增益向量  $\mathbf{L}(t)$  与标量新息  $e(t)$  的乘积来修正  $t-1$  时刻的辨识参数。

将其拓展, 根据式(21)、(22), 将标量新息  $e(t)$  扩展为向量形式, 便称为多新息。同时, 系统中其他的向量或者矩阵也要随之变化, 将增益向量  $\mathbf{L}(t) \in \mathbf{R}^n$  扩展为矩阵形式。由此得到了多新息辨识方法, 其算法就可以写作:

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \mathbf{I}'(p, t) \mathbf{E}(p, t) \quad (24)$$

其中:  $\mathbf{I}'(p, t) \in \mathbf{R}^{n \times p}$  为增益矩阵,  $\mathbf{E}(p, t) \in \mathbf{R}^p$  为新息向量,  $p \geq 1$  为新息长度。当  $p$  等于 1 时, 就退化为标量新息。

多新息理论将之前的部分有用信息充分利用, 能够有较

好的辨识效果。这种方法增加了算法本身的运算量, 但却提高了估计收敛速度, 以微小的时间复杂度换取实验中精度的提高。多新息理论的详细推导过程和证明可以查看参考文献 [13]。

### 2.2 结合多新息的 DBN 算法——MI-DBN

DBN 经过无监督和逐层的训练<sup>[14-15]</sup>, 使得在利用反向传播算法对网络进行全局优化之前, 网络参数能有一个好的初始值, 一个较好的局部极小点就在训练完成时达到。在学习系统中, 有可能数以百万计的样本和权值, 在最上层加上带有标签的样本, 用来训练机器。如图 1 所示, 为了正确地调整权值向量, 用反向传播算法计算每个权值的梯度向量, 表示如果权值增加了一个很小的值, 那么误差会增加或减少的量。反向传播算法通过多层网络的每一层可以被重复地用于传播梯度; 从该网络的最顶层的输出一直到最底层, 每层的输入导数求解完成之后, 就可以求解每一层权值的梯度矢量, 权值向量在梯度矢量的相反方向上进行调整。随机梯度算法的计算量小, 收敛速度慢, 为了提高随机梯度辨识方法的收敛速度, 引入新息长度。图 1 中, 在自顶向下过程中修正每层网络权值和偏置, 所有层更新完成时即是一个周期, 在此过程中每次的权值和偏置更新都是基于上一个周期的对应值, 在 DBN 原有更新权值和偏置的基础上, 对每层网络对应的前几个周期的值进行充分利用, 将单新息扩展为新息向量, 从而提高误差的收敛速率。

在对时变系统进行辨识时, 如果网络的权值按照某种规律动态变化, 那么网络的输入输出映射关系也必然会随着时间而变化, 从而有可能对时变系统进行辨识。由于权值随时间变化的规律很难获得, 故将网络微调的整个周期作为新息长度  $p$ , 周期数作为时间  $t$ , 定义正整数序列  $\{t_0, t_1, \dots, t_s\}$ , 满足  $0 < t_0 < t_1 < t_2 < \dots < t_s$ , 且  $1 \leq t_s^* = t_s - t_{s-1}$ 。在迭代间隔  $t_s^*$  秒内, 每  $1/N$  s 都可能得到第  $l$  层第  $i$  个节点的有用数据信息, 利用  $p$  组数据信息产生新息向量  $\mathbf{E}(p, t)$  对未知参数  $\boldsymbol{\theta}$  进行修正。此时,  $t = t_s - p + 1$  到  $t = t_s$  共有  $p$  组样本数据, 则第  $l$  层可视层的第  $j$  节点在  $t$  时刻的输出为:

$$\mathbf{A}_j^{(l)}(p, t_s) = [a_j^{(l)}(t_s), a_j^{(l)}(t_s - 1), a_j^{(l)}(t_s - 2), \dots, a_j^{(l)}(t_s - p + 1)]^T \quad (25)$$

第  $l$  层第  $i$  个节点在  $t$  时刻的误差为:

$$\bar{\delta}_i^{(l)}(p, t_s) = [\delta_i^{(l)}(t_s), \delta_i^{(l)}(t_s - 1), \delta_i^{(l)}(t_s - 2), \dots, \delta_i^{(l)}(t_s - p + 1)] \quad (26)$$

将式(25)、(26) 代入式(19) 中, 权值的修正式为:

$$W_{ij}^{(l)}(t_s) = W_{ij}^{(l)}(t_s - 1) - \alpha A_j^{(l)}(p, t_s) \bar{\delta}_i^{(l+1)}(p, t_s) \quad (27)$$

$$b_i^{(l)}(t_s) = b_i^{(l)}(t_s - 1) - \alpha \bar{\delta}_i^{(l+1)}(p, t_s) \quad (28)$$

将式(26) 代入式(19) 中, 偏置的修正式为:

$$b_i^{(l)}(t_s) = b_i^{(l)}(t_s - 1) - \alpha \bar{\delta}_i^{(l+1)}(p, t_s) \quad (29)$$

本文提出的 MI-DBN 与标准 DBN 相比, 具有以下优点:

- 1) 每次更新权值和偏置时, 标准 DBN 只是使用了当前周期的新息, 而本文提出的 MI-DBN 不仅使用了当前周期的新息, 还是用了过去几个周期的新息, 这样使得误差的收敛速度提高。
- 2) 在标准 DBN 算法中加入多新息能够提高误差的收敛速度, 能够改进算法的精度; 但并不是加入的多新息越多越

好,加入过多的多新息会增加算法计算量,使得实时性降低,所以要慎重考虑加入多新息的数量。

### 3 实验与分析

本文将用到两个数据集,分别是 MNIST 数据集<sup>[16]</sup>和 Caltech101 数据集<sup>[17]</sup>。用 Deep Learning 中的工具箱 DeeBNet toolbox<sup>[18]</sup>来实现实验结果。实验中,将 MI-DBN 与 K 最近邻算法(K-Nearest Neighbor, KNN)、支持向量机(Support Vector Machine, SVM)<sup>[19]</sup>、直推式支持向量机(Transductive Support Vector Machine, TSVM)<sup>[19]</sup>、神经网络(Neural Network, NN)<sup>[20]</sup>和深度信念网络-社区成分分析(Deep Belief Network-review Neighborhood Components Analysis, DBN-rNCA)<sup>[21]</sup>的实验结果进行比较。

#### 3.1 实验环境

进行实验的软件环境:Windows 7 系统,Visual Studio 2013,Matlab 2013。硬件平台:2.80 GHz Intel CPU E5-2680 V2,32 GB 内存。

#### 3.2 MNIST 数据集

本文实验所用数据是基于手写数字 MNIST 数据库的数据集,数据集中图像像素具有 0 到 255 之间的离散值<sup>[22]</sup>,图像的像素值归一化在 0~1。其所用网络结构为 784-500-500-2000-10,其中该数据集被划分为训练和测试两部分,分别包括 60 000 个训练样本和 10 000 个测试样本。如图 2 所示。



图 2 MNIST 数据集示例

本文实验所涉及的评判标准主要有以下几个方面:

1) 错误率。错误率指标主要用于衡量不同分类模型对于数据集分类的错误程度,该指标是用来衡量分类错误程度的一个重要指标,直接关系着算法的可靠性。在 Matlab 中使用公式计算错误率:

$$\text{err} = \text{mean}(\text{zeros}((\text{epoch} - \text{numbatches}); (\text{epoch} - 1))) \times 100\%$$

其中:epoch 表示周期数, numbatches 表示批次数。

表 1 给出了不同算法对 MNIST 数据集分类的错误率<sup>[4]</sup>,其中标准 DBN 算法、含有 FEPDC 的 DBN 改建算法和本文中的 MI-DBN 算法均使用 60 000 个样本训练,10 000 个样本测试,50 个周期训练,150 个周期测试,用相同的网络结构,并且其相关参数学习率等也相同的条件下进行的实验。显然,加入多新息的 DBN 算法运算量加大,运行速度下降,但改进算法以时间耗费为代价获得了数据集分类错误率的降低。表 1

中的时间是指每个周期所用时间,其中下划线的数据是本文改进算法对 MNIST 数据分类后的错误率,可以看出,MI-DBN 算法有更小的错误率。

表 1 不同算法对于 MNIST 数据集的分类错误率

分类算法	时间/s	错误率/%
SVM	46	1.40
KNN	48	1.60
标准 DBN	43	1.24
FEPDC-DBN	79	1.11
MI-DBN	<u>153</u>	<u>0.54</u>

2) 误差。对实验中误差数据进行仿真,实验过程中的训练周期为 50,进行微调的周期为 150。图 3 是 MI-DBN 算法和其他分类器对 MNIST 数据集分类的误差曲线对比分析。如图 3 中最下边曲线表示在 150 个周期中本文的改进算法 MI-DBN 对 MNIST 数据集进行分类时其误差的曲线,上边的曲线从上到下依次是 SVM、KNN、DBN 和 FEPDC 算法在同样条件下对 MNIST 数据集进行分类时其误差的曲线。可以看出 MI-DBN 中 MNIST 数据集的误差曲线收敛得更快。

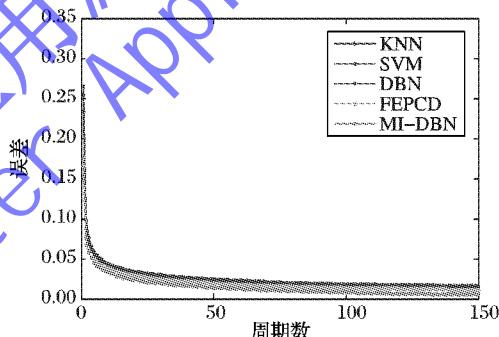


图 3 MNIST 数据集中误差对比分析

#### 3.3 Caltech 101 数据集

以上实验中验证了 MI-DBN 算法对 MNIST 数据集的分类时,其误差有明显地降低,本节则将 Caltech 101 数据集在 MI-DBN 算法和其他分类器算法上进行分类识别。Caltech 101 数据集包含有 101 类的图像,在本文的实验中,选用其中的 5 个类。在本次 Caltech 101 数据集实验中 MI-DBN 结构为 400-500-500-900-5,输入层是将图片转换为 20×20 像素即 400 个单元,输出层是 5 个单元,三个隐藏层分别是 500、500、900 个单元。如图 4 所示是 Caltech 101 数据集的不同类别。



图 4 Caltech 101 的 5 个不同类别

表 2 即是本次实验的误差分类错误率结果,由于本次实验的数据集样本数目少,类别小,因此实验结果与 MNIST 数据集相比较不是特别理想。

表2 不同算法对Caltech101数据集的分类错误率

分类算法	标签数	错误率/%
SVM	5	50.24
KNN	5	55.56
TSVM	5	50.13
NN	5	46.83
DBN-rNCA	5	56.89
DDBN	5	41.70
MI-DBN	5	40.21

如图5为MI-DBN算法和其他分类器对Caltech 101数据集分类的误差曲线对比分析。图5中最下边曲线表示本文的改进算法MI-DBN对Caltech 101数据集进行分类时其误差的曲线,上边的曲线从上到下依次是SVM、KNN、TSVM、NN、DBN-rNCA和DDBN算法在同样条件下对Caltech 101数据集进行分类时其误差的曲线。可以看出MI-DBN中Caltech 101数据集的误差曲线收敛得更快。

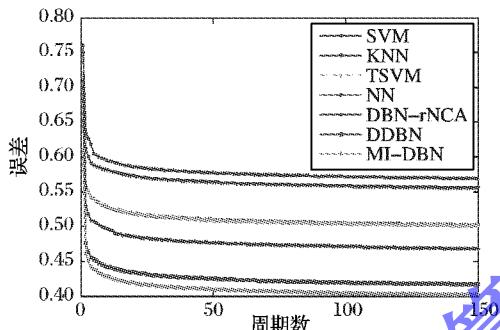


图5 Caltech101数据集中误差对比分析

## 4 结语

基于对DBN算法中的不足,本文提出了结合多新息随机梯度辨识的深度信念网络学习算法,在深度信念网络的基础上加上多新息理论算法。改进算法MI-DBN结合历史周期的数据信息,充分利用隐含在过去数据中的有用信息,使得每层网络的连接权值和偏置能够更快地达到预测值。仿真结果表明,误差的收敛速率提高,并且MI-DBN算法在对MNIST手写数据集和Caltech 101数据集的识别中,最终分类错误率有所降低。

## 参考文献:

- [1] LECUN Y, BENGIO Y, HINTON G E, et al. Deep learning [J]. Nature, 2015, 521(7553): 436–444.
- [2] HINTON G E, OSINDERO S, TEH Y W. A fast learning algorithm for deep belief nets [J]. Neural Computation, 2006, 18(7): 1527–1554.
- [3] LAROCHELLE H, ERHAN D, COURVILLE A, et al. An empirical evaluation of deep architectures on problems with many factors of variation [C]// ICML '07: Proceedings of the 2007 24th International Conference on Machine Learning. New York: ACM, 2007: 473–480.
- [4] KEYVANRAD M A, HOMAYOUNPOUR M M. Deep belief network training improvement using elite samples minimizing free energy [EB/OL]. [2015-11-22]. [http://xueshu.baidu.com/s?wd=paperuri%3A%282e0ed0ef0b45da606b7629105f1f7ed%29&filter=sc\\_long\\_sign&tn=SE\\_xueshusource\\_2kduw22v&sc\\_vurl=http%3A%2F%2Farxiv.org%2Fpdf%2F1411.4046v1&ie=utf-8&sc\\_us=](http://xueshu.baidu.com/s?wd=paperuri%3A%282e0ed0ef0b45da606b7629105f1f7ed%29&filter=sc_long_sign&tn=SE_xueshusource_2kduw22v&sc_vurl=http%3A%2F%2Farxiv.org%2Fpdf%2F1411.4046v1&ie=utf-8&sc_us=)
- [5] LIU Y, ZHOU S, CHEN Q. Discriminative deep belief networks for visual data classification [J]. Pattern Recognition, 2011, 44(10/11): 2287–2296.
- [6] HINTON G E, SALAKHUTDINOV R. Reducing the dimensionality of data with neural networks [J]. Science, 2016, 313(5786): 504–507.
- [7] 丁锋, 肖德云, 丁韬. 多新息随机梯度辨识方法[J]. 控制理论与应用, 2003, 20(6): 870–874. ( DING F, XIAO D Y, DING T. Multi-innovation stochastic gradient identification methods [J]. Control Theory and Applications, 2003, 20(6): 870–874. )
- [8] DING F. Several multi-innovation identification methods [J]. Digital Signal Processing, 2010, 20(4): 1027–1039.
- [9] HINTON G E. A practical guide to training restricted Boltzmann machines [C]// Neural Networks: Tricks of the Trade, LNCS 7700. Berlin: Springer, 2012: 599–619.
- [10] SWERSKY K, CHEN B, MARLIN B, et al. A tutorial on stochastic approximation algorithms for training restricted Boltzmann machines and deep belief nets [C]// Proceedings of the 2010 Information Theory and Applications Workshop. Piscataway, NJ: IEEE, 2010: 1–10.
- [11] 丁洁, 谢莉, 丁锋. 非均匀采样系统多新息随机梯度辨识性能分析[J]. 控制与决策, 2011, 26(9): 1338–1342. ( DING J, XIE L, DING F. Performance analysis of multi-innovation stochastic gradient identification for non-uniformly sampled systems [J]. Control and Decision, 2011, 26(9): 1338–1342. )
- [12] 丁锋. 系统辨识(6): 多新息辨识理论与方法[J]. 南京信息工程大学学报, 2012, 4(1): 1–28. ( DING F. System identification, Part F: multi-innovation identification theory and methods [J]. Journal of Nanjing University of Information Science and Technology, 2012, 4(1): 1–28. )
- [13] 丁锋, 杨家本. 衰减激励条件下确定性系统多新息辨识的收敛性分析[J]. 清华大学学报(自然科学版), 1998, 38(9): 111–115. ( DING F, YANG J B. Convergence of multi-innovation identification under attenuating excitation conditions for deterministic systems [J]. Journal of Tsinghua University (Science and Technology), 1998, 38(9): 111–115. )
- [14] LEE H, EKANADHAM C, NG A Y. Sparse deep belief net model for visual area V2 [EB/OL]. [2015-11-26]. <http://web.eecs.umich.edu/~honglak/nips07-sparseDBN.pdf>.
- [15] KRIZHEVSKY A. Learning multiple layers of features from tiny images [D]. Toronto: University of Toronto, 2009: 17.
- [16] LECUN Y, CORTEZ C. The MNIST database of handwritten digits [DB/OL]. [2011-12-20]. <http://yann.lecun.com/exdb/mnist/index.html>.
- [17] LI F F, FERGUS R, PERONA P. Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories [C]// CVPRW '04: Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2004, 12: 178.
- [18] KEYVANRAD M A, HOMAYOUNPOUR M M. A brief survey on deep belief networks and introducing a new object oriented Matlab toolbox (DeeBNetV2.2) [EB/OL]. [2015-11-06]. [https://www.researchgate.net/publication/264790642\\_A\\_brief\\_survey\\_on\\_deep\\_belief\\_networks\\_and\\_introducing\\_a\\_new\\_object\\_oriented\\_MATLAB\\_toolbox\\_DeeBNet\\_V20](https://www.researchgate.net/publication/264790642_A_brief_survey_on_deep_belief_networks_and_introducing_a_new_object_oriented_MATLAB_toolbox_DeeBNet_V20).

(下转第2534页)

目隐含因子向量维度  $f$  的选择对于实验精度有重大影响, 本文分别选取  $f$  值为 10、20、50、80、100 进行对比实验(UEITMF 算法中  $H_f$  取值均为 1.1), 这里训练集占 80%, 测试集占 20%, 实验结果如表 2 所示。

表 2 不同  $f$  值下算法对应的 RMSE 值

算法	$f$				
	10	20	50	80	100
BMF	0.8732	0.8730	0.8713	0.8705	0.8692
TACS	0.8682	0.8665	0.8641	0.8636	0.8629
UEITMF	0.8638	0.8640	0.8620	0.8610	0.8597

观察表 2 可得, 在相同  $f$  取值下, RMSE 取值由小到大依次为 UEITMF、TACS 和 BMF, 说明本文提出的 UEITMF 算法能有效提高推荐精度, 在  $f$  依次取值为 10、20、50、80、100 时, UEITMF 相对于 BMF 的精度提升依次为 1.08%、1.03%、1.07%、1.09%、1.09% (精度提升的计算公式为  $(\text{BMF} - \text{UEITMF}) / \text{BMF} \times 100\%$ , BMF 和 UEITMF 分别代表在同样的  $f$  下对应的 RMSE 值), 这些数值说明在不同的隐含因子维度  $f$  取值下 UEITMF 的精度提升在一定范围内是稳定的。进一步观察表 2 中数据可得, 在同等  $f$  取值下, UEITMF 相对于 TACS 的精度提升基本上都大于 TACS 相对于 BMF 的精度提升, 这说明噪声数据对于精度提升有更大的影响。

## 4 结语

本文提出了用户信息熵模型, 解决了协同过滤推荐中存在的噪声数据问题, 同时将用户信息熵模型和项目时效性模型相结合, 提出融合用户信息熵和项目时效性的矩阵分解算法, 实验结果表明本文提出的算法能有效提高推荐精度。

在过滤噪声数据的过程中, 采取的是直接删除噪声数据用户的方式, 但是其中不可避免地存在误分类的正常用户, 如何既为噪声数据用户产生推荐同时又消除噪声数据对于推荐结果的影响, 是进一步的研究方向。

## 参考文献:

- [1] 许海玲, 吴潇, 李晓东, 等. 互联网推荐系统比较研究 [J]. 软件学报, 2009, 20(2): 350–362. (XU H L, WU X, LI X D, et al. Comparison study of Internet recommendation system [J]. Journal of Software, 2009, 20(2): 350–362.)
- [2] RESNICK P, VARIAN H R. Recommender system [J]. Communications of the ACM, 1997, 40(3): 56–58.
- [3] CHU W, PARK S T. Personalized recommendation on dynamic con-
- [4] 孟祥武, 刘树栋, 张玉洁, 等. 社会化推荐系统研究 [J]. 软件学报, 2015, 26(6): 1356–1372. (MENG X W, LIU S D, ZHANG Y J, et al. Research on social recommender systems [J]. Journal of Software, 2015, 26(6): 1356–1372.)
- [5] WANG G, XIE S, LIU B, et al. Review graph based online store review spammer detection [C]// ICDM 2011: Proceedings of the 2011 International Conference on Data Mining. Washington, DC: IEEE Computer Society, 2011: 1242–1247.
- [6] SONG J, LEE S, KIM J. Spam filtering in twitter using sender-receiver relationship [C]// RAID '11: Proceedings of the 2011 14th International Conference on Recent Advances in Intrusion Detection. Berlin: Springer, 2011: 301–317.
- [7] CHIRITA P A, NEJDL W, ZAMFIR C. Preventing shilling attacks in online recommender systems [C]// WIDM '05: Proceedings of the 2005 7th Annual ACM International Workshop on Web Information and Data Management. New York: ACM, 2005: 67–74.
- [8] BILGE A, ÖZDEMİR Z, POLAT H. A novel shilling attack detection method [J]. Procedia Computer Science, 2014, 31: 165–174.
- [9] CAO J, WU Z, MAO B, et al. Shilling attack detection utilizing semi-supervised learning method for collaborative recommender system [J]. World Wide Web, 2013, 16(5/6): 729–748.
- [10] 刘江冬, 梁刚, 杨进. 基于时效性的冷启动解决算法 [J]. 现代计算机, 2016(2): 3–6. (LIU J D, LIANG G, YANG J. Timeliness-based algorithm for cold start [J]. Modern Computer, 2016 (2): 3–6.)
- [11] KOREN Y, BELL R, VOLINSKY C. Matrix factorization techniques for recommender systems [J]. Computer, 2009, 42(8): 30–37.

## Background

**LIU Jiangdong**, born in 1989, M. S. candidate. His research interests include machine learning, recommender system.

**LIANG Gang**, born in 1976, Ph. D., lecturer. His research interests include machine learning, intelligent computing.

**FENG Cheng**, born in 1991, M. S. candidate. His research interests include machine learning, rumor detection.

**ZHOU Hongyu**, born in 1990, M. S. candidate. His research interests include data analysis.

(上接第 2525 页)

- [19] COLLOBERT R, SINZ F, WESTON J, et al. Large scale transductive SVMs [J]. Journal of Machine Learning Research, 2006, 7: 1687–1712.
- [20] HAGAN M T, DEMUTH H B, BEALE M. Neural Network Design [M]. Beijing: China Machine Press, 2002: 357.
- [21] SALAKHUTDINOV R, HINTON G E. Learning a nonlinear embedding by preserving class neighbourhood structure [J]. Journal of Machine Learning Research, 2007, 2: 412–419.
- [22] TELEMAN T, HINTON G. Using fast weights to improve persistent contrastive divergence [C]// ICML '09: Proceedings of the 26th Annual International Conference on Machine Learning. New

York: ACM, 2009: 1033–1040.

## Background

This work is partially supported by the Natural Science Foundation of Shanxi Province (2015011045).

**LI Meng**, born in 1990, M. S. candidate. Her research interests include computer vision, digital image processing, deep learning.

**QIN Pingle**, born in 1978, Ph. D., associate professor. His research interests include big data, machine vision, 3D reconstruction.

**LI Chuanpeng**, born in 1991. M. S. candidate. His research interests include computer vision, digital image processing, deep learning.